Structure of an Antifreeze Polypeptide and Its Precursor from the Ocean Pout, Macrozoarces americanus*

Xiao-Mao Li, Khieth-Yen Trinh, and Choy L. Hew
From the Research Institute, Hospital for Sick Children and the Departments of Clinical Biochemistry and Biochemistry, University of Toronto, Toronto, Ontario M5G 1L5

Bert Buettner, John Baenziger, and Peter L. Davies
From the Department of Biochemistry, Queen’s University, Kingston, Ontario, K7L 3N6, Canada

Serum antifreeze polypeptides (AFP) from Newfoundland ocean pout have been resolved by ion exchange chromatography and reverse phase high performance liquid chromatography into at least 12 components. The protein sequences of three of the AFP were determined using a combination of protein Edman degradation and cDNA sequencing. The AFP precursor protein encodes for a preprotein of 87 amino acids with no obvious prosequences. Two of the AFP (SP1-A and SP1-C) were separate gene products with minor amino acid sequence differences. The protein structure of SP1-C precursor is MKSVIITGLLFVLL-

EXPERIMENTAL PROCEDURES

RESULTS

Microheterogeneity and Size of Ocean Pout AFP—From our earlier (7) and present investigations, it is clear that the ocean pout AFP can be fractionated into two separate groups based on their binding to ion exchange resins, namely, the QAE-, and SP- (or CM-) binding groups. The QAE-binding group eluted as a single peak from QAE-Sephadex (Fig. 1a), but the SP-binding group (Fig. 1b) was split into four peaks (SP1-4). On reverse phase HPLC, the AFP were resolved into 12 components (Fig. 2a). By running each of the peaks shown in Fig. 1 separately on the reverse phase column, it was possible to identify their constituents. Peak SP1 contained components HPLC 4, 5, and 6 (Fig. 2b). Peak SP2 contained HPLC 1, 2, 3, and 11, peak SP3 contained HPLC 8, 9, and 10, peak SP4 corresponded to HPLC 7, and the single QAE peak corresponded to HPLC 12 (not shown). Our earlier procedure uses QAE-, CM-, and SP-ion exchange chromatography and reverse phase HPLC to fractionate these ocean pout AFP (7). The present investigation has simplified the protocol by eliminating the need for CM-Bio-Gel chromatography.

Despite their chromatographic heterogeneity, the ocean pout AFP all have molecular weights in the range of 6,000-7,000 as determined by gel permeation HPLC in the presence of 0.1% trifluoroacetic acid and 45% acetonitrile (Fig. 3). This value, which is lower than that obtained by previous estimations.
tion on gel permeation HPLC in neutral buffers (M, 10,000–16,000), agrees with earlier molecular weight estimates based on electrophoresis in sodium dodecyl sulfate, analytical ultracentrifugation, and amino acid analysis (7).

Characterization of Three AFP Components: the SP-1 Group—To analyze the structure of ocean pout AFP, the SP-1 group (Fig. 1) was selected for sequencing. The three components in this group, corresponding to HPLC 4, 5, and 6 in Fig. 2a, are referred to here as SP1-A, SP1-B, and SP1-C, respectively (Fig. 2b). Not only do they co-elute on gel permeation and ion exchange chromatography but they have comparable thermal hysteresis activities and similar amino acid compositions (Table 1). Their structural homology was confirmed by tryptic peptide mapping (Fig. 4). SP1-B and SP1-C, which are almost identical, have the majority of tryptic peptides in common with SP1-A. A similar result was obtained from mapping chymotryptic peptides (not shown). The amino acid compositions of the major tryptic peptides from SP1-A, SP1-B, and SP1-C (Fig. 4) are listed in Table 2. Minor compositional differences occur in peptides 1, 1a, and 1b in each of the three components. In component SP1-A, there is an Ala to Thr substitution at position 21 of the mature AFP. In component SP1-B, there is an Ala to Thr substitution at position 21 of the mature AFP. In component SP1-C, there is an Ala to Thr substitution at position 21 of the mature AFP. However, from a composite of both SP1-A and SP1-C, it appears that the amino acid composition of the major tryptic peptides from SP1-A and SP1-C is almost identical. There are compositional differences in the C-terminal tryptic peptide of SP1-A with the peptides Thr-Tyr-Ala-Ala. Component SP1-B, there is an Ala to Thr substitution in the sequence to give peptide 1a (Thr-Tyr-Ala-Ala), and in component SP1-C the equivalent peptide 1b has an additional residue (Gly) at its C terminus. The presence of peptide 1a in the tryptic map of SP1-C (Fig. 4e) is due to contamination of the starting materials with SP1-B. The difference between peptide 8 in SP1-A and peptide 8a in SP1-B and SP1-C corresponds to the replacement of Ile by Leu in position 75 of the preprotein. N-terminal analyses indicated that the SP-1 group was blocked. Automatic Edman degradation reactions were done on chymotryptic peptides of SP1-A and tryptic peptides of SP1-B (Fig. 5). The large scale tryptic digest of SP1-B in Fig. 5b cannot be compared directly to the tryptic map in Fig. 4b because it was done with trypsin which was not treated with L-1-tosylamido-2-phenylethyl chloromethyl ketone. The amino acid compositions of the peptides used for automatic Edman degradation are presented in Table 3 and their sequences in Table 4. Subsequent successful treatment of SP1-C with pyroglutamate aminopeptidase indicates that pyroglutamic acid, which arises from the cyclization of terminal glutamine is the N-terminal residue. Once removed, the peptide is readily accessible to Edman degradation (Table 4).

