The Solution Structure of Type II Antifreeze Protein Reveals a New Member of the Lectin Family

Wolfram Gronwald,[‡] Michèle C. Loewen,[§] Bruce Lix,[‡] Andrew J. Daugulis,[||] Frank D. Sönnichsen,[⊥] Peter L. Davies,[§] and Brian D. Sykes*[,‡,#]

Protein Engineering Network of Centres of Excellence, University of Alberta, Edmonton, Alberta T6G 2S2, Canada, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada, Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, Ohio 44106-4970, Department of Biochemistry, Queen's University, Kingston, Ontario K7L 3N6, Canada, and Department of Chemical Engineering, Queen's University, Kingston, Ontario K7L 3N6, Canada

Received November 12, 1997; Revised Manuscript Received January 20, 1998

ABSTRACT: A recombinant form of the sea raven type II antifreeze protein (SRAFP) has been produced using the *Pichia pastoris* expression system. The antifreeze activity of recombinant SRAFP is indistinguishable from that of the wild-type protein. The global fold of SRAFP has been determined by two-dimensional ^{1}H homonuclear and three-dimensional ^{1}H−{^{15}N} heteronuclear NMR spectroscopy using 785 NOE distance restraints and 47 angular restraints. The molecule folds into one globular domain that consists of two helices and nine β-strands in two β-sheets. The structure confirms the proposed existence of five disulfide bonds. The global fold of SRAFP is homologous to C-type lectins and pancreatic stone proteins, even though the sequence identity is only approximately 20%.

Type II AFP[1] is one of four different antifreeze proteins (AFPs) isolated from fish (*1*). Type I AFPs are short (30−50 amino acids), alanine-rich, α-helical peptides found in flounder and sculpin (*2*). The globular type II AFPs, isolated from sea raven (SRAFP) (*Hemitripterus americanus*), herring (*Clupea harengus harengus*), and smelt (*Osmerus merdox*), are all greater than 120 amino acids in length and contain five disulfide bridges (*3*−*5*). Type III AFPs are globular proteins of intermediate size (>60 amino acids), rich in β-structure (*6*−*8*), found in eel pout and wolfish. Type IV AFP, recently isolated from shorthorn sculpin, is 108 amino acids in length, shows homology to the N-terminal domain of a human apolipoprotein, and is proposed to have a four-helix bundle structure (*9*). These AFPs protect fish from freezing in icy seawater by depressing the serum freezing point (*10*). Despite their structural diversity, these proteins are all believed to function in the same noncolligative manner, by binding to the surface of ice crystals and preventing their growth (*11*).

AFP types I and III are the only AFPs for which detailed structural information is available (*7*, *8*, *12*). This information has, in both cases, led to the development of models depicting the specific mode of binding to ice. The dissimilarities in the two proteins make it difficult to determine if there is a common theme used to achieve the same function. This emphasizes the importance of determining the precise three-dimensional (3D) structures of all these different AFPs and comparing their different modes of binding in completely understanding the mechanism of action of this unique group of proteins.

Three-dimensional models have been published for both the sea raven (*13*) and herring (*14*) type II AFPs. Both of these theoretical structures were derived by homology modeling of the carbohydrate recognition domain (CRD) of Ca^{2+}-dependent (C-type) lectins. These CRDs are part of a widespread superfamily of proteins that bind sugar specifically through contact with Ca^{2+} (*15*). The structures of two CRD domains have been solved by X-ray crystallographic analysis, including the CRD from rat mannose-binding protein (*16*) and E-selectin (*17*). SRAFP was modeled on the CRD of the rat mannose-binding protein, with which it shares 19% amino acid identity (*13*). Herring type II AFP was modeled on both CRD structures (*14*). As well, structural analysis of a number of other domains, including the hyaluronate-binding proteoglycan tandem repeat domain of link protein (*18*) and the pancreatic stone protein (PSP), lithostathine (*19*, *20*), indicate they also maintain the same overall fold as the CRD. Interestingly, PSP is known to bind to, and inhibit the growth of, calcium carbonate crystals (*21*), which is a function very similar to that of the AFPs.

The model of sea raven AFP indicates a fold virtually identical to that observed for the lectin CRDs, including conservation of two α-helices and two β-sheets and a high proportion of amino acids in loops and turns (*13*). One feature of the model was a revision of the disulfide bonding

* To whom correspondence should be addressed.
[‡] Protein Engineering Network of Centres of Excellence, University of Alberta.
[§] Department of Biochemistry, Queen's University.
[||] Department of Chemical Engineering, Queen's University.
[⊥] Case Western Reserve University.
[#] Department of Biochemistry, University of Alberta.

[1] Abbreviations: AFP, antifreeze protein; CRD, carbohydrate recognition domain; DQF-COSY, double quantum filtered correlation spectroscopy; DSS, 2,2-dimethyl-2-silapentane-5-sulfonic acid; DTT, dithiothreitol; HSQC, heteronuclear single quantum coherence; NOESY, nuclear Overhauser effect spectroscopy; PSP, pancreatic stone protein; RMBPA, rat mannose-binding protein; rSRAFP, recombinant sea raven antifreeze protein; SRAFP, sea raven antifreeze protein; TOCSY, total correlation spectroscopy.

S0006-2960(97)02788-8 CCC: $15.00
Published on Web 03/13/1998

arrangement to include Cys7−Cys18, Cys35−Cys125, Cys69−Cys100, Cys89−Cys111, and Cys101−Cys117 (*13*). While this disulfide map differs from that previously proposed on the basis of peptide mapping (*4*), it conserves the two bridges identified in the lectin structures. An ice-binding surface was also proposed in the sea raven model, comprising residues that correspond to the Ca^{2+}- and sugar-binding residues of the CRDs. While the herring and smelt isoforms require Ca^{2+} for activity, the sea raven AFP is Ca^{2+}-independent.

In this paper, we present the production and purification of a recombinant form of SRAFP (rSRAFP) using the *Pichia pastoris* expression system. The rSRAFP maintains activity and properties identical to those of wild-type AFP isolated from sea raven serum. With the application of fermentation techniques to this sytem, sufficient ^{15}N-isotopically labeled rSRAFP was obtained. We were then able to use heteronuclear, multidimensional NMR spectroscopy to assign backbone ^{1}HN, ^{15}N, and ^{1}H signals for 119 of the 129 residues, as well as 70% of the side chain ^{1}H resonances, and to determine the three-dimensional solution structure of SRAFP. This structure tests the proposed model and disulfide map, and provides a more detailed view of the proposed ice-binding surface. This will, in turn, allow us to better study the binding of this AFP to ice.

MATERIALS AND METHODS

Expression of rSRAFP. Recombinant type II AFP was produced using the methylotrophic yeast, *P. pastoris*, expression system (Invitrogen, San Diego, CA), essentially as described in Loewen et al. (*22*). The cDNA for SRAFP (*23*) was cloned downstream of the yeast α-mating factor signal peptide, under the control of the yeast alcohol oxidase promoter (AOX1), in the pPIC9 expression vector. A six-amino acid His tag was added at the C-terminal end to facilitate recovery of the secreted AFP from the medium using affinity chromatography. This cDNA construct (pPIC9−SRm−CTHT) was transformed into *P. pastoris* (GS115) by electroporation (*24*) and was integrated into the host genome by homologous recombination. It was targeted to replace the yeast alcohol oxidase gene site so that positive transformants were detected by a slow growth phenotype on medium containing methanol as the only carbon source (*22*). Positive transformants were tested for production and secretion of rSRAFP by shake flask induction. Cultures were grown from single colonies in 50 mL conical tubes at 30 °C, for 48 h in buffered minimal glycerol media. Cells were harvested by centrifugation and resuspended in the same volume of buffered minimal methanol media, which contained methanol for induction of the alcohol oxidase promoter, and again incubated for 48 h at 30 °C with shaking. Cells were removed by centrifugation, leaving rSRAFP in the medium. See Loewen et al. (*22*) for media recipes.

