Use of proline mutants to help solve the NMR solution structure of type III antifreeze protein

HEMAN CHAO,¹ PETER L. DAVIES,¹ BRIAN D. SYKES,² AND FRANK D. SÖNNICHSEN²

¹ Department of Biochemistry, Queen's University, Kingston, Ontario K7L 3N6, Canada

² Protein Engineering Network of Centres of Excellence and Department of Biochemistry,
University of Alberta, Edmonton, Alberta T6G 2S2, Canada

(RECEIVED March 29, 1993; REVISED MANUSCRIPT RECEIVED June 4, 1993)

Abstract

To help understand the structure/function relationships in antifreeze proteins (AFP), and to define the motifs required for ice binding, a Type III AFP suitable for two-dimensional (2D) NMR studies was produced in *Escherichia coli*. A synthetic gene for one of the Type III AFP isoforms was assembled in a T7 polymerase-directed expression vector. The 67-amino acid-long gene product differed from the natural AFP by inclusion of an N-terminal methionine but was indistinguishable in activity. The NMR spectra of this AFP were complicated by *cis-trans* proline isomerization from the C-terminal sequence YPPA. Substitution of this sequence by YAA eliminated isomer signals without altering the activity or structure of the mutant AFP. This variant (rQAE m1.1) was selected for sequential assignment and the secondary structure determination using 2D ¹H NMR spectroscopy. Nine β -strands are paired to form two triple-stranded antiparallel sheets and one double-stranded antiparallel sheet. Two further proline replacements, P29A and P33A, were made to delineate the role of conserved prolines in Type III AFP. These mutants were valuable in clarifying ambiguous NMR spectral assignments amongst the remaining six prolines of rQAE m1.1. In contrast to the replacement of the C-terminal prolyl residues, the exchange of P29 and P33 caused some structural changes and significantly decreased protein solubility and antifreeze activity.

Keywords: antiparallel; β -sheets; *cis*-peptide bond; *cis-trans* isomerization; *Escherichia coli* expression; synthetic gene

Certain marine teleosts are protected from freezing in ice-laden seawater below -0.7°C by the presence of macromolecular antifreezes in their blood circulation (DeVries, 1983; Davies et al., 1988). These are either antifreeze glycoproteins, built of a tripeptide repeat (Ala-Ala-Thr)_n with a disaccharide attached to the threonyl residue (Feeney & Yeh, 1978), or one of at least three very dif-

ferent types of antifreeze protein (Davies & Hew, 1990). Type I AFPs are alanine-rich, amphipathic α -helical proteins found in righteye flounders and sculpins, which range in size from 33 to 48 amino acids (Scott et al., 1987; Yang et al., 1988). Type II AFP, found in sea raven, is a larger protein (129 amino acids) containing 10 half-cystine residues that form five disulfide bridges (Ng et al., 1986; Hayes et al., 1989; Ng & Hew, 1992). Recently, homologs of this protein have been found in herring and smelt (Ewart & Fletcher, 1990). Type III AFPs are intermediate in size between the other AFPs and have been found only in different families of the Zoarcoidei (Davies & Hew, 1990). They have no extreme amino acid bias, although they are moderately rich in proline and methionine (Li et al., 1985; Hew et al., 1988).

Most of the Type III AFP isoform sequences determined directly from the proteins or indirectly from the DNA are 62–66 amino acids long (Davies & Hew, 1990). They are not only distinct from the other AFP types, but they have no obvious homologs in the protein sequence database. They can be divided into two homology groups.

Reprint requests to: Frank Sönnichsen, Protein Engineering Network of Centres of Excellence, University of Alberta, Edmonton, Alberta T6G 2S2, Canada.

Abbreviations: AFGP, antifreeze glycoprotein; AFP, antifreeze protein; dNTP, deoxyribonucleoside 5'-triphosphate; DQF-COSY, double quantum filtered correlation spectroscopy; DTT, dithiothreitol; FPLC, fast protein liquid chromatography; HPLC, high performance liquid chromatography; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; PAGE, polyacrylamide gel electrophoresis; PEG, polyethylene glycol; PMSF, phenylmethylsulfonyl fluoride; QAE, quaternary aminoethyl/AFP that binds QAE-Sephadex; SCUBA, stimulated cross peaks under bleached alphas; SDS, sodium dodecyl sulfate; SP, sulfopropyl; TEN, 50 mM Tris-HCl/1 mM EDTA (pH 8.0); 0.3 M NaCl, 0.1 mM PMSF; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy; 1D, one-dimensional; 2D, two-dimensional; 3D, three-dimensional.

One group, represented by HPLC-12 from ocean pout (*Macrozoarces americanus*), predominates in *Lycodes polaris*, *Rhigophila dearborni*, and *Austrolycithys brachycephalus* (Schrage et al., 1987; Cheng & DeVries, 1989). The members of this group have pIs just below neutrality and have been referred to as QAE components because of their binding to anion exchangers (Li et al., 1985). The known QAE components show 75% sequence identity. The other homology group predominates in ocean pout and Atlantic wolffish (*Anarhichas lupus*) (Li et al., 1985; Scott et al., 1988). Their pIs are basic, and because they were previously chromatographed on cation exchangers, they have been referred to as SP components. They show 90% sequence identity within the group and are 50% identical to the QAE components. Both components can occur in the same fish (Li et al., 1985).

Although ocean pout serum is a rich source of Type III AFPs, particularly in winter when circulating levels are approximately 20 mg/mL (Fletcher et al., 1985), it is a mixture of many components. At least 12 independently active components have been resolved by reversed-phase HPLC, most of which have internal sequence differences (Hew et al., 1988). These are the products of ~150 genes. Thus, the potential for sequence heterogeneity is extreme and is enhanced by variations in processing at both the N- and C-termini (Li et al., 1985). To simplify the purification of a single AFP from ocean pout, HPLC-12 (referred to here as QAE) was selected. It is the only major Type III AFP isoform in ocean pout that binds to QAE-Sephadex and the last to elute from reversed-phase HPLC.

Prior to this study no detailed secondary or tertiary structure information was available on Type II or Type III AFP. Circular dichroism studies only indicated that both AFPs have a well-defined tertiary structure with little, if any, α -helix present (Slaughter et al., 1981; Ananthanarayanan et al., 1986; Schrage et al., 1987). Type I AFP is the only macromolecular antifreeze for which a detailed atomic structure has been reported (Yang et al., 1988). X-ray crystallography determined the fold of this protein to be a single, straight helix with amphipathic character. This structure, in conjunction with studies of the ice crystal growth in the presence of AFP and AFGP, has yielded a detailed model of antifreeze action. The proposed adsorption-inhibition mechanism is generally accepted for all macromolecular antifreezes, despite their distinct primary structures (DeVries, 1984; Hew & Yang, 1992), because they have a similar effect on ice crystal growth (Raymond et al., 1989). Hydrophilic protein residues, primarily Thr, Asn, and Asp (Wen & Laursen, 1992a), or the hydroxyl groups of the sugar moieties of AFGP (Knight et al., 1993), are proposed to interact with the surface of ice by forming hydrogen bonds, while the hydrophobic residues are exposed to the solvent, thereby preventing water from joining the crystal. It has been observed that the binding of AFPs occurs primarily at the

prism faces of the hexagonal ice crystal, resulting in a change of the initial crystal growth toward bipyramidal crystals. This preferential binding is believed to be caused by defined spacing of the ice binding groups, matching the spacing of the water molecules at specific crystal facets. Initially, distances of 4–8 Å were suggested; more recently, molecular dynamics calculations (Chou, 1992; Jorgensen et al., 1993), studies with synthetic peptides (Wen & Laursen, 1992b), and ice etching studies by Knight et al. (1991) suggest that a spacing of 16.5 Å between hydrophilic residues is responsible for the observed preferential binding to nonbasal planes of the hexagonal ice crystal. However, these recent studies involved only Type I AFPs and the results might not be generally applicable.

Due to the lack of detailed information on the structure of the Type II or Type III AFP, it is not known whether all AFPs share a common mechanism and a common structural motif for ice binding. We therefore started structural studies on the Type III AFP using NMR spectroscopy, which in recent years has been shown to provide detailed 3D structural information for small proteins in solution, as well as information about protein hydration (Otting & Wüthrich, 1989; Otting et al., 1991). In this study we have used 2D NMR spectroscopy to sequence-specifically assign the ^1H NMR resonances and to determine the secondary structure of the protein. A description of the global fold has been reported elsewhere (Sönnichsen et al., 1993). We also describe the production and properties of several proline mutants that have been valuable in the verification of spectral assignments and that have indicated the importance of conserved prolyl residues for the structure, solubility, and function of Type III AFP.

Results

Production and purification of recombinant AFPs

After the synthetic gene for rQAE was assembled and cloned into pT7-7, it was sequenced on both strands. There were no changes from the sequence shown in Figure 1A. Expression of rQAE in *Escherichia coli* extracts was detected even before the end of the heat shock induction by both thermal hysteresis and SDS-PAGE (Fig. 2A,B). Recombinant QAE was almost the smallest protein in the *E. coli* extracts. It migrated between bovine trypsin inhibitor (M_r 6,000) and insulin (A and B chains) (M_r 3,000) with an apparent M_r of 4,500 consistent with previous observations (Hew et al., 1984). It was indistinguishable in its mobility from QAE protein purified from ocean pout serum (Fig. 2B, lane 13). Accumulation of the soluble protein leveled off after 45 min and remained stable for at least 2 h. At this point an equal quantity of rQAE was present as inclusion bodies in the pellet. The induction and distribution of rQAE m1.1 was similar to rQAE.

