ELSEVIER

Biochimica et Biophysica Acta 1247 (1995) 163–172

BBA Biochimica et Biophysica Acta

Antifreeze peptide heterogeneity in an antarctic eel pout includes an unusually large major variant comprised of two 7 kDa type III AFPs linked in tandem ☆

Xin Wang, Arthur L. DeVries, Chi-Hing C. Cheng *

Department of Physiology, University of Illinois, 524 Burrill Hall – 407 S. Goodwin, Urbana, IL 61801, USA

Received 26 April 1994; revised 23 August 1994; accepted 4 November 1994

Abstract

The structural heterogeneity of the major antifreeze peptides (AFPs) from the antarctic eel pout, *Lycodichthys dearborni* (formerly classified as *Rhigophila dearborni*) was characterized. Three major AFPs designated as RD1, RD2 and RD3, and five minor ones were isolated from the fish plasma. RD1 and RD2 are both 64 residues in length, about 7 kDa, and thus similar in size to all characterized type III AFPs, while RD3 is twice as large, about 14 kDa, and represents the first example of a disparately large size variant within the same fish for the three known types of antifreeze peptides. RD3 was found to be 134 residues in length, arranged as a 64-residue N-terminal half and a 61-residue C-terminal half of similar sequence to each other and to the 7 kDa type III AFPs, linked by a 9-residue connector of unmatched sequence. RD3 has slightly lower antifreeze activity than its 7 kDa counterparts, with a melting-freezing point difference of about 0.81°C at 10 mg/ml versus 0.95°C and 0.90°C for RD1 and RD2, respectively. RD1 and RD2 are 94% identical in sequence to each other. They are 98% and 94%, respectively identical to N-terminal half of RD3, and 85% and 77%, respectively, identical to C-terminal half of RD3. By sequence comparison, a previously characterized AFP from this fish [1] was identified to be RD2.

Keywords: Antarctic eel pout; Type III antifreeze peptide; Peptide heterogeneity; Large size variant; Amino acid sequence; Structure–function relationship; (*L. dearborni*)

1. Introduction

Many polar and cold-water marine fishes synthesize protein antifreeze molecules which circulate in their blood and body fluids to avoid freezing in their ice-laden habitats [2–4]. Antifreeze proteins act through an adsorption-inhibition mechanism: they adsorb to ice crystals that inadvertently enter the fish and inhibit ice growth to a temperature several tenths below that of the freezing point of seawater, enabling the fish body fluids to remain in the liquid state [3,5,6]. There are four known types of antifreeze proteins — one antifreeze glycopeptide (AFGP) and three different antifreeze peptides (AFP). AFGPs are found in the Antarctic cods (notothenioids) and a number of unrelated northern true cods. Within each fish, they exist as a family of discretely sized polymers composed of the glycotripeptide unit, (Ala-Ala-Thr-)$_n$, with the disaccharide galactose *N*-acetylgalactosamine attached to threonines, and $n = 4$ to 55, or a molecular mass range of 2600–34 000 Da [2,4,7]. The three types of antifreeze peptides (AFPs) differ from each other in composition and structure, and in contrast to the AFGPs, the AFPs within each type are similar in size. The alanine-rich, α-helical type I AFPs synthesized by several flounders, sculpins and Alaskan plaice are about 4000–4500 Da [4,8–11]. The cysteine-rich, β-structured type II AFP from sea raven [12] is about 14 000 Da, and about 16 500 and 17 500 Da, respectively, for those from herring and smelt [13]. Type III AFPs from Atlantic ocean pout and wolffish [14,15], and Arctic and Antarctic eel pouts [1,16] have unbiased amino-acid composition, and are all 63–66 residues in length or a mass of about 6500–7000 Da.

☆ The sequence data in this paper have been submitted to the SWISS-PROT data base under the accession numbers P35751, P35752 and P35753.

* Corresponding author. E-mail: cdevries@pop.life.uiuc.edu. Fax: +1 (217) 3331133.

SSDI 0167-4838(94)00205-3

In our previous characterizations of type III AFPs from Antarctic eel pouts, the single AFP from *Rhigophila dearborni* (reclassified as *Lycodicthys dearborni*) that was characterized was purified from the major fraction of five AFP-containing fractions obtained from anion-exchange column chromatography of fish plasma; the other four were not analyzed [1]. To include all the AFPs in the characterization of AFP structural heterogeneity in this fish, in this study, we first separated the AFPs from the large plasma proteins by size-fractionation column chromatography, followed by RP-HPLC (reverse-phase high-pressure liquid chromatography) to resolve the individual AFPs. Three major AFPs and at least five minor ones were identified, representing the full complement of AFPs in this fish. Two of the major AFPs are about 7 kDa in size based on gel mobilities, similar to all characterized type III AFPs, but the third one is twice as large, about 14 kDa. Such a disparately large AFP variant has not been observed within a single fish species for any of the three types of AFP. To ascertain the structure of this unusually large variant in order to gain an understanding of what functional implications there may be with respect to the large size, and also of the structure–function of type III AFPs in general, we have determined the primary structure of all the three major AFPs in this fish.

2. Materials and methods

2.1. Specimens and blood collection

Specimens of *L. dearborni* were caught from the bottom of McMurdo Sound, Antarctica, at depths of 500–700 meters with baited fish traps during austral summers (October–January). Blood was collected from the caudal blood vessels of anaesthetized fish, pooled and centrifuged. The plasma was removed and stored at −70°C until analyzed.

2.2. Peptide purification

Antifreeze proteins were purified from fish plasma by Sephadex G75 gel filtration column chromatography followed by RP-HPLC as previously described [2]. Briefly, 10–12 ml of plasma was applied to a Sephadex G75 column (2.7 cm i.d. × 150 cm) per run and the proteins were eluted with 50 mM NH_4HCO_3. Protein elution was monitored by absorbance at 230 nM, and fractions within each protein absorption peak were pooled and lyophilized to remove the volatile buffer. The lyophilized protein was assayed for antifreeze activity using the melting–freezing point difference method of DeVries [17]. Antifreeze activity was found in two separate peaks, peaks II and III (see Section 3). About 0.1 to 0.5 mg of the protein from each of the two peaks was analyzed by RP-HPLC with an analytical Ultrasphere (Altex) C18 column (4.6 × 250 mm), using a linear gradient of 0–90% acetonitrile (ACN) with 0.1 M triethylamine phosphate (TEAP) (pH 2.2) over 60 min. Protein were found to elute between 30–60% ACN. Larger quantities of proteins, about 5–6 mg per run, were then applied to a preparative Ultrasphere C18 column (10 × 250 mm) and resolved with a gradient of 30–60% ACN/0.1 M TEAP over 60 min. The corresponding protein peaks from separate runs were pooled and dialyzed in Spectrapor 3 (Spectrum Medical Industries) against distilled water. The dialyzed protein was then lyophilized and weighed, and its antifreeze activity reconfirmed.

