Acta Cryst. (1996). D**52**, 486–498

Structure Determination of a Lone α-Helical Antifreeze Protein from Winter Flounder

F. SICHERI[a] AND D. S. C. YANG[a,b]*

[a]*Department of Biochemistry, Faculty of Health Science, McMaster University, Hamilton, Ontario, Canada, L8N 3Z5, and* [b]*BioCrystallography Laboratory, PO Box 12055, VAMC, Pittsburgh, PA 15240, USA*

(*Received 14 September 1995; accepted 7 November 1995*)

Abstract

The X-ray crystal structure of a lone α-helical antifreeze protein from winter flounder has been determined to 1.5 Å using a combination of molecular-replacement and isomorphous-replacement techniques. Molecular replacement involved a multiparameter search using *X-PLOR* with two 37-mers of alanine in idealized α-helical conformations as the search models. Identified were a large number of potential solutions from which the correct solution was not distinguishable. Commitment of the top 1620 solutions to cycles of rigid-body, positional and simulated-annealing refinement identified the correct solution by a small margin in R factor. Low-resolution electron-density maps generated with phasing information from $TbNO_3$ and $LaNO_3$ derivatives were consistent with the top molecular-replacement solution. These derivatives also provided a means to filter and compare the large number of other molecular-replacement solutions with reasonable R-factor statistics. The structure-solution strategy described herein may prove useful for the determination of other relatively simple α-helical X-ray structures.

1. Introduction

Antifreeze proteins (AFP's) present in the blood of a number of species of fish provide protection against the freezing environments of polar salt waters. These proteins function by binding to ice surfaces to inhibit ice crystal growth. Understanding how AFP's bind to ice while other proteins are excluded from ice surfaces is critical for the unravelling of the protein's antifreeze mechanism. To date, the structures of AFP's from winter flounder (Yang, Sax, Chakrabartty & Hew, 1988), ocean pout (Sonnichsen, Sykes, Chao & Davies, 1993) and cod (Rao & Bush, 1987) have been analysed by NMR and X-ray crystallographic techniques. While these analyses were sufficient to define the gross folding architecture of each of the AFP's, the resolution of the structures were not sufficient to accurately define the crucial ice binding features. Previously, we reported the X-ray crystal structure of the 37-amino acid AFP from winter flounder which was determined in part through a visual interpretation of Patterson map features (Yang *et al.*, 1988). The crystal structure showed that the AFP molecule was composed of a single α-helix but the side chains and the helix termini of both AFP molecules in the asymmetric unit appeared disordered. In addition, the structure would not refine to better than the R factor of 27.4% (2.5–5 Å). Since solvent-exposed side chains are commonly found disordered in protein structures and lone α-helices are inherently unstable, particularly at the termini which show a great tendency to fray (Chakrabartty & Baldwin, 1996), we were not sure whether the status of the crystallographic refinement reflected intrinsic protein disorder or gross error in our model. We have resolved this problem by redetermining the AFP structure using a combination of molecular-replacement (MR) and isomorphous-replacement techniques. The new analysis shows that the AFP structure is in fact highly ordered and that although the original structure was correct in the tilt and the relative position of the two α-helical molecules in the asymmetric unit, significant errors were present including the direction and curvature of one of the molecules. In this work, we focus mainly on how the inherent difficulties of solving the AFP structure by molecular-replacement techniques were overcome. The difficulties arise from the helical symmetry of the AFP molecule, the packing arrangement of the two AFP molecules in the asymmetric unit and the deviation of the search model from the actual AFP structure. A detailed description of the AFP structure and a model which accounts for the protein's ice-binding ability and specificity has been presented elsewhere (Sicheri & Yang, 1995). Comparisons of the winter flounder AFP structure with the previous X-ray model and AFP models derived by molecular modelling are also made.

