Protein Science (1995), *4*:460-471. Cambridge University Press. Printed in the USA.
Copyright © 1995 The Protein Society

Comparative modeling of the three-dimensional structure of Type II antifreeze protein

FRANK D. SÖNNICHSEN,[1] BRIAN D. SYKES,[1] AND PETER L. DAVIES[2]

[1] Protein Engineering Network of Centres of Excellence and Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2S2, Canada
[2] Department of Biochemistry, Queen's University, Kingston, Ontario K7L 3N6, Canada

(RECEIVED October 31, 1994; ACCEPTED December 7, 1994)

Abstract

Type II antifreeze proteins (AFP), which inhibit the growth of seed ice crystals in the blood of certain fishes (sea raven, herring, and smelt), are the largest known fish AFPs and the only class for which detailed structural information is not yet available. However, a sequence homology has been recognized between these proteins and the carbohydrate recognition domain of C-type lectins. The structure of this domain from rat mannose-binding protein (MBP-A) has been solved by X-ray crystallography (Weis WI, Drickamer K, Hendrickson WA, 1992, *Nature 360*:127-134) and provided the coordinates for constructing the three-dimensional model of the 129-amino acid Type II AFP from sea raven, to which it shows 19% sequence identity. Multiple sequence alignments between Type II AFPs, pancreatic stone protein, MBP-A, and as many as 50 carbohydrate-recognition domain sequences from various lectins were performed to determine reliably aligned sequence regions. Successive molecular dynamics and energy minimization calculations were used to relax bond lengths and angles and to identify flexible regions. The derived structure contains two α-helices, two β-sheets, and a high proportion of amino acids in loops and turns. The model is in good agreement with preliminary NMR spectroscopic analyses. It explains the observed differences in calcium binding between sea raven Type II AFP and MBP-A. Furthermore, the model proposes the formation of five disulfide bridges between Cys 7 and Cys 18, Cys 35 and Cys 125, Cys 69 and Cys 100, Cys 89 and Cys 111, and Cys 101 and Cys 117. Based on the predicted features of this model, a site for protein-ice interaction is proposed.

Keywords: comparative modeling; C-type lectin; disulfide-bonding pattern; ice-binding site; NMR; Type II antifreeze protein

Marine teleosts from polar oceans and north temperate seas can be protected from freezing in icy seawater by serum antifreeze proteins or glycoproteins (DeVries, 1983; Davies et al., 1988). These macromolecular antifreezes function noncolligatively by binding to and preventing the growth of ice crystals within the fish. Despite their similar function, they are structurally diverse (Davies & Hew, 1990). The AFGP of nototheniids and cods are polymers of a tripeptide repeat, Ala-Ala-Thr, with a disaccharide attached to the threonine residue (Feeney & Yeh, 1978). There are at least three other types of nonglycosylated AFPs: Type I AFPs are alanine-rich, α-helical peptides found in flounder and sculpin (Yang et al., 1988); Type II AFPs of sea raven (*Hemitripterus americanus*), smelt (*Osmerus mordax*), and herring (*Clupea harengus harengus*) are larger (120+ amino acids), cysteine-rich proteins (Ewart et al., 1992); and Type III AFPs, found in eel pouts, are intermediate in size (60+ amino acids) and rich in β-structure (Sönnichsen et al., 1993). Of these four different types of antifreeze proteins, Type II AFP is the only class for which detailed structural information is not available. However, Type II AFP is also the only class for which a sequence similarity exists to other proteins in the current sequence database (Ewart et al., 1992; Ng & Hew, 1992). Its sequence is similar to the carbohydrate recognition domain of Ca^{2+}-dependent lectins. This 110+ amino acid domain is part of a widespread superfamily of proteins that bind sugars specifically through contact with a calcium ion (Drickamer, 1988). The CRD

Reprint requests to: Frank D. Sönnichsen, Protein Engineering Network of Centres of Excellence and Department of Biochemistry, Heritage Medical Research Centre 7-13, University of Alberta, Edmonton, Alberta T6G 2S2, Canada; e-mail: fds@saiph.biochem.ualberta.ca.

Abbreviations: AFP, antifreeze protein; SR AFP, sea raven antifreeze protein; AFGP, antifreeze glycoprotein; ASA, solvent-accessible surface area; CRD, carbohydrate-recognition domain; MBP-A, mannose-binding protein-A; PSP, pancreatic stone protein; SCR, structurally conserved regions; SVR, structurally variable regions; RMSD, RMS deviation; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; DQF-COSY, double quantum filtered correlation spectroscopy; CVFF, consistent valence force field; 1D, one-dimensional; 2D, two-dimensional; 3D, three-dimensional.

is generally associated with a variety of other domains, but can occur in isolation. The extent of sequence identity within the CRD superfamily is confined to short sequences and single amino acids at intervals throughout the protein (Weis et al., 1991). The 3D structure of the CRD from rat mannose-binding protein-A has been solved by X-ray crystallography (Weis et al., 1991, 1992). It is a compact domain cemented by two disulfide bridges in which the N- and C-termini are in close proximity. It has two α-helices, five β-strands, and a high percentage of amino acids in loops and turns. In light of this structural information, it is apparent that the dispersed sequence identities in CRDs correspond to key amino acids in the protein fold such as hydrophobic core residues, disulfide-bonded cysteines, or residues involved in binding Ca^{2+}. The sequence identity between the CRD of rat MBP-A and Type II AFP from sea raven is only 19%, but again the identical and similar amino acids tend to be key residues in the structure and common to the other CRDs (Ewart et al., 1992; Ng & Hew, 1992). The recent discovery of Type II AFP in herring and smelt (Ewart & Fletcher, 1990; Ewart et al., 1992) reinforces the link to the CRDs because these two AFPs have a Ca^{2+} requirement for ice binding, unlike sea raven Type II AFP. In the absence of a 3D structure, we have modeled the fold of Type II AFP based on the X-ray coordinates of the rat MBP-A CRD to see if it is consistent with the latter structure and to provide clues about the structure–function relationships in this antifreeze type.

Results and discussion

Sequence alignment

Drickamer and co-workers (Bezouska et al., 1991) have classified C-type lectins into four main types—selectins, proteoglycans, receptor-type, and adhesins—which show extensive sequence identity. Ewart et al. (1992) and Ng and Hew (1992) have recognized Type II AFPs and pancreatic stone proteins as homologues of the C-type lectins. An alignment of the three Type II AFPs, human PSP (DeCaro et al., 1988), and representatives of four different types of C-type lectins is shown in Figure 1. Pairwise comparisons for the three AFPs versus rat