Fig. 1. Synthetic QAE gene sequences. **A:** The sequences of the rQAE gene and its protein product are shown with the positions of unique restriction sites in italics. Amino acids are numbered starting with the initiating methionine as 0. The individual synthetic oligonucleotides used for gene assembly are boxed. **B:** Replacement oligonucleotides coding for C-terminal YAA that were inserted between *Bst*EII and *Eco*RI.

Routinely, 30–40 mg of recombinant AFP were produced from a 3-L culture. This corresponded to ~20% of the total *E. coli* protein. Gel permeation chromatography was an effective step for the isolation of this small protein from the soluble fraction (Fig. 3A). Most of the bacterial proteins eluted close to the void peak and well away from rQAE. In the subsequent ion-exchange chromatography, AFP was the main protein peak in the elution profile (Fig. 3B), and minor impurities were subsequently removed by HPLC (Fig. 3C). As much as 50% of the AFP in the soluble fraction was recovered.

The recovery of AFP from inclusion bodies by organic extraction circumvented the need for gel permeation chromatography. These extracts were of comparable purity to the G-75 Sephadex fractions applied to FPLC (not shown). After the HPLC step, rQAE and rQAE m1.1 recovered from the pellet were indistinguishable in their antifreeze activity and NMR spectra from protein purified from the soluble fraction (not shown).

Characterization of rQAE and rQAE m1.1

The amino acid composition of rQAE matched that of QAE from ocean pout except for the addition of one methionine (not shown). Also, the N-terminal sequence was

determined by Edman degradation as far as residue 35 and showed no difference between rQAE and QAE except for the presence of an N-terminal methionine in the former. When the antifreeze activities of QAE, rQAE, and rQAE m1.1 were compared by assessing their thermal hysteresis values over a concentration ranging from 0 to 0.29 mM, the activity profiles for the three AFPs were indistinguishable within the limits of standard errors for the measurements (Fig. 4). Moreover, all three proteins altered the direction of ice growth in their presence to produce the same bipyramidal crystals (not shown).

NMR characterization

One-dimensional ^1H NMR spectra were obtained for QAE, rQAE, and rQAE m1.1 in D_2O (Fig. 5). These proteins contain only one aromatic residue, Y63. Therefore, the aromatic region (6.5–7.5 ppm) of these spectra should show only one pair of doublets. However, the spectrum of QAE purified from ocean pout serum (Fig. 5C, left) consisted of several resonances in this region, indicating the presence of multiple forms or components in this sample. This was also visible in the regions of downfield-shifted αCH protons (Fig. 5C, right). The number of resonances in the aromatic region was significantly re-

Fig. 2. Induction of AFP. During heat shock and subsequent growth of *E. coli*, culture samples (5 mL) were taken at time intervals. Cells were recovered by centrifugation and frozen until all samples had been accumulated. Cell pellets were resuspended (5 mL) and broken as described in the Materials and methods. After centrifugation, the supernatant extracts were lyophilized and resuspended in 50 μ L of 10 mM Tris-HCl (pH 8.0). Aliquots (10 μ L) were retained for thermal hysteresis measurements. SDS-sample loading buffer (10 μ L of 62.5 mM Tris-HCl [pH 6.8], 10% [v/v] glycerol, 2% [w/v] SDS, 5% [v/v] β -mercaptoethanol, 0.001% [w/v] bromophenol blue) was added to the remainder of the samples (Laemmli, 1970). **A:** SDS-PAGE was done for 7 h at 25 mA. Lane 1 contained low molecular weight protein markers (GIBCO-BRL). Lane 2 contained *E. coli* extract immediately prior to heat shock. Lane 3 contained extract after 30 min of heat shock. Lanes 4–11 were loaded with extract recovered at 15-min intervals thereafter up to 120 min. Lanes 12 and 13 contained purified rQAE (20 μ g) and partially purified QAE from ocean pout serum ($\sim$ 50 μ g), respectively. **B:** Thermal hysteresis measurements of the same culture extracts are displayed as a histogram. Each value in $^{\circ}$ C is the average of two measurements. The time of sampling is displayed below where 0 represents the cessation of heat shock (equivalent to lane 3 in Fig. 2A).

Fig. 3. Purification of AFP. **A:** Concentrated *E. coli* extract from sonicated cells equivalent to 650 mL of culture was chromatographed on G-75 Sephadex as described in Materials and methods. Fractions (10 mL) were assayed for A_{280} (●—●), and after 20-fold concentration, for thermal hysteresis (○—○). Fractions 112–138 inclusive were pooled. **B:** FPLC chromatography step during rQAE m1.1 purification. The absorbance of the eluent was monitored at 280 nm (—). The NaCl gradient is represented by the dotted line, and the arrow indicated the peak that had thermal hysteresis activity. **C:** HPLC chromatography of AFP from B. The absorbance of the eluent was monitored at 230 nm. The acetonitrile/propanol gradient is represented by the dotted line.

duced for a sample of rQAE (Fig. 5B, left). Two doublets at 6.95 and 6.78 ppm established this sequence to be the major component in serum QAE. However, the spectrum still showed the presence of a second, minor component with a pair of aromatic doublets (6.96, seen as a shoulder under the main peak, and 6.69 ppm). The ratio of

these two components was approximately 5:1 and was only slightly affected by temperature, salt concentration or pH (data not shown). The NH- α CH fingerprint region of the DQF-COSY spectrum of rQAE (Fig. 6A) also revealed two forms because many of the intrasidue cross-peak multiplets that connect the NH and α CH resonances of a given residue showed a major and a minor component. The location of the structural difference between the two components can be inferred by the size of the chemical shift differences between the resonances of either

Fig. 4. Thermal hysteresis activity comparisons of natural and recombinant AFPs. The thermal hysteresis activities of natural QAE (●—●), rQAE (▲—▲), rQAE m1.1 (○—○), P33A (■—■), and P29A (□—□) were measured as a function of concentration. The activity of P29A was measured only up to a concentration of 0.19 mM because of its limited solubility. Each point is an average of three independent determinations. Standard deviations are shown by the vertical bars. A single line was drawn through the QAE, rQAE, and rQAE m1.1 data sets.

Fig. 5. One-dimensional ^1H NMR spectra of natural and recombinant AFPs. Portions of the aromatic region (6.5–7.4 ppm) and the region of downfield-shifted α CH (5.0–6.0 ppm) are on the left and right, respectively. The spectra for rQAE m1.1 (A), rQAE (B), and QAE (C) were acquired at 25 °C in D_2O . Conditions are as described in the Materials and methods.

Fig. 6. Comparison of parts of the fingerprint region of the DQF-COSY spectra. The spectra for rQAE (A) and rQAE m1.1 (B) were obtained in $\text{H}_2\text{O}/\text{D}_2\text{O}$ (80:20%) at 25 °C and pH 7.5 and 8.0, respectively. Cross peaks are labeled with their residue number. Solid lines connect equivalent cross peaks of the major and minor component. Dashed lines connect the NH- α CH1 and NH- α CH2 cross peaks for glycine residues. The vertical contours in B at 4.80 ppm are experimental artifacts from the H_2O resonance.

component. Although many cross peaks between α CH and NH protons of the two components were close or overlapped completely, the chemical shifts for protons of Y63 varied the most between these two forms; the shifts differ by 0.25 ppm for α CH and 0.30 ppm for the amide proton. On the assumption that the recombinant protein was homogeneous in sequence, this indicated the presence of a *cis-trans* isomerization of the C-terminal sequence YPPA.

Only one pair of doublets was exhibited in the aromatic region of rQAE m1.1 (Fig. 5A), a mutant with the C-terminal sequence YAA. Due to the sequence alteration, these doublets were observed at slightly different chemical shifts as compared to rQAE. The region of low-field α CH-resonances simplified as well. The fingerprint region of the DQF-COSY showed the expected number of peaks and no resonances of a second form or component (Fig. 6B), which confirmed that the structural heterogeneity in the rQAE sample originated in the YPPA sequence, thus implying a *cis-trans* isomerization.

Spin-type assignment

The NMR spectra obtained for rQAE m1.1 are of excellent quality, as shown by the representative 1D ^1H NMR spectrum of the amide and aromatic region (Fig. 7). This spectrum additionally shows the unusual dispersion of the amide resonances. A total of 59 amide protons were observed in this region because the protein contains six prolines, and the N-terminal residue cannot be detected due to chemical exchange. Most of the amide resonances are well resolved or overlap only with one other resonance. They are therefore easily assigned and are labeled with their residue number (Fig. 7). Spin-type as well as amino acid-type assignments relied primarily on the amide region of TOCSY spectra (Fig. 8). In this particular spectrum, 38 of the 59 observable amide resonances exhibit a transfer of magnetization throughout the complete spin system. By comparison with DQF-COSY and NOESY spectra at different temperatures, these and the remaining 21 residues (except Pro) could be assigned completely

Fig. 7. Amide and aromatic region of the 600-MHz ^1H NMR spectrum of rQAE m1.1 at 25 °C and pH 8.0. Some resolved backbone amide resonances and the aromatic ring protons of Y63 are labeled with their residue number.

and unambiguously. However, due to missing direct coupling information in the COSY spectra, a small number of the resonance assignments to specific atoms are tentative (Table 1).