^{15}N Isotopic Labeling of rSRAFP. A single positive transformant known to produce rSRAFP was grown in 10 mL of BMGY in a 50 mL conical tube at 30 °C with shaking for 24 h. This was used to innoculate a 100 mL culture (in a 250 mL Erlenmeyer flask) of MG media containing 3.6 g/L yeast nitrogen base (Difco, Detroit) without amino acids or ammonium sulfate and 10 g/L ^{15}N-labeled ammonium sulfate (CIL, Andover, MA). This culture was grown at 30 °C, with shaking at 250 rpm for 48 h, and then used to innoculate a 1.5 L Bioflo model C-30 fermentor (New Brunswick Scientific, Edison, NJ), containing 1 L of fermentation medium (*22*) with [^{15}N]ammonium sulfate as the sole nitrogen source. Cell culture conditions were set at 30 °C, 800 rpm, and 2.3 L of air per liter of culture per minute. The pH was maintained at 5.5 by the automatic addition of 5 M KOH. As described previously, a mixed-feed fermentation strategy was used to induce expression and secretion of the uniformly ^{15}N-labeled rSRAFP (*22*). Following completion of the fermentation, cells were removed by centrifugation and the labeled rSRAFP was purified from the supernatant medium.

Purification of Type II AFP. SRAFP-containing medium was saturated with EDTA [20 mL of 0.5 M EDTA (pH 9.0) in 100 mL of media]. The pH was increased to 8.0−8.5 with the addition of a 1 M Tris base. Residual EDTA was then saturated by the addition of excess $MgCl_2$ (20 mL of 2 M $MgCl_2$ in 100 mL of media), and more 1 M Tris base was added to maintain the pH. The resulting solution was centrifuged at 10 000 rpm for 20 min at 4 °C to remove precipitate prior to loading onto a nickel−agarose column (10 × 50 mm) (Novogen, Madison, WI). The loaded column was washed extensively with buffer N [20 mM Tris-HCl (pH 7.6), 0.5 M NaCl, 2% glycerol, and 5 mM imidazole] prior to elution of rSRAFP by an imidazole gradient (5 to 125 mM) over the course of 75 min at a flow rate of 1.5 mL/min. Fractions containing rSRAFP were pooled, dialyzed [using 6000−8000 molecular weight cutoff (MWCO) dialysis membrane (Fisherbrand)] against exchange buffer [25 mM Tris-HCl (pH 9.0)], and subsequently chromatographed on an FPLC (Pharmacia Biotech, Uppsala, Sweden) anion exchange column [Q-Sepharose 16/10 (16 × 100 mm), Pharmacia] pre-equilibrated in exchange buffer. Bound AFP was eluted by a linear 0.2 to 0.4 mM NaCl gradient in exchange buffer at a flow rate of 2.5 mL/min. Fractions containing AFP were pooled, dialyzed into 100 mM NH_4HCO_3, and lyophilized.

Analysis of rSRAFP. Samples were analyzed by electrophoresis on 17% polyacrylamide−SDS gels, containing 0.1 M sodium phosphate and 4 M urea at pH 6.8. These gels either were stained with Coomassie blue (0.25%) and amido black (0.1%) and then destained (10% methanol and 10% ethanol) or were Western blotted onto a nylon membrane (PVDF, NEN). The membrane was incubated with rabbit anti-SRAFP antiserum and then horseradish peroxidase-linked goat anti-rabbit IgG (Promega, Madison, WI). Detection was carried out using enhanced chemiluminescence (Amersham Canada, Oakville, ON). A_{280nm} (using the predetermined standard for type II AFP of 4.5 mg/mL = 10 optical density units) and Bradford protein detection (*25*) were used to quantify purified AFP.

Antifreeze Activity Measurements and Photomicroscopy. Thermal hysteresis was used as a quantitative measure of AFP activity. It is defined as the temperature difference (degrees Celsius) between the melting point and the non-equilibrium freezing point of a solution. This thermal hysteresis activity is the result of inhibition of crystal growth by the AFP (*11*). Thermal hysteresis can be measured by placing nanoliter volumes (1−10 μL) of an AFP solution on a cooling stage beneath a microscope for observation of

ice crystals. The temperature of the stage is controlled by a nanoliter osmometer (Clifton Technical Physics, Hartford, NY) (1 osM = 1.86 °C) (*26*). The growth of a single ice crystal in the presence of AFP can then be observed as the temperature is dropped from the melting point (where ice and its melt are at equilibrium) to the nonequilibrium freezing point (rapid ice growth). All measurements were carried out in 100 mM NH_4HCO_3, except for the type II reduction studies. Ice crystal morphology was observed through a Leitz Dialux 22 microscope and recorded by a Panasonic CCTV camera linked to a JVC Super VHS video recorder. Still images were obtained from a Silicon Graphics INDY terminal using IRIS Capture Version 1.2.

DTT Dependence Study. The antifreeze activity of a 1 mg/mL solution of wild-type SRAFP isolated from sea raven (*27*) was measured at three different temperatures in the presence of 10 mM dithiothreitol (DTT). The AFP was dissolved in 100 mM Tris-HCl (pH 8.0), and DTT was added from a freshly prepared 1 M stock solution. Samples were incubated for timed intervals at 0, 22, and 37 °C prior to the thermal hysteresis assay.

