Multiple Genes Provide the Basis for Antifreeze Protein Diversity and Dosage in the Ocean Pout, *Macrozoarces americanus*  

(Received for publication, April 28, 1987, and in revised form, April 13, 1988)

Choy L. Hew, Nam-C. Wang, and Shashikant Joshi  
From the Research Institute, Hospital for Sick Children, Toronto and the Departments of Clinical Biochemistry and Biochemistry, University of Toronto, Toronto, Ontario M5G 1L5, Canada

Garth L. Fletcher  
From the Marine Sciences Research Laboratory, Memorial University of Newfoundland, St. John's Newfoundland A1C 5S7, Canada

Gary K. Scott†, Pliny H. Hayes, Bert Buettner, and Peter L. Davies§  
From the Department of Biochemistry, Queen's University, Kingston, Ontario K7L 3N6, Canada

The ocean pout (*Macrozoarces americanus*) produces a set of antifreeze proteins that depresses the freezing point of its blood by binding to, and inhibiting the growth of, ice crystals. The amino acid sequences of all the major components of the ocean pout antifreeze proteins, including the immunologically distinct QAE component, have been derived by Edman degradation. In addition, sequences of several minor components were deduced from DNA sequencing of cDNA and genomic clones. Fifty percent of the amino acids are perfectly conserved in all these proteins as well as in two homologous sequences from the distantly related wolffish. Several of the conserved residues are threonines and asparagines, amino acids that have been implicated in ice binding in the structurally unrelated antifreeze protein of the righteye flounders. Aside from minor differences in post-translational modifications, heterogeneity in antifreeze protein components stems from amino acid differences encoded by multiple genes. Based on genomic Southern blots and library cloning statistics there are 150 copies of the 0.7-kilobase-long antifreeze protein gene in the Newfoundland ocean pout, the majority of which are closely linked but irregularly spaced. A more southerly population of ocean pout from New Brunswick in which the circulating antifreeze protein levels are considerably lower has approximately one-quarter as many antifreeze protein genes. Thus, there appears to be a correlation between gene dosage and antifreeze protein levels, and hence the ability to survive in ice-laden seawater. Southern blot comparison of the two populations indicates that the differences in gene dosage were not generated by a simple set of deletions/duplications. They are more likely to be the result of differential amplification.

The Newfoundland ocean pout, *Macrozoarces americanus*, produces a family of at least 10 active antifreeze polypeptides (AFP) to prevent it from freezing (1). These AFP occur in concentrations of approximately 20–25 mg/ml in the serum during the winter months and are retained at much lower concentrations during the summer (2). Studies from our laboratories (1) have shown that the ocean pout AFP (type III) are strikingly different in amino acid composition from the alamine-rich α-helical type I AFP isolated from winter founder (3–5) and shortnose sculpin (6, 7), and the cystine-rich type II AFP of the sea raven (8). Ocean pout AFP are all in the molecular weight range of 6000–7000 but can be fractionated into five distinct groups based on their behavior on ion exchange chromatography. One group binds to QAE-Sephadex (QAE-1) and four to SP-Sephadex (SP-1 to -4) (1). On reverse phase high performance liquid chromatography (HPLC) the QAE-1 group shows a single peak while each of the SP groups contains several components. Besides the difference in their binding properties on ion exchange chromatography, the QAE-1 group differs from all the SP components in immunological cross-reactivity. Antisera to QAE-1 react poorly to the SP components and vice versa (1).

Recently, the amino acid sequences of the three components in the SP-1 group have been deduced by a combination of protein and cDNA sequencing (9). Whereas the sequence of winter flounder AFP enabled its secondary structure and a mechanism of action to be predicted directly (10), the ocean pout AFP sequences do not provide any obvious clues to their secondary structure or to how they function as antifreezes. To see which amino acids are most conserved and might, therefore, play a key role in the structure and/or function of the ocean pout antifreezes we have determined the amino acid sequences of QAE-1 and the major components from SP-2, SP-3, and SP-4, and have derived other sequences from cDNA and genomic clones. In addition we have compared the ocean pout sequences to two AFP sequences derived from genomic clones of the wolfish (*Anarhichas lupus*), a related zoarcid from a different family (11). Approximately 50% of the 65

---

*This work was supported by grants from the Medical Research Council of Canada (to C. L. H. and P. L. D.) and from the Natural Sciences and Engineering Research Council of Canada (to G. L. F.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Current address: University of California School of Medicine, Cancer Research Institute, M-1282, San Francisco, CA 94143-0128.