cDNA Cloning and DNA Sequencing—Poly(A)+ RNA from the 9–10 S region of a sucrose density gradient was used for cDNA cloning as described under "Experimental Procedures" (Fig. 5). After sequencing by colony hybridization, three clones (36, 69, and 77) were selected for DNA sequence determination. Clone 69, which has an insert size of 0.3 kilobase, was sequenced by the chemical procedure of Maxam and Gilbert (Fig. 7). The two larger clones, 36 and 77, were sequenced using the M13-dideoxy chain terminator procedure (Fig. 8). The peptide sequences in Table 4 were used to establish the correct orientation and reading frame for the DNA sequences. The insert in clone 77 is 491 base pairs long. It codes for an 87-residue preprotein, the C-terminal portion of which matches in composition (Table 1) and sequence (Table 4) the SP-1-component of ocean pout AFP. The coding region is preceded by 57 base pairs of 5'-untranslated region which contains a 31-base pair stretch of AT residues and is followed by 170 base pairs of 3'-untranslated region. Clone 36 is very similar to clone 77. It is slightly truncated at both ends and contains a total of four single base changes. Only one of these changes occurs in the coding region. This change, at base 145, is a silent one. Neither sequence contains the polyadenylation signal AATAAA at the 3'-end, perhaps as a result of S1 nuclease digestion during cDNA cloning.

The insert in clone 69 is severely truncated at the 5'-end and lacks the DNA coding for the signal peptide plus the N-terminal portion of the mature AFP. However, from a composite of the remaining DNA sequence, the peptide sequences in Table 4, and amino acid compositional data (Tables 1 and 3), this clone can be matched to component 5 (T-A). Specifically, it contains the Leu to Ile replacement at position 75 and the Val to Ala replacement in the C-terminal tryptic peptide, but no other changes.

DISCUSSION

The primary structures of three of the 12 ocean pout AFP components have been derived here. These studies confirm our earlier suggestion that ocean pout AFP represent a new and distinct class of macromolecular antifreezes. Furthermore, computer search on a protein data bank does not indicate any sequence homology with other known proteins. In deducing the structure of these AFP proteins, we have provided some insight into the molecular basis of their microheterogeneity. First of all, both protein and DNA sequence analyses have established that SP1-A and SP1-C are products of separate genes. These two polypeptides differ in at least two positions in their amino acid sequences (positions 75 and 84). Second, SP1-B and SP1-C probably differ as a result of post-translational modification. All three SP1-components lack the C-terminal lysine which is encoded in the DNA but, in addition, both SP1-A and SP1-C, but not SP1-B, are missing the penultimate C-terminal residue, glycine. Peptide SP1-B is likely derived from SP1-C by post-translational processing. Thermal hysteresis measurements indicate that the loss of the C-terminal glycine has no obvious effect.
on antifreeze activity. Similar findings have been observed with AFP isolated from the sera of winter flounder (4). Last, although both clone 36 and clone 77 code for SP1-C, they are distinct sequences with at least four nucleotide differences. Thus, the same gene product might be derived from two or more genes. This suggestion is supported by preliminary results from genomic Southern blots of ocean pout DNA which indicate that the number of AFP genes is far in excess of 12, which is the number of separable protein components.