2.3. Polyacrylamide gel electrophoresis

RP-HPLC purified antifreeze peptides, about 30 μg each, were electrophoresed on a 15–20% SDS-polyacrylamide gradient gel and stained with a quick stain protocol with Coomassie brilliant blue G-250 (Bio-Rad). Approximate molecular mass was estimated from gel mobility of the AFPs.

2.4. Proteinase digestion

Trypsin (Sigma) digestion was done in 50 mM ammonium bicarbonate (pH 8.0) with an enzyme:substrate ratio of 1:50 (W:W) at 37°C for 4 h. *Staphylococcus aureus* (V8) proteinase (Pierce) digestion was done in 50 mM ammonium acetate (pH 8.0) with an enzyme:substrate ratio of 1:30 at 37°C for 24 h. Tryptic and V8 proteinase fragments were resolved on RP-HPLC using a linear gradient of 0–100% and 0–80% ACN/0.1% TFA (trifluoroacetic acid), respectively.

2.5. Molecular mass determination by mass spectrometry

Accurate molecular masses of native antifreeze peptides and peptide fragments were determined by fast atom bombardment mass spectrometry (FAB/MS) on a ZAB-SE spectrometer (VG Analytical, UK), or by electrospray ionization mass spectrometry (ESI/MS) on a Quattro spectrometer (VG Analytical, UK) at the Mass Spectrometry Laboratory, School of Chemical Sciences, University of Illinois.

2.6. Amino-acid analysis and peptide sequencing

Amino-acid analysis was carried out by pre-column PITC (phenylisothiocyanate) derivatization of acid hydrolysate of peptides followed by RP-HPLC resolution of PTC (phenylthiocarbamyl) amino acids, using the Waters PicoTag system. Automated N-terminal sequencing were carried out with the Applied Biosystem 470 A gas phase sequencer or a 477 A pulsed-liquid phase sequencer. Amino-acid analysis and automated sequencing were performed at the University of Illinois Biotechnology Center.

3. Results

3.1. Purification

Fig. 1 shows the separation of the antifreeze proteins from the other plasma proteins using a G75 Sephadex sizing column. Peaks II and III were found to contain antifreeze activity, and their yields were about 12–14 mg and 24–29 mg, respectively, per 10 ml plasma. Reverse-phase HPLC of 0.25 mg of peak II and 0.5 mg of peak III proteins further resolved them into three major AFP variants designated as RD1, RD2, and RD3 (Fig. 2A). RD3 was from G75 peak II (Fig. 2B), and RD1 and RD2 from G75 peak III. There are two minor AFPs that eluted right before RD1, and one after RD2; the small single peak that eluted at 31 min. (or about 40% ACN) did not have antifreeze activity (Fig. 2A). Fig. 2B is the elution profile of a large amount (5 mg) of peak II material on preparative RP-HPLC. Besides a major peak corresponding to RD3 in Fig. 2A, at least two other variants indicated by the incompletely resolved peak on both the ascending and descending limbs were revealed (Fig. 2B). The material from the ascending, central and descending portion of this composite RD3 peak was collected separately, and the protein from the central portion which represents the major RD3 AFP was used for the characterizations in this study.

The AFP RD2 co-eluted with the AFP designated as RD previously characterized by Schrag et al. [1] when the two were co-chromatographed (unpublished results), and automated sequencing of RD2 confirmed that the two are indeed the same AFP (see Section 3.6). At 20 mg/ml (approximates physiological concentration of AFPs in blood) in 0.1% TFA, the antifreeze activity, i.e. melting–freezing (mf–fp) point difference for RD1, RD2, and RD3 is 1.23°C, 1.19°C, and 0.94°C, respectively. At this concentration, RD3 was close to but not completely soluble which might have contributed to its lower activity. Activity

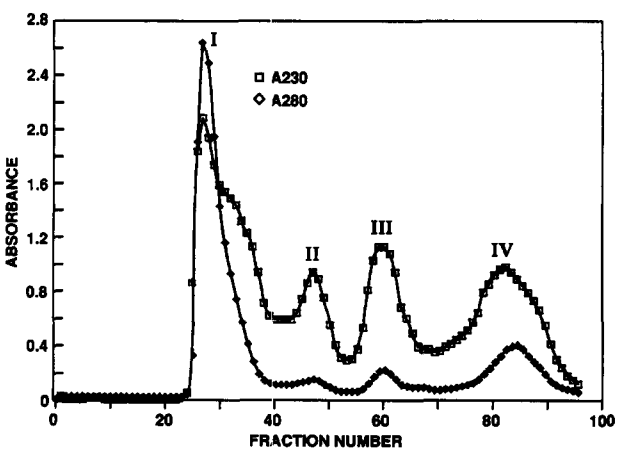

Fig. 1. Elution profile of proteins from 10 ml of plasma from the antarctic eel pout, *L. dearborni* fractionated on a Sephadex G-75 gel filtration column (2.7 cm i.d. × 150 cm). Antifreeze activity was found in peak II and peak III material.