NMR Spectroscopy. For the heteronuclear two-dimensional (2D) and 3D NMR experiments, 5 mg of the recombinant ^{15}N-labeled protein was dissolved in 300 μL of 90% H_2O/10% D_2O (v/v) to give a 1.2 mM protein solution. A symmetrical susceptibility-matched NMR microtube (Shigemi, Allison Park, PA) was used for this sample. The pH of the sample was adjusted to 5.6. For the homonuclear ^{1}H 2D NMR experiments, 7 and 3.5 mg of the native protein were dissolved in 500 μL of 90% H_2O/10% D_2O (v/v) and 500 μL of 100% D_2O to give 1.0 and 0.5 mM protein solutions, respectively. 2,2-Dimethyl-2-silapentane-5-sulfonic acid (DSS) was added to all samples (0.1 mM) for internal referencing. NMR experiments were carried out at 25, 35, and 45 °C on Varian Unity and Varian UnityPlus 600 MHz NMR spectrometers equipped with three channels, pulsed-field gradient triple-resonance probes with a *z* gradient, and gradient amplifier units. The following 2D ^{1}H homonuclear experiments were carried out: DQF-COSY in H_2O at 35 °C (*28*), TOCSY in H_2O and D_2O at 35 and 45 °C, respectively (*29*, *30*), and NOESY in H_2O at 35 °C (*31*, *32*). Mixing times of 30 and 50 ms were used for the TOCSY spectra in D_2O and H_2O, respectively, while a mixing time of 150 ms was used for the NOESY spectrum. The acquired data for the 2D spectra consisted of 2048 (all H_2O spectra) or 1024 (TOCSY in D_2O) complex data points in the acquisition domain and of 300 (DQF-COSY) or 256 (TOCSY and NOESY in H_2O) or 280 (TOCSY in D_2O) complex data points in the indirectly detected domain. The corresponding spectral widths for both dimensions were either 11.76 ppm (DQF-COSY) or 12.50 ppm (TOCSY in D_2O) or 16.76 ppm (TOCSY and NOESY in H_2O). Two-dimensional ^{1}H−^{15}N HSQC spectra were recorded at 35 °C, consisting of 800 (HN) × 256 (^{15}N) complex data points. ^{15}N-edited 3D NOESY (*33*) and ^{15}N-edited 3D TOCSY (*34*) experiments were recorded using the sensitivity-enhanced pulsed-field gradient method. The NOESY and TOCSY spectra were recorded at 25 and 35 °C using mixing times of 150 and 50 ms, respectively. The acquired data consisted of 512 (HN) × 128 (^{1}H) × 32 (^{15}N) complex data points. Thirty-two complex points were added to the ^{15}N dimensions by linear prediction to obtain 1024 × 256 × 64 point spectra after zero filling. The ^{15}N-edited 3D HNHA (*35*) spectra acquired at 35 °C consisted of 512 (HN) × 56 (^{1}H) × 48 (^{15}N) complex data points. Forty-eight additional complex points were predicted in the ^{15}N dimension to finally generate, after zero filling, a 1024 × 112 × 96 point spectrum. The corresponding spectral widths for all of the heteronuclear experiments were 13.34 ppm in the HN acquisition domain, 11.17 (NOESY at 35 °C), 12.53 (NOESY at 25 °C), and 8.01 ppm (HNHA) in the indirectly detected proton dimension, and 32.90 ppm in the indirectly detected nitrogen domain. In all homo- and heteronuclear 2D and 3D experiments, the ^{1}H acquisition domains were centered at the water frequency at 4.68 (for all experiments at 35 °C), 4.80 (for all experiments at 25 °C), and 4.56 ppm (for the 2D TOCSY experiments in D_2O at 45 °C). For the ^{15}N-edited experiments, the indirectly detected nitrogen frequencies were centered at 119.50 (35 °C) and 119.58 ppm (25 °C).

Processing of the 2D and 3D data sets was accomplished using either the VNMR software (VNMR 5.1, Varian, Palo Alto, CA) or NMRPIPE (*36*). For most experiments, a postacquisition solvent suppression by convolution of the time domain data was applied prior to Fourier transformation (*37*). In most of the cases, a 66°-shifted sine bell function was applied in the ^{15}N dimension, while 90°-shifted squared sine bell functions were applied in the other dimensions. After Fourier transformation of the acquisition dimension, the portions of the spectra without resonances were discarded prior to the processing of the indirectly detected dimensions, thus reducing the size of the final spectra by 50−75%. The ^{1}H chemical shifts were directly referenced using DSS, while the ^{15}N shifts were indirectly calibrated using the procedures of Live et al. (*38*) and Bax and Subramanian (*39*). Peak picking was performed using the interactive graphic-based program PIPP (*40*). Both processing and peak picking were performed on either a Sun Sparc5 or a Silicon Graphics Indigo 2 workstation.

Structure Calculations. Calculations were performed using the program X-PLOR 3.1 (*41*) and the dynamic simulated annealing protocol for extended-strand starting structures (*42*). The initial structure was energy minimized with 2000 cycles of Powell energy minimization. High-temperature dynamics were run for 30 ps at an initial temperature of 1000 K. The system was then slowly cooled to a temperature of 300 K in 50 K steps over a period of 20 ps. At 300 K, a final stage of 400 steps of Powell minimization was performed to yield the final simulated annealing structures.

RESULTS

Production and Purification of ^{15}N-Labeled rSRAFP. The SRAFP expression construct was designed to link the yeast α-mating factor signal sequence directly to the N terminus of the 135-amino acid mature His-tagged AFP. N-Terminal sequencing and mass spectrometry of rSRAFP purified from the fermentation medium by His tag affinity chromatography showed that the yeast leader sequence had been removed along with one or two N-terminal residues of the mature AFP (*43*). A total of 12 mg of ^{15}N-labeled rSRAFP was purified from two bioreactor preparations. The labeled

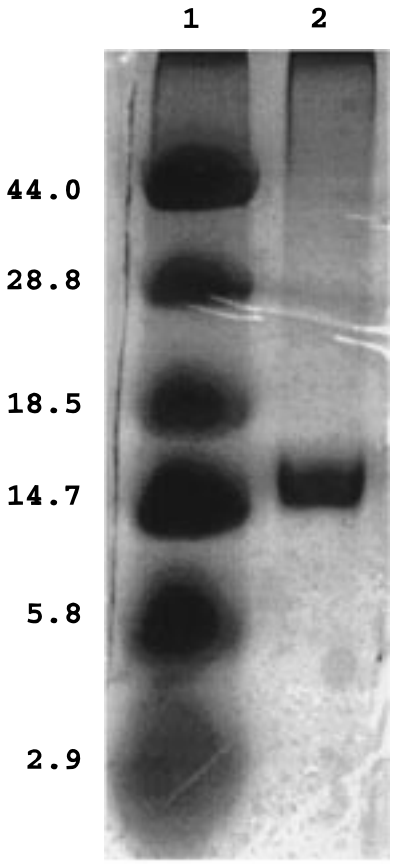

FIGURE 1: SDS−PAGE of rSRAFP: lane 1, low-molecular mass prestained markers (kilodaltons) (GIBCO/BRL); and lane 2, purified rSRAFP (from pPIC9−SRm−CTHT) after nickel−agarose column and ion exchange purification steps.

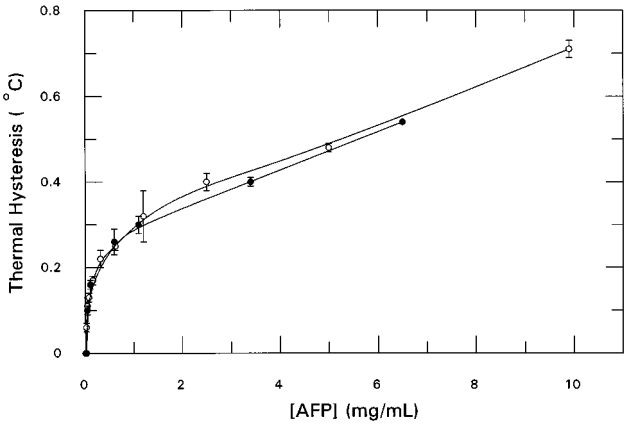

FIGURE 2: Thermal hysteresis activity of rSRAFP. Thermal hysteresis values for rSRAFP (recombinant, ●) purified from *P. pastoris* and SRAFP (wild-type, ○) purified from sea raven serum were compared at various concentrations. Each point represents the mean of at least three determinations. Standard deviations are shown as vertical bars.

protein was expressed to a level of approximately 12−15 mg/L of medium as estimated by Western blotting, and the net yield after purification was 40−50%. The most problematic step in the purification was raising the pH of the supernatant medium from 5.5 to 8.5 to facilitate nickel−agarose column affinity chromatography, which functions optimally at pH values between 8 and 9. Raising the pH by simple addition of a base or concentrated buffer led to significant precipitation, which caused a substantial loss of

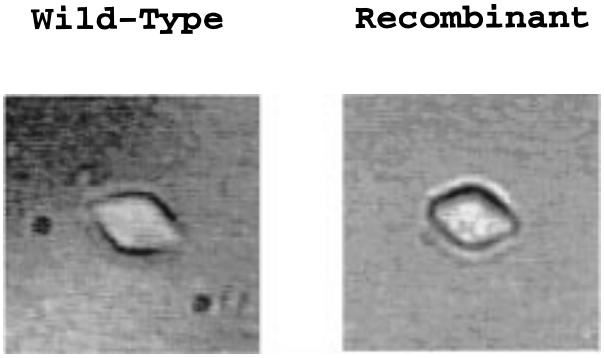

FIGURE 3: rSRAFP ice crystal morphology. The rounded hexagonal bipyramidal ice crystal produced by purified rSRAFP (Recombinant) is compared to that produced by SRAFP (Wild-Type) purified from sea raven.