Sequential assignment

The sequential resonance assignment was carried out by standard procedures (Wüthrich, 1986) using the fingerprint region of the NOESY spectrum (Fig. 9). Due to the high dispersion of the amide resonances, only a few degenerate cross peaks for backbone atoms are observed. In cases of weak or ambiguous $d_{\alpha\text{N}}(i, i+1)$ connectivities, sequential assignments could be made due to the presence of medium or strong NH to NH connectivities ($d_{\text{NN}}(i, i+1)$) or additional side-chain proton-NH cross peaks, primarily $d_{\beta\text{N}}(i, i+1)$. Therefore, the sequence-specific assignment was straightforward and was accomplished almost independently of the initial spin-type assignment, leading to the assignment of seven contiguous sequence regions that were interrupted only by prolyl residues.

Proline assignments

Five out of six prolines (12, 29, 33, 38, 50, 57) give rise to strong $d_{\alpha\text{N}}(i, i+1)$ cross peaks to the respective subsequent residues, which identify the Pro αCH resonances sequence specifically. For P12, P50, and P57, sequential connectivities were also observed between their

Fig. 8. Amide region from a 2D TOCSY ^1H NMR spectrum of rQAE m1.1 at 10 °C and pH 4.8. This region illustrates the good dispersion of the signals from amide protons. Sequence-specific assignments are indicated for some residues, which show a magnetization transfer through the complete spin system. The horizontal lines connect the relayed peaks arising from the same spin system.

Table 1. ^1H resonance assignments for rQAE m1.1^a

Residue	NH	αCH	βCH	γCH	Other
0 M		4.13	2.19, 1.83	2.63	
1 N	9.01	4.75	2.78, 2.69		NH ₂ 7.62, 7.03
2 Q	8.48	4.32	2.26, 2.12	2.42, 2.42	NH ₂ 8.12, 7.62
3 A	9.06	4.50	1.33		
4 S	9.19	5.56	3.65, 3.65		OH 7.35
5 V	8.20	4.21	1.82	0.82, 0.78	
6 V	9.07	4.54	1.49	0.77, 0.64	
7 A	8.40	4.48	1.67		
8 N	9.11	4.74	2.71, 2.02		NH ₂ 7.52, 6.88
9 Q	7.83	4.43	2.07, 1.99	2.32, 2.32	NH ₂ 7.50, 6.94
10 L	8.46	4.73	1.69, 1.69*	1.31*	$\delta\delta'$ 0.86*, 0.80*
11 I	9.31	4.57	2.02	1.05, 0.94*	γCH_3 2.16*, δCH_3 1.26*
12 P	—	4.42	2.37, 1.67	2.12, 2.05	$\delta\delta'$ 4.12, 3.66
13 I	8.46	3.33	1.63	0.93*, 0.91*	γCH_3 0.86, δCH_3 1.13*
14 N	7.35	4.56	3.11, 3.00		NH ₂ 7.85, 7.11
15 T	7.45	4.00	3.94	1.19	
16 A	8.54	4.29	1.25		
17 L	9.17	4.57	2.28, 2.28*	1.94	$\delta\delta'$ 0.93, 0.84
18 T	8.01	4.72	4.54	1.27	
19 L	8.92	3.83	1.74, 1.56	1.69	$\delta\delta'$ 0.92, 0.77
20 V	7.55	4.14	2.24	0.96, 0.96	
21 M	7.61	4.00	2.20, 2.17	2.86, 2.45	
22 M	7.70	4.87	2.10, 1.69*	2.32, 2.32*	
23 R	9.08	4.84	1.85, 1.60	1.40, 1.17	$\delta\delta'$ 3.15, 3.08, ϵNH 7.42
24 S	8.39	5.22	3.63, 3.53		
25 E	9.08	4.61	1.90, 1.78	2.23, 2.16	
26 V	9.14	4.11	2.08	0.88, 0.80	
27 V	8.28	4.75	2.08	0.86, 0.76	
28 T	8.40	4.41	3.93	1.17	
29 P	—	4.80	2.45, 2.12	1.96, 1.77	$\delta\delta'$ 3.64, 3.47
30 V	8.35	3.87	2.24	1.11, 1.01	
31 G	8.28	5.00, 3.52			
32 I	8.76	3.94	1.28	1.20, 0.36	γCH_3 1.01, δCH_3 0.08
33 P		4.80	2.41, 2.01	2.14, 2.12	$\delta\delta'$ 3.82, 3.78
34 A	8.29	3.84	1.23		
35 E	9.11	4.06	2.32, 2.08	2.46, 2.46	
36 D	8.22	4.92	2.96, 2.83		
37 I	7.58	3.69	1.92	1.20*, 0.81*	γCH_3 0.83*, δCH_3 0.78
38 P		4.18	2.34, 1.87	2.10, 1.96	$\delta\delta'$ 3.69, 3.54
39 R	7.33	4.26	1.94, 1.94	1.78, 1.67	$\delta\delta'$ 3.25, 3.17, ϵNH 7.21
40 L	7.76	4.13	1.81*, 1.72*	1.42	$\delta\delta'$ 0.85, 0.80
41 V	6.77	3.23	1.97	1.05, 0.97	
42 S	9.41	4.05	4.29, 4.29		
43 M	8.80	4.49	2.23, 2.08	2.76, 2.42	ϵCH_3 2.09
44 Q	8.96	5.32	2.09, 1.76	2.29, 2.29	NH ₂ 7.61, 6.97
45 V	7.54	5.35	2.48	1.12, 1.02	
46 N	8.58	4.73	3.27, 2.40		NH ₂ 7.40, 7.29
47 R	7.85	4.41	2.09, 1.85	1.58, 1.52	$\delta\delta'$ 3.21, ϵNH 7.18
48 A	8.55	4.77	1.31		
49 V	9.30	4.48	2.24	1.02, 0.87	
50 P	—	4.31	2.34, 1.68	2.03, 1.82	$\delta\delta'$ 3.88, 3.65
51 L	8.26	3.38	1.56, 1.32	1.29	$\delta\delta'$ 0.84, 0.84
52 G	8.93	4.24, 3.48			
53 T	7.79	4.26	4.10	1.27	
54 T	8.83	4.29	3.94	1.11	
55 L	9.14	4.36	1.47, 1.16	1.41	$\delta\delta'$ 0.78, 0.65
56 M	7.90	5.09	2.22, 1.84	2.90, 2.58	ϵCH_3 2.04
57 P		3.72	2.48, 2.11	2.30, 2.30	$\delta\delta'$ 3.91, 3.59
58 D	8.05	4.47	2.92, 2.67		
59 M	8.16	4.59	2.16, 2.11	2.67, 2.31	ϵCH_3 2.22
60 V	7.10	4.42	1.78	0.89, 0.60	
61 K	9.12	4.18	1.86, 1.78*	1.47, 1.31	$\delta\delta'$ 1.72, 1.61, $\epsilon\epsilon'$ 2.98, 2.88, ξNH 7.56
62 G	9.09	4.12, 3.73			
63 Y	7.60	4.00	3.10, 2.76		$\delta\delta'$ 6.88, $\epsilon\epsilon'$ 6.64
64 A	7.86	4.23	1.26		
65 A	7.96	3.97	1.33		

^a The chemical shifts are given in ppm and are relative to the methyl resonance of DSS (sodium-4,4-dimethyl-4-silapentane-1-sulfonate) at 0 ppm. Tentative assignments are marked with an asterisk. The data were obtained at 10 °C and pH 4.8.

Fig. 9. Two regions of a 2D SCUBA-NOESY ^1H NMR spectrum of rQAE m1.1. The spectrum was obtained at 10°C and pH 4.8. The fingerprint (αCH to NH) region and the region of NH-NH connectivities are shown. Cross peaks that arise from sequential αCH -NH connectivities and lead to the sequential assignment of residues 38–49 are labeled and connected with lines with the intrasidue αCH -NH cross peaks of the residues involved. In cases where only weak $d_{\alpha\text{N}}(i, i+1)$ NOEs are observed, the corresponding $d_{\text{NN}}(i, i+1)$ NOEs are labeled in the amide-amide portion of the spectrum.

δCH -protons and the αCH of the preceding residues ($d_{\alpha\delta\delta'}(i, i+1)$). These cross peaks allowed the assignment of the δCH resonances and subsequently facilitated the complete side-chain resonance assignment. Additionally, the observation of these sequential connectivities indicated that the respective X-Pro peptide bonds are in *trans* conformation.