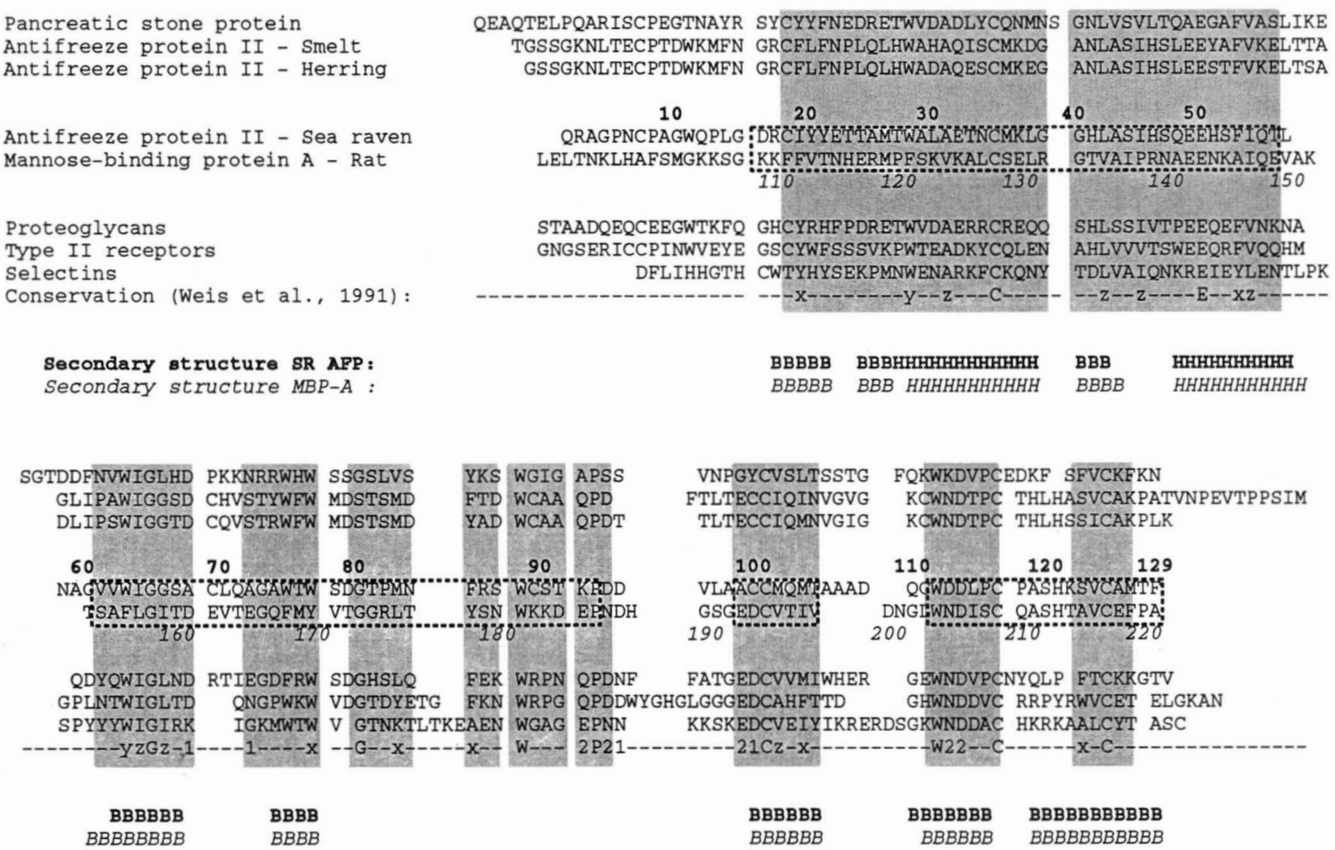

Fig. 1. Representative alignment obtained from multiple sequence alignments from up to 50 lectin and homologous sequences. Sequences of PSP, three known Type II AFP sequences, MBP-A, and one representative of the other three groups of lectins (Drickamer et al., 1986) are shown: rat proteoglycan core protein (Halberg et al., 1988) from group 1 (Proteoglycans), rat hepatic lectin 1 (Leung et al., 1985) from group 2 (Type II receptors), and lymphocyte homing receptor (Lasky et al., 1989) from group 4 (Selectins). Shaded areas represent sequence regions that are always consistently aligned. Gaps indicate regions of deletions or insertions caused by any sequence included in the alignment. Boxes (dashed lines) around the SR AFP and MBP-A sequences represent regions that were defined as SCRs for this modeling study. Sequence numbers for the two proteins are shown above and below their sequences in bold and italics, respectively. Conservation indicates invariant or conserved residues in the C-type lectin family as determined by Weis et al. (1991), with the following code: x, aliphatic or aromatic; y, aromatic; z, aliphatic; 1, ligands for Ca^{2+}-binding site 1; 2, ligands for Ca^{2+}-binding site 2. Secondary structure classification was obtained with the program SEQSEE, using a consensus classification that differs slightly from the classification of Weis et al. (1991).

Table 1. *Comparison of sequence identities (%) between AFP and lectins*

	Type II AFP		
	Sea raven	Smelt	Herring
SR AFP	100	41	39
Proteoglycan[a]	29	28	30
Hepatic lectin[a]	29	28	28
PSP	29	22	26
MBP-A[a]	19	20	20
Lymphocyte receptor[a]	19	19	19

[a] Representative lectin sequences chosen for determination of sequence identities are identical to those shown in Figure 1. Sequence identities of AFP to other members of the four CRD types vary significantly and can be significantly higher (up to 33%).

MBP-A are given in Table 1. The sea raven AFP is 19% identical to the lectin CRD, which is below the threshold value (25%) for a certain structurally significant sequence similarity (Sander & Schneider, 1991). However, SR AFP shows sequence identities of more than 25% to other C-type lectin sequences, such as the large aggregating proteoglycan (33%, Doege et al., 1991), mouse immunoglobulin E receptor (30%, Bettler et al., 1989), and barnacle lectin (29%, Muramoto & Kamiya, 1986). Modeling the structure of SR AFP on the basis of its homology to MBP-A therefore seemed feasible. Although it would have been clearly preferable to base the modeling on more than one atomic structure, the number of known C-type lectin sequences was advantageous. We used multiple sequence alignments on a variety of different C-type lectins (as described in the Methods section) and their analysis to obtain reliably aligned sequence regions, which are indicative of the structurally conserved regions of the C-type lectin fold. Despite the diversity of functions in these proteins and the presence of four subgroups of lectins, the sequences could be reliably aligned over wide regions (Fig. 1). The alignment is very similar to those of Weis et al. (1991), Ewart et al. (1992), and Ng and Hew (1992) over most of the protein. The identified SCRs contain almost all regular secondary structure elements as well as much of the loop regions. This observation was pointed out by Weis et al. (1991) for the various lectins, and remains valid after including the AFP sequences in the alignments. Excluding the Ca^{2+}-coordinating ligands, Weis et al. (1991) identified 28 invariant or highly conserved residues (defined as having no more than one mismatch) in the C-type lectin family. Of these structurally characteristic residues in the lectins, all but three are present in the AFP sequences. These include *cis*-Pro *186*,[3] the four disulfide-bonded cysteines, and most hydrophobic or aromatic core residues. At the positions of the nine highly conserved Ca^{2+}-coordinating ligands, five noncoordinating residues are present in the SR AFP sequence. This agrees with the observation that SR AFP does not bind Ca^{2+}. Herring and smelt AFP, however, show the required coordinating residues at Ca^{2+}-binding site 2 in this alignment, which is consistent with the experimentally observed Ca^{2+} binding of these proteins (Ewart et al., 1992; Ng & Hew, 1992) and further supports the lectin–AFP structural similarity. Also noticeable in Figure 1 is the reliable alignment of the sequences over much of the loop regions. This remarkably consistent alignment could reflect their importance for the function of lectins, and their steric fixation to the β-strand and helical framework by disulfide bridges.

The alignment from residues *111* to *150* is very consistent (one minor insertion was predicted at residue *132*), indicating that this part of the protein fold is highly conserved throughout the members of the CRD family. This agrees well with the high proportion of regular secondary structure elements (81%), i.e., two β-strands and the two helices. Within the latter half of the protein sequence, only shorter stretches of SCRs were identified. From residues *155* to *188*, several short SCRs were predicted, interrupted by only one- or two-residue insertions in some of the lectin sequences, indicating some structural flexibility in this predominantly coiled region. Due to the high level of sequence similarity between MBP-A and SR AFP and the absence of any deletions or insertions between these two proteins in this region, we classified this region as one continuous SCR for the purpose of model building. The same arguments were used for the C-terminal sequence from *205* to *221*. One further short SCR was predicted from residues *194* to *199* due to the stable sequence alignment and the conserved cysteine residue in this β-strand.

Major insertions were encountered only in three regions in which various CRD types exhibited longer, significant sequence insertions. Thus, these regions were defined as structurally variable regions: one loop of identical length in SR AFP and MBP-A (57–66 [*150–154*]) and two regions of insertions and deletions (94–98 [*187–192*] and 105–111 [*200–204*]).

Comparison to E-selectin

During the preparation of this manuscript, Graves et al. (1994) reported the structure of E-selectin solved by X-ray crystallography. This protein is a representative member of the cytokine-inducible, endothelial cell-specific membrane glycoprotein family and contains one CRD. This domain of E-selectin shares 30% sequence identity with MBP-A, and its structure was shown to be very similar to that of MBP-A. The two proteins have the same overall fold and the same number of regular secondary structure elements. Also, a best molecular fit between the two structures calculated on the basis of 109 topologically equivalent Cα positions yielded an RMS deviation of only 1.9 Å, which shows that a sequence identity of about 30% is structurally significant in this protein class. This new structure also provided the opportunity to critically analyze and compare the SCRs and SVRs as determined from the two experimental protein structures to those determined here by sequence alignment. Inspection of the sequence alignment derived from the best molecular fit of MBP-A and E-selectin identifies several regions of insertions and deletions, which are completely consistent with our alignment. Two major SVRs are located at residues *150–153* and *199–204*, caused by three- and five-residue insertions, respectively, in the structure of E-selectin. The third region of significant structural change between E-selectin and MBP-A occurred rather surprisingly in a region of pronounced sequence and functional similarity between these two proteins (Weis, 1994). The best fit also proposes one- or two-residue insertions in the vicinity of residues *162*, *165*, *172*, and *177*. Multiple sequence

[3] Residues of lectins throughout this paper are numbered according to the mannose binding protein A sequence (in italics), whereas antifreezes are numbered relative to their respective sequences.

alignments, however, indicated some sequence length variability in all these areas of the CRD family (Fig. 1).