2. Material and methods

2.1. *Purification and crystallization*

Lyophilized semi-purified AFP, prepared from fish sera using Sephadex G75 exclusion chromatography (Hew & Yip, 1976), was kindly provided by Dr C. Hew and Dr G. Fletcher. Isoelectric focusing using a Biorad Rotofor unit was employed to fractionate AFP isoforms. 50–150 mg of semi-purified protein was applied in 6 M urea, 5%(w/v) ampholyte (pH 3–10), and focused for 5.5 h at 12 W and 278 K. Fractions from pH 5.2 to 8.5

Table 1. *Heavy-atom derivative data collection and phasing statistics*

Derivative/ Crystal No.	$LaNO_3$/1	$LaNO_3$/2	$LaNO_3$/3	$TbNO_3$/1	$TbNO_3$/2
Resolution (Å)	3.0	3.0	3.0	3.0	3.0
R_{merge}* (%)	10.2	8.2	9.2	10.0	6.5
R_{iso}† (%)	18.0	17.8	17.8	14.0	13.2
No. of observed reflections	4448	2316	2782	4869	4479
No. of unique reflections	1222	889	978	1200	1253
Completeness (%)	91.0	68.2	75.0	90.6	93.8
R_{cullis}‡ (%)	57	70	59	60	64
Phasing power§	1.82	1.5	1.55	1.54	1.52
FOM¶	0.38	0.33	0.34	0.41	0.41
No. of phased reflections	1087	845	922	1075	1158

* $R_{merge} = \sum |I_i - \langle I \rangle| / \sum I_i$ where I_i is the intensity of an individual reflection and $\langle I \rangle$ is the mean intensity of that reflection. † $R_{iso} = \sum ||F_{PH}| - |F_P|| / \sum |F_P|$, fractional isomorphous difference, where F_{PH} and F_P are the structure factors of the derivative and native data sets, respectively. ‡ $R_{cullis} = \sum ||F_H| - (|F_{PH}| - |F_P|)| / \sum |F_H|$, where F_H is the calculated heavy-atom structure factor; F_{PH} and F_P are the structure factors of the derivative and native data sets, respectively. § Phasing power = r.m.s. $|F_H|/E$. ¶ FOM = figure of merit.

The overall mean figure of merit for 1184 phased reflections is 0.76.

Table 2. *Native data collection and refinement statistics*

Space group $P2_1$		
Cell parameters (Å, °)		
277 K	$a = 38.17(7)$ $b = 37.01(6)$	$c = 21.84(2)$ $\beta = 101.1(1)$
93 K	$a = 37.41(3)$ $b = 36.56(4)$	$c = 21.35(2)$ $\beta = 103.34(6)$
Data-collection statistics	277 K	93 K
Resolution (Å)	1.7–40	1.5–40
No. of observed reflections ($>1\sigma$)	16614	28251
No. of unique reflections ($>1\sigma$)	5505	8329
Completeness (%)		
overall	81.7	89.2
1.8–1.7 Å	60.5	—
1.6–1.5 Å	—	79.3
R_{merge} (%)	4.4	5.4
Refinement parameters		
No. of molecules per asymmetric unit	2	2
No. of protein atoms	454	454
No. of water molecules	39	130
Resolution (Å)	1.7–8.0	1.5–8.0
No. of reflections ($>1\sigma$)	5429	8218
R_{factor}/R_{free}*	17.9/22.6	17.7/19.6
Deviation from ideal bond distances (Å)	0.014	0.013
Deviation from ideal bond angles (°)	2.2	1.8
Disorder regions	Asn27 molecule 1	None

* $R_{factor} = \sum |F_o| - |F_c| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure-factor amplitudes, respectively. Before commencing extensive crystallographic refinement, 10% of F_o were randomly selected and set aside for monitoring R_{free}.

which were shown by analytical reverse-phase high-pressure liquid chromatography (HPLC) to contain the two main AFP isoforms, HPLC6 and HPLC8, were then pooled. The two isoforms were resolved by preparative reverse-phase HPLC and lyophilized (analytical column = Waters μBondapak C_{18}, 300 × 7.1 mm internal dimensions; preparative column = Waters DeltaPak C_{18}, 300 × 19 mm internal dimensions; gradient = 15–40% acetonitrile against 0.5% aqueous trifloroacetic acid). The HPLC6 AFP isoform was crystallized at 277 K as reported (Yang, Chung, Chen, Rose & Hew, 1986). Many crystals diffracting to greater than 1.5 Å were obtained attributable in part to increased protein purity.