Both V30 and A34 show strong connectivities between their NH resonances and protons resonating at 4.80 ppm, indicating a degeneracy of the P29 αCH and P33 αCH resonances. Whereas only one αCH - βCH cross peak was found in the DQF-COSY, the TOCSY exhibits several connectivities, thus allowing only a tentative assignment of the two spin systems. A strong NOE was observed between T28 αCH and one of the Pro αCH , which was interpreted as a $d_{\alpha\alpha}(i, i+1)$ between T28 and P29. This cross peak, in conjunction with the lack of $d_{\alpha\delta\delta'}$ cross peaks, determined the respective peptide bond to be *cis*. For two reasons P33 is assumed to be in a *trans* conformation. First, no $d_{\alpha\alpha}(i, i+1)$ could be observed between I32 and P33 and secondly, the expected $d_{\alpha\delta\delta'}(32, 33)$ cross peak is degenerate with an intrasidue Pro connectivity due to the proximity of the I32 αCH and the P33 δCH resonance. The P29 and P33 assignments have been confirmed subsequently by the analysis of two mutants, in which either proline was replaced by an alanine (see below).

The resonance assignment of P38 was accomplished by DQF-COSY and TOCSY spectra. The sequential assignment is unambiguous by exclusion as well as by the observation of some weak sequential connectivities. Despite the fact that sequential $d_{\alpha\delta\delta'}(37, 38)$ could not be established due to degeneracy between Ile 37 αCH and P38 δCH , weak Ile 37 NH-Pro 38 $\delta'\text{CH}$ and Pro 38 $\delta'\text{CH}$ -Arg 39 NH cross peaks could be observed. These peaks are the proline equivalent to d_{NN} cross peaks. Because no strong cross peak is present that could be interpreted as a $d_{\alpha\alpha}(i, i+1)$ connectivity between Ile 37 and Pro 38, the peptide bond is assumed to be in a *trans* conformation. The above proline assignments fill all gaps in the sequential assignments except for the Ile 32-Pro 33 sequence, the only two adjacent residues for which unambiguous sequential cross peaks could not be observed.

Secondary structure determination

Secondary structure elements were deduced from distinctive sequential and medium-range NOE patterns (Fig. 10). The β -strands were identified on the basis of strong $d_{\alpha\text{N}}(i, i+1)$ and no (or weak) $d_{\text{NN}}(i, i+1)$ connectivities as well as the lack of any medium range cross peaks. Generally, this coincides with the observation of increased values for NH- αCH coupling constants and positive secondary shifts, i.e., αCH resonances being downfield

Fig. 10. Summary of sequential connectivities, $^3J_{\text{NH}\alpha\text{CH}}$ and amide exchange data. The regular secondary structure elements as deduced from the NMR data are also shown. NOEs involving the NH, αCH , and βCH protons, as well as the δCH protons of proline, are presented as bars; the intensity of the NOE cross peaks is indicated by the heights of the bars. Sequential connectivities involving the C δH of proline are indicated by cross-hatched boxes along the same line as the corresponding connectivities involving NH protons. Slowly exchanging amide protons are indicated by filled circles. The size of the open squares indicates the size of the $^3J_{\text{NH}\alpha\text{CH}}$, grouped into $J = <6$ Hz, 6–8 Hz, >8 Hz. Secondary shifts of αCH protons relative to their respective random coil values (Wishart et al., 1991) are depicted as bar graph. The values given for each residue are an average with the values of its ± 1 neighbors (Pastore & Saudek, 1990).

shifted relative to random coil values (Wishart et al., 1991). Seven β -strands of a length between three and five residues were identified (Fig. 11). These regions, as well as two more sequences, residues P33–A34 and residues M59–V60, exhibit long-range connectivities, in particular, strong $d_{\alpha\alpha}$ and adjacent $d_{\alpha\text{N}}$ cross peaks. In conjunction with the observation of lowered amide exchange rates, these NOEs establish the arrangement of the strands into antiparallel sheets and the concurrent formation of interstrand hydrogen bonds. These secondary structure assignments differ slightly from previous ones (Sönnichsen et al., 1993). The C-terminal strand is defined here from M59 to V60, strictly on the basis of the observed long-range NOEs and hydrogen bonds. Apart from the question of whether or not a two-residue β -strand is meaningful, in structure calculations this strand extends

to K61 as previously reported. Similarly, the region P33–A34 was not defined in the description of the global fold (Sönnichsen et al., 1993). With the use of the mutants P29A and P33A, it was possible to distinguish NOE connectivities to the αCH proton of P29 and P33. This unambiguously assigned a strong NOE as a $d_{\alpha\alpha}$ between P33 and S4. Because a close proximity between αCH -protons is indicative of an antiparallel sheet, P33–A34 was identified as part of the β -sheet structure. To reflect this, Figure 11A and B includes these residues as a short, two-residue strand. Whether this is justified will be resolved when a high-resolution structure becomes available.

Between segments of β -strands, four short regions were found showing medium range and sequential d_{NN} connectivities. For three of these (N8–Q9, I13–N14, L51–G52), the NOE patterns are characteristic for turn conformations, although they do not agree with the classical, regular β -, or γ -turn NOE patterns. Only one residue divides the strands M43–N46 and A48–P50; the conformation can be described as a β -bulge at R47. Some residues within the turns still exhibit long-range NOEs; $d_{\alpha\alpha}$ connectivities were observed for Q9–L51 and I13–R47. Therefore, the backbone conformation of residues A3–T18 and M43–L55 might be best explained as strong twists or bends in two long, paired antiparallel β -strands, leading to a β -bulge and local, turnlike conformations for the residues mentioned above.

Two loop regions were identified on the basis of medium-range cross peaks and the lack of long-range connectivities involving backbone atoms. The first loop, residues V27–I32, includes the *cis* Thr 28–Pro 29 sequence and connects two strands on opposite sides of the first N-terminal strand (Fig. 11). For residues E35–S42, which form the second loop, several $d_{\alpha\text{N}}(i, i+2)$ and $d_{\alpha\text{N}}(i, i+3)$ connectivities were observed, indicating the presence of several turns. Alternatively, these NOE patterns can be interpreted as a 3_{10} -helix or α -helix, which would be supported by the presence of contiguous sequential $d_{\text{NN}}(i, i+1)$ cross peaks and, from residues P38 to S42, by the observation of upfield-shifted αCH resonances. With a length of more than five residues, this segment would be expected to favor an α -helical conformation; however, the presence and intensity of the $d_{\alpha\text{N}}(i, i+2)$ connectivities makes this unlikely.

The sequential assignment of rQAE m1.1 can be transferred easily to rQAE, allowing the separation of the resonances of the two isoforms and their sequential assignment. A detailed comparison of the proton resonances of the three species, the two isoforms of rQAE and the mutant rQAE m1.1, showed that they have almost identical chemical shifts throughout the protein. Exceptions were the C-terminal residues V60–A65/66, i.e., the residues close in sequence to the site of mutation or *cis-trans* isomerization. Additionally, differences in chemical shift were observed for other residues, most notably

A3, S4, I32, L40–V45, and T53–L55 (Fig. 6). These residues were also similarly affected by the isomerization and by the change in the C-terminal sequence.

Mutation of internal prolines

Type III AFPs are moderately rich in proline. The different isoforms contain from five to eight prolyl residues, of which four are conserved in all known sequences (P12, P29, P33, and P57). The conserved prolines may be required for structural reasons, such as governing the stability or folding of the protein. The secondary structure assignments of rQAE m1.1 reinforce the possibility of a structural role. Most of the prolines are observed at or near the end of β -strands, indicating that they are required to terminate the strands or facilitate turn conformations. To test this hypothesis, two further mutants of rQAE m1.1 were produced, in which either P29 or P33 were changed to alanine (P29A and P33A). These conserved prolyl residues were selected for mutation because they are located close to one another in a region that generally shows greater sequence variations within, as well as between, the homology groups of Type III AFPs (Davies & Hew, 1990). Moreover, the T28–P29 peptide bond was observed to be in a *cis* conformation, which further underlines a possible structural importance. An additional benefit expected of these mutations was the clarification of some ambiguous NMR resonance assignments, caused

by the presence of six prolines in rQAE m1.1 and, in particular, by the degeneracy of the α CH resonances of P29 and P33.

In the *E. coli* host, mutant AFPs P29A and P33A were present almost exclusively in inclusion bodies, although small quantities of P33A were recovered from the soluble fraction. The proline to alanine mutations also notably lowered the thermal hysteresis activity of P29A and P33A relative to QAE by 50% and 20%, respectively (Fig. 4). Furthermore, the solubility of P29A decreased significantly, so that its activity was not measurable at concentrations above 0.19 mM.

One-dimensional ^1H NMR spectra were acquired for P29A, P33A, and rQAE m1.1 at 3.5 °C and pH 7.0 (Fig. 12). Under these conditions all exchangeable side-chain protons as well as many backbone amide protons of rQAE m1.1 exchanged with the solvent within 1 h of being dissolved and were not detected any more. The remaining amide resonances were observed with undiminished intensity and did not exchange over several days. These amide protons belong to residues in the β -sheet regions of the protein. Their exchange rates are lowered due to their involvement in hydrogen bonds between the β -strands. The chosen conditions therefore facilitated a simple way to detect any changes in the structure of the protein, looking selectively at changes of the protein backbone by 1D NMR. The spectra of P33A (Fig. 12C) and rQAE m1.1 (Fig. 12B) are very similar. After 2 days, iden-

Fig. 11. Schematic representation of the antiparallel β -sheet structure of Type III AFP. **A:** The arrangement of the polypeptide chain in rQAE m1.1 as deduced from a qualitative analysis of the NMR data. The observed interstrand connectivities between backbone NH and/or α CH are indicated by arrows. Hydrogen bonds deduced from NOE patterns as well as slowly exchanging amide protons are shown as dashed lines. **B:** Schematic presentation defining the secondary structure of rQAE m1.1. The β -strands are presented as arrows. Connecting loops are represented by lines and are not to scale. The rectangle depicts a series of turns or a 3_{10} -helix.