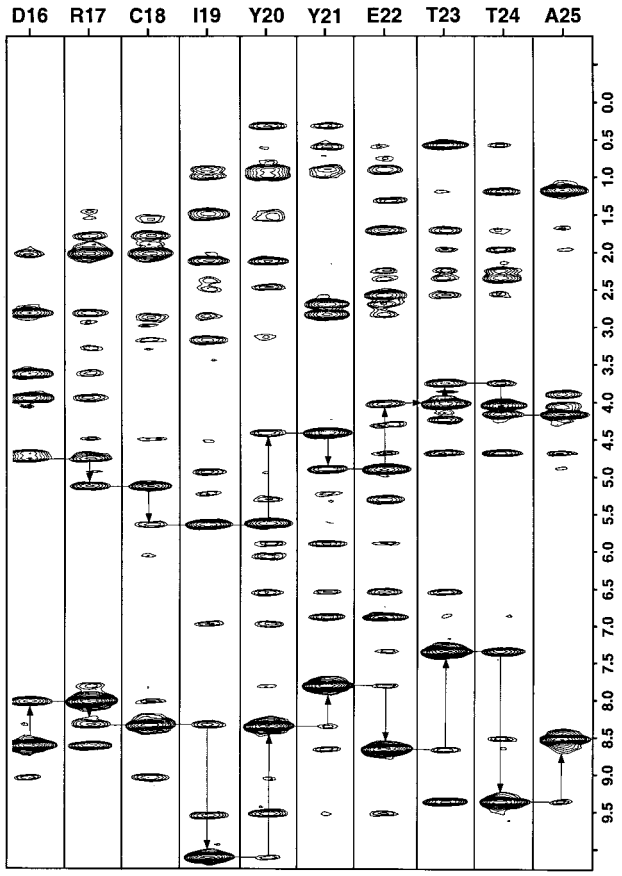

FIGURE 4: Strip plots extracted from a 150 ms mixing time ^{15}N-edited 3D NOESY spectrum of SRAFP (residues Asp16−Ala25). The lines connecting the amide and Hα proton cross-peaks demonstrate the assignment strategy employed for SRAFP. Proton chemical shifts in parts per million are indicated on the *y*-axis. The *x*-axis indicates the residue name and the location in the sequence.

rSRAFP. The precipitate was very fine rather than flocculent and occurred between pH 7 and 8. It adhered to the sides of glassware and was only removed by using a calcium−lime remover (CLR Jamie Industries). This suggested the precipitate might contain metal ions, possibly as hydroxides formed at the alkaline pH, resulting in losses of rSRAFP due to adsorption to or entrappment in the precipitate. To counter these possibilities, excess EDTA was added to chelate divalent ions prior to the pH change. Subsequent addition of Tris base to pH 8.5 was not accompanied by precipitation. To avoid excess EDTA stripping Ni^{2+} from

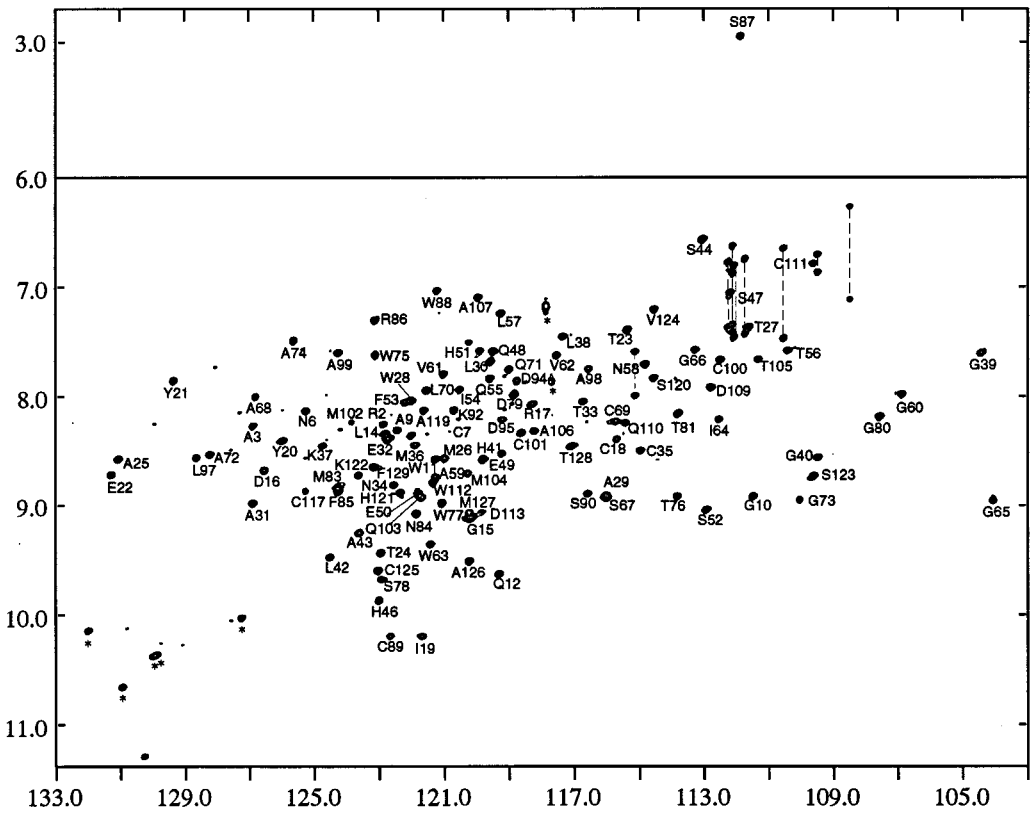

FIGURE 5: $^{15}N-^{1}H$ HSQC NMR spectrum of SRAFP at 35 °C and pH 5.6. The assignments are shown by the residue name (one-letter code) and number. Cross-peaks connected by dotted lines on the right side of the spectrum are derived from the side chain amide groups of Asn and Gln residues. Cross-peaks marked with an asterisk correspond to Trp side chain amide groups.

the nickel−agarose column, the chelating agent was saturated by the addition of excess $MgCl_2$ prior to chromatography. Any residual precipitate was removed by centrifugation before loading it onto the nickel−agarose column. Recombinant SRAFP eluted from the nickel column at 60 mM imidazole in a single peak. This partially purified rSRAFP was loaded onto the Q-Sepharose FPLC column, from which a single protein peak eluted at approximately 0.27 M NaCl. A Coomassie blue-stained SDS gel of the FPLC product (Figure 1) indicated that the product was pure. The antibody to SRAFP cross-reacted with the Western-blotted recombinant protein (*43*), producing a band with a mobility similar to that of wild-type SRAFP.

Characterization of Purified rSRAFP. The antifreeze activities of purified rSRAFP (unlabeled) and natural SRAFP from sea raven serum were compared by assessing their thermal hysteresis values over a range of concentrations. The activity profile for the rSRAFP could not be distinguished from that of the SRAFP (*44*) (Figure 2). This indicated that neither the loss of one or two N-terminal residues nor the six-amino acid C-terminal addition interfered with rSRAFP activity. As well, purified rSRAFP altered the ice crystal morphology to the rounded hexagonal bipyramidal form characteristic of SRAFP (*45*) (Figure 3).

