

An antifreeze glycopeptide gene from the antarctic cod *Notothenia coriiceps neglecta* encodes a polypeptide of high peptide copy number

(antarctic nototheniid/DNA sequence/tandem repeats/signal peptide)

KU-CHUAN HSIAO*, CHI-HING C. CHENG*, INDIRA E. FERNANDES[†], H. WILLIAM DETRICH[†],
AND ARTHUR L. DeVRIES*

*Department of Physiology and Biophysics, University of Illinois, Urbana, IL 61801; and [†]Department of Biology, Northeastern University, Boston, MA 02115

Communicated by George N. Somero, September 4, 1990 (received for review July 10, 1990)

ABSTRACT The antarctic fish *Notothenia coriiceps neglecta* synthesizes eight antifreeze glycopeptides (AFGP 1–8; M_r 2600–34,000) to avoid freezing in its ice-laden freezing habitat. We report here the sequence of one of its AFGP genes. The structural gene contains 46 tandemly repeated segments, each encoding one AFGP peptide plus a 3-amino acid spacer. Most of the repeats (44/46) code for peptides of AFGP 8; the remaining 2 code for peptides of AFGP 7. At least 2 of the 3 amino acids in the spacers could act as substrate for chymotrypsin-like proteases. The nucleotide sequence between the translation initiation codon (ATG) and the first AFGP-coding segment is G+T-rich and encodes a presumptive 37-residue signal peptide of unusual sequence. Primer extension establishes the transcription start site at nucleotide 43 upstream from ATG. CAAT and TATA boxes begin at nucleotides 53 and 49, respectively, upstream from the transcription start site. The polyadenylation signal, AATAAA, is located \approx 240 nucleotides downstream from the termination codon. A mRNA (\approx 3 kilobases) was found that matches the size of this AFGP gene. Thus, this AFGP gene encodes a secreted, high-copy-number polypeptide that is processed posttranslationally to produce active AFGPs.

To survive the freezing temperatures of the ice-laden polar and north temperate oceans, many fish inhabiting these waters synthesize biological antifreezes (AFs), which lower the freezing point of their body fluids below that of the ambient seawater (1, 2). These AFs, either peptides (AFP) or glycopeptides (AFGP), are synthesized by the liver and secreted into the blood from which they become distributed to almost the entire extracellular space (3, 4).

There are three classes of AFPs: the alanine-rich α -helical AFPs from winter flounder (5, 6), alaskan plaice (A.L.D., unpublished data), and short-horn sculpin (7); the cysteine-rich β -structured AFP from sea raven (8); and the eelpout AFPs, which have no biased amino acid composition or defined secondary structure (9–11). The AFGPs are found in the antarctic nototheniid fish commonly known as antarctic cods (1, 12), and several northern true cods (13–15) in at least eight molecular sizes (M_r 2600–34,000), composed of various numbers of repeats of the basic glycotripeptide unit, Ala-Ala-Thr, with the disaccharide, galactose-*N*-acetylgalactosamine attached to the threonine residues (16–18). Minor variations in structures are found in the smaller forms (AFGP 6, 7, and 8), in that proline and arginine replace one or more of the alanine and threonine residues, respectively (13, 19, 20).

A striking feature of AFs is their marked molecular heterogeneity. Each fish synthesizes multiple AF components,

which are either compositionally nearly identical but size variant, as in the case of AFGPs, or compositionally variant ($\leq 20\%$) but similar in size, as in the case of AFPs. The AFPs of several northern fishes have been shown to be encoded by multigene families of various numbers of members (30 to 150) that are closely linked, mostly in direct repeats (21–25).