Fig. 12. One-dimensional ^1H NMR spectra of rQAE m1.1 and proline replacement mutants. The region of amide and aromatic protons is shown. Spectra for P29A (A), rQAE m1.1 (B), and P33A (C) were acquired at 3.5 °C in D_2O . Conditions were as described in the Materials and methods. The side-chain protons of Y63 and slowly exchanging amide protons of rQAE m1.1 are labeled with their residue number. Amide protons that exhibit different chemical shifts in P33A are indicated by lines.

tical amide protons remained unexchanged as judged by the similarity of the spectra. This has been confirmed by 2D NMR spectra (not shown), in particular NOESY spectra, in which all connectivities characteristic for the fold of the protein were also observed. Furthermore, the intensities of the amide resonances were comparable between rQAE m1.1 and P33A, indicating similar amide exchange rates for the observed protons. The only differences were changes in the chemical shifts observed for amide protons G31, I32, and A34, close to the site of mutation. The amide protons of S4 to V6, I37, as well as the protons of the Y63 side chain were also slightly perturbed.

Due to the low solubility of P29A, a significantly lower signal-to-noise ratio was obtained in 1D ^1H NMR spectra of this protein (Fig. 12A). The ^1H NMR spectrum was acquired 2–4 h after dissolving the protein in D_2O . In contrast to the other mutants, all amide protons had completely exchanged with the solvent after 8 h (not shown). The amide resonances present in the illustrated spectrum aligned better with the ones in rQAE m1.1 than with those in P33A. However, the assignments of the amide protons have not been confirmed by 2D spectra. In conjunction with the limited solubility of the protein, the protons exchanged too rapidly to be detected in these longer experiments. Nevertheless, 2D spectra were used to assign the nonexchangeable proton resonances of this mutant. Generally, chemical shift values for P29A were found to be very similar to those for rQAE m1.1. Additionally, many NOE connectivities indicative of the protein fold were observed—in particular, the majority of

$d_{\alpha\alpha}$ connectivities, which establish the assembly of the β -strands into sheets. One difference between the spectra of P29A and rQAE m1.1 concerns the resonances observed at 7.06 and 6.90 ppm (Fig. 12A). They align almost perfectly with the amide resonances of V60 and V41 in the spectrum of rQAE m1.1. However, in a 2D-TOCSY spectrum of P29A these resonances are connected by a cross peak, linking them to the same spin system. In conjunction with identical coupling constants being present in both signals, these resonances were assigned as protons of a tyrosyl side chain, thus indicating the presence of a second, minor isoform in P29A.

Discussion

Although QAE isolated from ocean pout serum appeared to be pure by HPLC and N-terminal sequencing, its 1D ^1H NMR spectra showed some heterogeneity. The source of this heterogeneity was not apparent. However, because there are up to 150 copies of the AFP gene in ocean pout coding for the 12 isoforms separable by HPLC, it is possible that minor sequence variants of QAE exist that were not resolved from the main QAE component. Proteolysis or covalent modification of a fraction of the QAE are other explanations for the heterogeneity.

Production of rQAE in *E. coli* resulted in a homogeneous protein but with the addition of an N-terminal methionine. A comparison of the N-terminal sequences of Type III AFPs (Fig. 13) showed there is considerable variation up to S4, the first conserved amino acid. Li et al. (1991) reported that the inclusion of an N-terminal Met-Lys sequence had little effect on the activity of an SP-Sephadex-binding component of the ocean pout AFP. Similarly, the N-terminal extension of rQAE did not significantly change its thermal hysteresis activity when compared to the serum QAE.

The homogeneity of rQAE improved the NMR spectra. This was particularly evident in the aromatic region of the 1D spectra (Fig. 5) where several peaks had been removed. However, some extraneous signals remained. These were correctly attributed to *cis-trans* isomerization of the proline peptide bond adjacent to Y63. The isomerization causes serious problems for NMR structure determination by complicating the sequential assignments when using standard procedures. Moreover, it creates immense difficulties for the determination of accurate distances from NOESY spectra, which are required for a 3D structure determination by NMR (Kördel et al., 1990). A comparison of the sequences for Type III AFP isoforms showed that there was also considerable sequence variability in the C-terminal region (Fig. 13). The last conserved amino acid was Y63. The sequence YAA, used by several other Type III AFP isoforms, was substituted for YPPA to cure the isomerization problem. Conversion of the C-terminal sequence to give rQAE m1.1 had no adverse effects. The protein had comparable activity and solubil-

to the changes in the C-terminal sequence as well. As indicated by the folding pattern (Fig. 11) and shown in the solution structure (Sönnichsen et al., 1993), these regions are in close proximity to the C-terminus in the folded protein. NOEs are observed between Y63 and I32, and L55 and V60, the residues involved therefore being within 5 Å of the C-terminal residues. The remaining affected residues (A3–S4 and L40–V45) are part of the β -sheet structure of the protein; they are within a 10-Å distance of Y63. This suggests that the isomerization as well as a sequence change in the C-terminus slightly affect the position or conformation of Y63. Due to the strong anisotropy and the involvement of the aromatic side chain in the hydrophobic core, this change causes small chemical shift changes for residues within a range of 10 Å, changes that are not representative of structural changes in the fold of the protein but rather indicate changes in the local environment of the aromatic side chain of Y63.

Figure 11 illustrates the arrangement of β -strands into sheets. Two sets of strands contiguous in sequence, from residues A3 to T18 and M43 to L55, arrange to form antiparallel sheets. Bends or turns occur simultaneously in both strands in position N8-Q9/L51-G52 and I13-N14/R47, creating three double-stranded sheets of various lengths. Three more strands fold back onto this structure. An antiparallel triple-stranded sheet results from the arrangement of M59-V60 with V45-N46. Similarly, the β -strand from M22 to V26 assembles with the N-terminal strand (A3-A7) as well as residues P33 and A34, which exhibit backbone-backbone connectivities to the opposite

side of the N-terminal strand. The resulting topology of this sheet structure is best described as an antiparallel triple-stranded sheet, in which one strand is discontinuous (formed by two different sequences). The unusual topology of the N-terminal triple sheet implies that the loop connecting V26 and I32 crosses the N-terminal strand. Because both strands are parallel with respect to each other, and the loop consists of only five residues, the backbone conformation in the loop is expected to be very constrained. This might explain the observation of a *cis* peptide bond in this loop.

The majority of residues in this protein are involved in the β -sheet structure and associated turns, only two short loop regions were found. The protein is therefore expected and has been shown to fold into a single domain (Sönnichsen et al., 1993). This is supported by the observation of several NOE connectivities for the side chains of Y63 to different regions in the protein, e.g., I32, I37, L55, V60. The aromatic side chain seems to be involved in the hydrophobic core of the protein. This compares well with the results of the CD studies (Schrag et al., 1985; Ananthanarayanan et al., 1986), which indicated an asymmetric environment for this side chain and the presence of defined secondary and tertiary structure. However, the CD spectra did not find evidence for the presence of a β -sheet structure, which is in contrast to the observation of extended β -sheet areas in this study. This discrepancy could originate in the fact that all β -strands are rather short.

When the QAE components are aligned and compared (Fig. 13), 23 out of the 31 β -sheet residues are identical, whereas only 13 of the other 35 residues are the same throughout. Thus, the β -sheet regions are better conserved than the turns, loops, and termini. Based on sequence comparisons alone, the equivalent β -sheets can be identified readily in the SP components (Fig. 6). Within this group, 24 of the 31 β -sheet residues are identical. This strongly suggests that the secondary structure of rQAE m1.1 is representative of all the Type III AFPs and that they will take up a similar fold.

Internal proline mutants

Two further mutants were produced in order to delineate the role of internal prolines in the structure and activity of Type III AFP. The two prolines selected for mutation, P29 and P33, are completely conserved in all known Type III AFPs amidst a more variable region of the sequence (Davies & Hew, 1990). Both are located in or near a short loop spanning the N-terminal strand (A3–A7). As judged by the similarity of proton chemical shifts as well as NOESY spectra, in both mutants the overall fold of the protein stays intact. Therefore, the mutant proteins were used in comparing spectral assignments with rQAE m1.1. This resulted in the unambiguous assignment of P29 and P33 resonances in rQAE m1.1 and allowed the distinction

of several NOE connectivities involving these residues. In particular, the assignment of NOEs involving the degenerate P29 α CH and P33 α CH was facilitated, resulting in an improvement of the structural refinement in this region of the protein.