NMR Resonance Assignment. The assignment strategy for SRAFP was based on conventional methods using 2D DQF-COSY, 2D TOCSY, ^{15}N-edited 3D HNHA, and ^{15}N-edited 3D TOCSY experiments to identify spin systems and 2D NOESY and ^{15}N-edited 3D NOESY experiments to identify inter-residue connectivities along the polypeptide chain (*46*). From these experiments, the backbone ^{1}HN, ^{15}N, and $^{1}H\alpha$ assignments of 119 of the 129 SRAFP residues (excluding the flexible C-terminal His tag extension which was not assigned), and assignments for 70% of the side chain ^{1}H resonances were made. Due to overlap in the Hα and Hβ region of the homonuclear 2D NMR spectra, it was not possible to unambiguously assign the seven proline residues. Further, it was not possible to identify resonances for Gly4, Thr92, and Val96. Figure 4, which shows 10 strip plots extracted from a ^{15}N-edited 3D NOESY spectrum, illustrates the general quality of the spectra. Displayed is the region from Asp16 to Ala25, which corresponds to the first and parts of the second β-strand. As expected for β-strands, strong H$\alpha-$HN(i, $i+1$) and relatively weak HN−HN(i, $i+1$) NOEs are observed from Arg17 to Thr23, while the strong HN−HN(i, $i+1$) NOEs between Thr23 and Thr24 indicate the end of this strand. The second β-strand starts at Ala25 and continues to Thr28.

Figure 5 shows the assignment of the $^{15}N-^{1}H$ HSQC spectrum of SRAFP at pH 5.6 and 35 °C. Cross-peaks in the lower left of the spectrum annotated with an asterisk indicate tryptophan indole NH resonances. Signals connected by dotted lines on the right side of Figure 5 are from the side chain amide groups of Asn and Gln residues. Several cross-peaks derived from backbone amide groups (such as Cys111, Ser44, Ser47, and Thr27) are also located in this region of the spectrum. The most striking feature of this spectrum is the backbone amide signal of Ser87 with an amide proton shift of 2.97 ppm. Inspection of the structure shows that this unusual shift is caused by the ring current effect of Trp88 on the Ser87 amide proton, which is located above the center of the aromatic ring. This is similar to the dramatic shifts observed for Ala68, which takes a similar orientation relative to Trp75 (*13*). Due to fast

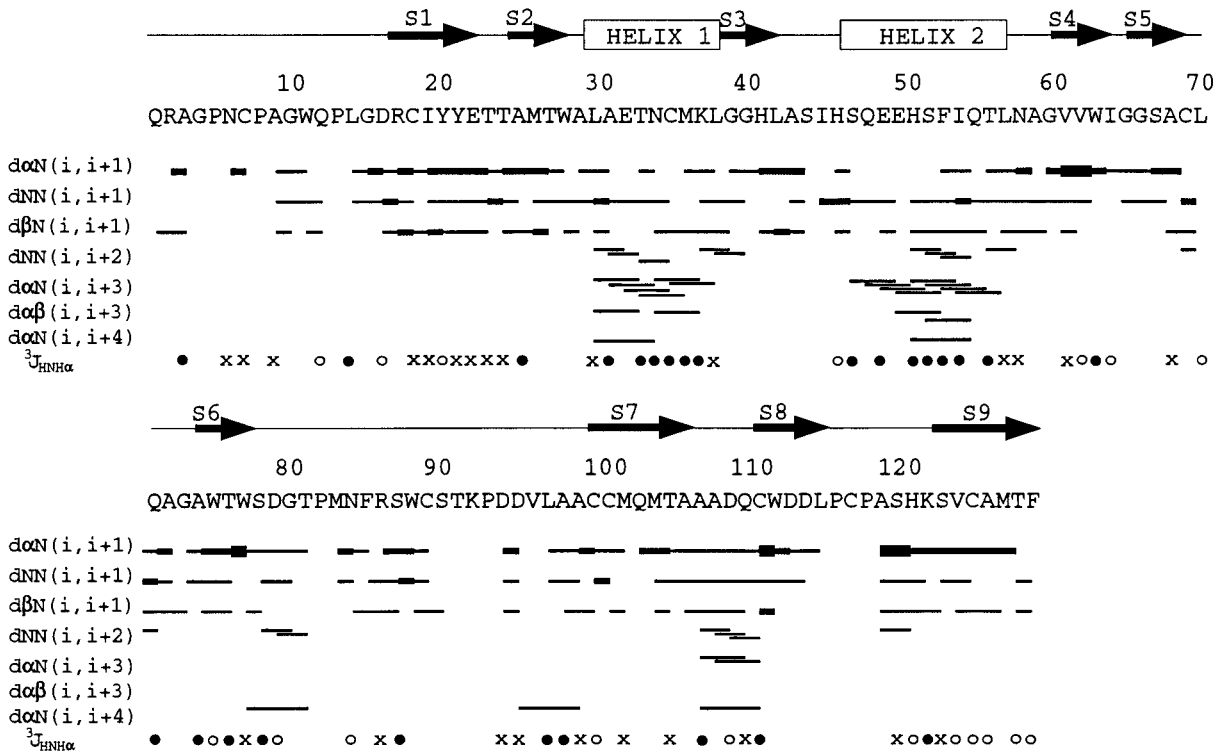

FIGURE 6: Summary of short- and medium-range ($|i - j| \leq 4$) NOE connectivities and $^3J_{\alpha N}$ coupling constant data for SRAFP. Below the one-letter code for the amino acids, the relative strengths of $d_{\alpha N}$ and d_{NN} NOE cross-peak intensities are indicated by the bar thickness. Various medium-range NOE connectivities ($2 \leq |i - j| \leq 4$) observed between specific proton pairs are indicated by a solid line between two residues. Filled circles represent $^3J_{\alpha N}$ coupling constants of <6 Hz, whereas open circles represent $^3J_{\alpha N}$ coupling constants of >8 Hz. A cross indicates a coupling constant in the range of 6−8 Hz.

exchanging amide protons, Ile45, Asp114, and Leu115 are not present and were identified from ^{1}H 2D DQF-COSY, TOCSY, and NOESY spectra.

Overall, the HSQC spectrum is well-resolved with little overlap present in the central region of the spectrum. Further, all of the ^{1}H 2D spectra are well-dispersed and showed surprisingly little overlap for a molecule of this size (14 kDa) (*13*). This is due to the presence of 12 aromatic residues in the protein (Trp11, Trp28, Trp63, Trp75, Trp77, Trp88, Trp112, Tyr20, Tyr21, Phe53, Phe85, and Phe129).

Secondary Structure. Figure 6 shows a summary of short- and medium-range ($|i - j| \leq 4$) NOE connectivities and $^3J_{\alpha HNH}$ coupling constant data together with the secondary structure elements for SRAFP. In all, the molecule contains two α-helices and nine β-strands. The helices are located between strands 2 and 3 and strands 3 and 4. The presence of these helices from Leu30 to Leu38 and from Ser47 to Leu57 is clearly indicated by $d_{\alpha N}(i, i + 3)$, $d_{\alpha\beta}(i, i + 3)$, and $d_{NN}(i, i + 2)$ NOEs together with small $^3J_{\alpha N}$ coupling constants ($^3J_{\alpha N} < 6$ Hz). The regions of the nine β-strands are characterized by the presence of strong $d_{\alpha N}(i, i + 1)$ and weak $d_{NN}(i, i + 1)$ NOEs, together with medium to large $^3J_{\alpha N}$ coupling constants ($^3J_{\alpha N} > 6$ Hz). Furthermore, the β-strands are defined by long-range NOEs between the strands. The molecule contains a three-strand and a four-strand antiparallel β-sheet where the first (Arg17−Glu22), last (Ser123−Phe129), and third β-strands (Gly39−Leu42) form the first sheet. The four-stranded β-sheet is formed by strands 4 (Val61−Ile64), 5 (Gly66−Cys69), 6 (Ala74−Trp77), 7 (Cys100−Ala106), and 8 (Gln110−Leu115). Contiguous strands 4 and 5, interrupted by one flexible glycine, run antiparallel to the same side of strand 7. This is clearer in Figure 7, which shows a schematic view of the β-sheet structure.