The genomic organization of AFGP genes has not been characterized to date. The heterogeneity of the AFGPs implies that they may be encoded by a multigene family as well. However, since all the AFGPs are composed of the same basic glycotripeptide unit, it is also possible that they are the products of posttranslational cleavage of a large precursor. To distinguish between these two possibilities, we have characterized a genomic subclone that contains the gene for AFGP 8 from the antarctic cod, *Notothenia coriiceps neglecta*.[‡]

MATERIALS AND METHODS

Library Construction and Screening. *N. coriiceps neglecta* genomic library was constructed with testes DNA from a single fish in the λ vector Charon 35 (26). The DNA was partially digested with *Mbo* I and 15- to 20-kilobase (kb) fragments obtained from sucrose gradient size fractionation were ligated to *Bam*HI sites of the λ vector (27). Recombinant phage DNA was packaged *in vitro* (Packagene; Promega), and the library was amplified. The library was screened for AF genes (28) with an end-labeled oligonucleotide probe, 5'-GCIGCIACIGCIGCIACIGCIGCIA-3', constructed according to the amino acid sequence AATAATAAT.

DNA Sequence Analysis. Positive genomic clones were characterized by Southern hybridization (29), and appropriate restriction fragments were subcloned into either pTZ19U or pTZ19R (gifts from Byron Kemper, University of Illinois, Urbana). Nested sets of unidirectional deletions in the inserts were generated by the exonuclease III digestion method (30). Double-stranded DNA templates generated from these constructs were sequenced by the dideoxynucleotide chain-termination procedure (31) using modified T4 DNA polymerase (Sequenase; United States Biochemical).

Purification and Analysis of RNA. RNA was isolated from fish liver by homogenization in the presence of guanidinium isothiocyanate followed by centrifugation of the homogenate in CsCl solution (32). Poly(A)⁺ RNA was isolated by oligo(dT)-cellulose affinity chromatography (33). For Northern analysis, 50 µg of poly(A)⁺ RNA was electrophoresed on agarose gels with formaldehyde (32) and then blotted onto Hybond-N membranes (Amersham) and probed with a *Pst* I

Abbreviations: AF, antifreeze; AFP, AF peptide; AFGP, AF glycopeptide; ORF, open reading frame.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M55000).

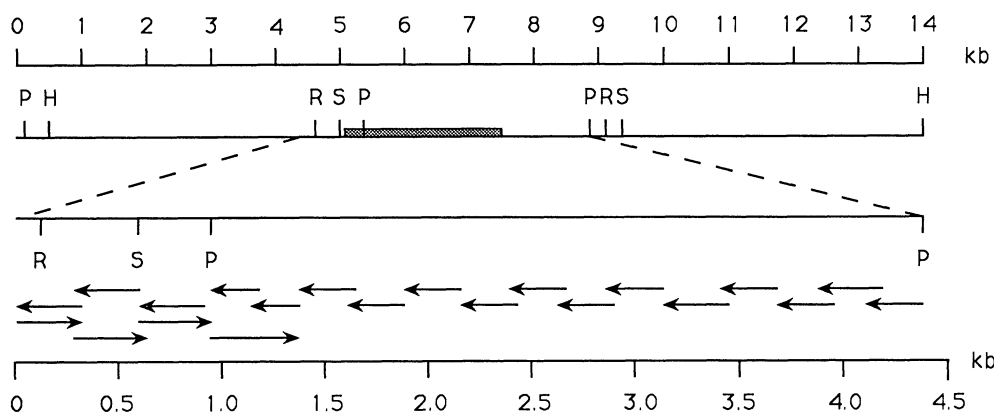


FIG. 1. Restriction map and DNA sequencing strategy for a genomic clone containing an AFGP gene. Arrows indicate direction and extent of sequencing. H, *HindIII*; P, *PstI*; R, *EcoRI*; S, *SalI*. The location of the AFGP gene is indicated by the stippled box.