Despite the overall similarity of the NMR data, both mutations caused significant changes in the respective proteins. In the *E. coli* host, both mutants were found almost exclusively in inclusion bodies. This indicated both prolyl residues to be required for the proper protein folding in vivo. After refolding during purification, P33A completely recovered its solubility. In the NMR spectra of P33A, only small changes were observed as compared to rQAE m1.1 (Fig. 12). The slowly exchanging amide protons of several residues, as well as side-chain protons of Y63 are affected by the mutation, exhibiting small changes in chemical shifts. All these residues are close to P/A33 in sequence or space in the folded protein. The shift changes might indicate a slight conformational change around A33 as compared to P33 in rQAE m1.1. Additionally, I32, the residue next to the mutation, is involved in the hydrophobic core of the protein as shown by NOEs between I32 side-chain protons and the side chain of Y63. Thus, slight changes in the conformation of I32 might cause changes in the conformation of Y63. As seen in the case of the C-terminal mutation described above, this can cause small, wide-reaching chemical shift changes. All characteristic spatial connectivities are still present in NOESY spectra of P33A. In particular, a strong NOE observed between S4 α CH and P33 α CH in rQAE m1.1 was similarly observed between S4 α CH and A33 α CH in this mutant. Thus, qualitatively, the loop and the sheet close to the site of mutation were not significantly altered. However, compared to rQAE, the mutant P33A is 20% less active in thermal hysteresis measurements. The sensitivity of the antifreeze activity to minor structural changes might reflect either the involvement of residues that are proximal to P33 in the interaction with ice or a fine tolerance in the spatial organization of distal ice-binding residues, which is upset by a slight conformational change mediated through the hydrophobic core.

The T28–P29 peptide bond in rQAE m1.1 was determined to be in a *cis* conformation. The mutation of this prolyl residue was expected to cause severe changes in P29A because *cis* peptide bonds containing nonprolyl residues are energetically unfavored and are rarely observed in proteins. Indeed, all experimental results indicate that changes occurred upon mutation. P29A was difficult to dissolve. The solubility decreased to at least one tenth of that found for rQAE or rQAE m1.1. As indicated by the observation of two Y63 side-chain spin systems in the NMR spectra of P29A, two isomers are present. The major isomer seems to have a fold similar to rQAE m1.1. The chemical shifts of the aromatic side-chain protons, as well as the appearance of the slowly exchanging amide protons, are comparable to rQAE m1.1 as far as this

can be judged in a spectrum with an inferior signal-to-noise ratio. Very few chemical shift changes could be observed for the amide protons. However, this could reflect the fact that no amide protons close to the site of mutation are observed under these conditions. Secondly, the identities of the observed amide protons were not confirmed by 2D NMR spectra. Nevertheless, the aliphatic region of 2D NOESY and TOCSY spectra is similar to spectra of rQAE m1.1. In particular, NOEs characteristic of the protein fold are still present. The loop involving the mutation connects two β -strands on either side of the N-terminal strand. Several $d_{\alpha\alpha}$ connectivities define this part of the secondary structure in rQAE m1.1. These peaks (S4-P33; A3-V26) are still observed in P29A, indicating that the mutation does not interrupt the fold or arrangement of secondary structure elements in the major isomer. It is therefore very likely that the T28-A29 peptide bond is in a *cis* conformation, although a characteristic NOE between the α CH of these two residues was not observed. Consequently, the minor form observed for P29A appears to be an isomer with a *trans* T28-P29 peptide bond, where the protein avoids the high-energy backbone conformation but interrupts the arrangement of secondary structure elements in this region. However, due to the low percentage of the *trans* isomer in conjunction with the poor solubility of the protein, structural information on the minor isomer was not available.

The 1D ^1H NMR spectrum of P29A (Fig. 12) was acquired 2 h after dissolving the protein in order to retain some amide proton resonances in the spectrum. In contrast to the other proteins, P33A and rQAE m1.1, all amide protons in P29A were completely exchanged after 8 h. The increase in all amide exchange rates indicated that the stability of the fold was significantly decreased. This correlates well with the assumption that the major isoform contains a *cis* peptide bond. A replacement of a proline by any residue favoring a *trans* bond will significantly reduce the free energy of folding if the peptide bond assumes a *cis* conformation in the folded protein. Interestingly, P29A exhibits not only a decrease in protein stability, but also a twofold decrease in activity. Although it might be surprising that this mutation does not affect the antifreeze activity more severely, the intact global fold of P29A and a comparable structure at the site of the mutation would argue for a similar activity to P33A.

Both internal mutants have caused interesting changes in the protein structure, stability, and antifreeze activity. It will be important to study these mutants more closely to delineate the reasons for these changes.

Materials and methods

AFP gene synthesis

A gene coding for the QAE-Sephadex-binding component of ocean pout AFP (rQAE) was assembled from five

pairs of complementary oligonucleotides (Oligonucleotide Synthesis Laboratory, Queen's University) (Fig. 1A). The open reading frame of this gene corresponds to the HPLC-12 sequence (Davies & Hew, 1990) with the addition of an initiating methionine. The nucleotide sequence of the synthetic gene was chosen to incorporate unique restriction sites at intervals and accommodate commonly used *E. coli* codons (de Boer & Kastelein, 1986). All but the two outer oligonucleotides were phosphorylated at their 5' ends. Complementary oligonucleotide pairs (50 pmol each) were annealed separately in 50 mM Tris-HCl (pH 7.6), 10 mM MgCl_2 (15 μL) during cooling from 95 to 23 $^\circ\text{C}$ over 3 h. The five pairs were combined and were ligated overnight at 14 $^\circ\text{C}$ using 2 μL of T4 DNA ligase in 100 μL of 50 mM Tris-HCl (pH 7.6), 10 mM MgCl_2 , 1 mM ATP, 1 mM DTT, 5% PEG 8000 (w/v) (Sambrook et al., 1989). The ligation products (0.6 pmol) were mixed with 0.6 pmol of *Nde*I/*Hind*III-digested pT7-7 vector (Tabor, 1990), ligated overnight at 14 $^\circ\text{C}$, and transformed into *E. coli* JM 83. Putative clones were checked by restriction analysis and sequencing by the di-deoxy method (Sanger et al., 1977).

AFP expression

Plasmid from positive clones was prepared by alkaline lysis and PEG precipitation (Sambrook et al., 1989) and was transformed into *E. coli* K 38 that contained the vector pGP1-2 (Tabor & Richardson, 1985). Expression of AFP was induced according to Tabor (1990). Harvested cells from a 3-L culture were broken by sonication in 10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 0.1 mM PMSF (50 mL).

AFP purification

The *E. coli* sonicate was centrifuged at $12,000 \times g$ for 30 min at 4 $^\circ\text{C}$. The supernatant fraction containing soluble AFP (50 mL) was chromatographed on a G-75 Sephadex column (50 $\times$ 1,000 mm) in sonication buffer at a flow rate of 1.5 mL/min. Fractions containing rQAE were pooled (200 mL), acidified to pH 3.7 with HCl, clarified by centrifugation at $10,000 \times g$ for 10 min, and loaded at a flow rate of 1.5 mL/min onto an FPLC cation-exchange column (S-Sepharose HP 16/10) pre-equilibrated in 50 mM Na acetate (pH 3.7). Bound AFP was eluted by a linear 0–0.3 M NaCl gradient in 50 mM Na acetate (pH 3.7) at a flow rate of 2.5 mL/min. Fractions (2.5 mL) containing AFP were pooled and applied to a C18 reversed-phase semipreparative HPLC column (Vydac 10 $\times$ 250 mm). The column was equilibrated in 10% organic buffer (acetonitrile/isopropanol 2:1, v/v)/90% H_2O , containing 0.1% TFA. The rQAE was eluted by a gradient of acetonitrile/isopropanol at a flow rate of 3 mL/min.

Insoluble AFP was extracted as inclusion bodies from the *E. coli* sonicate pellet. The pellet was resuspended in 4 mL of TEN buffer. Each milliliter of the resuspended pellet was layered onto a 40% (w/v) sucrose cushion (10 mL) and centrifuged at $13,000 \times g$, for 30 min at 4 °C. The resultant pellet was then washed in TEN buffer (2 mL) containing 1.0% Triton X-100 for 30 min on ice. Partially pure inclusion bodies were collected again by centrifugation ($10,000 \times g$, for 10 min at 4 °C). The AFP was extracted from this pellet by resuspension in 10 mL of organic extraction buffer: 70% acetonitrile/isopropanol (2:1, v/v)/30% H₂O, containing 0.1% TFA. The mixture was gently swirled for 30 min at 22 °C. Solubilized AFP was recovered after removal of insoluble material by centrifugation ($16,000 \times g$ for 10 min at 4 °C). The organic solvent content of the supernatant was reduced to $\leq 5\%$ v/v by dilution with water. When necessary, the volume was reduced by lyophilization. The sample was then chromatographed by cation-exchange FPLC after pH adjustment and clarification as described above.

The QAE-Sephadex-binding AFP (QAE) from ocean pout serum (2 mL) was purified as described previously (Li et al., 1985) with the following modifications. The AFP-containing fractions from the G-75 Sephadex column were applied to a CM-52 (Whatman) column (25 $\times$ 200 mm) in 10 mM Tris-HCl (pH 8.0). The unbound fraction, which contained mostly QAE, was chromatographed on reversed-phase HPLC as described above.