The ocean pout AFP precursor has a typical signal sequence at its N terminus, but the precise point of cleavage has not yet been established. According to the observations of Perlman and Halverson (24), processing is likely to occur after the alanine at position 21. However, amino acid analyses on the whole components (Table 1) and N-terminal peptides of the deblocked SP1-C suggest that Ser-22 is also missing. The figure in b denotes the junction between the pro segment and the mature AFP.

The isolation of cDNA clones for ocean pout AFP has given us access to their genes and the opportunity to study both the organization of this multigene family and its regulation.

Acknowledgments—We wish to thank the diving facility, Marine Sciences Research Laboratory, Memorial University of Newfoundland, St. John’s, Newfoundland, for the collection of experimental animals and S. B. Joshi for protein sequence determination. We thank Drs. V. S. Ananthanarayanan and Garth Fletcher for helpful discussions.

REFERENCES

1. Holmes, W. N., and Donaldson, E. M. (1979) in Fish Physiology (Hoar, W. S., Randall, D. S., eds) Vol. 1, pp. 1–83, Academic Press, New York
2. Hew, C. L., and Yip, C. (1976) Biochem. Biophys. Res. Commun. 71, 845–850
3. Duman, J. G., and DeVries, A. L. (1976) Comp. Biochem. Physiol. A Comp. Physiol. 73, 545–555
4. Fourney, R. M., Hew, C. L., Joshi, S. B., and Kao, M. S. (1984) Can. J. Zool. 62, 20–23
5. Hew, C. L., Fletcher, G. L., and Ananthanarayanan, V. S. (1980) Can. J. Biochem. 58, 377–383
6. Slaughter, D., Fletcher, G. L., Ananthanarayanan, V. S., and Hew, C. L. (1981) J. Biol. Chem. 256, 2022–2026
7. Hew, C. L., Slaughter, D., Joshi, S. B., Fletcher, G. L., and Ananthanarayanan, V. S. (1984) J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 155, 81–88
8. Scholander, P. F., VanDane, L., Kanwisher, J. W., Hammel, H. T., and Gordon, M. S. (1957) J. Cell Comp. Physiol. 49, 5–24
9. Feehery, R. E., and Yeh, Y. (1978) Adv. Protein Chem. 32, 191–292
10. DeVries, A. L. (1982) Comp. Biochem. Physiol. 73, 627–640
11. Ananthanarayanan, V. S., and Hew, C. L. (1977) Biochem. Physiol. Res. Commun. 74, 685–690
12. Ananthanarayanan, V. S., Slaughter, D., and Hew, C. L. (1985) Biochim. Biophys. Acta, in press
13. DeVries, A. L., and Lin, Y. (1977) Biochim. Biophys. Acta 495, 388–392
14. Davies, P. L., Roach, A. H., and Hew, C. L. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 335–339
15. Pickett, M. H., Scott, G., Davies, P., Wang, N., Joshi, S., and Hew, C. L. (1984) Eur. J. Biochem. 136, 35–38
16. Hew, C. L., Joshi, S., Wang, N. C., Kao, M. H., and Ananthanarayanan, V. S. (1985) Eur. J. Biochem., in press
17. Swergold, G. D., and Rubin, C. S. (1983) Anal. Biochem. 131, 295–300
18. Podell, D. N., and Abraham, G. N. (1978) Biochem. Biophys. Res. Commun. 81, 176–185
19. Davies, P. L., and Hew, C. L. (1980) J. Biol. Chem. 255, 8729–8734
20. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499–560
21. Sanger, F., and Coulson, A. R. (1975) J. Mol. Biol. 94, 441–448
22. Messing, J. M., Groenbom, B., Muller-Hill, B., and Hofschnieder, P. H. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 3642–3646
23. Chou, P. Y., and Fasman, G. D. (1974) Biochemistry 13, 222–245
24. Perlman, D., and Halvorson, H. O. (1983) J. Mol. Biol. 167, 391–409

\* G. Scott, C.-L. Hew, and P. L. Davies, unpublished results.

\* N. Ng, and C. L. Hew, unpublished results.
Structure of an Antifreeze Polypeptide

12907

By: Lin-Yun Li, Chien-Yon Trisch, Chong Lee, and Peter L. Davies

**MATERIALS**

- Cysteic acid, chromatography, thermolysin and L-1-Phenylmethyl-3-thiopropionic acid-2-carboxylic acid (TPP) were purchased from Sigma. Sephadex G-100, Sephadex G-100, and phosphatase was from Pharmacia, the Tissue Extract of the Thermotoga maritima N4 was obtained from the Marine Biological Laboratory. Coomassie Brilliant Blue R-250 was purchased from Sigma.