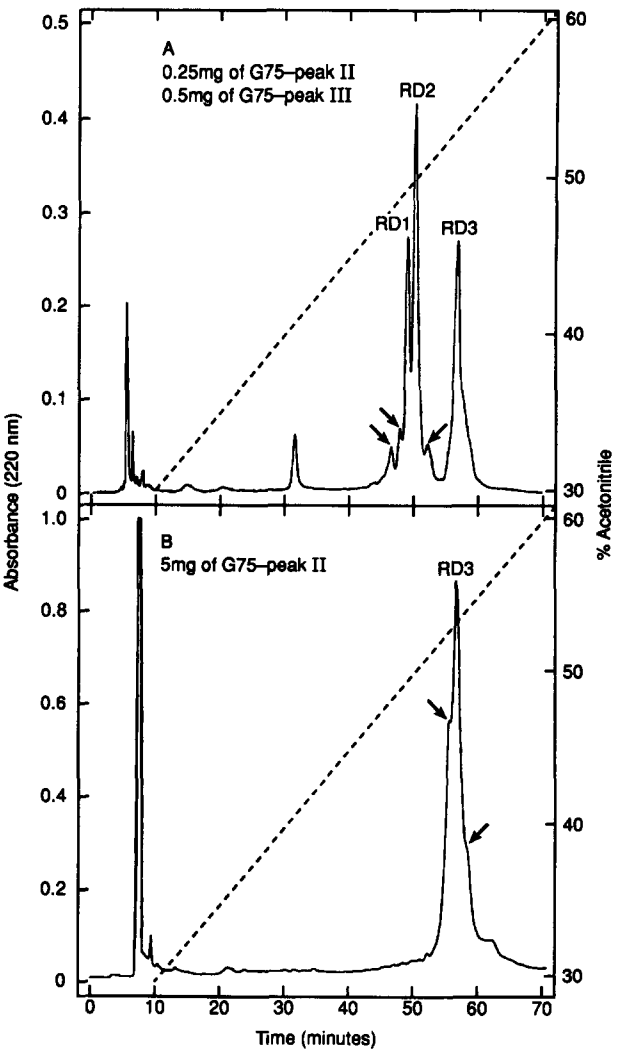

Fig. 2. RP-HPLC profile of antifreeze proteins in peaks II and III from the Sephadex G-75 column. (A) shows all the major AFPs-RD1, RD2, and RD3, and 3 minor variants of RD1 and RD2 indicated by arrow heads. (B) shows a large amount of peak II proteins resolved on preparative RP-HPLC. The composite peak contains 3 components — the major RD3 in the central portion, and one variant each on the ascending and descending slope indicated by arrow head.

was assayed again at 10 mg/ml at which RD3 was completely soluble. The activity at 10 mg/ml is 0.95°C for RD1, 0.90°C for RD2, and 0.81°C for RD3. There was some variability in the activity when the measurement was repeated for the same sample, but RD3 consistently has slightly lower activity. The buffer, 0.1% TFA was used to facilitate dissolving the peptides, but makes no contribution to the mp–fp difference.

3.2. SDS-polyacrylamide gel electrophoresis

Fig. 3 shows the approximate size of the HPLC-purified AFPs on SDS-PAGE. RD1 and RD2 are both about 7 kDa while RD3 is twice as big, about 14 kDa. All three AFPs were apparently purified to homogeneity as each appeared as a single band.

3.3. Molecular mass and sequencing of tryptic and V8 proteinase fragments of RD3

To elucidate the structure of the large AFP, i.e. the 14 kDa RD3, overlapping peptide fragments were generated by trypsin and *Staphylococcus aureus* (V8) proteinase digestions for sequence and mass analyses.

Trypsin digestions of RD3 consistently yielded 10 distinct fragments, T1–T10, which resolved completely on RP-HPLC except for T6 and T7 which co-eluted in a single peak with a 'shoulder' on the descending slope and were thus collected as one sample (Fig. 4A). V8 proteinase digestions on the other hand yielded a less consistent HPLC profile, probably due to resistance of RD3 to digestion by the enzyme. Several of the V8 proteinase digestions yielded 10 peaks on RP-HPLC, V1-V10 (Fig. 4B). Peaks V2 and V3 were collected as one sample because of the closeness of the their elution times. Subsequent sequence analysis also showed carry-over of peak V4 material into its close neighbor peak V5.

A single molecular ion by FAB/MS was obtained for each of the 10 tryptic fragments as shown in their respective spectrum (Fig. 5). T6 and T7 which were unresolvable on RP-HPLC appeared as two separate molecular ions within the same spectrum. Single molecular ions were obtained for V8 proteinase fragments V1, V2, V3, V5, V6, V8 and V9. Fragments V2 and V3 applied as a single sample produced 2 separate molecular ions, at mass number 704.4 (V3) and 1042.5 (V2), in the same spectrum

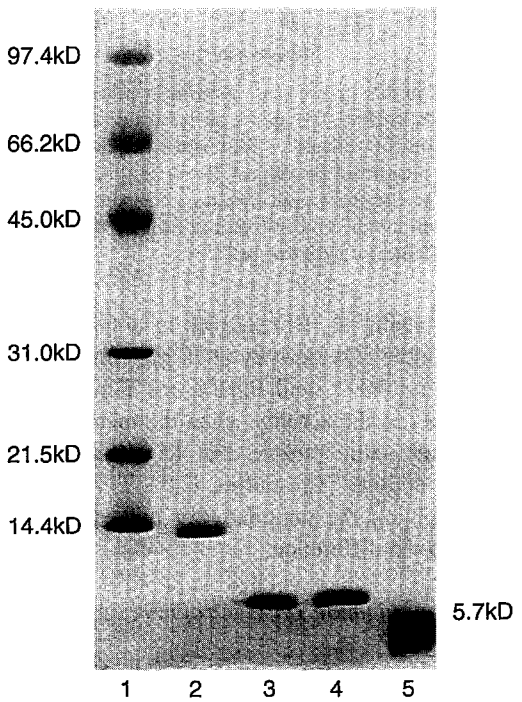

Fig. 3. SDS-polyacrylamide gel electrophoresis of HPLC-purified AFPs on a 15–20% gradient gel. Lane 1, Bio-Rad low molecule weight standards; lane 2, RD3; lane 3, RD1; lane 4, RD2; and lane 5, insulin standard.

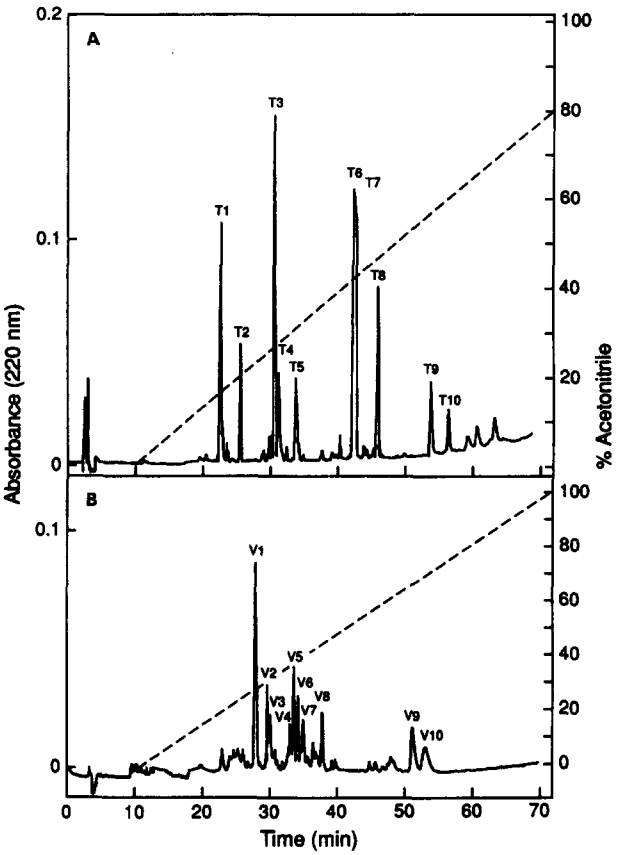

Fig. 4. RP-HPLC separation of peptide fragments of RD3, after digestion with trypsin and *Staphylococcus aureus* (V8) proteinase. (A) Elution profile of tryptic fragments, T1–T10 on a linear gradient of 0–80% acetonitrile /0.1% TFA. (B) Elution profile of V8 proteinase fragments, V1–V10 on a 0–100% linear gradient.