Tertiary Structure. A total of 785 NOE distance restraints were obtained from the NOESY spectra and were used in the final calculations. These were distributed as 288 intraresidue restraints, 252 sequential restraints, 76 short-range restraints ($|i - j| \leq 4$), and 169 long-range restraints ($|i - j| > 4$). In addition, 47 ϕ angle restraints were identified from the 3D HNHA spectrum and included in the structure calculations. For a molecule of this size, this is a relatively small set of restraints. Without ^{13}C labeling, it was not possible to unambiguously identify a greater number of NOEs. Therefore, the goal of the following structure calculations was to establish the global fold of the molecule. In all, 50 structures were calculated. On the basis of the total and NOE energies, five structures were selected for further analysis. The structures were pairwise superimposed using residues 17−44, 51−69, 75−79, and 99−129 (i.e. the regions comprising the regular secondary structure elements). On average, the rmsd in the superimposed regions is 1.8 Å for the backbone atoms. The quality of the calculated structures was sufficient to define the global fold of the molecule. Figure 8 shows a schematic presentation of this fold in two orientations. The molecule folds into one globular domain that consists of the two helices and two β-sheets. The two helices of the molecule are oriented perpendicular to one another framing the four β-strands (β1, β2, β3, and β9) in the lower half of the molecule (Figure 8B). The upper half of the molecule includes the second β-sheet with β-strands 4−8 (Figure 8B). SRAFP contains five disulfide bonds. The first disulfide bond (Cys7−Cys18) forms a loop that connects the N-terminal residues with the

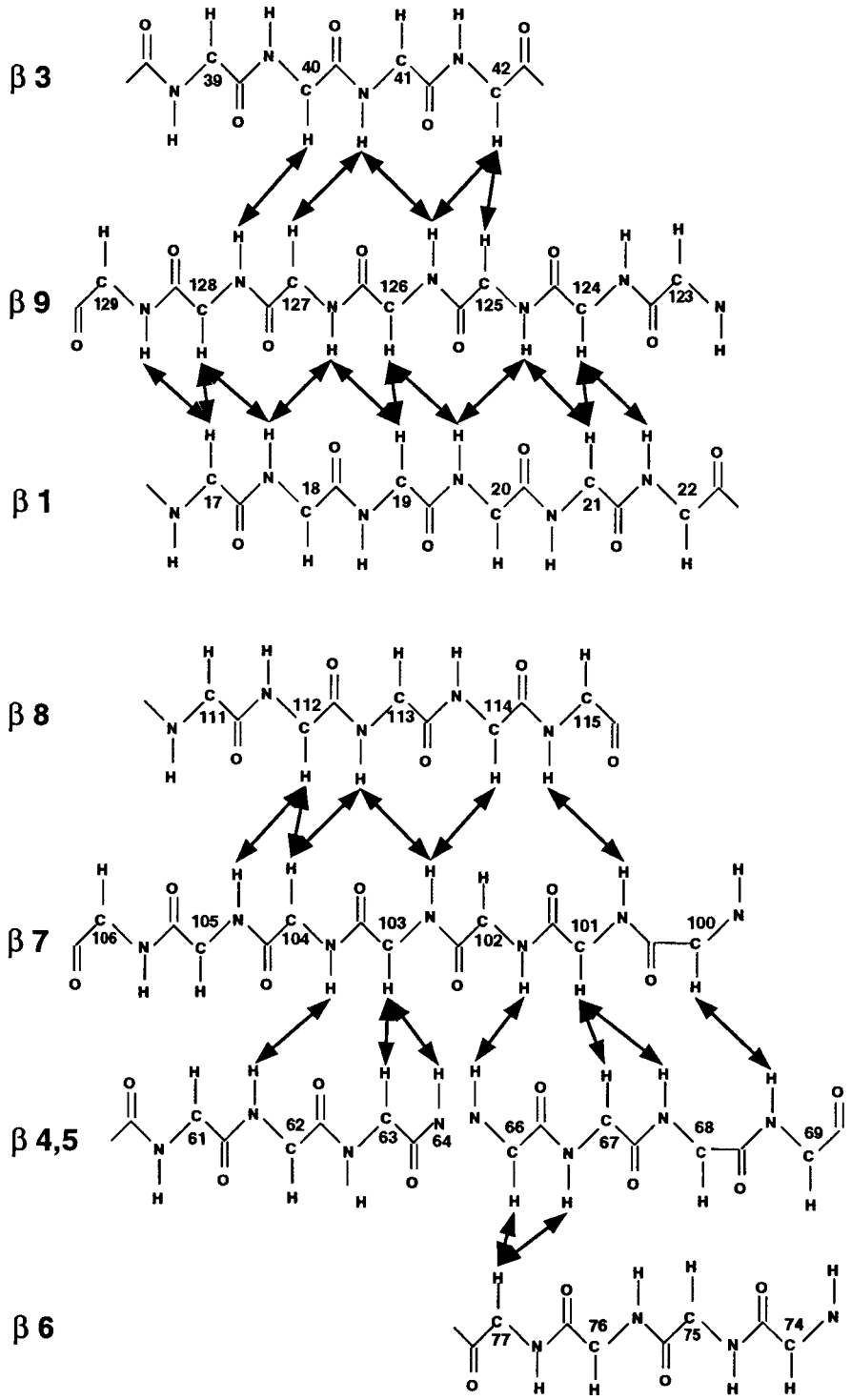

FIGURE 7: Schematic representation of the antiparallel β-sheet structure of SRAFP deduced from a qualitative analysis of the NMR data. The observed interstrand NOE connectivities between backbone NH and αH protons are indicated by arrows. The first β-sheet is formed by strands 3, 9, and 1, while the second β-sheet is formed by strands 8, 7, 4, 5, and 6. Strands 4 and 5 are both aligned to one side of strand 7 so they almost form one strand with a break at residue 65. Interstrand connectivities such as NH−βH NOEs, which define the β-sheets even further, are not shown.

beginning of the first β-strand. The first helix is connected to the last β-strand by the disulfide bond from Cys35 to Cys125. Within the second β-sheet, β-strands 5 and 7 are connected by the disulfide bond from Cys69 to Cys100. Cys89 and Cys111 form the next disulfide bond that connects an extended loop region with the start of β-strand 8. The last disulfide bond from Cys101 to Cys117 makes a connection within the second β-sheet from the end of β-strand 7 to the region following β-strand 8. Each disulfide bond was confirmed by a number of local NOEs.