‡ To whom correspondence should be addressed.

1 The abbreviations used are: AFP, antifreeze polypeptide(s); HPLC, high performance liquid chromatography.
The amino acid sequences of all the major ocean pout AFP components (HPLC-1, -4, -6, -7, -9, -11, and -12) are shown aligned with the two sequences derived from cDNA cloning, two sequences predicted from genomic cloning (this study), and two from the related wolffish (31) (Fig. 3). The sequences range from 62 to 69 amino acids in length. Most of the length variation occurs at the termini and is due to post-translational processing (9). The overall sequence identity among the 13 components aligned is approximately 50%. Many of the non-identical residues include sets of conservative substitutions such as isoleucine for methionine or leucine, and valine for alanine. Sequence identity increases to 90% among the ocean pout SP components, and the two wolffish sequences are far more similar to these sequences (85% identical) than ocean pout SP sequences are to the QAE sequences (55% identical).

Four of the perfectly conserved residues (14, 18, 47, and 55) are threonines or asparagines. In addition, position 8 is occupied by threonine or asparagine, and position 15 by threonine, except in cDNA c10 where this residue is replaced by serine. The potential significance of this observation lies in the fact that the structurally different AFP of the flounder (type 1 AFP) also has six or seven conserved threonine and asparagine residues (23) that are believed to be responsible for binding to ice crystals (10).

**DISCUSSION**

**Size and Organization of the AFP Multigene Family—**An estimate of the number of AFP genes in the ocean pout genome can be made on the basis of the frequency with which AFP genomic clones appeared in the library screenings. A total of 20 clones was detected in 1.2 × 10^6 recombinant phage screened. Given a haploid genome size for the ocean pout of 1 × 10^9 base pairs, an observed average phage insert length of 1.7 × 10^7 base pairs, and an average of 1.5 AFP genes/clone, the estimated AFP gene copy number for the Newfoundland fish is 150. This is not an unreasonable figure in light of the number and intensity of hybridization signals seen on the genomic Southern blot (Figs. 6 and 10). Since none of the enzymes used to prepare this blot cut within the genes or cDNAs analyzed, it appears that each band on the genomic blot represents at least one gene. In effect the genes are so closely linked that several may be present in some of the longer hybridizing restriction fragments. Indeed, based on the maps of XOP 6, 12, 21, and 23, there are genomic EcoRI and BamHI fragments with at least two AFP genes apiece. Also, as mentioned previously several of the hybridizing bands in the genomic blot appear to be comprised of signals from multiple restriction fragments. For example, the maps in Fig. 7 show that λOP 5, 12, 19, and 23 each contain a 2-kilobase pair HindIII fragment that hybridizes to the cDNA probes and that could contribute to the extensive band of hybridization in this size range in the HindIII lane of the genomic Southern blot. These observa-
**Antifreeze Protein Multigene Family in the Ocean Pout**

**Fig. 3. Compendium of sequences for type III AFP.** Sequences were derived from ocean pout HPLC components, cDNA clones, and genomic clones as indicated, and from two wolffish genomic clones (W1.5 and W1.9). Regions of identity are boxed. Spaces were introduced to optimize the alignment. Asterisks mark conserved threonines or asparagines, while asterisks in parentheses indicate positions occupied by either of these residues or aspartic acid.

**Fig. 9. DNA sequence of ocean pout AFP genes.** A, the gene from OIP 3; B, the gene from OIP 5. Bases are numbered from the 5' end HindIII site. The intervening sequence is written in lowercase letters, and the putative "CAAT" and "TATA" sequences are boxed.
the AFP multigene family. Southern blot analysis of testis DNA from two Newfoundland (Nfd) and two New Brunswick (NB) ocean pout digested with BamHI (B) or SstI (S) and probed with cloned AFP cDNA (9). B, the same blot reprobed with cloned chicken \( \beta \)-tubulin cDNA (30).

**Fig. 10.** A, population differences in the AFP multigene family. Southern blot of testis DNA from two Newfoundland (Nfd) and two New Brunswick (NB) ocean pout. B, the same blot reprobed with cloned chicken \( \beta \)-tubulin cDNA (30).