(Fig. 6). A third molecular ion appeared at 1027.6, but did not agree with the mass calculated from the sequence of any of the V8 proteinase fragments when the sequences were subsequently obtained. Its possible origin is discussed in the Section 4. FAB/MS of V4 did not produce a recognizable molecular ion probably because of insufficient material. FAB/MS mass spectrum was not obtained for V10 because there was only sufficient material for peptide sequencing.

Automated sequencing produced complete sequences for T2–T8 and T10, and partial sequence for T9. T6 and T7 which were applied as a single sample produced two distinct peptide sequences, distinguishable by the PTH-amino acid yields in the successive cycles of Edman degradation. The residues of T1 were determined by amino-acid composition which showed 3 residues, Asx, Glx and Tyr in a 1:1:1 ratio (Table 1).

Complete sequences for V2, V3, V5, V6, V8, and V9, and partial sequence for V10 were obtained. The combined sample of V2 and V3 produced two distinct sequences distinguishable by PTH-amino-acid yields. Sequencing of V5 produced one major sequence of 11 residues, and two similar minor ones of 6 residues each. These two minor

sequences were most likely from two co-eluted fragments in peak V4 carried over into peak V5 due to the close proximity of their elution times (Fig. 4B). They were thus designated as V4* − 1 and V4* − 2. The residues in V1 were determined by amino-acid composition which showed 6 residues, Asx, Glx, Tyr, Met, Val, Lys in a ratio of 1 to each other (Table 1). Fragment V7 appeared to be partially digested material based on amino-acid analysis and thus was not further characterized.

3.4. Construction of the complete RD3 sequence

The sequences of native RD3 and peptide fragments as well as their alignments are shown in Table 2. N-terminal sequencing of native RD3 (N) produced 64 residues which were unambiguously assigned except residues number 54 and 55 due to sequencer problems during those two cycles. Alignment of the sequences of the peptide fragments V10, T10, T8, V5, V4* − 1, V8, T6 with the first 64 residues of the native peptide not only confirmed the sequence, but allowed assignment of residues 54 and 55 as Thr and Leu. V4* − 1 was one of the two minor sequences that appeared in the sequencing of V5 due to carry-over of peak V4 material into peak V5. The position of V4* − 1 sequence, residues 37–42, was corroborated by the overlapping sequence of T8 and of the native peptide. The 6 residues of V1 appeared in positions 59–64, and their order, -Met-Val-Lys-Asn-Tyr-Glu-, was established according to the partial native RD3 sequence and the overlapping upstream T6 and downstream T3 sequences.

The first 3 of the 12-residue T3 fragment overlapped residues 62–64 of the native peptide sequence, extending it to residue 73. The overlap of V9 with T3, T9 and T2,

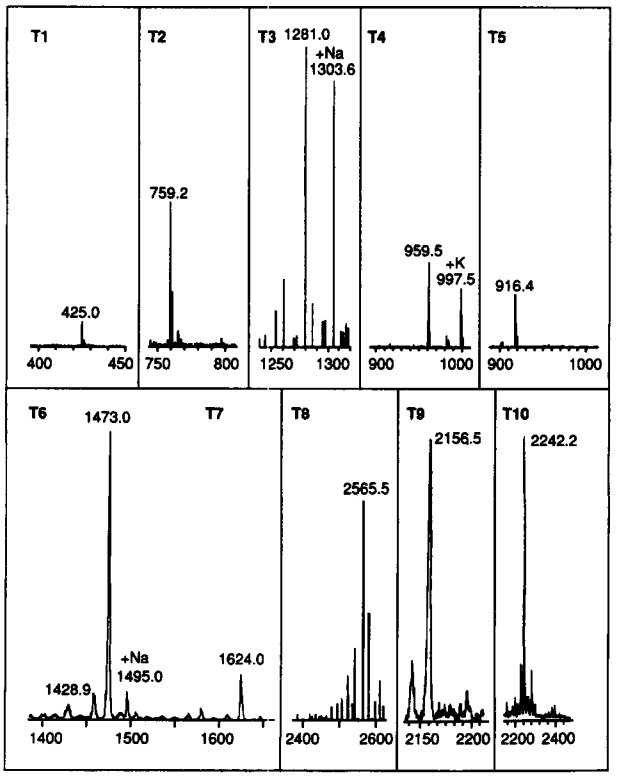

Fig. 5. FAB/MS mass spectra showing the molecular ions ($M \cdot H^+$) of the tryptic fragments T1–T10 of RD3. The mass numbers are: T1, 425.0; T2, 759.2; T3, 1281.0; T4, 959.5; T5, 916.4; T6, 1473.0; T7, 1624.0; T8, 2565.5; T9, 2156.5, and T10, 2242.2. The mass of the fragment in Da equals the mass number of $M \cdot H^+$ minus 1, the mass of the proton.

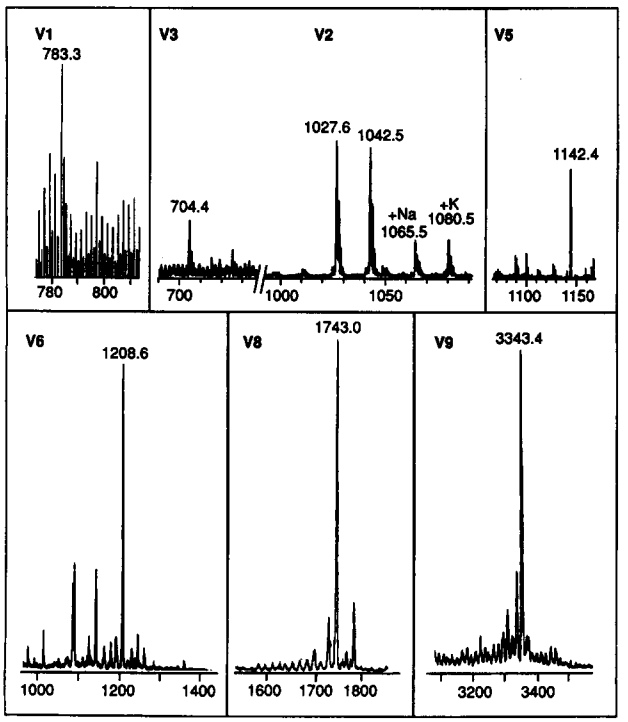

Fig. 6. FAB/MS mass spectra showing the molecular ions ($M \cdot H^+$) of V8 proteinase fragments, V1–V3, V5, V6, V8 and V9. The mass numbers are: V1, 783.3; V2, 1042.5; V3, 704.4; V5, 1142.4; V6, 1208.6; V8, 1743.0; V9, 3343.4. The mass of the fragment in Da equals the mass number of $M \cdot H^+$ minus 1, the mass of the proton.