Analysis of AFP

AFP samples were analyzed by electrophoresis on an 18% SDS-polyacrylamide gel (Laemmli, 1970) stained with both Coomassie blue (0.25%) and amido black (0.1%). The identity of the AFP was confirmed by amino acid composition analysis and N-terminal sequencing. Measurements of the antifreeze activity were made in 100 mM NH₄HCO₃ (pH 7.9) using a nanoliter osmometer (Clifton Technical Physics, Hartford, New York) as described by Chakrabarty and Hew (1991). Wells in a silver sample holder were filled with mineral oil and placed on a thermoelectrically controlled microscope stage. A single drop-let of AFP solution was placed in a well and rapidly frozen. Subsequently, the temperature was raised and the sample thawed until only a single crystal remained. Antifreeze activity was defined as thermal hysteresis, the difference between the temperatures at which the seed crystal would grow or shrink.

Cassette mutagenesis

The C-terminal sequence of rQAE, YPPA, was changed to YAA by cassette mutagenesis. The section of the synthetic gene in pT7-7 between *Bst*EII and *Eco*RI was replaced by the oligonucleotides shown in Figure 1B.

Oligonucleotide annealing and ligation to the vector were done as described previously. Mutants derived from transformation of *E. coli* JM83 were identified by *Afl*III digestion and confirmed by dideoxy sequencing.

Primer-directed mutagenesis

Mutants P29A and P33A were made by primer-directed mutagenesis (Kunkel et al., 1987). To generate a single-stranded, uracil-containing template, the rQAE m1.1 gene was excised from the pT7-7 vector by *Xba*I and *Eco*RI digestion, inserted into the phagemid pTZ-19r (Mead et al., 1986), and transfected into *E. coli* CJ236. Single-stranded DNA template was produced using the helper phage M13 K07, recovered by standard procedures (Mead et al., 1986; Kunkel et al., 1987) and further purified on a hydroxyapatite column (Richards, 1989). Mutagenic primer 5'-pGAAGTTGTTACTGCAGTTGGTATCCCCG-3' (for P29A) or 5'-pGGTTGGTATCGCGCCGAAG-3' (for P33A) (20 pmol) and single-stranded DNA template (1 μ g) were combined in 20 μ L of annealing buffer (20 mM Tris-HCl [pH 7.5], 50 mM NaCl, 2 mM MgCl₂). The mixture was placed in a 95 °C water bath that was cooled to 22 °C over 2 h. Following the addition of 10 mM dNTPs (2.5 μ L each), 10 mM ATP (5 μ L), synthesis buffer (100 mM Tris-HCl [pH 8.0], 50 mM MgCl₂, 20 mM DTT [5 μ L]), and water (4 μ L), the mixture was left on ice for 5 min. At the end of this incubation, 1 μ L each of gene 32 protein (3 μ g), T4 DNA polymerase (6.4 U), and T4 DNA ligase (4 U) were added sequentially. The mixture was incubated on ice for a further 5 min, then at 25 °C for 5 min, and finally at 37 °C for 2 h. Twenty microliters of this incubation mixture was used to transform *E. coli* K38. Mutants were identified by dideoxy DNA sequencing. The mutant rQAE m1.1 (P33A and P29A) genes were returned to the pT7-7 vector following removal from pTZ-19r by *Nde*I/*Eco*RI digestion.

NMR experiments

All experiments were carried out on 0.5-mL samples in a Varian Unity 600 spectrometer at a ¹H frequency of 600 MHz. For 1D spectra at room temperature, the different AFPs were dissolved in 100% D₂O to yield 1 mM (QAE, rQAE m1.1), and 0.1 mM (rQAE) solutions with pH values adjusted to 7.0. For spectra at 3.5 °C, samples of rQAE m1.1, P33A, and P29A (which were previously lyophilized from aqueous solutions), were redissolved in precooled D₂O in an ice bath. The pH was adjusted to 7.0. While the concentrations for rQAE m1.1 and P33A were 1 mM, a sample of ~ 0.2 mM was prepared for P29A owing to its limited solubility. For structure determination, NMR experiments were carried out with 1 mM rQAE m1.1 dissolved in either 100% D₂O or 80% H₂O/20% D₂O solution at various pH values between 4.8 and 8.0 (not corrected for deuterium isotope effects).

and at temperatures between 10 and 25 °C. All 1D spectra were usually acquired using 30,016 data points, a spectral width of 7,500 Hz, 256–1,024 transients, with a 2.0-s presaturation delay and a 90° pulse width of 10.8 μ s. The data were processed with zero filling to 64K points and with exponential line broadening of 0.5 Hz. All 2D experiments were acquired in the phase-sensitive mode using the method of States et al. (1982), in 256–400 t_1 increments, 32–64 scans per increment, with a presaturation delay of 2.0 s, and 2,048 data points per scan with a spectral width of 7,500 Hz. The DQF-COSY spectra (Piantini et al., 1982; Rance et al., 1983), NOESY spectra with mixing times of 50-, 75-, 100-, and 150-ms (Jeener et al., 1979; Kumar et al., 1980; Macura et al., 1981), and TOCSY spectra (Braunschweiler & Ernst, 1983; Davis & Bax, 1985) were acquired, the latter using isotropic mixing times between 50 and 80 ms and an MLEV-17 pulse sequence (Bax & Davis, 1985) to produce a spin lock field of 7.3 kHz. Additionally, NOESY and DQF-COSY spectra were obtained using a SCUBA-pulse sequence (Brown et al., 1988) with a relaxation delay of 60–110 ms inserted immediately after the presaturation delay to allow for the recovery of magnetization of protons that resonate at or close to the decoupler frequency. In these NOESY spectra, a 180° pulse was applied at the center of the mixing time in order to prevent the recovery of solvent magnetization. The 2D data were processed on a SUN-IPC workstation using the Varian VNMR software package. The data were zero filled routinely to 4K $\times$ 4K data points, and shifted sinebells were used for resolution enhancement in both dimensions. Coupling constants were determined using DQF-COSY spectra at 5–25 °C. For this purpose the DQF-COSY data were processed 16K $\times$ 2K using a shifted sinebell weighting function. Traces of NH- α CH cross peaks were obtained along F2. The antiphase signals were then fitted to two Lorentzian functions. These calculations were performed using the program Jfit, written by R. Boyko, University of Alberta.

Determination of slowly exchanging amides

Protein, which was lyophilized from aqueous solution, was redissolved in ice-cold D₂O and the pH adjusted from approximately 4 to 7.0. The sample was warmed to 10 °C and kept at this temperature. Within 1 h all exchangeable side-chain protons and the majority of amide protons completely exchanged with the solvent. Amide resonances, which were still present, were undiminished in intensity over a period of several days at this temperature.

Acknowledgments

We thank Claude Lazure for amino acid composition and protein sequencing, Garth Fletcher for the gift of ocean pout serum, Stan Tabor for the expression system, Scott Bunnell for participating in the P33A mutagenesis, Sherry Gauthier and Gerry

McQuaid for technical assistance, and Choy Hew for helpful discussions. This work was supported by research grants from the Medical Research Council of Canada (B.D.S., P.L.D.), the Protein Engineering Network of Centres of Excellence (B.D.S., F.D.S.), and the Alberta Heritage Foundation for Medical Research (F.D.S.). H.C. is the recipient of an E.G. Baumann predoctoral studentship.

References

- Ananthanarayanan, V.K., Slaughter, D., & Hew, C.L. (1986). Antifreeze proteins from the ocean pout, *Macrozoarces americanus*: Circular dichroism spectral studies on the native and denatured states. *Biochim. Biophys. Acta* 870, 154–159.
- Bax, A., & Davis, D.G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* 65, 355–360.
- Braunschweiler, L. & Ernst, R.R. (1983). Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J. Magn. Reson.* 53, 521–528.
- Brown, S.C., Weber, P.L., & Mueller, L. (1988). Toward complete ¹H NMR spectra in proteins. *J. Magn. Reson.* 77, 166–169.
- Chakrabarty, A. & Hew, C.L. (1991). The effect of enhanced α -helicity on the activity of a winter flounder antifreeze polypeptide. *Eur. J. Biochem.* 202, 1057–1063.
- Cheng, C.-H.C. & DeVries, A.L. (1989). Structure of antifreeze peptides from the Antarctic eel pout, *Austrolyctichthys brachycephalus*. *Biochim. Biophys. Acta* 997, 55–64.
- Chou, K.-C. (1992). Energy-optimized structure of antifreeze protein and its binding mechanism. *J. Mol. Biol.* 223, 509–517.
- Davies, P.L. & Hew, C.L. (1990). Biochemistry of fish antifreeze proteins. *FASEB* 4, 2460–2468.
- Davies, P.L., Hew, C.L., & Fletcher, G.L. (1988). Fish antifreeze proteins: Physiology and evolutionary biology. *Can. J. Zool.* 66, 2611–2617.
- Davis, D.G. & Bax, A. (1985). Assignment of complex ¹H NMR spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy. *J. Am. Chem. Soc.* 107, 2820–2821.
- de Boer, H. & Kastelein, R.A. (1986). Biased codon usage: An exploration of its role in optimization of translation. In *Maximizing Gene Expression* (Reznikoff, W.S. & Gold, L., Eds.), pp. 225–285. Butterworth-Heinemann, Stoneham, Massachusetts.
- DeVries, A.L. (1983). Antifreeze peptides and glycopeptides in cold-water fishes. *Annu. Rev. Physiol.* 45, 245–260.
- DeVries, A.L. (1984). Role of glycopeptides and peptides in inhibition of crystallization of water in polar fishes. *Philos. Trans. R. Soc. Lond. B* 304, 575–588.
- Ewart, K.V. & Fletcher, G.L. (1990). Isolation and characterization of antifreeze proteins from smelt (*Osmerus mordax*) and Atlantic herring (*Clupea harengus harengus*). *Can. J. Zool.* 68, 1652–1658.
- Feeney, R.E. & Yeh, Y. (1978). Antifreeze proteins from fish bloods. *Adv. Protein Chem.* 32, 191–282.
- Fletcher, G.L., Hew, C.L., Li, X., Haya, K., & Kao, H.M. (1985). Year-round presence of high level of plasma antifreeze peptide in a temperate fish, ocean pout (*Macrozoarces americanus*). *Can. J. Zool.* 63, 488–493.
- Grathwohl, C. & Wüthrich, K. (1976). The X-Pro peptide bond as an NMR probe for conformation studies of flexible linear peptides. *Biopolymers* 15, 2025–2041.
- Hayes, P., Scott, G.K., Ng, N.F.L., Hew, C.L., & Davies, P.L. (1989). Cysteine-rich type II antifreeze protein precursor is initiated from the third AUG codon of its mRNA. *J. Biol. Chem.* 264, 18761–18767.
- Hew, C.L., Slaughter, D., Joshi, S.B., Fletcher, G.L., & Ananthanarayanan, V.S. (1984). Antifreeze polypeptides from the Newfoundland ocean pout, *Macrozoarces americanus*: Presence of multiple and compositionally diverse components. *J. Comp. Physiol. B* 155, 81–88.
- Hew, C.L., Wang, N.C., Joshi, S., Fletcher, G.L., Scott, G.K., Hayes, P.H., Buettner, B., & Davies, P.L. (1988). Multiple genes provide the basis for antifreeze protein diversity and dosage in the ocean pout, *Macrozoarces americanus*. *J. Biol. Chem.* 263, 12049–12055.
- Hew, C.L. & Yang, D.S.C. (1992). Protein interaction with ice. *Eur. J. Biochem.* 203, 33–42.