Analysis of the model structure

The ribbon presentation of the refined model of SR AFP (Fig. 2A) includes the first 15 residues in one possible conformation in order to permit complete visualization of the disulfide bonding pattern (see Kinemage 1). However, the model was generated without these residues because no template coordinates were available for this region. Different secondary structure prediction algorithms consistently indicated the presence of coil structures in this region owing to the high number of prolines and glycines. Thus, we felt that the conformation of this region could not be predicted with sufficient reliability, and we excluded residues 1–15 from the model building and structural analysis.

The protein folds into one compact globular domain (Kinemage 1). It consists of two helices and eight β-strands, the latter forming two antiparallel β-sheets. Based on a consensus secondary structure definition (Colloc'h et al., 1993) as incorporated in SEQSEE (Wishart et al., 1994a), the positions of the secondary structure elements are very similar to those in MBP-A, differing at most by one or two residues. The largest changes are observed among the β-turns. All regular β-hairpin turns in MBP-A change upon refinement of the AFP structure into irregular turns, two type III and one type I turn. The RMSD between coordinates of 104 equivalent residues in SR AFP and MBP-A (Fig. 2B) is 0.8 Å for backbone atoms. This low deviation is only partly due to the use of tethering during the modeling procedure. It also is a consequence of the adequate side-chain packing obtained after the sequence replacement. This is reflected in the high quality of the main- and side-chain geometries of the model structure. Deviations from the ideal are small and often comparable to the deviations observed for experimental structures of 1.5 Å resolution or better. Other important criteria for the quality of the model are the distribution of ϕ/ψ angles and the hydrogen bonds. In total, 84% of non-glycine residues are in the Ramachandran core region, whereas 97% are in the allowed areas (Morris et al., 1992). With only three threonines and four glycines in the disallowed or generous regions of the Ramachandran plot (Fig. 3), the number of residues with positive ϕ angles is actually lower than in MBP-A. The number of hydrogen bonds and their length variation are also close to the expected values for high-resolution structures and thus confirm the high stereochemical quality of our model.

In addition to these local, stereochemical criteria, the model also fits all global folding expectations (Table 2). The free energy of folding, the numbers of buried residues and packing defects, and the model volume are all close to or better than expected values. Also, the distribution of $\chi 1$ angles is close to expected values based on high-resolution structures of correctly folded proteins. The model is compact, with a slightly smaller solvent-accessible surface area (93% of the expected value) being characteristic of a structure generated by in vacuo modeling techniques (Chiche et al., 1989). To compensate for this tendency, main-chain atoms were tethered for most of the calculations. Although tethering helped to reduce this effect, it also attenuated the RMS differences between this model and MBP-A, as pointed out above. The limitations of this compromise could be overcome with dynamics and minimization calculations of this model in a water shell, but these would be beyond the scope of this study. The compactness (97% ASA compared to MBP-A) does not significantly reduce the information content of this model.

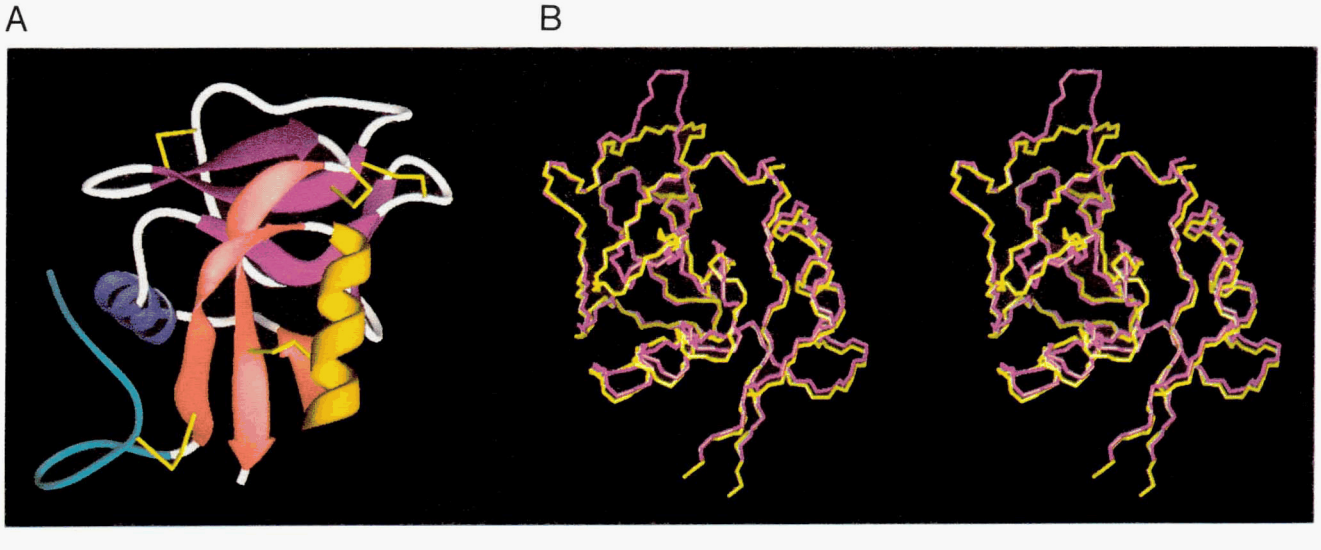

Fig. 2. Backbone presentations of sea raven Type II AFP model. **A:** Richardson presentation created with the program RIBBONS (Carson, 1987). Secondary structure elements are highlighted using the following code: red, N/C-terminal sheet; purple, antiparallel triple-stranded sheet; dark yellow, helix 1; and blue, helix 2. Disulfide bridges are represented by yellow cylinders. The first 15 residues (turquoise) are included in this presentation in one possible conformation to indicate the N-terminal disulfide bridge between Cys 7 and Cys 18. Their coordinates, however, could not be modeled with sufficient reliability. **B:** Stereo presentation of the superposition of SR AFP (yellow, residues 16–129) and MBP-A (purple). Only backbone atoms are shown. The best-fit superposition was obtained using 104 equivalent residues (C, Cα, and N).

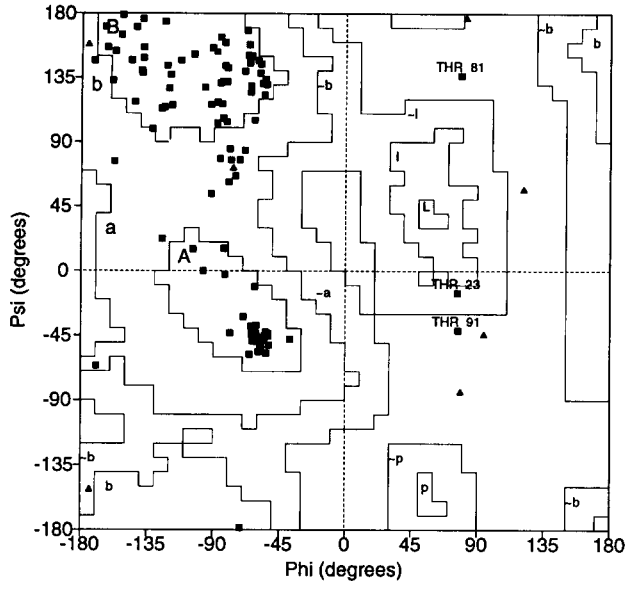

Fig. 3. Distribution of backbone dihedral angles of residues in SR AFP. The ϕ/ψ space is divided into four regions: core, allowed, generously allowed, and non-allowed regions. Residues in non-allowed areas are labeled with their type and sequence number. Glycines are represented by triangles. The plot was generated using the program PROCHECK (Laskowski et al., 1993).

In order to further test the quality of the model, we performed an unrestrained and untethered molecular dynamics calculation for 100 ps. After the initial equilibration (20 ps), during which the potential energy decreased by approximately 6%, the model stayed stable for the remainder of the calculation. This is reflected in the small RMSD between the structures taken at regular intervals over the calculation and either the starting or final structure, which averaged to 2.4 ± 0.1 and 1.5 ± 0.1 Å, respectively. The stability can also be deduced from the per-residue RMSD between all individual molecular dynamics structures and their average (Fig. 4). Most regions of the protein did not change during the calculation, which resulted in an average RMSD of just 1.3 Å. The largest structural fluctuations were observed at the N/C-termini and at residues 70–74, which were part of the first defined SCR. The modeled loops (57–60, 94–98, and 106–111) exhibited RMSD close to the average, with values of 1.2, 1.2, and 1.5 Å, respectively. Overall, the individual and average fluctuations during this calculation were small and reemphasize the stability of the fold of the SR AFP model.