Table 1
Amino-acid composition analyses of peptide fragments T1 and V1

Amino acid	T1 (mol%)	No. [a]	V1 (mol%)	No. [a]
Asx	33.9	1.4	19.1	1.2
Glx	31.6	1.3	16.7	1.0
Ser	0.6	–	0.0	–
Gly	1.8	–	1.5	–
Arg	2.3	–	0.0	–
Thr	0.0	–	0.0	–
Ala	0.0	–	1.8	–
Pro	2.3	–	0.8	–
Tyr	23.3	1	16.0	1
Val	3.0	–	13.1	0.8
Met	0.0	–	10.6	0.7
Cys	0.0	–	0.0	–
Ile	0.2	–	0.6	–
Leu	0.3	–	0.7	–
Phe	0.0	–	0.0	–
Lys	0.5	–	17.4	1.1

[a] Number of residues estimated by normalizing to mol% of Tyr.

Table 2
Amino-acid sequences and alignment of peptide fragments of RD3

Peptide or Fragment	M.H^{+a} mass	Calculated^{b} mass	Residue No.	Sequence^{c}
				1 10 20 30 40 50 60 70 80 90 100 110 120 130
N	14430.0	-	1- 64	(N) NKASVVANQLIPINTALTLIMMKAEVVTPMGIPAEEIPNLVGMQVNRAVPLGT..MPDMVKNYE
V10	-	2684.3	1- 25	(V10) NKASVVANQL.....
T10	2242.2	2241.8	3- 23	(T10) -ASVVANQLIPINTALTLIMMK-
T8	2565.5	2564.0	24- 47	(T8) -AEVVTPMGIPAEEIPNLVGMQVNR-
V5	1142.4	1142.3	26- 36	(V5) -VVTPMGIPAEE-
V4*-1	-	611.4	37- 42	(V4*-1) -IPNLVG-
V8	1743.0	1741.9	43- 58	(V8) -MQVNRAVPLGTTLMPD-
T6	1473.0	1472.8	48- 61	(T6) -AVPLGTTLMPDMVK-
V1	783.3	782.9	59- 64	(V1) -MVKNYE-
T3	1281.8	1281.3	62- 73	(T3) -NYEDGTTSPGLK-
V9	3343.4	3342.9	65- 96	(V9) -DGTTSPGLKSVVANQLIPINTALTLVMMKAEE-
T9	2156.5	2156.7	74- 93	(T9) -SVVANQLI.....
T2	759.2	758.8	94-100	(T2) -AEEVSPK-
V2	1042.4	1041.2	97-106	(V2) -VSPKGIPSEE-
T4	959.5	959.1	101-109	(T4) -GIPSEEISK-
V4*-2	-	615.3	106-112	(V4*-2) -ISKLVG-
T5	916.4	916.1	110-117	(T5) -LVGMQVNR-
V6	1208.6	1207.5	113-122	(V6) -MQVNRAVYLD-
T7	1624.0	1624.0	118-131	(T7) -AVYLDQTLMPDMVK-
V3	704.4	703.3	123-128	(V3) -QTLMPD-
V1	783.0	782.9	129-134	(V1) -MVKNYE
T1	425.0	424.4	132-134	(T1) -NYE
				1 10 20 30 40 50 60 70 80 90 100 110 120 130
RD3	14330.6	14480.2	1-134	NKASVVANQLIPINTALTLIMMKAEVVTPMGIPAEEIPNLVGMQVNRAVPLGTTLMPDMVKNYEDGTTSPGLKSVVANQLIPINTALTLVMMKAEEVSPKGIPSEEISKLVGMQVNRAVYLDQTLMPDMVKNYE

[a] Mass number of protonated molecular ions from FAB/MS for peptide fragments, and mass of native RD3 from ESI/MS.
[b] Mass calculated based on full sequence of peptide fragments and reconstructed complete sequence of RD3.
[c] Sequences obtained from automated sequencing. Dashed line indicates ambiguous residues or incomplete sequence.
N – Native RD3.
T – Tryptic fragments.
V – V8 proteinase fragments.
RD3 – Reconstructed complete sequence of RD3. The 9-amino acid connecting sequence (residues 65–73) between the N-terminal and C-terminal is underlined.

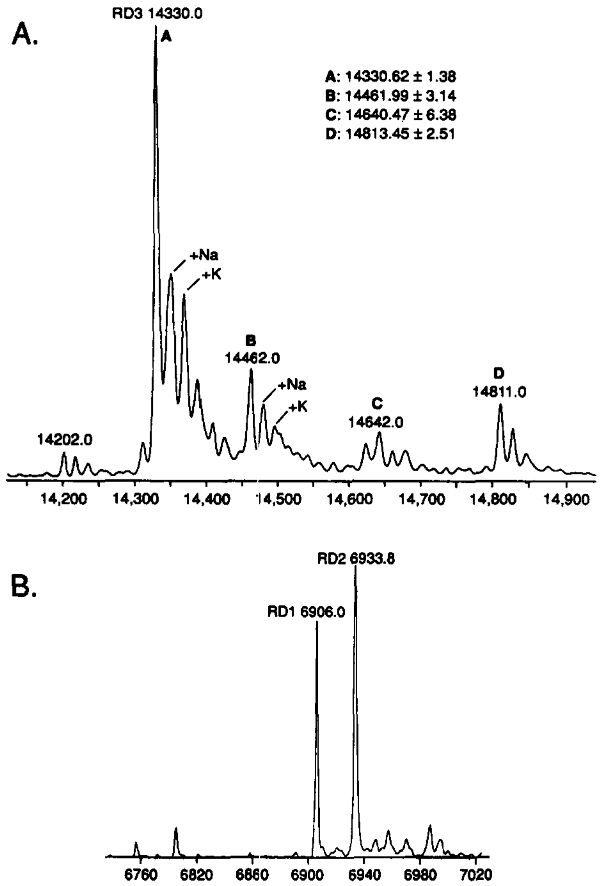

Fig. 7. (A) ESI/MS mass spectrum of native RD3 showing 1 major peptide A, and 3 minor ones, B, C and D. The actual masses and error margin for each are indicated. (B) ESI/MS mass spectrum of native RD1, mass 6906.0 Da, and of RD2, mass 6933.8 Da.

followed by the staggered overlaps between T2, V2, T4, V4* − 2 (the other 6-residue fragment that co-eluted in peak V4), T5, V6, T7 and V3 established the sequence to residue 131.