2.2. *Data collection*

Data sets were collected at 277 K on crystals mounted in capillary tubes using a Rigaku R-AXIS IIc area detector with Cu $K\alpha$ radiation (Rigaku RU200 generator, 0.2 mm cathode, operating at 3 kW with Supper double focusing mirror optics). Data was processed using software provided by Rigaku. Isomorphous $LaNO_3$ and $TbNO_3$ heavy-atom derivatives were prepared by soaking crystals in 50 m*M* acetone solutions for 4–12 h. Both derivatives were found to bind to the same two sites with the same relative occupancies. A 93 K native data set diffracting to greater than 1.5 Å was collected using a rayon loop mounted crystal with butanol as the cryoprotectant. The crystal temperature was maintained using a Molecular Structure Corp. cooling device. Data-collection parameters and statistics are provided in Tables 1 and 2.

2.3. *Initial phasing*

Structure determination was approached using both molecular-replacement and isomorphous-replacement techniques. Due to problems with heavy-atom solubility, non-isomorphism and the limited number of targetable groups on the AFP, good heavy-atom derivatives proved difficult to obtain. Two suitable derivatives were found late in the structure determination after screening 20 different compounds. In the end, the structure was solved using combined information from molecular-replacement and isomorphous-replacement techniques although it appears that phasing information from the former source alone could have been used to obtain the correct structure.

Molecular-replacement searches with *X-PLOR* (Brünger, 1992*a*) and *MERLOT* (Fitzgerald, 1988) failed to produce unique rotation and translation solutions using a 37-mer of alanine in an α-helical conformation ($\varphi = -57°$, $\psi = 47°$) as a search model. In addition, self-rotation studies failed to identify a non-crystallographic symmetry axis relating the two AFP molecules in the asymmetric unit. In the rotation searches using *X-PLOR*, Patterson correlation (PC) refinement of the top rotation-function peaks revealed 22 strong orientations ($>7\sigma$), all of which possessed the same helix axis tilt with respect to the crystallographic axes. Two groups of 11 solutions differed in the direction of the AFP molecules (referred to as amino terminus up or down with respect to the y axis) and the 11 solutions in each group were related by a 31° rotational periodicity about the helix axis. None of the rotation solutions were

discernibly better at any of the resolution shells tested and this rotational ambiguity could not be broken by adding side chains in their statistically favoured conformations to the search model.

Using many of the rotation solutions, translation searches (on the xz plane) were carried out. Rather than two solutions corresponding to the positions of the two molecules in the asymmetric unit, families of solutions appearing as two distinct ridges ($z = -x + \frac{1}{2}$ and $z = -x + \frac{1}{4}$) parallel to the projection of the helix axis on the xz plane were observed (Fig. 1). The relative height of the two ridges and the position of peaks on each ridge varied significantly with the resolution limits used.