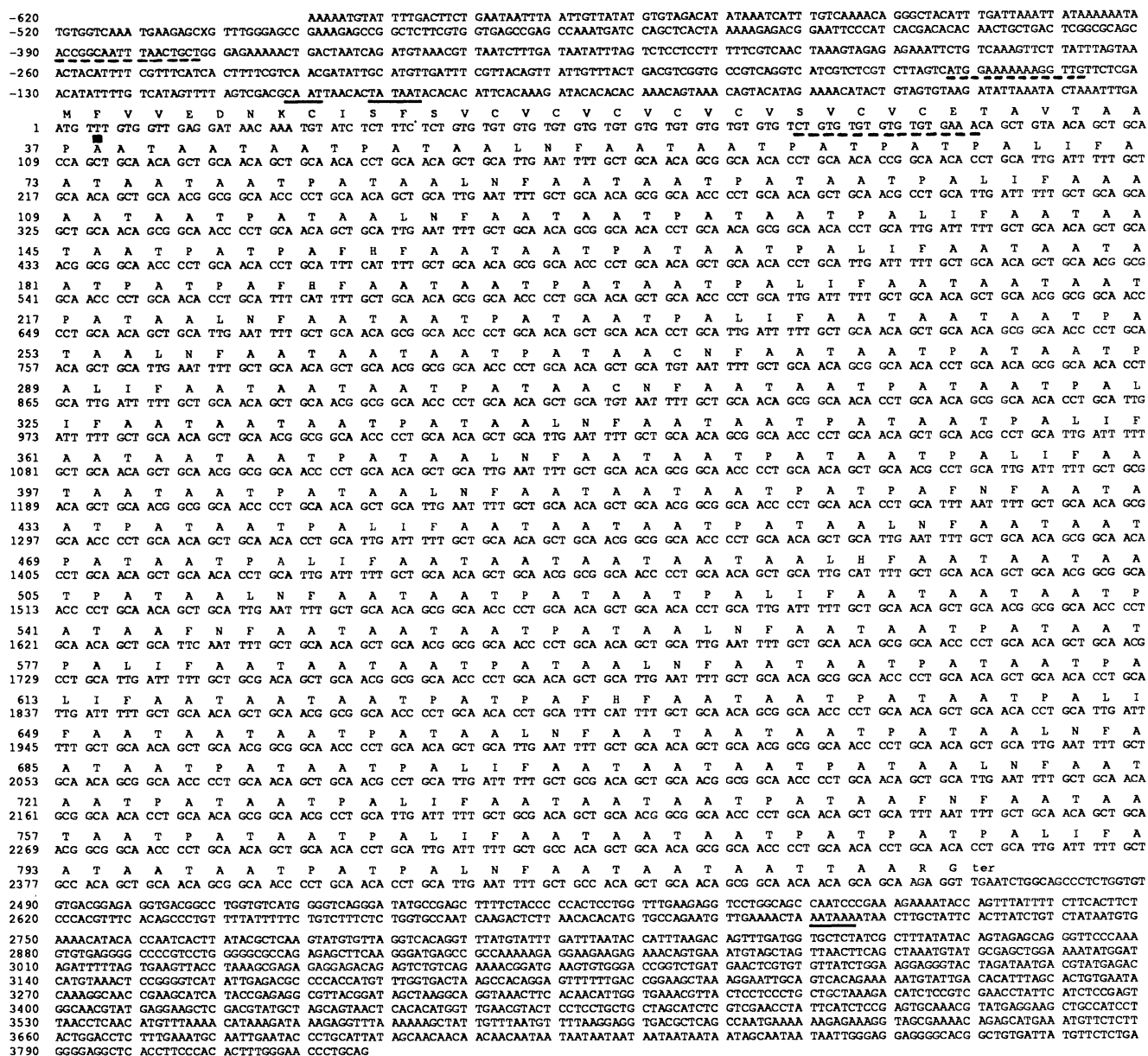


FIG. 2. Nucleotide sequence of an AFGP gene and amino acid sequence of the encoded protein. The cap site of the transcript is denoted by a solid circle. The CAAT and TATA boxes as well as the polyadenylation signal, AATAAA, are underlined. The apparent start of the mature AFGP 8 polypeptide is marked by a solid square. The synthetic oligonucleotides (complementary to the sense strand) used in the primer-extension experiments are marked by dashed lines.

restriction fragment of the AFGP gene labeled with ^{32}P by the random-primer method (34). The membrane was hybridized at 65°C and subsequently washed twice in $1\times$ SSC/1% SDS and once in $0.1\times$ SSC/0.1% SDS at 65°C ($1\times$ SSC = 0.15 M NaCl/0.015 M sodium citrate).