T1 was readily assigned as the C-terminal tryptic fragment, since it was the only tryptic fragment without a lysine or arginine. Its 3 residues, Asx, Glx, Tyr (Table 1) were assigned the order, -Asn-Tyr-Glu, since it is the conserved C-terminal sequence of all characterized Antarctic eel pout AFPs [1,2]. The 6 residues of V1 which appeared as residues 59–64, reappeared in the same order as residues 129–134, with its first 3 residues overlapping with the last three of T7, and its second 3 residues with T1 confirming the order of the 3 residues in T1. V1 is thus the C-terminal fragment of RD3 from V8 proteinase digestion. The RP-HPLC peak V1 (Fig. 4B) in fact then contained two identical sequence fragments originating from two different locations in the native AFP.

The FAB/MS mass of peptide fragments and the calculated mass based on sequence are also given in Table 2. The two values agreed for all peptide fragments whether their complete sequence was obtained from automated sequencing or deduced from corroborating overlapping sequences, confirming the accuracy of these sequences.

The ten tryptic fragments, in the order of T10, T8, T6, T3, T9, T2, T4, T5, T7 and T1, accounted for the entire RD3 peptide except for the first 2 residues at the N-terminus, which due to its small size, probably eluted during the isocratic equilibration before the acetonitrile gradient started. Its absence however is inconsequential because the N-terminal sequence was very clear from those of the native peptide and fragment V10.

The calculated mass of the reconstructed RD3 sequence is 14 480 Da, but the mass of native RD3 obtained by ESI/MS is 14330 Da (14 430.62 ± 1.38 Da) (Table 2, Fig. 7A), thus the two differ by 150 Da (see Section 4 for mass discrepancy). The mass spectrum of RD3 also showed minor molecular ions at 14 202.0, 14 462.0, 14 642.0 and 14 811.0, indicating presence of minor RD3 variants in the sample.

It should be pointed out that V8 proteinase digestion in this study caused non-specific cleavage at the C-terminal side of two Gly, residues 42 and 112. Such non-specific cleavage by V8 proteinase is not uncommon, and has been reported elsewhere [18].

3.5. Sequence of RD1 and RD2

The complete sequences for the 7000 Da AFPs, RD1 and RD2 were obtained by automated N-terminal sequencing of the native peptides. The calculated mass of RD1 is 6906.0 Da and 6934.4 Da for RD2 which agreed precisely with their respective ESI/MS mass, 6906.4 Da and 6933.5 Da (Fig. 7B). The accuracy of the two sequences was thus confirmed and no further analyses were carried out. The sequences of RD1, RD2, and RD3 are shown in Fig. 8.

```
               1                                                    64
7 kDa RD1:     NKASVVANQLIPINTALTLIMMKAEVVTPMGIPAEEIPKLVGMQVNRAVPLGTTLMPDMVKNYE

7 kDa RD2:     NKASVVANQLIPINTALTLIMMKAEVVTPMGIPAEDIPRIIGMQVNRAVPLGTTLMPDMVKNYE

               1                                                                  (connector)  74                                                        134
14 kDa RD3:    NKASVVANQLIPINTALTLIMMKAEVVTPMGIPAEEIPNLVGMQVNRAVPLGTTLMPDMVKNYE - DGTTSPGLK - SVVANQLIPINTALTLVMMKAEEVSPKGIPSEEISKLVGMQVNRAVYLDQTLMPDMVKNYE

                                                                                               1                                                         63
                                                                                  7 kDa AB2:   TKSVVANQLIPINTALTLVMMKAEEVSPKGIPAEEIPRLVGMQVNRAVYLDETLMPDMVKNYE
```

Fig. 8. Protein sequence of the three major AFPs of *L. dearborni*, RD1, RD2 and RD3, and sequence comparisons. The 4 different residues between RD1 and RD2 are double-underlined at corresponding positions in the two AFPS. RD3 is composed of two similar halves, residues 1–64 and residues 74–134, joined by a 9-amino-acid connector (residues 65–73) as indicated. The 10 different residues between the two halves are underlined at corresponding positions. The first half differs from RD1 and RD2 by 1 and 4 residues respectively, and the second half differs from AB2 [2] by 4 residues. The dissimilar residues are indicated by a dot.

3.6. Sequence comparisons

RD1 and RD2 are both 64 residues in length and differ by only 4 residues, all of which are conservative substitutions (Fig. 8). Sequence analysis of RD3 shows that the N-terminal half (residues 4 to 64) is largely repeated in the C-terminal half (residues 74 to 134), differing by only 10 residues. A 9-residue segment, residues 65–73, with no matching sequence in the peptide, connects the two halves (Fig. 8). The amino half of RD3 is very similar to RD1 and RD2, differing by only 1 and 4 amino acids, respectively. By contrast, the carboxyl half differs from RD1 and RD2 by 9 and 14 amino acids respectively, but closely resembles AB2, the minor AFP from a sympatric eel pout [16], differing by only 4 amino acids.

4. Discussion

Antifreeze peptide heterogeneity in the Antarctic eel pout, *Lycodichthys dearborni* (formerly *Rhigophila dearborni*), is exhibited in the presence of 3 major AFPs, RD1, RD2 and RD3, and at least 5 of minor variants. We have determined the complete protein sequence of the three major AFPs. RD1 and RD2 are 94% identical in sequence, and like all previously characterized type III AFPs are about 7 kDa in size. Sequence comparisons showed that the AFP 'RD' reported previously by Schrag et al. [1] is in fact RD2. By contrast, RD3 is 134 amino acids in length, and the calculated mass is 14471 Da, and thus RD3 is twice as large. This is the first example of a disparately large AFP from a single fish species for any of the 3 known types of AFPs. Additionally, RD3 is not a minor AFP in *L. dearborni*. The proteins in peaks II and III from Sephadex G75 chromatography of fish plasma proved to be almost exclusively antifreeze proteins. The protein yield (per 10 ml plasma) of peak II which contained solely RD3 was about 12–14 mg, and that of peak III which contained RD1 and RD2 were about 24–29 mg. Thus RD3 constitutes close to one-third by weight of all the AFPs in the blood, and RD1 and RD2 the remaining two-thirds.