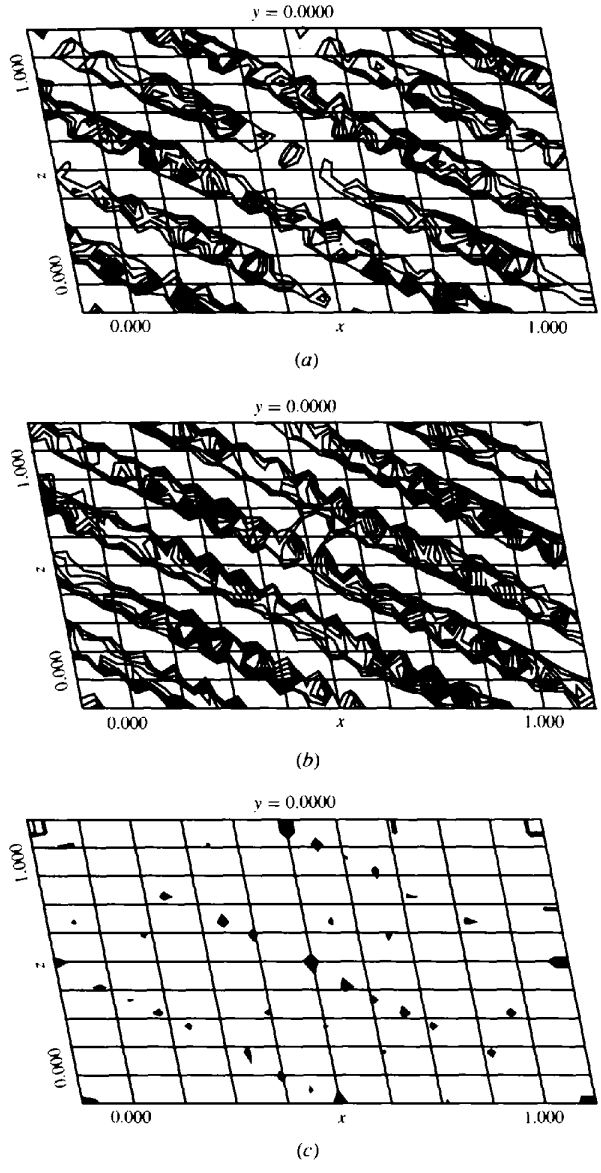

Fig. 1. Representative contour plots of single-molecule translation searches using *X-PLOR*. (*a*) 277 K native data set, 3–15 Å resolution, 37-mer of alanine as a search model. (*b*) Ideal data set, 3–15 Å resolution, 37-mer of alanine as a search model. (*c*) Ideal data set, 3–8 Å resolution, ideal search model. The ideal data set (F_{calc}) was generated for two idealized AFP molecules (all side chains included) with center of masses positioned at fractional coordinated (0,0,0) and (0,0,0.5). Plots are contoured at 70 to 100% of the total PC (Patterson or standard correlation coefficient using the square of the normalized structure factors) variance.

In order to determine the origin of the observed rotation and translation ambiguities and possible means to circumvent them, structure-factor amplitudes were calculated for two AFP molecules (built with extended side-chain conformations) positioned in the experimental unit cell with rotations and translations selected from our single-molecule search results. Rotation and translation searches were then performed using the polyalanine search model and the same ambiguities found with the experimental structure-factor amplitudes were observed. Correct rotation and translation solutions were only distinguishable at certain resolution ranges when all side chains were included in the search model (Fig. 1). Since few of the winter flounder AFP side-chain conformations (only threonine residues) could be confidently predicted based on statistical preferences in protein structures (McGregor, Islam & Sternberg, 1987) or conformational energy calculations (Piela, Nemethy & Scheraga, 1987), we concluded that the rotation and translation ambiguities observed with the experimental structure-factor amplitudes could not be resolved using a conventional molecular-replacement approach.

A combined two molecule search was then attempted. At this stage, the helical tilt of the AFP molecules were confidently known but ambiguity still existed in the rotation of each AFP molecule about its helical axis, the direction of each molecule (up or down), the translation of each molecule on the xz plane (restricted to the diagonal ridges) and the relative positioning of the two AFP molecules along the y axis. To reduce the number of search parameters further, we decided to hold fixed the rotation of each AFP molecule about its helical axis. Since side chains were absent from our search model and translation and rotation are coupled in a α-helical structure, we reasoned that an error in the rotation parameters would result in small errors in backbone register that could be corrected during structure refinement.

Three families of packing arrangements are possible for two rod-like molecules (55 Å long and 8 Å in diameter) considering the rotation and translation results of the one molecule searches. These are shown in Fig. 2. In case *A*, the two AFP molecules of the asymmetric unit lie on the diagonal ridges containing the crystallographic twofold axes while in cases *B* and *C*, the AFP molecules lie on the ridges between the axes. The 8 Å diameter of an α-helical structure rules out a packing arrangement in which one AFP molecule lies on the diagonal ridge containing the twofold axes while the other lies on the diagonal ridge between the twofold axes. The possible helix direction combinations are up–up,

up–down, down–up and down–down. Since the up–down and down–up orientations of the AFP molecules are equivalent, related by a shift in origins on the xz plane, the latter orientation of AFP molecules was neglected. Combination of the three packing arrangements and the three distinct helix-direction combinations leads to nine distinct subfamilies.