Table 2. *Folding characteristics as determined with VADAR*[a]

	SR AFP Type II	MBP-A	Expected
Resolution (Å)		1.7	
Packing defects	2	2	<7
95% buried residues	28	28	>23[b]
Free energy of folding (kcal/mol)	−96.6	−81.1	−96.8[c]
Fractional volume	0.99 ± 0.09	0.96 ± 0.10	1.00 ± 0.13
Total ASA (Å^2)	5,748	5,944	6,200[d]
Fraction nonpolar (%)	65	58	61 ± 3[d]
Fraction polar (%)	24	21	21 ± 5[d]
Fraction charged (%)	11	20	19 ± 5[d]
Number of hydrogen bonds	73 (64%)	80 (71%)	85 (75%)
Average H-bond distance (Å)	2.26 ± 0.38	2.15 ± 0.33	2.20 ± 0.4
% Helix	18	20	
% Beta	38	39	
% Coil	44	41	

[a] Wishart et al. (1994b). Expected values are taken from: [b] Janin (1976); [c] Chiche et al. (1990), calculated as described by Eisenberg and McLachlan (1986); [d] Miller et al. (1987).

Prediction of disulfide bridges

Ten cysteine residues (7, 18, 35, 69, 89, 100, 101, 111, 117, and 125) are present in Type II AFP (Fig. 1; see Kinemage 1); all were determined by chemical methods to be involved in the formation of disulfide bonds (Ng & Hew, 1992). Four of these are absolutely conserved in lectins as well as Type II AFP and PSP and they are aligned in the SCRs, which suggested the formation of equivalent disulfide bonds as defined by the crystal coordinates of MBP-A. These cysteines are therefore identically linked in this model (C35–C125 and C101–C117). Two more cysteine residues present in Type II AFP are located in the N-terminal region. These cysteines (7, 18) are, however, present in various other lectins and in PSP. Chemically, they have been shown to be bonded to one another (Muramoto & Kamiya, 1986; Rouimi et al., 1987; Ng & Hew, 1992). Furthermore, in the template structure, it is sterically unlikely that both could concurrently be bonded to other cysteines.

Four more cysteines (69, 89, 100, 111) are unique to Type II AFPs and are part of or directly adjacent to the reliably aligned sequence regions. Pairs of these cysteines were in ideal positions to selectively form disulfide bonds, with Cβ distances of less than 4 Å and SS distances of less than 3 Å between the cysteine residues involved. This established the following bonding pattern: C7–C18, C35–C125, C69–C100, C89–C111, and C101–C117, which was introduced in the molecular structure of the template and used in all subsequent steps.

This bonding pattern (Fig. 5A) differs from that established by Ng and Hew (1992) using chemical methods (Fig. 5B), where C117 was linked to C89, and C101 was linked to either C69 or C111. The reason for uncertainty in the latter assignment was that the cysteine adjacent to C101 (C100) could not be physically separated from C101 and, therefore, the peptide containing C100 and C101 was linked to two peptides, those containing C69 and C101. Our model is consistent with a disulfide bond between C100 and C69, which would leave C101 paired to C111. However, linkage of C101 to C111 (or to C69) is inconsistent with the homology between Type II AFPs and the C-type lectins. C101 is one of the four absolutely conserved cysteines in the lectins, where it is linked to the equally conserved C117. The resulting disulfide bond fixes the start of β-strand 3 (C101) to the end of the antiparallel β-strand 4 (C117). A bond between C101 and C111 (which would link the start of β-strand 3 to the start of β-strand 4) would be incompatible with the antiparallel arrangement of these two strands and would require a major rearrangement of the protein fold. Similarly, the other disulfide bond (C89–C117) that is inconsistent with our model would suggest a linkage between the end of β-strand 4 and the middle of loop 3. The formation of such a bond would also re-

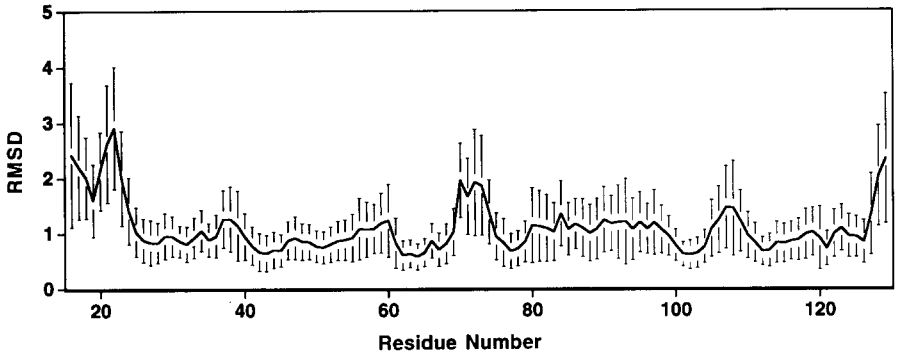

Fig. 4. Dynamic stability of the Type II model. Average per-residue RMSDs of individual molecular dynamics structures to their average structure. During the 100-ps molecular dynamics calculations, snapshots of the protein were taken every 0.5 ps. The best-fit superposition of structures after equilibration (20–100 ps) to their average structure was obtained for all backbone atoms (C, Cα, N). The line represents per-residue average values in Å with error bars indicating standard deviation.

quire a major rearrangement, i.e., the presence of a different fold. However, in a fully disulfide-bonded protein, a misalignment of one disulfide bond must be accompanied by another. The fact that C89–C117 and C101–C111 are both inconsistent with our model based on the CRD fold, whereas C89–C111 and C101–C117 (as well as the three other disulfide bonds) fit the model perfectly well, argues strongly for a realignment of the bonds during the chemical determination.

Confirmation of the model

The predicted structure for SR AFP is consistent with the interpretations based on its CD spectra by Slaughter et al. (1981). These were that the structure had little α-helix content and some β-structure, and that there were aromatic residues in an asymmetric environment that might be the protein core. In order to get additional support for this model, we evaluated preliminary structural evidence obtained by NMR experiments. Structure determination of this protein is currently under way in our laboratory using protein samples isolated from sea raven serum. Because it is a natural sample, NMR analysis is restricted to homonuclear methods, leading to difficulties due to the complexity of the spectra of proteins of this size. The 1D ^{1}H-NMR spectrum (Fig. 6A) is typical for folded proteins. The resonances are well dispersed over the entire ^{1}H-resonance range (from 12 to −0.8 ppm). Several resonances in the range from 5 to 6 ppm indicate a substantial amount of β-sheet structure. This was confirmed using a simple technique for secondary structure determination on the basis of ^{1}H chemical shifts (Wishart et al., 1991a). Evaluation of the fingerprint region of a DQF-COSY spectrum of SR AFP acquired at 35 °C suggested the presence of 15% helix, 32% coil, and 53% β-sheet, which compares reasonably well with values deduced from the model of SR AFP (residues 16–129) (Table 2).

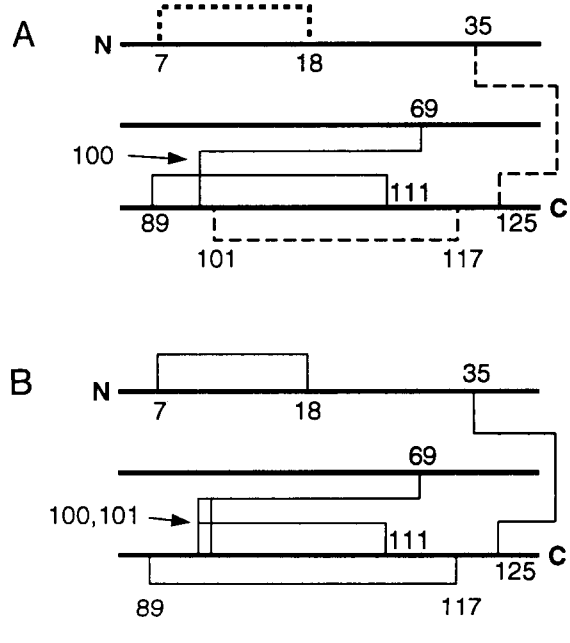

Fig. 5. Disulfide bond patterns in sea raven Type II AFP. **A:** Pattern based on comparative modeling. Links established by the X-ray structure of rat MBP-A between four cysteines that are invariant in all homologues are illustrated by dashed lines. Dotted line represents linkage of the third cysteine pair found in a subset of C-type lectins and in PSP. **B:** Pattern proposed based on chemical methods (Ng & Hew, 1992).