The 5' terminus of the AFGP mRNA was mapped by primer extension (35). DNA oligonucleotide primers were end-labeled, and each primer (1 pmol, 10^6 cpm) was then hybridized to 50 μg of total RNA and extended with AMV reverse transcriptase (Life Sciences, Saint Petersburg, FL). The length of the extension product in bases was determined by coelectrophoresis next to an independent sequence reaction, which gives a reference ladder of known sequence.

RESULTS

From an initial screen of 3.6×10^5 plaques of the *N. coriiceps neglecta* genomic library, three positive clones were obtained. All three clones exhibit the same restriction map and therefore are most likely identical clones. One of the clones was analyzed by Southern analysis using the oligonucleotide probe, and the restriction map is shown in Fig. 1. The two *Pst* I fragments, ≈ 5.4 and ≈ 3.3 kb, were recovered and subcloned into pTZ19. The sequencing strategy for the AFGP gene is shown in Fig. 1.

A contiguous sequence of 4381 nucleotides containing the AFGP gene sequence along with the deduced amino acids is shown in Fig. 2. It contains an open reading frame (ORF), starting with the methionine codon at nucleotides 1–3, with 822 codons up to an opal termination codon TGA (nucleotides 2467–2469). Also shown are 620 and 1358 nucleotides of the 5' and 3' untranslated sequences, respectively. The putative regulatory elements, TATAAT and CAAT, are located at –92 to –87 and –102 to –99, respectively, and the polyadenylation signal sequence AATAAA is at nucleotides 2710–2716.

Starting at nucleotide 112, 46 direct repeats are tandemly arranged until the termination codon is reached. The first 43 repeats are 51 nucleotides long, the next 2 are 60 nucleotides, and the last one is 42 nucleotides. A comparison of the deduced amino acid sequence with the published peptide sequence of several AFGPs (16, 19, 20) shows that each of these repeats encodes the peptide backbone of an AFGP, plus 3 additional amino acids located at the C terminus (absent in the last repeat). The 51- and 42-nucleotide repeats (44 total) encode the smallest AFGP—i.e., AFGP 8. The two 60-nucleotide repeats (nucleotides 2254–2313 and 2314–2373) encode two copies of AFGP 7, which is 3 residues longer than AFGP 8.

There are eight distinct AFGP forms coded in this gene—six of AFGP 8, and two of AFGP 7 (Fig. 3). The majority of AFGP 8 peptides (20/44; 45.5%) have a single proline at position 10, 17 (38.6%) have 2 prolines at positions 7 and 13, and 4 (9.1%) have 2 prolines at positions 10 and 13. Two other unique AFGP 8 appear once (2.3%) in this gene sequence. One (nucleotides 163–213) has 3 prolines at positions 7, 10, and 13, and the other (nucleotides 316–366) has a single proline at position 10, but the threonine at position 3 is replaced by an alanine. Both of these single copy forms are the result of point mutations in the DNA sequence. A third single copy form is encoded by the 42-nucleotide segment immediately before the termination codon; it has no proline but an arginine and a glycine at positions 13 and 14, respectively, and its fourth tripeptide unit (Ala-Ala-Thr) is turned around and has become Thr-Ala-Ala.

There are a total of six different sets of the 3-residue segment (spacer) separating each AFGP (Fig. 3). The majority of these are either Leu-Asn-Phe (LNF) or Leu-Ile-Phe (LIF). In all cases but two, the first amino acid is either

	No. of Copies
AATAATAATPATAA <i>LNF</i>	20
<i>LHF</i>	
<i>FNF</i>	
<i>CNF</i>	
AATAATPATAATPA <i>LIF</i>	17
AATAATAATPATPA <i>FHF</i>	4
<i>FNF</i>	
AATAATPATPATPA <i>LIF</i>	1
AAAAATAATPATAA <i>LNF</i>	1
AATAATAATTAARG	1
AATAATAATPATAATPA <i>LIF</i>	1
AATAATAATPATPATPA <i>LIF</i>	1

FIG. 3. The different AFGPs and their copy number in the AFGP 8 polypeptide. The 3-amino acid segments flanking the mature AFGP are italicized. The numbers above various amino acids indicate their positions in the AFGP.

leucine or phenylalanine, while the third amino acid is always phenylalanine.