The dramatic difference in the size of the AFP gene family between the Newfoundland and New Brunswick ocean pout populations can largely account for the order of magnitude difference in their circulating AFP concentrations (2). A similar correlation between gene dosage and AFP levels has been noted before in closely related species of righteye flounders (26). However, in this instance the difference occurs within a single species, which shortens the potential time frame for its development. As pointed out above, a comparison of the patterns of AFP gene hybridization between the Newfoundland and New Brunswick ocean pout (Fig. 10) suggests that one has not been derived from the other by a simple set of deletions/duplications since so many of the gene fragments are of different lengths in the two populations, and even between the two New Brunswick individuals. Instead, the AFP gene locus appears to be subject to extensive rearrangement. We suggest that the periodicity of Cenozoic glaciation in the Northern hemisphere has provided intense but sporadic selection for ocean pout that produce high levels of AFP. This challenge has been met on a number of occasions by AFP gene amplification through some form of disproportionate DNA replication (27, 28). Subsequently, during glacial minima, a net loss of AFP genes that can be predicted from the potential for internal recombination (29) would not be selected against until the onset of another glacial episode. Thus, the combination of an AFP gene locus predisposed to amplification and a periodic selection pressure might account for the current geographical and individual variation in the locus.

**Functionality of the Genes—**It is difficult to say how many of the 150 AFP genes are functional. Protein sequence analysis indicates that of the eight major HPLC components analyzed here and in Ref. 9, only one, HPLC-5, can be produced as a post-translational modification of another. In addition there are several minor components not yet sequenced, such as HPLC-2 and -3, that could potentially correspond to the cDNA sequences of clones 7 and 10 or the genomic sequence in XOP 5. Other minor components flank HPLC-7, -9, and -11. Based on these estimates there are 10–15 different AFP components. However, the number of functional genes is likely to be larger since sequence analysis of cDNA clones 36 and 77 obtained from an individual fish has shown that the same mRNA (HPLC-6) is produced from at least two genes that differ by silent base changes (9). Some of the genes may even be exact duplicates. To obtain a more accurate figure for the number of functional AFP genes it would be necessary to sequence a large number of them. The two genes (AOP 3 and 5) that have been sequenced both appear to be functional although they do not code for any of the major components. They show extensive DNA sequence conservation around the putative control elements and even in their intervening sequences.

**The Type III Antifreeze—**Through a combination of protein and DNA sequencing a fairly complete picture of ocean pout AFP component heterogeneity has emerged (Fig. 3). Perhaps the most surprising feature is the extensive sequence difference between the SP and QAE components. This difference is greater than might be predicted from a comparison of their amino acid compositions and chromatographic properties but is in line with immunological data (1). Their comparison does, however, serve to identify conserved residues that might be important in ice binding or in the folding of the type III AFP in such a way as to present the ice-binding residues in the correct configuration. This information will be especially valuable when the first type III AFP tertiary structure is derived by physical methods.

**Acknowledgments—**We thank the diving facilities at the Marine
**Antifreeze Protein Multigene Family in the Ocean Pout**

B. Fletcher, C. L. Hew, C. L. Slaughter, D. Joshi, N. C. Wang, and Ananthanarayan, V. S. (1984) *J. Comp. Biochem.* 54, 375-380

**EXPERIMENTAL PROCEDURES**

**Antifreeze Protein Multigene Family in the Ocean Pout**

B. Fletcher, C. L. Hew, C. L. Slaughter, D. Joshi, N. C. Wang, and Ananthanarayan, V. S. (1984) *J. Comp. Biochem.* 54, 375-380

- **Collection of tissue.** Ocean pout, *M. americanus*, were collected from the ocean off the coast of Newfoundland, and from Pennsylvania Bay, New Brunswick. The fish were kept in aquaria at ambient temperature and photoperiod prior to sampling. The tissue was prepared for centrifugation (5,000 g for 30 min and stored at -80 °C).
- **Liver and testes were removed from intact fish and stored at -80 °C prior to use.**

**Isolation of Ocean Pout AFP.** The procedure for the purification of AFP is described in Table 1. Briefly, the serum was initially chromatographed on a Sephadex G-75 column (2 x 80 cm). After gel filtration, the AFP was focussed by reverse phase HPLC using a Wasp-Bondapak C-18 column (5 μm x 25 cm) with a 0.1% trifluoroacetic acid-methanol gradient, and their molecular masses were estimated with a laser light scattering detector. The AFP was eluted from the HPLC column with a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.), and the apparent molecular mass was determined for each peak using a 10% sucrose gradient. The AFP was then purified further by reverse phase HPLC using a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.).