The appearance of a ^{1}H-resonance line at −0.8 ppm is noteworthy (Fig. 6A). Its intensity and line pattern characterize the involved protons as an Ala-methyl group, which was readily confirmed in 2D DQF-COSY and TOCSY spectra (data not shown). Partial assignments have been obtained independently of the model. Using 2D NOESY spectra, the residue of interest could be sequentially assigned to Ala 68. A comparison of its ^{1}H-resonance shifts with random-coil values (Wishart et al., 1991b) revealed an unusual upfield shift of 2.1 ppm for the methyl group. Deviations of this magnitude in proteins are caused by strong anisotropic effects of aromatic residues such as Tyr, Phe, or Trp. Several crosspeaks in the 2D NOESY spectra gave direct evidence for close proximity of Ala 68 to a Trp side chain (Fig. 6B). Moreover, the direction of the shift and the presence of NOEs of similar intensity to all protons of the aromatic ring system indicate that Ala 68 is closely positioned above the plane of such a side chain rather than to its side.

The unique environment of Ala 68 as proposed by the model structure is shown in Figure 7 and Kinemage 2. Consistent with the interpretation of the NMR data, the methyl group of Ala 68 is placed directly above the center of the aromatic ring of Trp 75. The short distance (3.5 Å between Ala 68 Cβ and the plane of the aromatic side chain) is appropriate to explain the magnitude of the observed chemical shift deviation. No other alanine methyl group is in a comparable position in the model, which similarly can be deduced from the absence of any other significantly upfield-shifted resonances in that region of the NMR spectrum.

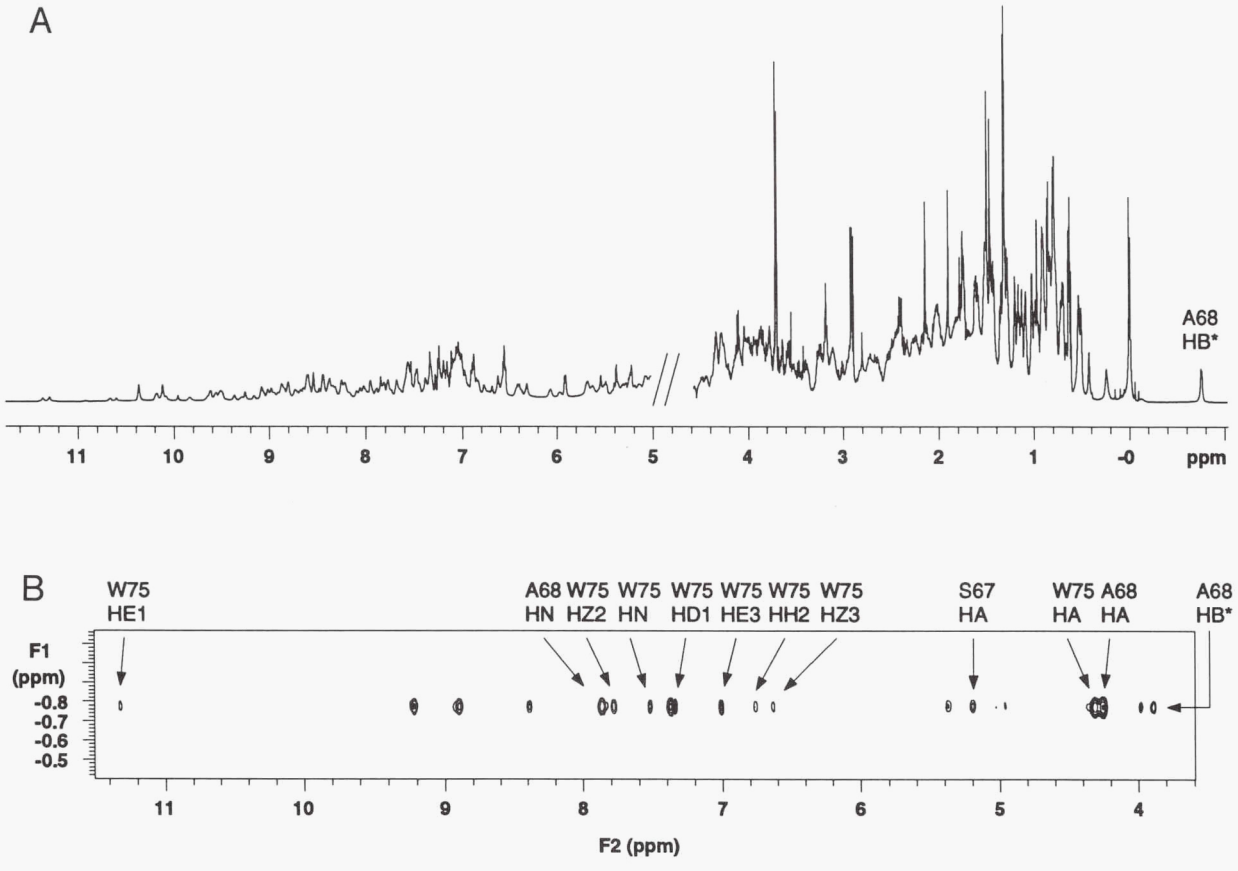

Fig. 6. ^{1}H-NMR spectra of sea raven Type II AFP at 35 °C in H_2O at pH 7.1. **A:** One-dimensional NMR spectrum of the protein showing the entire ^{1}H-resonance range. Double line at the center indicates the position of the strong water signal. **B:** Small portion of a 2D ^{1}H-^{1}H-NOESY spectrum showing the connectivities observed for the upfield-shifted resonance at −0.77 ppm. Assigned crosspeaks are labeled with the residue and proton type, as well as the sequence number of the protons involved.

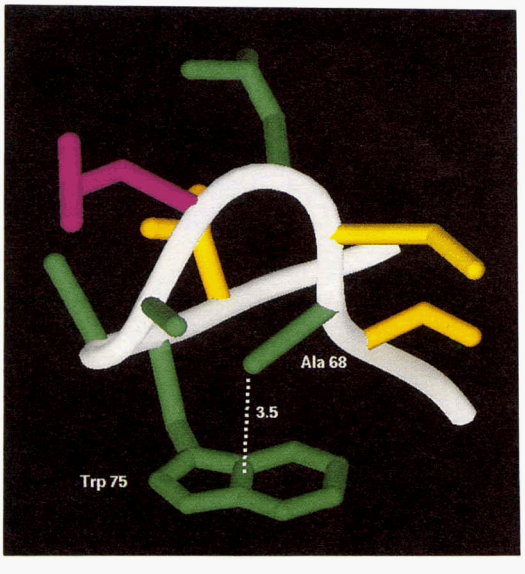

Fig. 7. Presentation of a portion of the Type II model. Backbone is shown as a smoothed tube from residues 66 to 77; all side chains are presented in a stick presentation prepared by the program RIBBONS (Carson, 1987). Hydrophobic residues are colored green, Ser and Thr orange, and Gln purple. Residues 68 and 75 are labeled close to their respective side chains, with the distance between them given in Å.

Further assignments have been made, none of which, however, show properties of the model as clearly as the example above. Without referring further to the ongoing NMR analysis, we should add that all structural information obtained so far is consistent with the model presented here.

Implications of the model

Because of the homology of Type II AFP and rat MBP-A and the small size of the proteins, their different functions (ice-binding versus sugar-binding) are more likely to be due to variations of an active site than to the evolution of distinct and separate sites in these homologues. Thus, the ice-binding site is suggested to be in the region of the triple-stranded sheet, corresponding roughly to Ca^{2+}-site 2, where sugars are bound to the lectin via Ca^{2+} coordination (see Kinemage 3). One argument to support this hypothesis is that the Type II AFPs from smelt and herring have a Ca^{2+} dependency for antifreeze activity and have retained the critical Ca^{2+}-coordinating residues (92, 94, 99, 113, and 114) for Ca^{2+}-site 2. These have been lost in the SR AFP and with them the requirement for Ca^{2+} binding. Another argument might be that ice and sugars are similar hydroxyl-rich ligands that are bound by hydrogen bonds and could therefore be accommodated with relatively minor variation of the active site residues.