- Jeener, J., Meier, B.H., Bachmann, P., & Ernst, R.R. (1979). Investigation of exchange processes by two dimensional NMR spectroscopy. *J. Chem. Phys.* 71, 4546-4553.
- Jorgensen, H., Mori, M., Matsui, H., Kanaoka, M., Yanagi, H., Yabasaki, Y., & Kikuzono, Y. (1993). Molecular dynamics simulation of winter flounder antifreeze protein variants in solution: Correlation between side chain spacing and ice lattice. *Protein Eng.* 6, 19-27.
- Knight, C.A., Cheng, C.C., & DeVries, A.L. (1991). Adsorption of α -helical antifreeze peptides on specific ice crystal surface planes. *Biophys. J.* 59, 409-418.
- Knight, C.A., Driggers, E., & DeVries, A.L. (1993). Adsorption to ice of fish antifreeze glycopeptides. *Biophys. J.* 64, 252-259.
- Kördel, J., Forsén, S., Drakenberg, T., & Chazin, W.J. (1990). The rate and structural consequences of proline *cis-trans* isomerization in calbindin D_{9k}: NMR studies of the minor (*cis*-Pro43) isoform and the Pro43Gly mutant. *Biochemistry* 29, 4400-4409.
- Kumar, A., Ernst, R.R., & Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton relaxation networks in biological macromolecules. *Biophys. Res. Commun.* 95, 1-6.
- Kunkel, T.A., Roberts, J.D., & Zarkour, R.A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* 154, 367-382.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
- Li, X., Trinh, K.Y., & Hew, C.L. (1991). Expression and characterization of an active and thermally more stable recombinant antifreeze polypeptide from ocean pout, *Macrozoarces americanus*, in *Escherichia coli*: Improved expression by the modification of the secondary structure of the mRNA. *Protein Eng.* 4, 995-1002.
- Li, X.M., Trinh, K.Y., Hew, C.L., Buettner, B., Baezinger, J., & Davies, P.L. (1985). Structure of an antifreeze polypeptide and its precursor from the ocean pout, *Macrozoarces americanus*. *J. Biol. Chem.* 260, 12902-12909.
- Macura, S., Huang, Y., Suter, D., & Ernst, R.R. (1981). Two-dimensional chemical exchange and cross-relaxation spectroscopy of coupled nuclear spins. *J. Magn. Reson.* 43, 259-281.
- Mead, D.A., Szczesna-Skorupa, E., & Kemper, B. (1986). Single-stranded DNA 'blue' T7 promoter plasmids: A versatile tandem promoter system for cloning and protein engineering. *Protein Eng.* 1, 67-74.
- Ng, N.F., Trinh, K.Y., & Hew, C.L. (1986). Structure of an antifreeze polypeptide precursor from the sea raven, *Hemitripterus americanus*. *J. Biol. Chem.* 261, 15690-15695.
- Ng, N.F.L. & Hew, C.L. (1992). Structure of an antifreeze polypeptide from the sea raven. *J. Biol. Chem.* 267, 16069-16075.
- Otting, G., Liepinsh, E., & Wüthrich, K. (1991). Protein hydration in aqueous solution. *Science* 254, 974-980.
- Otting, G. & Wüthrich, K. (1989). Studies of protein hydration in aqueous solution by direct NMR observation of individual protein-bound water molecules. *J. Am. Chem. Soc.* 111, 1871-1875.
- Pastore, A. & Saudek, V. (1990). The relationship between chemical shift and secondary structure in proteins. *J. Magn. Reson.* 90, 165-176.
- Piantini, U., Sørensen, O.W., & Ernst, R.R. (1982). Multiple quantum filters for elucidating NMR coupling networks. *J. Am. Chem. Soc.* 104, 6800-6801.
- Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., & Wüthrich, K. (1983). Improved spectral resolution in COSY ¹H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* 117, 479-485.
- Raymond, J.A., Wilson, P., & DeVries, A.L. (1989). Inhibition of growth of nonbasal planes in ice by fish antifreezes. *Proc. Natl. Acad. Sci. USA* 86, 881-885.
- Richards, E.J. (1989). Chromatography of nucleic acids. In *Current Protocols in Molecular Biology* (Ausubel, F.A., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., & Struhl, K., Eds.), pp. 2.10.1-2.10.3. Greene Publishing & Wiley-Interscience, New York.
- Sambrook, J., Fritsch, E.F., & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sanger, F., Nicklen, S., & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
- Schrag, J.D., Cheng, C.-H.C., Panico, M., Morris, H.R., & DeVries, A.L. (1987). Primary and secondary structure of antifreeze peptides from Arctic and Antarctic zoarcid fishes. *Biochim. Biophys. Acta* 915, 357-370.
- Scott, G.K., Davies, P.L., Shears, M.A., & Fletcher, G.L. (1987). Structural variations in the alanine-rich antifreeze proteins of the Pleuronectinae. *Eur. J. Biochem.* 168, 629-633.
- Scott, G.K., Hayes, P.H., Fletcher, G.L., & Davies, P.L. (1988). Wolf-fish antifreeze protein genes are primarily organized as tandem repeats that each contain two genes in inverted orientation. *Mol. Cell. Biol.* 8, 3670-3675.
- Slaughter, D., Fletcher, G.L., Ananthanarayanan, V.S., & Hew, C.L. (1981). Antifreeze proteins from the sea raven *Hemitripterus americanus*: Further evidence for diversity among fish polypeptide antifreezes. *J. Biol. Chem.* 256, 2022-2026.
- Sönnichsen, F.D., Sykes, B.D., Chao, H., & Davies, P.L. (1993). The nonhelical structure of antifreeze protein type III. *Science* 259, 1154-1157.
- States, D.J., Haberkorn, R.A., & Ruben, D.J. (1982). A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J. Magn. Reson.* 48, 296-292.
- Tabor, S. (1990). Expression using the T7 RNA polymerase/promotor system. In *Current Protocols in Molecular Biology* (Ausubel, F.A., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., & Struhl, K., Eds.), pp. 16.2.1-16.2.11. Greene Publishing & Wiley-Interscience, New York.
- Tabor, S. & Richardson, C.C. (1985). A bacteriophage T7 RNA polymerase/promotor system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. USA* 82, 1074-1078.
- Wen, D. & Laursen, R.A. (1992a). A model for binding of an antifreeze polypeptide to ice. *Biophys. J.* 63, 1659-1671.
- Wen, D. & Laursen, R.A. (1992b). Structure-function relationships in an antifreeze polypeptide. *J. Biol. Chem.* 267, 14102-14108.
- Wishart, D.S. & Sykes, B.D. (1993). Chemical shifts as a tool for structure determination. *Methods Enzymol.*, in press.
- Wishart, D.S., Sykes, B.D., & Richards, F.M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* 222, 311-333.
- Wüthrich, K. (1986). NMR of proteins and nucleic acids. In *NMR of Proteins and Nucleic Acids*. Wiley, New York.
- Yang, D.S.C., Sax, M., Chakrabarty, A., & Hew, C.L. (1988). Crystal structure of an antifreeze polypeptide and its mechanistic implications. *Nature* 333, 232-237.