Peptide sequencing of AFGP 8 (mixture) isolated from *N. coriiceps neglecta* serum (data not shown) shows proline present at positions 7, 10, and 13, while the 3 residues flanking the C terminus deduced from the gene sequence are absent.

Between the presumptive initiation codon and the start of the AFGP 8 structural gene, there are 111 nucleotides encoding 37 amino acids, which include a long stretch of alternating valine and cysteine residues (15 total). The location of this region corresponds to the coding region of the signal peptide. However, the sequence is G+T-rich and therefore potentially Z-DNA forming, which has not been observed in known signal sequences. To determine whether this region is actually a part of the message, primer extensions were performed with total RNA using 3 different oligonucleotides, oligo-1, -2, and -3, complementary to nucleotides 74–91, –153 to –158, and –392 to –373, respectively, of the sense strand of the AFGP gene (Fig. 2, dash line). Only oligo-1 produced an extension product (Fig. 4), and it is 134 nucleotides long. This confirms that the presumptive signal sequence region is indeed a part of the mature mRNA and identifies the T at nucleotide –43 upstream of the initiation codon to be the transcription start site.

Northern analysis using the 3.3-kb *Pst* I fragment of the AFGP clone (contains 85% of the AFGP structural gene) as a probe reveals the presence of several different sized messages (Fig. 5). A strongly hybridizing band of ≈ 3.0 kb is present that corresponds to the expected size for the message produced by this gene (2.8 kb from cap site to polyadenylation signal sequence). Because AFGPs all contain the tripeptide repeat Ala-Ala-Thr, other hybridizing bands presumably represent the mRNAs for other AFGPs that cross-hybridize with the probe due to high sequence homology.

DISCUSSION

We have characterized a genomic subclone from the antarctic fish *N. coriiceps neglecta*, which contains a gene that encodes 44 copies of AFGP 8 and 2 copies of AFGP 7 plus a potential signal peptide. Also obtained from this genomic clone are 620 nucleotides upstream and 1358 nucleotides downstream from the ORF.

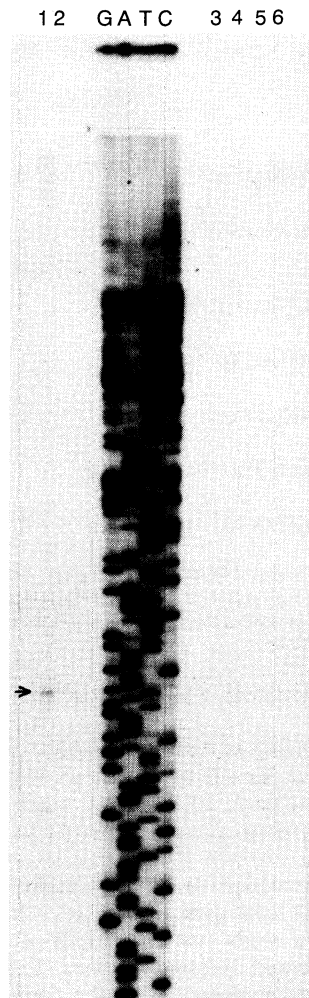


FIG. 4. Mapping of 5' ends of AFGP 8 transcripts by primer extension. Lanes: 1, 3, and 5, extensions of oligo-1, -2, and -3, respectively, as primer with *N. coriiceps neglecta* RNA; 2, 4, and 6, control lanes of the same primers extended without RNA. The extension product of oligo-1 is marked with an arrow. The sequence ladder is from an independent sequencing reaction.

The ATG at the beginning of the ORF is very likely the true initiation codon based on the following evidence. First, there are stop codons in all three reading frames within the 620 nucleotides upstream of this ATG. Second, the 111 nucleotides inclusive of this ATG encode 37 amino acids, over half of them hydrophobic, which could constitute a signal peptide. Third, primer extension identified the T at nucleotide -43 upstream of this ATG as the transcription start site. Finally, the proximal sequence upstream from this transcription start site contains the putative TATA and CAAT boxes (start at 49 and 59 nucleotides, respectively, upstream from the transcription start site).