**METHODS**

- The polypeptide was isolated from the Thermotoga maritima N4 by ion exchange chromatography on a Sephadex G-100 column. The fractions were then subjected to high-performance liquid chromatography (HPLC) on a Waters 600 high-performance liquid chromatography pump. The elution profile was monitored at 215 nm using a Waters 490 variable wavelength detector. The fractions were then subjected to SDS-PAGE, the intact polypeptide was visualized by Coomassie Brilliant Blue R-250 staining. The purified polypeptide was then subjected to amino acid analysis and amino acid sequencing to confirm the purity.

**RESULTS**

- The polypeptide was found to contain 12% cysteic acid and 2% phenylpropionic acid. The molecular weight of the polypeptide was determined to be approximately 4 kDa. The amino acid sequence of the polypeptide was determined to be: Cys-Cys-Asp-Leu-Asp-Leu-Cys-Cys.

**DISCUSSION**

- The antifreeze activity of the polypeptide was tested by freezing intensity assay. The polypeptide was able to reduce the freezing point of water by 1.2°C. This activity was stable over a wide range of pH and temperature.

**CONCLUSIONS**

- The antifreeze polypeptide isolated from Thermotoga maritima N4 is a novel antifreeze protein with potential applications in ice nucleation inhibition and cryopreservation.
Structure of an Antifreeze Polypeptide

Fig. 2d: Isolation of chromotryptic peptides of SPF-1 on reverse phase HPLC. Sample was loaded on a Vydac adsorbed C18 column (7.8 mm i.d. × 6 cm) in 0.05N trifluoroacetic acid solvent A gradient with a flowrate of 1 ml/min.

Fig. 2d: Isolation of tryptic peptides of SPF-1 on reverse phase HPLC. Sample was loaded on an Alumina Spherisil 800 column in 0.02 M triethylammonium acetate-acetonitrile gradient. Flowrate was 0.5 ml/min.

Fig. 3. Partial sequence of an ocean quahog ACP enzyme derived from clone 69. The partial amino acid sequence of the clone corresponds to the C-terminal portion of SPF-1.

Table 1: Amino Acid Analyses of the Ocean Quahog ACP Components

| Amino Acid | SPF-1-A | SPF-1-B | SPF-1-C |
|------------|---------|---------|---------|
| Asp        | 4.8(1)  | 4.9(1)  | 4.8(1)  |
| Thr        | 5.8(1)  | 5.6(1)  | 5.8(1)  |
| Ser        | 5.8(2)  | 5.8(2)  | 5.8(2)  |
| Glu        | 5.8(7)  | 5.8(7)  | 5.8(7)  |
| Gly        | 5.8(5)  | 5.8(5)  | 5.8(5)  |
| Val        | 5.8(1)  | 5.8(1)  | 5.8(1)  |
| Leu        | 5.8(1)  | 5.8(1)  | 5.8(1)  |
| Ile        | 5.8(1)  | 5.8(1)  | 5.8(1)  |
| Pro        | 5.8(1)  | 5.8(1)  | 5.8(1)  |
| Ala        | 5.8(1)  | 5.8(1)  | 5.8(1)  |
| Lys        | 5.8(1)  | 5.8(1)  | 5.8(1)  |

Table 2: Amino Acid Composition of Major Tryptic Peptide

| Amino Acid | SPF-1-A | SPF-1-B | SPF-1-C |
|------------|---------|---------|---------|
| Asp        | 4.85(1) | 4.91(1) | 4.81(1) |
| Thr        | 5.91(1) | 5.85(1) | 5.85(1) |
| Ser        | 5.86(1) | 5.86(1) | 5.86(1) |
| Glu        | 5.80(1) | 5.80(1) | 5.80(1) |
| Gly        | 5.80(1) | 5.80(1) | 5.80(1) |
| Val        | 5.80(1) | 5.80(1) | 5.80(1) |
| Leu        | 5.80(1) | 5.80(1) | 5.80(1) |
| Ile        | 5.80(1) | 5.80(1) | 5.80(1) |
| Pro        | 5.80(1) | 5.80(1) | 5.80(1) |
| Ala        | 5.80(1) | 5.80(1) | 5.80(1) |
| Lys        | 5.80(1) | 5.80(1) | 5.80(1) |