**Detection of AFP component(s).** The AFP component(s) were detected using reverse phase HPLC using a Wasp-Bondapak C-18 column (5 μm x 25 cm) with a 0.1% trifluoroacetic acid-methanol gradient, and their molecular masses were estimated with a laser light scattering detector. The AFP was eluted from the HPLC column with a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.), and the apparent molecular mass was determined for each peak using a 10% sucrose gradient. The AFP was then purified further by reverse phase HPLC using a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.).

**Analysis of AFP component(s).** The AFP component(s) were detected using reverse phase HPLC using a Wasp-Bondapak C-18 column (5 μm x 25 cm) with a 0.1% trifluoroacetic acid-methanol gradient, and their molecular masses were estimated with a laser light scattering detector. The AFP was eluted from the HPLC column with a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.), and the apparent molecular mass was determined for each peak using a 10% sucrose gradient. The AFP was then purified further by reverse phase HPLC using a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.).

**RESULTS**

The pepsin-EDLC-4 family of AFP was detected in crude extracts of *M. americanus*, and the presence of the AFP component(s) was confirmed using reverse phase HPLC and laser light scattering detectors. The AFP component(s) were detected using reverse phase HPLC using a Wasp-Bondapak C-18 column (5 μm x 25 cm) with a 0.1% trifluoroacetic acid-methanol gradient, and their molecular masses were estimated with a laser light scattering detector. The AFP was eluted from the HPLC column with a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.), and the apparent molecular mass was determined for each peak using a 10% sucrose gradient. The AFP was then purified further by reverse phase HPLC using a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.).

**AFD-CPD clones.** A polypeptide that was purified from crude extracts of *M. americanus* and identified by reverse phase HPLC was isolated and purified further by reverse phase HPLC using a Wasp-Bondapak C-18 column (5 μm x 25 cm) with a 0.1% trifluoroacetic acid-methanol gradient, and their molecular masses were estimated with a laser light scattering detector. The AFP was eluted from the HPLC column with a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.), and the apparent molecular mass was determined for each peak using a 10% sucrose gradient. The AFP was then purified further by reverse phase HPLC using a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.).

**AFD-CPD clones.** A polypeptide that was purified from crude extracts of *M. americanus* and identified by reverse phase HPLC was isolated and purified further by reverse phase HPLC using a Wasp-Bondapak C-18 column (5 μm x 25 cm) with a 0.1% trifluoroacetic acid-methanol gradient, and their molecular masses were estimated with a laser light scattering detector. The AFP was eluted from the HPLC column with a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.), and the apparent molecular mass was determined for each peak using a 10% sucrose gradient. The AFP was then purified further by reverse phase HPLC using a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.).

**AFD-CPD clones.** A polypeptide that was purified from crude extracts of *M. americanus* and identified by reverse phase HPLC was isolated and purified further by reverse phase HPLC using a Wasp-Bondapak C-18 column (5 μm x 25 cm) with a 0.1% trifluoroacetic acid-methanol gradient, and their molecular masses were estimated with a laser light scattering detector. The AFP was eluted from the HPLC column with a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.), and the apparent molecular mass was determined for each peak using a 10% sucrose gradient. The AFP was then purified further by reverse phase HPLC using a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.).

**AFD-CPD clones.** A polypeptide that was purified from crude extracts of *M. americanus* and identified by reverse phase HPLC was isolated and purified further by reverse phase HPLC using a Wasp-Bondapak C-18 column (5 μm x 25 cm) with a 0.1% trifluoroacetic acid-methanol gradient, and their molecular masses were estimated with a laser light scattering detector. The AFP was eluted from the HPLC column with a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.), and the apparent molecular mass was determined for each peak using a 10% sucrose gradient. The AFP was then purified further by reverse phase HPLC using a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.).

**AFD-CPD clones.** A polypeptide that was purified from crude extracts of *M. americanus* and identified by reverse phase HPLC was isolated and purified further by reverse phase HPLC using a Wasp-Bondapak C-18 column (5 μm x 25 cm) with a 0.1% trifluoroacetic acid-methanol gradient, and their molecular masses were estimated with a laser light scattering detector. The AFP was eluted from the HPLC column with a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.), and the apparent molecular mass was determined for each peak using a 10% sucrose gradient. The AFP was then purified further by reverse phase HPLC using a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.).