Beginning from the transcription start site, and ending at the polyadenylation signal sequence plus a poly(A) tail, a mRNA ≈ 3 kb long can be predicted. Northern analysis showed a strongly hybridizing band matching this size. This 3-kb message, along with multiple copies of the mature peptide and 3-residue spacer, indicates that the product of this AFGP 8 gene (so called because it encodes mostly AFGP 8) is a monofunctional polyprotein.

A number of polyproteins have been reported, the majority of which are polyfunctional mammalian neuroendocrine peptide precursors (for review, see ref. 36). Polyproteins can be grouped into three classes based on the peptides they generate (36). The AFGP 8 polyprotein carries repetitions of

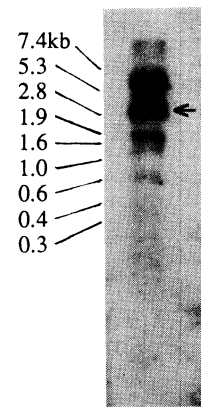


FIG. 5. Northern blot hybridization analysis of poly(A)⁺ with [α -³²P]ATP-labeled 3.3-kb *Pst* I restriction fragment, which contains 85% of the AFGP 8 structural gene. The RNA molecular size standards are shown on the left. The 3-kb mRNA corresponding to the AFGP 8 polyprotein gene is marked with an arrow.

specific and near-identical peptide and spacer sequences. These characteristics place it in the same class as the amphibian preprocaerulein (37), and the yeast prepro- α -mating factor (38), which carry two and five identical repeats, respectively. There are two unique features in the AFGP 8 polyprotein. First, the polyproteins identified to date are of low peptide copy number, up to 8 (36). The AFGP 8 polyprotein contains 46 copies, the highest peptide copy number known so far. Second, most of the bioactive peptide domains in the known polyproteins are flanked by pairs of basic amino acids (Lys-Arg, Lys-Lys, or Arg-Arg), suggesting that trypsin-like proteases are involved in the cleavage reaction (36). Of the 45 3-residue spacers in the AFGP 8 polyprotein, 43 are Leu/Phe-Xaa-Phe, and 2 are Cys-Asn-Phe, suggesting that a chymotrypsin-like protease is the cleavage enzyme. Presumably, a carboxypeptidase-like protease can remove the remaining single (Leu or Phe) or two (Cys-Asn) residues to produce the mature AFGP peptides.

The significance of the 2 copies of AFGP 7 in the midst of AFGP 8 is unclear. AFGP 7 is the second most abundant AFGP, next to AFGP 8, in the serum of all antarctic nototheniids studied (39). This implies that if these 2 copies of AFGP 7 represent the only AFGP 7 genes present in the fish genome, there must be some special differential processing mechanism that can dramatically increase the amount of their protein products. It is more likely that other AFGP polyprotein genes are present that code predominantly for AFGP 7. In any event, the presence of the AFGP 8 polyprotein gene shows that the low molecular weight forms of AFGP in fish serum are not the result of posttranslational cleavage of high molecular weight AFGPs.

In another antarctic nototheniid, *Pagothenia borchgrevinki*, serum AFGP 8 consists of a mixture of three components in a ratio of 7:2:1 (proline at 7 and 13/proline at 10 and 13/proline at 13) (20). The ratio of the three major AFGP 8 components coded in the *N. coriiceps neglecta* AFGP 8 polyprotein gene, and the placement of proline in one of them, are quite different (proline at 7 and 13, 38.6%/proline at 10 and 13, 9.1%/proline at 10, 45.5%). This may mean that different antarctic nototheniids possess AFGP 8 polyprotein genes with peptide domains that differ in composition and ratio. It is unknown whether the serum ratio of AFGP 8 components in *P. borchgrevinki* reflects the products of a single or several AFGP 8 genes. In *N. coriiceps neglecta*, partial sequencing of clones obtained from a subsequent library screening showed that there is at least one other AFGP 8 gene (data not shown).