Table 3: Comparison of Proteins and ACP Sequences

| Composition from proteins and ACP sequences | SPF-1-A | SPF-1-B | SPF-1-C |
|--------------------------------------------|---------|---------|---------|
| ACP                                          | 4.85(1) | 4.91(1) | 4.81(1) |
| Thr                                          | 5.91(1) | 5.85(1) | 5.85(1) |
| Ser                                          | 5.86(1) | 5.86(1) | 5.86(1) |
| Glu                                          | 5.80(1) | 5.80(1) | 5.80(1) |
| Gly                                          | 5.80(1) | 5.80(1) | 5.80(1) |
| Val                                          | 5.80(1) | 5.80(1) | 5.80(1) |
| Leu                                          | 5.80(1) | 5.80(1) | 5.80(1) |
| Ile                                          | 5.80(1) | 5.80(1) | 5.80(1) |
| Pro                                          | 5.80(1) | 5.80(1) | 5.80(1) |
| Ala                                          | 5.80(1) | 5.80(1) | 5.80(1) |
| Lys                                          | 5.80(1) | 5.80(1) | 5.80(1) |

Yields are in mg. The residues in brackets indicate the calculated numbers of amino acid residues.
Structure of an Antifreeze Polypeptide

Table 3. Amino Acid Compositions of some Dynamic Peptides from SP1-A and SP1-B used for Automatic Protein Edman Degradation Studies

|     | AC-3 | AC-8 | AC-5 | AC-6 | AC-9 | AC-13 | BT-3 | BT-6 | BT-8 | BT-10 |
|-----|------|------|------|------|------|-------|------|------|------|-------|
| Arg | 1.15(1) | 0.86(1) | 1.06(1) | 6.36(1) | 1.20(1) | 27.53(1) | 5.68(1) | 9.34(1) | 9.76(1) | 3.76(1) |
| Thr | 10.46(1) | 3.85(1) | 4.16(1) | 2.16(1) | 29.38(1) | 6.86(1) | 9.36(1) | 7.09(1) | 7.96(1) |         |
| Ser | 0.85(1) | 0.97(1) | 0.78(1) | 0.78(1) | 0.85(1) | 1.49(1) |         |         |         |         |
| Glx | 22.78(2) | 0.95(1) | 26.12(1) | 19.59(1) | 0.15(1) | 9.45(1) | 6.76(1) | 14.00(1) | 5.59(1) |         |
| Pro | 12.92(1) | 3.12(1) | 5.26(1) | 6.72(1) | 61.27(2) | 7.93(1) | 3.00(1) | 15.44(2) | 12.02(2) |         |
| Gly | 9.19(3) | 2.56(2) | 10.72(1) | 9.30(2) | 0.20(1) | 20.69(2) | 2.21(1) | 5.08(1) | 4.98(1) |         |
| Ala | 11.23(1) | 0.49(1) | 1.56(1) | 0.47(1) | 0.25(1) | 2.07(1) | 0.20(1) | 1.25(1) | 0.21(1) |         |
| Val | 21.76(2) | 9.26(1) | 8.47(1) | 11.84(1) | 1.00(1) | 25.98(1) | 13.52(2) | 5.27(1) | 7.79(1) | 20.52(2) |
| Met | 1.30(1) | 4.19(1) | 3.29(1) | 1.00(1) | 8.16(1) | 0.20(1) |         |         |         |         |
| Lip | 9.07(1) | 3.50(1) | 1.00(1) | 96.19(1) | 9.08(1) | 19.74(2) | 24.23(1) |         |         |         |
| Leu | 2.97(1) | 0.40(1) | 16.07(1) | 8.93(1) | 0.25(1) |         |         |         |         |         |
| Phe | 1.27(1) | 3.78(1) | 9.07(1) | 9.30(2) | 1.18(1) | 20.56(1) | 8.20(1) | 7.20(1) | 5.20(1) |         |
| No of amino acids: | (18) | (16) | (9) | (14) | (10) | (12) | (8) | (9) | (18) |         |

Table 4: Automatic Degradation of some Dynamic Peptides from Goose Point AFP

The numbers in brackets correspond to the portions of the amino acid sequence based on the peptide’s composition.

SP1-A

AC-3: Gin Val Thr Pro Val Ala Lys Gly Gin

AC-6: 79 63

AC-9: 57

AC-12: 57

SP1-B

BT-2: 64 71

BT-5: 55 63

BT-8: 46 48

BT-10: 46

Deblinded SP1-A

38 29

Sar Val Val Ala Thr Gin Lys Ile Pro Ile Asn