The protein sequence of RD3 contains an interesting architecture of two similar 7 kDa type III AFP sequences linked in tandem. The 64-residue amino half (residues 1–64) and the 61-residue carboxyl half (residues 74–134) are joined head-to-tail by a 9-residue connecting sequence. The amino half closely resembles RD1 and RD2, with a sequence identity of 98% and 94% respectively. Greater sequence heterogeneity is observed between the amino half and the carboxyl half of RD3 (84% identity), as well as between the carboxyl half of RD3 and RD1 and RD2 (85% and 77% identity respectively). Interestingly, the carboxyl half of RD 3 closely resembles AB2, the minor 7 kDa AFP from a sympatric Antarctic eel pout, *Pachycara brachycephalum* (formerly *Austrolycicthys brachycephalus*) [16], with a difference of only 4 residues, or 94% sequence identity (Fig. 8). The 9-residue connector between the two halves of RD3, -DGTTSPGLK-, has no matching sequence in any of the characterized type III AFPs.

There is a discrepancy of 150 Da between the mass of RD3 calculated from its reconstructed complete sequence (14 480 Da) and that obtained from electrospray ionization mass spectrometry (14 330.0 Da). ESI/MS is a very accurate method of mass determination, and the mass of the major molecular ion in the mass spectrum of RD3 is $14\,330.62 \pm 1.38$ Da (Fig. 7A), and should be accurate within the error margin indicated. One possible explanation for the mass discrepancy is that the RD3 preparation used for peptide fragment characterizations was not fully homogeneous for the major RD3 AFP, but contained some percentage of other RD3 variants which could not be completely separated from the major one. As seen in Fig. 2B, RP-HPLC revealed at least two other RD3 variants indicated by the incompletely resolved peak on both the ascending and descending slope of the composite RD3 peak. The closeness in elution times precluded complete separation of material from these peak components during manual collection of eluting proteins. Accordingly, the ESI/MS mass spectrum of RD3 showed minor molecular ions at both lower and higher mass numbers (Fig. 7A), presumably those of minor RD3 variants. Thus the possibility exists that one or more of the peptide fragments in the reconstructed RD3 complete sequence came from a minor RD3 variant. However, this is unlikely, since the minor variant and thus its peptide fragments would be much less abundant than those of the major RD3 to appear as substantial peaks on the tryptic and V8 proteinase RP-HPLC profiles to warrant selection for characterization, unless a particular fragment from the minor variant is similar enough to the corresponding one from the major RD3 that they co-eluted on RP-HPLC. An indication of the latter possibility is the molecular ion at mass number 1027.6 in the FAB/MS mass spectrum of V2 (Fig. 6), but had no matching calculated sequence mass. The molecular ion at 1042.5 in the same spectrum was chosen to be that of V2, because it agreed with the calculated mass of V2 from the sequence of the fragment. The signal intensity of a FAB/MS molecular ion does not depend on the abundance of the peptide fragment, but the PTC-amino-acid yields in protein sequencing does. Thus it is almost certain that the peptide fragment sequences obtained were of those originating from the most abundant AFP, the major RD3. In support of this, if a fragment does come from a minor variant and contains residues that would account for the mass discrepancy, these residues would differ from the ones at corresponding positions in the upstream and downstream overlapping sequences. This situation did not occur, and additionally FAB/MS masses of the peptide fragments agreed precisely with the calculated masses from fragment sequences verifying the accuracy of these sequences. In any event, though the precise cause of the small discrepancy between ESI/MS mass and calculated mass of RD3 cannot be fully explained at this time, the

distinctive structural architecture of RD3, i.e., two 7 kDa AFPs of similar amino-acid sequence linked in tandem by a short connector of unmatched sequence, is very clear.

It is not obvious what functional advantage there may be for the fish to have a large peptide as a major AFP since purified RD3 in solution appeared to be no more effective in freezing point depression on a weight basis (10 mg/ml), than the other major AFPs, RD1 and RD2 half its size. Studies have shown that all of the antifreeze peptides and glycopeptides with a molecular weight above 3300 Da depress the freezing point non-colligatively to the same extent on a weight basis, but on a molar basis, the larger-sized antifreezes are more effective [19,20]. Since RD3 is twice as large as RD1 and RD2 but comparable in activity on a weight basis, numerically it would have twice the activity of the latter two on a molar basis. It is almost certain that the two halves of RD3 both have antifreeze activity because of the high degree of sequence identity of each half to known functional AFPs. RD1 and RD2 which are near identical to the amino half, and AB2 which is near-identical to the carboxyl half all produce a non-colligative freezing point depression of about 1°C (this study, [16]). However, the collective antifreeze activity of all the AFPs in the blood of *L. dearborni* remains at about 1.1°C [7], comparable to that of the sympatric eel pout *P. brachychepahlum* which synthesizes only 7 kDa AFPs [16]. Perhaps when serum abundance is taken into consideration, the weighted contribution of RD3 does not augment the overall freezing protection to a level higher than that generally observed in Antarctic fishes.

All antifreeze-bearing fish synthesize not just one but a number of structurally related antifreeze molecules. The protein heterogeneity of antifreezes is a manifestation of many distinct antifreeze protein genes in the fish. A multigene family has been described for the type I AFPs from the flounders [21,22], type III AFPs from the atlantic ocean pout [14] and wolffish [15], and for the AFGPs from the antarctic cod [23]. AFPs of *L. dearborni* are no exception. Southern blot analysis of *L. dearborni* genomic DNA showed many positive hybridization bands, indicating that its AFPs are encoded by a multigene family as well (unpublished results). It is possible that RD3 was a structural variant that evolved during amplification of antifreeze genes and was maintained in the genome, since the contribution of this large AFP to freezing avoidance is no worse or no better than its 7 kDa counterparts. An interesting and obvious question that follows is whether there are distinct genes that separately encode the 14 kDa RD3 and the 7 kDa RD1 and RD2, or are they the products of the same genes capable of giving rise to both size variants by post-transcriptional processing. We have investigated the genomic basis for the RD AFP heterogeneity and our results indicated that there are distinct genes for the two size variants (unpublished results).

All protein antifreezes achieve non-colligative freezing point depression by binding to ice, and inhibiting ice growth likely through the Kelvin effect [5]. The first antifreezes that were characterized, the AFGPs and type I AFPs, were elongated molecules [24] with repetitive sequences [2,8]. The affinity of these antifreezes for ice was thus ascribed to a matching of the repeat spacings of polar residues or side chains along the elongated antifreeze molecule and the periodicity of water molecules in the ice lattice, allowing hydrogen bonding and thus adsorption to occur [6,8]. Type I AFP of winter flounder has been confirmed to be a single helix by X-ray crystallography [25], and NMR (nuclear magnetic resonance) evidence indicated a left-handed helical configuration similar to the polyproline II helix for the small AFGP 8 [26]. Recent ice-crystallographic studies by Knight et al. which showed adsorption of the type I AFP [27] and the small AFGPs 7 and 8 [28] to specific crystallographic planes in hexagonal ice with the antifreeze molecules aligned along a specific direction lend support to the structural matching theory of antifreeze adsorption. For these helical antifreezes, the alignment direction can be deduced because the geometric constraints imposed by the axial asymmetry of the molecules on how they would be accommodated on a growth step on their adsorption plane produced a specific pattern of adsorption regions amenable to crystallographic analyses. Knowing the alignment direction, the polar residues or side chains that match the periodicity along that direction, and thus likely involved in ice-binding, were identified [27,28].