**AFD-CPD clones.** A polypeptide that was purified from crude extracts of *M. americanus* and identified by reverse phase HPLC was isolated and purified further by reverse phase HPLC using a Wasp-Bondapak C-18 column (5 μm x 25 cm) with a 0.1% trifluoroacetic acid-methanol gradient, and their molecular masses were estimated with a laser light scattering detector. The AFP was eluted from the HPLC column with a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.), and the apparent molecular mass was determined for each peak using a 10% sucrose gradient. The AFP was then purified further by reverse phase HPLC using a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.).

**AFD-CPD clones.** A polypeptide that was purified from crude extracts of *M. americanus* and identified by reverse phase HPLC was isolated and purified further by reverse phase HPLC using a Wasp-Bondapak C-18 column (5 μm x 25 cm) with a 0.1% trifluoroacetic acid-methanol gradient, and their molecular masses were estimated with a laser light scattering detector. The AFP was eluted from the HPLC column with a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.), and the apparent molecular mass was determined for each peak using a 10% sucrose gradient. The AFP was then purified further by reverse phase HPLC using a Beckman 500 Bio-Log detector (Clifton Technical Institute, Clifton, N.J.).
Antifreeze Protein Multigene Family in the Ocean Pout

Table 1. Automated Edman degradation of HPLC-12 and its chymotryptic digests. The numbers in parentheses represent the number of residue/peptide pairs in each digest.

Table 2. Automated Edman degradation of HPLC-12 and its CNBr and chymotryptic digests. The numbers in parentheses represent the number of residue/peptide pairs in each digest.

Table 3. Automated Edman degradation of HPLC-12 and its CNBr and chymotryptic digests. The numbers in parentheses represent the number of residue/peptide pairs in each digest.

Table 4. Automated Edman degradation of HPLC-12 and its CNBr and chymotryptic digests. The numbers in parentheses represent the number of residue/peptide pairs in each digest.

Fig. 2. HPLC analysis of the tryptic, chymotryptic, and CNBr digests of HPLC-12. Chromatography was performed as described in the legend to Fig. 1. (A) Tryptic digest; (B) chymotryptic digest; (C) CNBr digest.
Antifreeze Protein Multigene Family in the Ocean Pout

Flq.

ReScr1CtlOn masts Of Ocean pout AFP genomic clones. The single horizontal lines denote the single inserts in the clones and the parallel lines the adjacent DNA sequences. Vertical lines indicate the placement of restriction sites. The vertical dotted lines in AFP 12 and AFP 21 represent the two possible locations of a single HindIII site. The stippling between restriction sites identifies DNA fragments that hybridize strongly to the ocean pout AFP cdna.

Fig. 4 Trypsin peptide mapping of different ocean pout AFP. Trypsin digests of (a) AFP-12; (b) AFP-4; (c) AFP-11; (d) AFP-9; (e) AFP-C7 were chromatographed as described in the legend to Fig. 1.

Fig. 5 Sequence of ocean pout AFP cDNA clones #7 and #10. The continuous DNA sequence is that of the #7 cDNA insert with the corresponding amino acid sequence above. The cDNA of #10 is indicated below by base changes from the #7 sequence, four of which lead to the amino acid changes indicated. The #10 DNA sequence begins at base 301.

Fig. 6 Genomic Southern blots. Genomic DNA (15 µg) from ocean pout was digested with restriction enzymes BglII (B), BglII (R), EcoRI (E) and HindIII (H). The left-hand blot was probed with cDNA clone #7, the right-hand blot was probed with ocean pout antifreeze protein cDNA clone #10.

Fig. 7 Restriction maps of ocean pout AFP genomic clones. The single horizontal lines denote the single inserts in the clones and the parallel lines the adjacent DNA sequences. Vertical lines indicate the placement of restriction sites. The vertical dotted lines in AFP 12 and AFP 21 represent the two possible locations of a single HindIII site. The stippling between restriction sites identifies DNA fragments that hybridize strongly to the ocean pout AFP cDNA.

Fig. 8 Strategy for sequencing the AFP genes in AFP 12 and AFP 32. Within the restriction maps of the ocean pout AFP gene regions the hatched rectangles denote exons. The length of the single intervening sequence is indicated in each case. Arrow and attendant numbers indicate the direction and extent of DNA sequence determined from the various restriction sites. Library when using AFlP cDNA clone #14 as a probe.