For each of the nine subfamilies, a multiparameter search was performed with *X-PLOR* using the PC and *R*-factor target function (15–3 Å). The parameters searched included the diagonal slide of each AFP molecule along the xz plane and the translation of one molecule along the y axis. The translation space sampled on the xz plane and the starting positions of the two molecules after each was tilted appropriately are shown in Fig. 2. In all searches, AFP molecule one was translated along its diagonal for a distance covering the asymmetric unit (24 Å) while molecule two was translated along its diagonal for a distance covering the full unit cell (48 Å). Molecule two was also translated along the y axis for a full unit-cell length (37 Å). These search lengths covered all possible translation space for a particular packing arrangement. A 1 Å grid was used for all parameters searched.

The result of each subfamily search was a featureful three-dimensional map from which a clear solution was not apparent (Fig. 3). Using the *R*-factor target function rather than the PC function gave qualitatively similar three-dimensional maps but map features were less pronounced. A large number of the multiparameter solutions were then refined against *R*-factor in the hope that this would help to differentiate which was correct. For this purpose, the three-dimensional map (PC target function) from each subfamily search was clustered by treating them as electron-density maps and performing peak searches with *PSRCH* from the *PHASES* package (Furey & Swaminathan, 1990).

The top 180 clustered solutions from each of the nine subfamilies were taken and committed to 140 cycles of rigid-body minimization, 40 cycles of positional refinement and one cycle of simulated annealing (SA). Results and statistics after each stage are shown in Fig. 4 and Table 3. After the SA stage, one solution from subfamily 3 was discernibly better with an *R* factor of 30.2%. In comparison, the next best solutions possessed *R* factors of 31.4, 31.7 and 32.1%. Initially, and after rigid-body and positional-refinement, the top solution was ranked 40th, 11th and second overall. While the top SA solution proved to be the correct solution (proof follows in subsequent paragraphs), the margin in *R* factor was not very large. Within the *R*-factor range 30.2–33.2%, there were 44 solutions representing all nine packing subfamilies. The electron-density maps of a number of very different solutions were equally reasonable and in all cases, including the top solution, side-chain features were not apparent. As a result, we were not confident that the top solution was in fact the correct solution.

At this time, two related isomorphous derivatives were found (Table 1). Since both the $LaNO_3$ and $TbNO_3$ derivatives bound to the same two sites in the asymmetric unit (major site coordinates = 0.37, 0.17,