Unlike the AFGPs and type I AFPs, the 7 kDa type III AFPs are non-helical, lack repetitive sequences, and are globular in tertiary structure, and how these AFPs bind to ice remains an unanswered question. Two-dimensional NMR studies of the solution structure of a recombinant AFP from Atlantic ocean pout revealed it to be a small globular protein within which there are three sheets of anti-parallel β-strands [29]. Our preliminary adsorption crystallographic studies of the 7 kDa antarctic eel pout AFPs (AB1, RD1 and RD2) indicated adsorption to the prism planes (Knight and DeVries, unpublished data), but the alignment direction of the molecules could not be readily deduced because the pattern of adsorption regions was non-specific, consistent with a globular or lack of axial asymmetry in these molecules.

By contrast, the 14 kDa type III AFP, RD3, is likely to have axial asymmetry, with two globular moieties linked by a short sequence segment (the 9-residue connector) indicated to be a random chain by protein secondary structure predictions [30], which may allow its molecular alignment on its adsorption plane to be deduced. As discussed above, both halves of RD3 are likely active antifreezes, and thus both would be expected to bind to ice, independent of each other but likely in a similar way because of the similarity in sequence. Studies on the ice crystallography of adsorption of this newly discovered 14 kDa type III AFP variant may help advance the understanding of how type III AFPs bind to ice in general.

Acknowledgements

This work was supported in part by grant NSF DPP 90-19881 to A.L.D. The ZAB-SE mass spectrometer for FAB/MS mass spectra, and the Quattro mass spectrometer for ESI/MS spectra were equipments at the U. of I. Mass Spectrometry Center purchased in part with grants from NIH RR 01575, NSF PCM 8121494, and NIGMS GM 27029, and NIH RR 07141. We thank Dr. Richard Milberg for his valuable assistance in obtaining and interpreting the mass spectra.

References

[1] Schrag, J.D., Cheng, C.C., Panico, M., Morris, H.R. and DeVries, A.L. (1987) Biochim. Biophys. Acta. 717, 322–326.
[2] DeVries, A.L., (1971) Science 172, 1152–1155.
[3] DeVries, A.L. and C.C. Cheng (1992) in Water and Life: Comparative Analysis of Water Relationships at the Organismic, Cellular, and Molecular Levels (Somero, G.N. and Bolis, C.L., eds.), pp. 717–720, Springer, Berlin.
[4] Davies, P.L. and C.L. Hew, (1990) FASEB J. 4, 2460–2468.
[5] Raymond, J.A. and DeVries, A.L. (1977) Proc. Natl. Acad. Sci. USA 74, 2589–2593.
[6] DeVries, A.L., (1984) Phil. Trans. R. Soc. Lond. B304, 575–588.
[7] DeVries, A.L., (1988) Comparative Biochem. Physiol. 90B, 611–621.
[8] DeVries, A.L. and Y. Lin. (1977) Biochim. Biophys. Acta 495, 388–392.
[9] Hew, C.L., Joshi, S., Wang, N.-C., Kao, M.-H. and Ananthanarayanan, V.S. (1985) Eur. J. Biochem. 151, 167–172.
[10] Pickett, M., Scott, G., Davies, P., Wang, N., Joshi, S. and Hew, C.L. (1982) Eur. J. Biochem. 143, 35–38.
[11] Cheng, C.C. and A.L. DeVries (1991) in Life Under Extreme Condition (DePrisco, G., ed.), pp. 1–14, Springer, Berlin.
[12] Ng, N.F.L. and Hew, C.L. (1992) J. Biol. Chem. 267, 16069–16075.
[13] Ewart, K.V. and Fletcher, G.L. (1993) Mol. Mar. Biol. Biotechnol. 2, 20–27.
[14] Hew, C.L., Wang, N.C., Joshi S., Fletcher, G.L., Scott, G.K., Hayes, P.H., Buettner, B. and Davies, P.L. (1988) J. Biol. Chem. 263, 12049–12055.
[15] Scott, C.K., Hayes, P.H., Fletcher, G.L. and Davies, P.L. (1988) Mol. Cell Biol. 8, 3670–3675.
[16] Cheng, C.C. and A.L. DeVries, (1989) Biochim. Biophys. Acta 997, 55–64.
[17] DeVries, A.L., (1986) Methods Enzymol. 127, 293–303.
[18] Dognin, M.J. and Wittman-Liebold, B. (1977) FEBS Lett. 84, 343–346.
[19] Kao, M.H., Fletcher, G.L., Wang, N.C. and Hew, C.L. (1985) Can. J. Zool. 64, 578–582.
[20] Schrag, J.D., O'Grady, S.M. and DeVries, A.L. (1982) Biochim. Biophys. Acta 717, 322–326.
[21] Gourlie, B., Lin. Y., Powers, D., DeVries, A.L. and Huang, R.C. (1984) J. Biol. Chem. 264, 11313–11316.
[22] Scott, G.K., Hew, C.L. and P.L. Davies, (1985) Proc. Natl. Acad. Sci. USA 82, 2613–2617.
[23] Hsiao, K.-C., Cheng C.C., Indira, E.F., Detrich, H.W. and DeVries, A.L. (1990) Proc. Natl. Acad. Sci. USA 87, 9265–9269.
[24] Raymond, J.A., Radding, W. and DeVries, A.L. (1977) Biopolymers 16, 2575–2578.
[25] Yang, D.S.C., Sax. M., Chakrabartty, A. and Hew, C.L. (1988) Nature 333, 232–237.
[26] Rao, B.N.N. and Bush, C.A. (1987) Biopolymers 26, 1227–1244.
[27] Knight, C.A., Cheng, C.C. and DeVries, A.L. (1991) Biophys. J. 59, 409–418.
[28] Knight, C.A., Driggers, D. and DeVries, A.L. (1993) Biophys. J. 64, 252–259.
[29] Sonnichsen, F.D., Sykes, B.D., Chao, S. and Davies, P.L. (1993) Science 259, 1154–1157.
[30] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.