Based on this assumption, it is of interest to examine the fold of the Type II model and, in particular, the differences in residues surrounding the putative ice-binding site (Kinemage 3). The structural change to loop residues 93–98 (as compared to MBP-A; Fig. 1) gives the protein an extensive, flat, exposed molecular surface. Sterically, this region seems to be ideally suited for an interaction with an ice-crystal surface plane. The loss of Ca^{2+}-site 1 (in all three AFPs) might have been expected to cause some loosening or added flexibility to the loop regions, except that the two disulfide bridges that are unique to the AFPs appear to anchor these loops to the rest of the fold. Specifically, C89–C111 links loop 3 to β-strand 4, and C69–C100 links loop 1 to the junction between β-strand 3 and loop 4. In this regard, it is significant that C100 in Type II AFP replaces one of the Ca^{2+}-site 1 residues of the lectins. Figure 8 illustrates the positions of surface-accessible side chains in this region. The majority of these 13 residues are capable of hydrogen bonding, and some, such as Thr, Asn, Gln, are residue types that have been implicated in ice-binding (Wen & Laursen, 1992; Chao et al., 1994). Ten hydrophilic residues (90, 91, 92, 94, 103, 105, 109, 113, 120, and 122) are found in a somewhat planar arrangement. They are spaced semi-regularly along all three strands with approximate inter-side-chain distances of 4–8 Å, creating an interaction area with maximum dimensions 30 × 16 Å. Due to the uncertainty in the side-chain position of the surface residues, it is not possible to deduce probable ice crystal planes that could sterically match hydrogen bonding partners at the protein surface based on residue spacing alone. However, there seem to be sufficient hydrophilic residues present to account for AFP activity. In Type I AFP, seven hydrophilic residues linearly spaced along a straight α-helix are inferred to be responsible for inhibiting ice growth. In the globular, β-sheeted Type III AFP, the mutation of three residues is sufficient to eliminate ice binding (Chao et al., 1994). There are other moderately hydrophilic surfaces elsewhere in the model, such as helix 2, but none of these other sites are as planar, nor do they have a semi-regular arrangement of hydrophilic residues.

The striking difference between SR AFP and MBP-A is in their overall hydrophobicity: 65% of the accessible surface area of SR AFP is apolar compared to 58% in MBP-A (Table 2). This makes Type II a rather hydrophobic protein with a nonpolar surface area close to those observed for water-insoluble proteins, such as crambin. However, this nonpolar surface area is not so much a consequence of the particular fold or the modeling procedure, but rather reflects the amino acid composition (Table 3). A comparison of its sequence with those of other antifreeze proteins reveals that this hydrophobicity might be a general feature of fish antifreeze proteins. The role of the hydrophobic residues is unclear. According to the adsorption–inhibition mechanism (DeVries, 1984; Knight et al., 1991), only hydrophilic residues and their hydrogen bonds are important for binding to ice. Hydrophobic residues could retard the ice from growing around the protein by repelling water from the ice lattice in their vicinity; however, it is not obvious to what extent, if at all, these residues could contribute to the affinity between protein and ice other than sterically.

This model for Type II AFP will be invaluable in assisting the structure determination by homonuclear NMR methods. It provides a working hypothesis for the location and geometry of the ice-binding site and proposes corrections to the disulfide bonding pattern, which can be tested by site-directed mutagenesis in conjunction with production of the mutant protein in an expression system. It also facilitates the design of experiments aimed at probing the structure–function relationships between C-type lectins and this class of antifreeze proteins.

Methods

Sequence alignments

Multiple sequence alignments were performed on proteins of the Ca^{2+}-dependent lectin family, Type II AFPs, and other homologues. Related sequences were systematically identified by an extensive homology search of protein sequences in the PIR data-

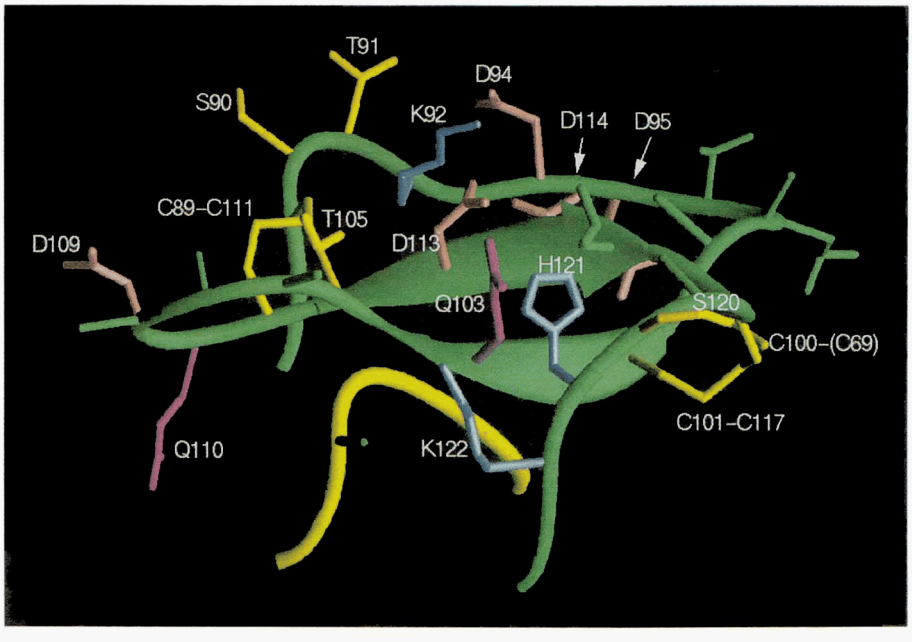

Fig. 8. Presentation of the region of the SR AFP model that is equivalent to the Ca^{2+}-binding site 2 in MBP-A. Smoothed backbone (green) and all surface-accessible side chains in this region are depicted. Loop encompassing residues 57–64 is also shown (yellow) due to its proximity and possible steric importance for this interface. Amino acid side chains are colored using the code: green, hydrophobic; (light) yellow, Cys; yellow or orange, Ser and Thr; purple, Gln; blue, Lys, His, and Arg; red, Asp and Glu.

Table 3. *Composition and ASAs of antifreeze proteins*[a,b]

	MBP-A	Type II	Type I	Type III	Average
Amino acids	112	129	37	66	
Nonpolar	58 (52%)	78 (60%)	25 (68%)	43 (65%)	(55%)
Polar	26 (23%)	30 (23%)	8 (21%)	15 (23%)	(21%)
Charged	28 (25%)	21 (16%)	4 (11%)	8 (12%)	(23%)
Ratio nonpolar to polar and charged residues	0.93	0.65	0.48	0.53	0.80
Average hydrophobicity[c]	−4.59	−0.98	5.43	4.98	−2.67
Ratio of hydrophilicity versus hydrophobicity[c]	1.41	1.09	0.59	0.69	1.22
Total ASA[a] (Å^2)	5,944	5,748	3,049	4,484	
ASA of backbone (Å^2)	734	762	197	641	
ASA of side chains (Å^2)	5,210	4,983	2,853	3,843	
Fraction nonpolar ASA	0.58	0.65	0.78	0.68	0.62 ± 0.03
Fraction polar ASA	0.21	0.24	0.09	0.21	0.21 ± 0.05
Fraction charged ASA	0.20	0.11	0.13	0.11	0.19 ± 0.05

[a] ASAs were obtained using the program ANAREA (Richmond, 1984).
[b] Sequences used in this table are the fragment MBP-A (Weis et al., 1991), Type II SR AFP (Ng & Hew, 1992), Type I AFP from winter flounder (HPLC-6), and Type III AFP (rQAEm1.1) (Chao et al., 1993). Expectation values are based on average composition of soluble proteins. Coordinates used are the crystal structure of MBP-A (Weis et al., 1992), the model structure of sea raven Type II as generated in this paper, the crystal structure of Type I AFP (Yang et al., 1988), and the NMR structure of Type III AFP (Sönnichsen et al., 1993).
[c] Determined using the Kyte and Doolittle (1982) hydropathy scale.

base (version 39). The 100 highest ranking sequence matches to each of the five query sequences (all three Type II proteins, human PSP, and rat MBP-A) were retained and combined. Approximately 90% of similar sequences obtained were identified as lectin-type proteins or protein domains. Non-lectin-type sequences and different lectin entries with only one or two mutations were excluded, as were lectins that after initial alignments did not possess a proline equivalent to Pro *186*, which has been shown to be in a *cis*-conformation in MBP-A. Subsequently, various multiple sequence alignments of 10–50 of the retained sequences were performed (by changing gap or extension penalties) and analyzed. Reliably aligned regions were defined as those where the alignment of the five query sequences was not affected by the choice of the sequences in the multiple alignments or (within limits) by the alignment parameters. All computed alignments were performed with a modified version of the Needleman and Wunsch (1970) algorithm (NW-ALIGN for exhaustive alignment searches and MULT_ALIGN for alignment of two or more sequences) as implemented in the program SEQSEE (Wishart et al., 1994a).

Modeling structurally conserved regions

The coordinates of MBP-A were obtained from the Brookhaven Data Bank (file pdb2msb.ent). This structure of MBP-A complexed with an oligosaccharide (Weis et al., 1992) is basically identical to the uncomplexed MBP-A (Weis et al., 1991), but the structure is better refined. Differences between the two copies of MBP-A in the unit cell were observed only for loop regions and the termini, i.e., regions outside of the reliably aligned areas or expected SCRs. The coordinates of copy A were arbitrarily chosen for the template structure. The model was generated using InsightII and Homology (BioSym Technologies, Inc., San Diego, California) on a Silicon Graphics workstation. Sequence stretches that did not encompass any sequence insertions or deletions, and that were always identically aligned between MBP-A and SR AFP, were defined as SCRs for this modeling study. The side chains of amino acids in these SCRs were mutated to reflect the sequence of Type II from sea raven (Hayes et al., 1989). Upon side-chain replacement, all required χ angles were initially retained as in the X-ray structure. For replacements into longer side chains, default values were used for new χ angles. Steric clashes were eliminated by either manual intervention or by an iterative side chain conformational search of neighboring side-chains based on minimizing VDW energy.