DISCUSSION

C-Type lectins are an abundant superfamily of proteins containing domains that bind saccharides in a calcium-dependent manner (*15*). High-resolution structures have been experimentally determined for rat mannose-binding protein

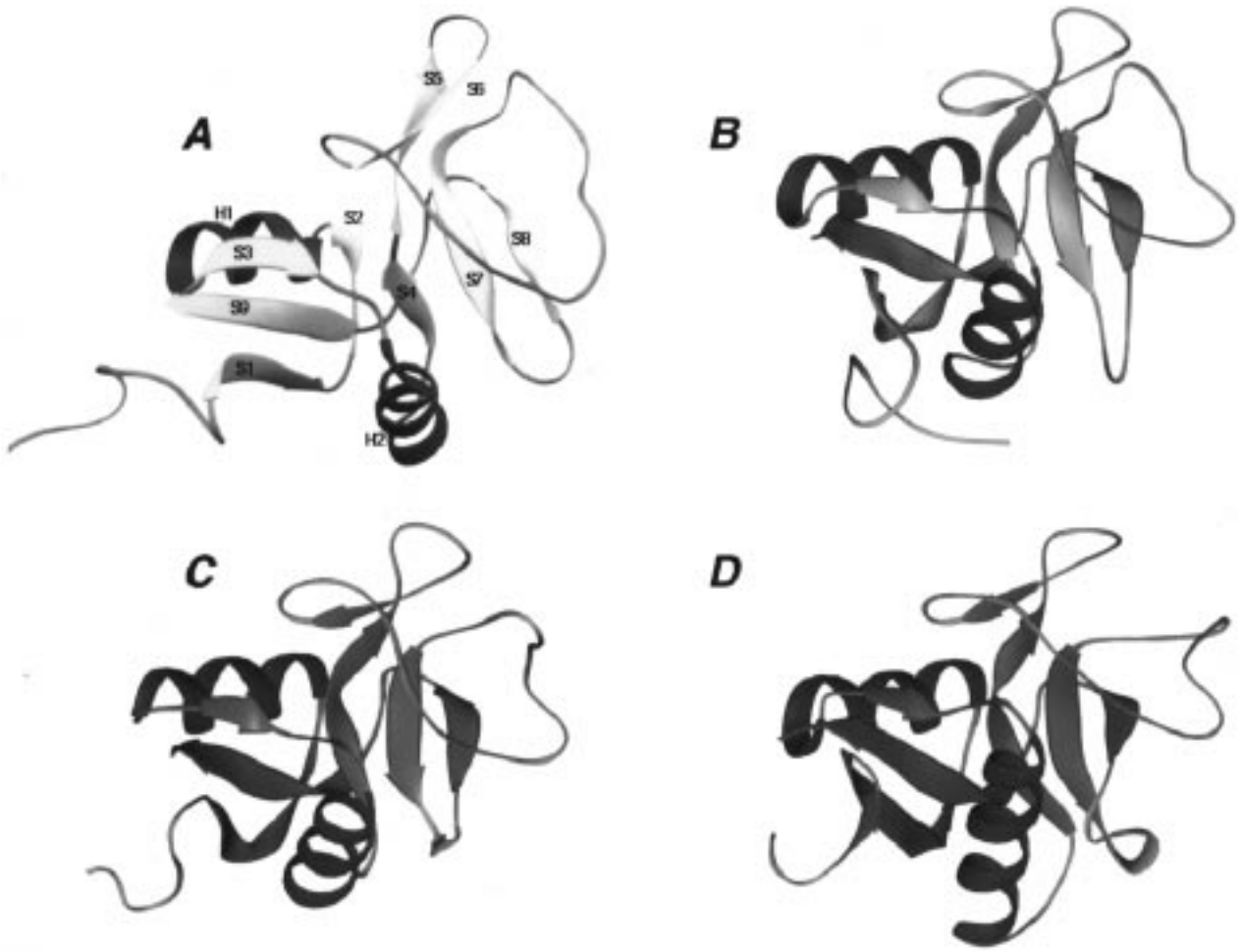

FIGURE 8: Superimposed and separated ribbon diagrams of the solution structure of SRAFP (A, yellow), with the SRAFP model structure (B, gold) and the X-ray structures of rat mannose-binding protein (C, red) and pancreatic stone protein, lithostathine (D, green).

(*16*) and E-selectin (*17*). These protein domains exhibit basically identical folds with small deviations in the extended loop regions. The latter differences are associated with their calcium and ligand binding (*47*) and result in quite distinct saccharide binding specificities. Recently, Bertrand et al. (*19*) determined the structure of human lithostathine and showed that it has a lectin-type fold. The high conservation of the coordinates in these domains (RMBP-A, E-selectin, and PSP) is highlighted by a best-fit superposition of equivalent Cα atomic coordinates (data not shown) which includes the SRAFP structure determined herein. rmsds for Cα coordinates are below 2.5 Å for all pairwise lectin−SRAFP superpositions. This confirms previous predictions that the SRAFP belongs structurally to the C-type lectin family (*13*, *48*). Further, the slight deviations in secondary structure assignments between SRAFP and C-type lectins, namely, one additional strand in each β-sheet, do not indicate significant differences in the fold.

While the direct sequence identity between RMBP-A or E-selectin and SRAFP is relatively low (approximately 20%), structurally important residues such as aromatic core residues, glycines, and prolines tend to be conserved and are found in spatially equivalent positions. This also includes the disulfide bonds. As proposed by the model (*13*), the disulfide bridges from Cys35 to Cys125 and from Cys101 to Cys117 in SRAFP are found in equivalent positions in PSP and RMBP-A. The N-terminal disulfide bond from Cys7 to Cys18 in SRAFP is also present in PSP, while the fourth and fifth bonds between Cys69 and Cys100, and Cys89 and Cys111, are unique to type II AFPs.

The presence of the two additional disulfide bridges in type II AFPs reopens some interesting evolutionary questions. If type II AFP preexisted in the common ancestor of the three widely divergent fishes where it is found today (herring, smelt, and sea raven), then its gene would be present in virtually all teleosts. This is difficult to reconcile with the observation that at least four other types of distinct, completely unrelated AFPs are found in other fishes, some of which are close relatives of the sea raven (*49*). Thus, the derivation of type II AFPs is most likely to have resulted from separate events in which a lectin-like precursor assumed the new role of binding to ice. However, one needs to be cautious about this interpretation because a similar argument was advanced to account for the striking similarity of the antifreeze glycoproteins (AFGPs) found both in northern cods and in the notothenioids from Antarctic waters, fish that are not at all closely related (belonging to different superorders) (*50*). These AFGPs are a mixture of similarly sized polymers of Ala-Ala-Thr repeats that have the same disaccharide attached to the Thr and are derived from polyprotein precursors by cleavage of simple amino acid spacers. Despite these remarkable similarities, Chen et al. (*51*) have recently presented evidence that the AFGPs in the cods and notothenioid fishes evolved recently from completely dif-

ferent precursors, one of them being a trypsinogen gene (*52*). If these arguments are valid, the AFGPs have attained a high amino acid sequence identity (well above the cutoff for homology) through convergent evolution, while the type II AFPs have ended up with a much lower sequence identity through divergence from their putative C-type lectin homologues. In the latter case, the resemblance of the 3D structure obtained by NMR to the lectin carbohydrate recognition domain belies the low sequence identity and provides reassurance about their homology. Even so, the occurrence of the extra two disulfide bridges is puzzling because their placement is identical in all three type II AFPs. It is not clear if the additional disulfide bridges in AFP evolved independently (convergence) or if they derived from C-type lectin progenitors in fish that had these extra bridges already in place.