The AFGP 8 polypeptide gene consists of multiple direct repeats of highly homologous nucleotide sequences, indicating that it may be the result of genetic duplications of one of the sequence repeats. The obvious advantages of having genes in a polypeptide form are (i) increase of peptide copy number per gene and (ii) economy of cellular energetics such as high efficiency of the transcription and translation machinery in terms of final product output. The increase in number of copies can be escalated if a family of genes encodes these proteins, creating an enormously high "gene dosage." Since only AFGPs 7 and 8 are present in the AFGP 8 gene, other genes must be present in the fish genome to encode the remaining molecular weight forms. This suggests that AFGPs are encoded by a multigene family. Both Northern (Fig. 5) and genomic Southern (data not shown) analyses support the presence of multiple genes. A very high gene dosage may be an important factor in providing the constant high levels of the AFGPs (35 mg/ml) in the serum of antarctic nototheniids to allow survival in their perennially freezing habitats.

The AFPs from the northern fishes are also encoded by multigene families (21–25), but the organization and structure of the AFP genes are different. First, the AFP genes in each family are closely linked, some irregularly spaced (23, 25), some tandemly arrayed in direct repeats (22) or in reverse repeats (24), but none of them encodes a polypeptide. There is a single copy of AFP per gene, and therefore the gene dosage (known AFP multigene families have 30–150 member genes) is likely to be lower than that of AFGPs in antarctic fishes (the AFGP 8 polypeptide gene alone is equivalent to 46 single-copy genes). If gene dosage is related to serum level of AFPs, which has been suggested to be the case of northern flounders (23) and eelpouts (25), it may be the reason why peak AFP levels (10–25 mg/ml) in northern fishes (40–42) are lower than AFGP levels (35 mg/ml) in antarctic fishes. Second, some AFP genes contain introns (24, 25), while the AFGP 8 gene contains no intron. And third, there have been no reports of potential Z-DNA-forming sequence in any AFP genes. It would be interesting to compare the genetic structure and organization of the antarctic AFGP genes with those of northern cods, which also synthesize AFGPs, and between the antarctic and northern eelpouts, which synthesize AFPs of substantial homology. The results may shed light on the phylogenetic relationships between the genes of each type of AF.

Lastly, the potential signal sequence of the AFGP 8 polypeptide gene is very peculiar. Primer-extension experiments established it to be a portion of the transcript. However, the three basic building blocks characteristic of a secretory signal peptide (43) are not obvious. Neither does it obey the (–3, –1) rule (44), which predicts the signal peptide cleavage site. In addition, the long segment of alternating valine and cysteine, and the alternating G and T in the DNA sequence of this segment, which can potentially form Z-DNA, have not been observed in known signal sequences.

This work was supported in part by National Science Foundation Grants DPP87-16296 to A.L.D. and DPP83-17724 and DPP86-14788 to H.W.D.