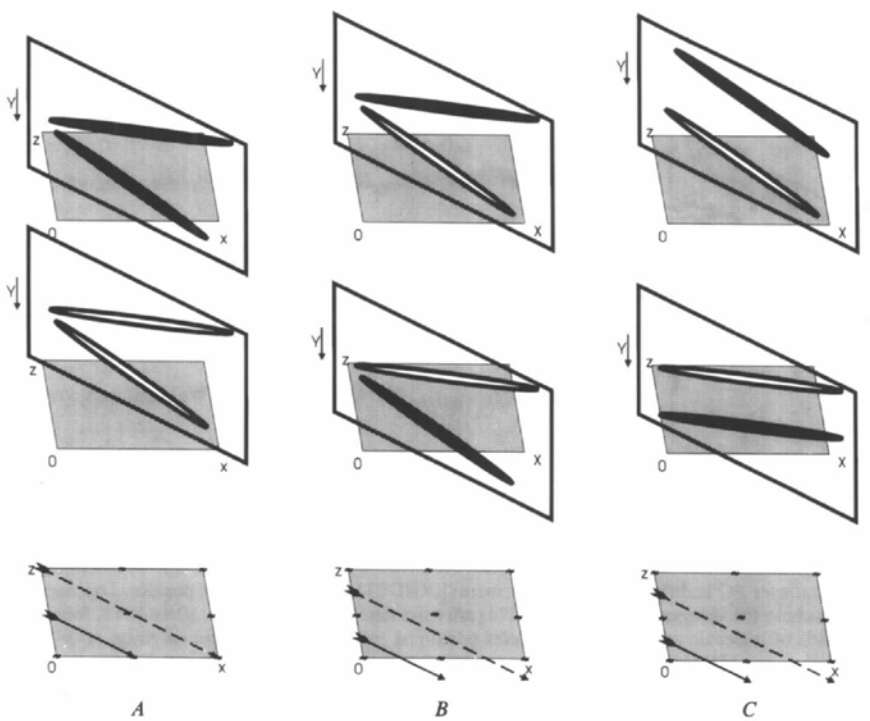

Fig. 2. Hypothetical packing families *A*, *B* and *C*. The three families are consistent with the packing constraints imposed by two rod-like structures and the single-molecule rotation- and translation-search results. The top two rows of panels display the packing arrangements of the two molecules in the asymmetric unit (shown as black and white rods) in cross sections perpendicular to the xz plane and containing the $\langle 10\bar{1} \rangle$ vector. Lower row of panels displays the translation space searched for the two AFP molecules on the xz plane. Arrow heads and tails correspond to the beginning and the end positions searched with a 1 Å grid. Molecules 1 and 2 are denoted by solid and dashed arrows, respectively.

0.55; minor site coordinates = 0.06, 0.00, 0.18) with the same relative occupancies (1.0 and ~ 0.6), the phase information provided by each was not significantly different. Heavy-atom positions were identified using the program *HASSP* (Terwilliger, Kim & Eisenberg, 1987) and SIRAS phases were calculated and refined using the *PHASES* package (Furey & Swaminathan, 1990). This was followed by 16 cycles of solvent flattening and then phase extension as described by Wang (1985) and implemented in the *PHASES* package. An electron-density map was produced which clearly demonstrated the packing arrangement of the AFP molecules. As shown in Fig. 5, this packing arrangement was consistent with the top molecular-replacement solution. However, since the quality of the electron-density map was limited, no additional features could be identified.

The finding of a reasonable heavy-atom derivative also provided a convenient and powerful method to filter and analyse the large number of SA-refined solutions that possessed reasonable *R*-factor statistics. The top ten solutions of each subfamily after SA refinement were used to phase the amplitudes of one of the $TbNO_3$ derivatives. The solutions that revealed the heavy-atom positions as major peaks in difference Fourier maps were origin correlated and viewed. Five of the ten solutions from subfamily 5, including the top solution, and one solution from each of subfamilies 2, 7 and 9 identified the heavy-atom positions. In all cases, the backbone structure of the AFP molecules superimposed exceptionally well but variability existed in backbone register (Fig. 6). Although the solutions from subfamily 7 and 9 changed packing arrangements through a 4 Å shift of AFP molecules during rigid-body refinement, other solutions from the same two subfamilies maintained incorrect packing arrangements, displayed reasonable refinement statistics but were effectively filtered by the heavy-atom method. The superimposed structures also revealed that one of the two AFP molecules in the asymmetric unit possessed significant curvature and that all but one were oriented in the same direction (up–up).