Due to the lower definition of the backbone fold in loop areas of SRAFP, it is difficult to draw new conclusions regarding its proposed ice-binding site, i.e. the surface responsible for adsorption to ice. Nevertheless, previous suggestions based on the homology model still seem to apply in light of recent ideas about AFPs binding to ice (*7*, *8*, *53*). The protein surface involving the second β-sheet and extensive regions of the loop structure is the most likely candidate for the protein's ice-binding site. Both its surface planarity and the presence of polar residues on the plane are properties that seem to be shared with other AFPs (*7*, *8*, *12*, *53*). This site comprises 10 hydrophilic residues (90−92, 94, 103, 105, 109, 113, 120, and 122), although according to a recent modeling study (*54*), not all of these residues would make hydrogen bonds to the ice surface. The site on PSP that is suggested to be involved in calcium carbonate crystal growth inhibition abuts and overlaps this site, but also includes the loop (between the first two β-strands) containing T23 and T24 (*19*). It remains to be seen which, if either, of these sites can be substantiated as ice-binding surfaces by site-directed mutagenesis studies similar to those used to identify the ice-binding site of type III AFP. In conclusion, SRAFP is homologous to C-type lectins and PSP and presents yet another activity which is manifested by this widely distributed folding motif.

REFERENCES

1. Yeh, Y., and Feeney, R. E. (1996) *Chem. Rev. 96*, 601−617.
2. Davies, P. L., and Hew, C. L. (1990) *FASEB J. 4*, 2460−2468.
3. Ng, N. F. L., Trinh, K.-Y., and Hew, C. L. (1986) *J. Biol. Chem. 261*, 15690−15695.
4. Ng, N. F. L., and Hew, C. L. (1992) *J. Biol. Chem. 267*, 16069−16075.
5. Ewart, K. V., and Fletcher, G. L. (1993) *Mol. Mar. Biol. Biotechnol. 2*, 20−27.
6. Chao, H., Davies, P. L., Sykes, B. D., and Sönnichsen, F. D. (1993) *Protein Sci. 2*, 1411−1428.
7. Sönnichsen, F. D., DeLuca, C. I., Davies, P. L., and Sykes, B. D. (1996) *Structure 4*, 1325−1337.
8. Jia, Z., DeLuca, C. I., Chao, H., and Davies, P. L. (1996) *Nature 384*, 285−288.
9. Deng, G., Andrews, D. W., and Laursen, R. A. (1997) *FEBS Lett. 402*, 17−20.
10. DeVries, A. L. (1982) *Comp. Biochem. Physiol. 73A*, 627−640.
11. Raymond, J. A., and DeVries, A. L. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2589−2593.
12. Sicheri, F., and Yang, D. S. C. (1995) *Nature 375*, 427−431.
13. Sönnichsen, F. D., Sykes, B. D., and Davies, P. L. (1995) *Protein Sci. 4*, 460−471.
14. Ewart, K. V., Yang, D. S. C., Ananthanarayanan, V. S., Fletcher, G. L., and Hew, C. L. (1996) *J. Biol. Chem. 271*, 16627−16632.
15. Drickamer, K. (1988) *J. Biol. Chem. 263*, 9557−9560.
16. Weis, W. I., Drickamer, K., and Hendrickson, W. A. (1992) *Nature 360*, 127−134.
17. Graves, B. J., Crowther, R. L., Chandran, C., Rumberger, J. M., Li, S., Huang, K.-S., Presky, D. H., Familletti, P. C., Wolitsky, B. A., and Burns, D. K. (1994) *Nature 367*, 532−538.
18. Brisset, N. C., and Perking, S. J. (1996) *FEBS Lett. 388*, 211−216.
19. Bertrand, J. A., Pignol, D., Bernard, J.-P., Verdier, J.-M., Dagorn, J.-C., and Fontecilla-Camps, J. C. (1996) *EMBO J. 15*, 2678−2684.
20. Patard, L., Stoven, V., Gharib, B., Bontems, F., Lallemand, J.-Y., and DeReggi, M. (1996) *Protein Eng. 9*, 949−957.
21. Bernard, J. P., Adrich, Z., Montalto, D., DeCaro, A., DeReggi, M., Sarles, H., and Dagorn, J. C. (1992) *Gastroenterology 103*, 1277−1284.
22. Loewen, M. C., Liu, X., Davies, P. L., and Daugulis, A. J. (1997) *Appl. Microbiol. Biotechnol. 48*, 480−486.
23. Hayes, P. H., Scott, G. K., Ng, N. F. L., Hew, C. L., and Davies, P. L. (1989) *J. Biol. Chem. 264*, 18761−18767.
24. Scorer, C. A., Clare, J. J., McCrambie, W. K., Ramanos, M. A., and Sreekrishna, K. (1994) *Bio/Technology 12*, 181−184.
25. Bradford, M. M. (1976) *Anal. Biochem. 72*, 248−254.
26. Chakrabartty, A., and Hew, C. L. (1991) *Eur. J. Biochem. 202*, 1057−1063.
27. Duncker, B. P., Gauthier, S. Y., and Davies, P. L. (1995) *Biochim. Biophys. Acta 1292*, 312−316.
28. Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.
29. Braunschweiler, L., and Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521−528.
30. Davies, D. G., and Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820−2821.
31. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.
32. Macura, S., and Ernst, R. R. (1980) *Mol. Phys. 41*, 95−117.
33. Kay, L. E., Marion, D., and Bax, A. (1989) *J. Magn. Reson. 84*, 72−84.
34. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry 28*, 6150−6156.
35. Vuister, G. W., and Bax, A. (1993) *J. Am Chem. Soc. 115*, 7772−7777.
36. Delagio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
37. Marion, D., Ikura, M., and Bax, A. (1989) *J. Magn. Reson. 84*, 425−430.
38. Live, D. H., Davies, D. G., Agosta, W. C., and Cowburn, D. (1984) *J. Am. Chem. Soc. 106*, 1939−1941.
39. Bax, A., and Subramanian, S. (1986) *J. Magn. Reson. 67*, 565−569.
40. Garrett, D. S., Powers, R., Gronenborn, A. M., and Clore, G. M. (1991) *J. Magn. Reson. 95*, 214−220.
41. Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science 235*, 458−460.
42. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) *FEBS Lett. 239*, 129−136.
43. Loewen, M. C. (1997) Thesis, Queen's University, Kingston, ON.
44. Kenward, K. D. (1995) Ph.D. Dissertation, Queen's University, Kingston, ON.
45. Chao, H., DeLuca, C. I., and Davies, P. L. (1995) *FEBS Lett. 357*, 183−186.
46. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
47. Weis, W. I. (1994) *Structure 2*, 147−149.

48. Ewart, K. V., Rubinsky, B., and Fletcher, G. L. (1992) *Biochem. Biophys. Res. Commun. 185*, 335−340.
49. Davies, P. L., Ewart, K. V., and Fletcher, G. L. (1993) in *Biochemistry and Molecular Biology of Fishes* (Mommsen, T. P., and Hochachka, P. W., Eds.) Vol. 2, pp 279−291, Elsevier Science Publishers, Amsterdam.
50. Scott, G. K., Fletcher, G. L., and Davies, P. L. (1986) *Can. J. Fish. Aquat. Sci. 43*, 1028−1034.
51. Chen, L., DeVries, A. L., and Cheng, C. H. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 3811−3816.
52. Chen, L., DeVries, A. L., and Cheng, C. H. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 3817−3822.
53. Chao, H., Houston, M. E., Hodges, R. S., Kay, C. M., Sykes, B. D., Loewen, M. C., Davies, P. L., and Sönnichsen, F. D. (1997) *Biochemistry 36*, 14652−14660.
54. Wierzbicki, A., Madura, J. D., Salmon, C., and Sönnichsen, F. D. (1997) *J. Chem. Inf. Comput. Sci. 37*, 1006−1010.

BI972788C