Replacement of variable loop regions

Three short loops (57–60, 94–98, and 106–111) remained as variable regions and were modeled using a knowledge-based procedure (Greer, 1990). Possible loop conformations were extracted from a structurally nonredundant selection of high-resolution atomic structures with less than 30% sequence homology (Hobohm et al., 1992) using a loop search algorithm in InsightII. Suitable loops were ranked using RMSD of the Cα positions of at least two SCR residues flanking the loop between the model and the loop candidate. Best candidates were manually evaluated using Van der Waals sphere overlap and packing errors as criteria, and the most suitable candidate was selected.

No coordinates were available in the template structure for the first 15 residues of Type II AFP because there is no counterpart in the MBP-A X-ray structure. An analysis of the structure and the presence of the disulfide bridge indicated that this region might be part of the fold (rather than an N-terminal extension or hinge region). Sterically, it could form an additional antiparallel strand in the N/C-terminal sheet. The sequence of Type II in this region, which is rich in Pro and Gly residues, does not exhibit any β-strand-forming propensities in secondary struc-

ture predictions (not shown) and could not be consistently aligned with MBP-A. Therefore, no attempt was made to provide structural information on the fold of this region.

Structural analysis

At every stage of the modeling procedure, the effects of modifying the structure (i.e., changing rotamers of side chains in the protein core, selection of loop candidates, and refinement) were analyzed using standard protein folding criteria. In particular, the fold was analyzed for the quality of packing in an attempt to reduce poor contacts or resulting cavities in the protein fold. This was accomplished by calculating residue volumes with the program VOLUME (Richards, 1974). Fractional volumes, i.e., the ratios between observed and residue-specific reference volumes, of more than 1.2 or less than 0.8 were defined as packing errors. Additionally, other criteria, such as ASAs, distribution of ϕ/ψ angles, and the number of buried residues, were evaluated and compared to expected values determined from a number of experimental high-resolution protein structures. The analysis was performed using the in-house-written software package VADAR, version 1.2 (Wishart et al., 1994b). The final structure was also evaluated with PROCHECK (Laskowski et al., 1993), using comparison values typical for proteins of 1.7 Å resolution.

Refinement of the model structure

After modifications to the protein core and modeling of the loops had been completed, the resulting model structure was subjected to a refinement procedure using a series of energy minimization routines. A procedure was developed after studying the effects of various force fields (CVFF, CFF91, and AMBER) and minimization parameters on the starting coordinates. The crystal structure of MBP-A is a high-resolution structure (1.7 Å, $R = 0.174$) and is of a very high quality when analyzed with PROCHECK or VADAR. This structure, however, was not stable during energy minimizations. We observed significant differences between the crystal structure and the resulting minimized structures (up to 1.5 Å RMSD for backbone atoms), together with significant decreases in structural quality. Possible reasons for this behavior are the simplification of calculating in vacuo in the absence of water molecules, or packing effects on the protein structure in the crystal. A comparison of identical minimizations employing different force fields, however, resulted in quite different final structures, which indicated that the force fields themselves are a major contributor to the coordinate changes. Faced with the task of having to refine the model structure, we employed the minimization procedure that best preserved the quality of the MBP-A crystal structure and showed the smallest deviations in atomic position before and after the minimization. This series of minimizations used tethering applying a harmonic potential on selected atoms positions (as described below) with a force constant of 100 kcal/Å * mol. Each minimization step was performed in vacuo, using the AMBER force field in Discover (BioSym Technologies, Inc.). After an initial 100 iterations with a steepest descent algorithm, a conjugate gradient algorithm was employed until convergence was reached. A distance-dependent dielectric constant ($\epsilon = 4 \times r$) was used to simulate the dielectric effect of water. A cutoff distance of 11 Å was used for nonbonded interactions.

In order to reduce artifacts in the calculation due to initial high local energies from the manual manipulation of the structure, the refinement was performed in several steps. The model was divided into three regions: core regions (most of the SCRs), the three modeled loops and two residues on either side, and two regions in the SCRs where the backbone conformation was expected to change due to the insertion of proline residues (80–83 and 115–119). Starting with the loops and fixing all atoms in the other regions, a series of minimizations was performed by (1) fixing backbone atoms and tethering heavy side-chain atoms to the initial coordinates, (2) free minimization of the side-chain atoms with fixed backbone, (3) minimization with tethered backbone atoms, and finally, (4) a free minimization of all atoms. Subsequently, the same procedure was repeated including the regions 80–83 and 115–119, and then for the whole model, including all core residues. In the latter, however, tethering of the backbone atoms of residues in the core regions was always employed. This was found to be necessary to avoid unduly compressed structures and to maintain proper main-chain geometry and side-chain packing.

Molecular dynamics

A molecular dynamics calculation of the refined model was performed using the AMBER force field in BioSym software, a distance-dependent dielectric constant ($\epsilon = 4 \times r$), and a nonbonded cutoff of 11 Å. No tethering or peptide bond forcing was employed. Calculations were initialized for 200 steps of 1 fs at 300 K before 100 ps of dynamics with coupling to a temperature bath and an integration time step of 1 fs were performed. Coordinates of the protein were stored at each 0.5-ps interval.

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. Type II AFP (4–12 mg) isolated from sea raven serum as described earlier (Ng & Hew, 1992) was dissolved in 100% D_2O or 90% H_2O/10% D_2O to yield 0.3–0.8 mM solutions. All spectra were acquired at 25 °C and pH 7.1 (not corrected for deuterium isotope effects). One-dimensional spectra were acquired using 24,000 data points, a spectral width of 10,000 Hz, 2,048 transients with 2.0-s presaturation delay, and a 90° pulse width of 11 μs. Data were processed with zero filling to 64K points and with exponential line broadening of 0.5 Hz. Two-dimensional experiments were acquired in the phase-sensitive mode using the methods of States et al. (1982), using 300–400 t_1 increments, 32 scans per increment, and 2,048 data points per scan with a spectral width of 10,000 Hz. DQF-COSY spectra (Piantini et al., 1982; Rance et al., 1983), NOESY spectra with a mixing time of 100 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 of 30 and 60 ms and an MLEV-17 pulse sequence (Bax & Davis, 1985) to produce a spin lock field of 7.8 kHz. The 2D data were processed on a SUN-IPC workstation using the Varian VNMR software package. Routinely, the data were zero filled to 4K × 4K data points, and shifted sine-bells were used for resolution enhancement in both dimensions.

Acknowledgments

We thank R. Boyko, S. Gauthier, G. McQuaid, and L. Willard for technical assistance, and Dr. C. Hew for the gift of sea raven serum. We are indebted to Dr. D. Wishart for many inspiring discussions and for help with the protein fold analysis. We also appreciate many helpful suggestion from Drs. K.S. Kim and K. Rajarathnam. This work was supported by research grants from the Medical Research Council of Canada (B.D.S., P.L.D.) and the Protein Engineering Network of Centres of Excellence (B.D.S., F.D.S.). Parts of this work were presented at the "Symposium on Macromolecular Structure and Function," Toronto, Canada, June 1993, and at the 14th GDCH-Hauptversammlung, Hamburg, Germany, September 1993.

References

Bax A, Davis DG. 1985. MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy. *J Magn Reson 65*:355–360.

Bettler B, Hofstetter H, Rao M, Yokoyama WM, Kilchherr F, Conrad DH. 1989. Molecular structure and expression of the murine lymphocyte low-affinity receptor for IgE (F_c-epsilon-RII). *Proc Natl Acad Sci USA 86*: 7566–7570.

Bezouska K, Crichlow GV, Rose JM, Taylor ME, Drickamer K. 1991. Evolutionary conservation of intron position in a subfamily of genes encoding carbohydrate-recognition domains. *J Biol Chem 266*:11604–11609.

Braunschweiler L, Ernst RR. 1983. Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J Magn Reson 53*:521–528.

Carson M. 1987. RIBBONS models of macromolecules. *J Mol Graphics 5*:103–106.

Chao H, Davies PL, Sykes BD, Sönnichsen FD. 1993. Use of proline mutants to help solve the NMR solution structure of Type III antifreeze protein. *Protein Sci 2*:1411–1428.