1. DeVries, A. L. (1988) *Comp. Biochem. Physiol.* **90B**, 611–621.
2. Davies, P. L. & Hew, C. L. (1990) *FASEB J.* **4**, 2460–2468.
3. O'Grady, S. M., Clarke, A. & DeVries, A. L. (1982) *J. Exp. Zool.* **220**, 179–189.
4. Ahlgren, J. A., Cheng, C. C., Schrag, J. D. & DeVries, A. L. (1988) *J. Exp. Biol.* **137**, 549–563.
5. DeVries, A. L. & Lin, Y. (1977) *Biochim. Biophys. Acta* **495**, 380–392.
6. Scott, G. K., Davies, P. L., Shears, M. A. & Fletcher, G. L. (1987) *Eur. J. Biochem.* **168**, 629–633.
7. Hew, C. L., Joshi, S., Wang, N. C., Kao, M. H. & Ananthanarayanan, V. S. (1985) *Eur. J. Biochem.* **151**, 167–172.
8. Ng, N., Trinh, K. Y. & Hew, C. L. (1986) *J. Biol. Chem.* **261**, 15690–15696.
9. Li, X. M., Trinh, K. Y., Hew, C. L., Buettner, B., Baenziger, J. & Davies, P. L. (1985) *J. Biol. Chem.* **260**, 12904–12909.
10. Schrag, J. D., Cheng, C. C., Panico, M., Morris, H. R. & DeVries, A. L. (1987) *Biochim. Biophys. Acta* **915**, 357–370.
11. Cheng, C. C. & DeVries, A. L. (1989) *Biochim. Biophys. Acta* **997**, 55–64.
12. DeVries, A. L. (1971) *Science* **172**, 1152–1155.
13. Raymond, J. A., Lin, Y. & DeVries, A. L. (1975) *J. Exp. Zool.* **193**, 125–130.
14. Osuga, D. T. & Feeney, R. E. (1978) *J. Biol. Chem.* **253**, 5338–5343.
15. Fletcher, G. L., Hew, C. L. & Joshi, S. (1982) *Can. J. Zool.* **60**, 348–355.
16. Komatsu, S. K., DeVries, A. L. & Feeney, R. E. (1970) *J. Biol. Chem.* **245**, 2909–2913.
17. Shier, W. T., Lin, Y. & DeVries, A. L. (1972) *Biochim. Biophys. Acta* **263**, 406–413.
18. Shier, W. T., Lin, Y. & DeVries, A. L. (1975) *FEBS Lett.* **54**, 135–138.
19. Lin, Y., Duman, J. G. & DeVries, A. L. (1972) *Biochem. Biophys. Res. Commun.* **46**, 87–92.
20. Morris, H. R., Thompson, M. R., Osuga, D. T., Ahmed, A. I., Chan, S. M., Vandenheede, J. R. & Feeney, R. E. (1978) *J. Biol. Chem.* **253**, 5155–5162.
21. Gourlie, B., Lin, Y., Price, J., DeVries, A. L., Powers, D. & Huang, R. C. (1984) *J. Biol. Chem.* **259**, 14960–14965.
22. Scott, G. K., Hew, C. L. & Davies, P. L. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2613–2617.
23. Scott, G. K., Davies, P. L., Kao, M. H. & Fletcher, G. L. (1988) *J. Mol. Evol.* **27**, 29–35.
24. Scott, G. K., Hayes, P. H., Fletcher, G. L. & Davies, P. L. (1988) *Mol. Cell Biol.* **8**, 3670–3675.
25. Hew, C. L., Wang, N.-C., Joshi, S., Fletcher, G. L., Scott, G. K., Hayes, P. H., Buettner, B. & Davies, P. L. (1988) *J. Biol. Chem.* **263**, 12049–12055.
26. Loenen, W. A. M. & Blattner, F. R. (1983) *Gene* **26**, 171–179.
27. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
28. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
29. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
30. Henikoff, S. (1987) *Methods Enzymol.* **155**, 156–165.
31. Sanger, F. S., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
32. Davis, L. G., Dibner, M. D. & Battey, J. F. (1986) *Basic Methods in Molecular Biology* (Elsevier, New York), pp. 143–146.
33. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
34. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
35. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1987) *Current Protocols in Molecular Biology* (Wiley, New York), Vol. 1.
36. Douglass, J., Civelli, O. & Herbert, E. (1984) *Annu. Rev. Biochem.* **53**, 665–715.
37. Hoffmann, W., Bach, T. C., Seliger, H. & Kreil, G. (1983) *EMBO J.* **2**, 111–114.
38. Kurjan, J. & Herskowitz, I. (1982) *Cell* **30**, 933–943.
39. Ahlgren, J. A. & DeVries, A. L. (1984) *Polar Biol.* **3**, 93–97.
40. Petzel, D. H., Reisman, H. & DeVries, A. L. (1980) *J. Exp. Zool.* **211**, 63–69.
41. Fletcher, G. L., Hew, C. L., Li, X. M., Haya, K. & Kao, M. H. (1984) *Can. J. Zool.* **63**, 488–493.
42. Fletcher, G. L., Haya, K., King, M. J., & Reisman, H. M. (1985) *Mar. Ecol. Prog. Ser.* **21**, 205–212.
43. von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21.
44. von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690.