Chao H, Sönnichsen FD, Sykes BD, Davies PL. 1994. Structure-function relationship in the globular Type III antifreeze protein: Identification of a cluster of surface residues required for binding to ice. *Protein Sci 3*:1760–1769.

Chiche L, Gaboriaud C, Heitz A, Mornon JP, Castro B, Kollman P. 1989. Use of restrained molecular dynamics in water to determine three-dimensional protein structure: Prediction of the three-dimensional structure of *Ecballium elaterium* trypsin inhibitor II. *Proteins Struct Funct Genet 6*:405–417.

Chiche L, Gregoret LM, Cohen FE, Kollman PA. 1990. Protein model structure evaluation using the solvation free energy of folding. *Proc Natl Acad Sci USA 87*:3240–3243.

Colloc'h N, Etchebest C, Thoreau E, Henrissat B, Mornon JP. 1993. Comparison of three algorithms for the assignment of secondary structure in proteins: The advantages of a consensus alignment. *Protein Eng 6*: 377–382.

Davies PL, Hew CL. 1990. Biochemistry of fish antifreeze proteins. *FASEB J 4*:2460–2468.

Davies PL, Hew CL, Fletcher GL. 1988. Fish antifreeze proteins: Physiology and evolutionary biology. *Can J Zool 66*:2611–2617.

Davis DG, Bax A. 1985. Assignment of complex ^{1}H NMR spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy. *J Am Chem Soc 107*:2820–2821.

DeCaro A, Multigner L, Dagorn JC, Sarles H. 1988. The human pancreatic stone protein. *Biochimie 70*:1209–1214.

DeVries AL. 1983. Antifreeze peptides and glycopeptides in cold water fishes. *Annu Rev Physiol 45*:245–260.

DeVries AL. 1984. Role of glycopeptides and peptides in inhibition of crystallization of water in polar fishes. *Philos Trans R Soc Lond B Biol Sci B 304*:575–588.

Doege KJ, Sasaki M, Kimura T, Yamada Y. 1991. Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. Human-specific repeats, and additional alternatively spliced forms. *J Biol Chem 266*:894–902.

Drickamer K. 1988. Two distinct classes of carbohydrate recognition domains in animal lectins. *J Biol Chem 263*:9557–9560.

Drickamer K, Dordal MS, Reynolds L. 1986. Mannose-binding proteins isolated from rat liver contain carbohydrate-recognition domains linked to collagenous tails. Complete primary structures and homology with pulmonary surfactant apoprotein. *J Biol Chem 261*:6878–6887.

Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. *Nature 316*:199–203.

Ewart KV, Fletcher GL. 1990. Isolation and characterization of antifreeze proteins from smelt (*Osmerus mordax*) and Atlantic herring (*Clupea harengus harengus*). *Can J Zool 68*:1652–1658.

Ewart KV, Rubinsky B, Fletcher GL. 1992. Structural and functional similarity between fish antifreeze proteins and calcium-dependent lectins. *Biochem Biophys Res Commun 185*:335–340.

Feeney RE, Yeh Y. 1978. Antifreeze proteins from fish bloods. *Adv Protein Chem 32*:191–282.

Graves BJ, Crowther RL, Chandran C, Rumberger JM, Li S, Huang KS, Presky DH, Familetti PC, Wolitzky BA, Burns DK. 1994. Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains. *Nature 367*:532–538.

Greer J. 1990. Comparative modeling methods: Application to the family of mammalian serine proteases. *Proteins Struct Funct Genet 7*:317–334.

Halberg DF, Proulx G, Doege K, Yamada Y, Drickamer K. 1988. A segment of the cartilage proteoglycan core protein has lectin-like activity. *J Biol Chem 263*:9486–9490.

Hayes PH, Scott GK, Ng NFL, Hew CL, Davies PL. 1989. Cystine-rich Type II antifreeze protein precursor is initiated from the third AUG codon of its mRNA. *J Biol Chem 264*:18761–18767.

Hobohm U, Scharf M, Schneider R, Sander C. 1992. Selection of a representative set of structures from the Brookhaven Protein Data Bank. *Protein Sci 1*:409–417.

Janin J. 1976. Surface area of globular proteins. *J Mol Biol 105*:13–14.

Jeener J, Meier BH, Bachmann P, Ernst RR. 1979. Investigation of exchange processes by two-dimensional NMR spectroscopy. *J Chem Phys 71*:4546–4553.

Knight CA, Cheng CC, DeVries AL. 1991. Adsorption of α-helical antifreeze peptides on specific ice crystal surface planes. *Biophys J 59*:409–418.

Kumar A, Ernst RR, 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.

Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol 157*:105–132.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.

Lasky LA, Singer MS, Yednock TA, Dowbenko D, Fennie C, Rodriguez H, Nguyen T, Stachel S, Rosen SD. 1989. Cloning of a lymphocyte homing receptor reveals a lectin domain. *Cell 56*:1045–1055.

Leung JO, Holland EC, Drickamer K. 1985. Characterization of the gene encoding the major rat liver asialoglycoprotein receptor. *J Biol Chem 260*:12523–12527.

Macura S, Huong Y, Suter D, Ernst RR. 1981. Two-dimensional chemical exchange and cross-relaxation spectroscopy of coupled nuclear spins. *J Magn Reson 43*:259–281.

Miller S, Janin J, Lesk AM, Chothia C. 1987. Interior and surface of monomeric proteins. *J Mol Biol 196*:641–656.

Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. 1992. Stereochemical quality of protein structure coordinates. *Proteins Struct Funct Genet 12*:345–364.

Muramoto K, Kamiya H. 1986. The amino-acid sequence of a lectin of the acorn barnacle *Megabalanus rosa*. *Biochim Biophys Acta 874*:285–295.

Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequences of two proteins. *J Mol Biol 48*:443–453.

Ng NFL, Hew CL. 1992. Structure of an antifreeze polypeptide from the sea raven. *J Biol Chem 267*:16069–16075.

Piantini U, Sorensen OW, Ernst RR. 1982. Multiple quantum filters for elucidating NMR coupling networks. *J Am Chem Soc 104*:6800–6801.

Rance M, Sorensen OW, Bodenhausen G, Wagner G, Ernst RR, Wüthrich K. 1983. Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. *Biochem Biophys Res Commun 117*: 479–485.

Richards FM. 1974. The interpretation of protein structures: Total volume, group volume distribution and packing density. *J Mol Biol 82*:1–14.

Richmond TJ. 1984. Solvent accessible surface area and excluded volume in proteins. *J Mol Biol 178*:63–89.

Rouimi P, Bonicel J, Rovery M, DeCaro A. 1987. The disulfide bridges of the immunoreactive forms of the human pancreatic stone protein isolated from pancreatic juice. *FEBS Lett 229*:171–174.

Sander C, Schneider R. 1991. Database of homology derived protein structures and the structural meaning of sequence alignment. *Proteins Struct Funct Genet 9*:56–68.

Slaughter D, Fletcher GL, Ananthanarayanan VS, Hew CL. 1981. Antifreeze proteins from sea raven *Hemitripterus americanus*: Further evidence for diversity among fish polypeptide antifreezes. *J Biol Chem 256*:2022–2026.

Sönnichsen FD, Sykes BD, Chao H, Davies PL. 1993. The nonhelical structure of antifreeze protein Type III. *Science 259*:1154–1157.

States DJ, Haberkorn RA, Ruben DJ. 1982. A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J Magn Reson 48*:286–292.

Weis WI. 1994. Lectins on a roll: The structure of E-selectin. *Structure 2*: 147–149.

Weis WI, Drickamer K, Hendrickson WA. 1992. Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature 360*:127–134.

Weis WI, Kahn R, Fourme R, Drickamer K, Hendrickson WA. 1991. Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science 254*:1608–1615.

Wen D, Laursen RA. 1992. Structure-function relationships in an antifreeze polypeptide. *J Biol Chem 267*:14102–14108.

Wishart DS, Boyko R, Willard L, Richards FM, Sykes BD. 1994a. SEQSEE: A comprehensive program suite for protein sequence analysis. *CABIOS 10*:121–132.

Wishart DS, Sykes BD, Richards FM. 1991a. Simple techniques for the quantification of protein secondary structure by ^{1}H NMR spectroscopy. *FEBS Lett 293*:72–80.

Wishart DS, Sykes BD, Richards FM. 1991b. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol 222*:311–333.

Wishart DS, Willard L, Richards FM, Sykes BD. 1994b. *VADAR: A comprehensive program for protein structure evaluation. Version 1.2*. Edmonton, Alberta, Canada: University of Alberta.

Yang DSC, Sax M, Chakrabartty A, Hew CL. 1988. Crystal structure of an antifreeze polypeptide and its mechanistic implications. *Nature 333*: 232–237.