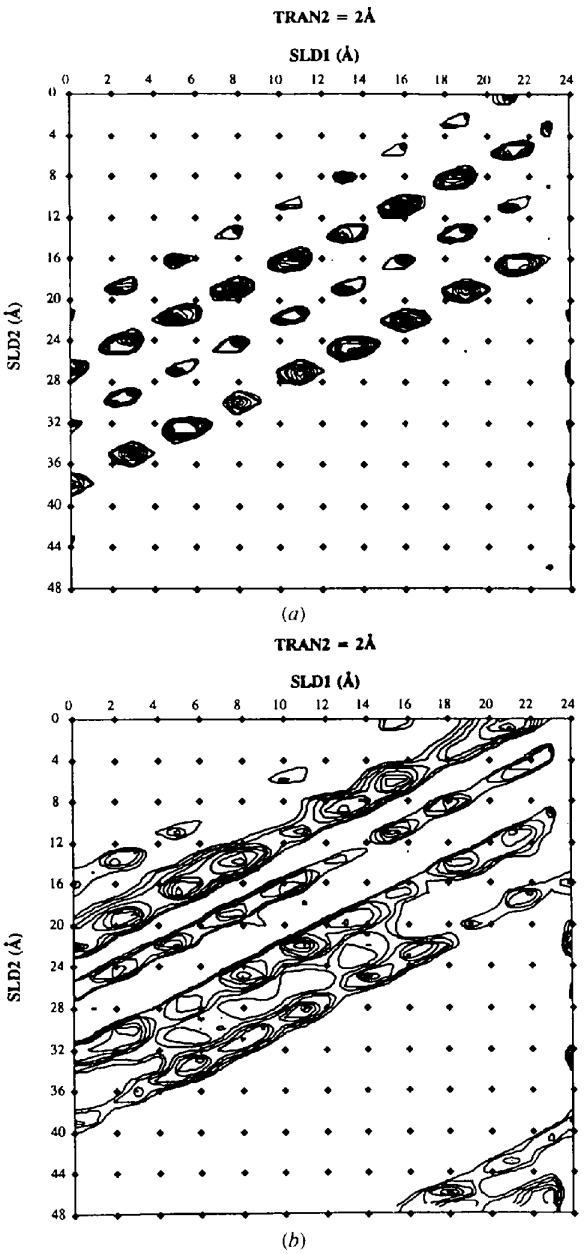

Fig. 3. Representative contour plot of (*a*) multiparameter PC and (*b*) *R*-factor search results. Parameters searched include the diagonal slide of AFP molecule one (SLD1), diagonal slide of molecule two (SDL2), and the translation of molecule 2 along the *y* axis (TRAN2). Distances searched were 24, 48 and 37 Å, respectively, using 1 Å increments.

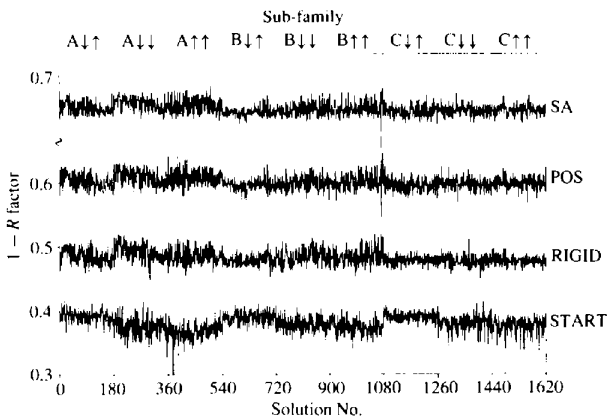

Fig. 4. Refinement results of the top subfamily solutions. From bottom to top, respectively, plots of $1 - R$ factor *versus* solution number at stages (1) before refinement (START), (2) after 140 cycles of rigid-body refinement (RIGID), (3) after 40 cycles of positional refinement (POS) and (4) after simulated annealing [(SA), 3000–300 K followed by 40 cycles positional refinement]. A resolution range of 3–15 Å with a 2σ cut off was employed. Subfamilies are denoted by their packing family *A*, *B* or *C* (see Fig. 2) and the helix directions ↑↑, ↑↓ or ↓↓. *R*-factor statistics are given in Table 3.

Table 3. *R-factor statistics following various stages of subfamily solution refinement*

Ave., average; SD, standard deviation; and Best, best *R*-factor statistics following the stages of subfamily solution refinement. See also Fig. 4.

		Subfamily								
		A↓↑	A↓↓	A↑↑	B↓↑	B↓↓	B↑↑	C↓↑	C↓↓	C↑↑
START	Ave.	0.607	0.626	0.631	0.609	0.622	0.625	0.609	0.62	0.624
	SD	0.011	0.017	0.016	0.008	0.013	0.013	0.01	0.013	0.015
	Best	0.578	0.584	0.597	0.59	0.573	0.588	0.584	0.562	0.597
RIGID	Ave.	0.514	0.508	0.515	0.522	0.516	0.516	0.521	0.523	0.523
	SD	0.013	0.014	0.015	0.007	0.013	0.015	0.01	0.009	0.008
	Best	0.485	0.478	0.468	0.502	0.484	0.47	0.482	0.488	0.497
POS	Ave.	0.397	0.391	0.392	0.405	0.402	0.4	0.404	0.403	0.402
	SD	0.013	0.013	0.014	0.01	0.012	0.015	0.011	0.009	0.01
	Best	0.366	0.363	0.351	0.375	0.373	0.356	0.369	0.372	0.369
SA	Ave.	0.353	0.347	0.349	0.361	0.358	0.357	0.361	0.359	0.359
	SD	0.011	0.011	0.014	0.01	0.011	0.013	0.008	0.008	0.008
	Best	0.327	0.327	0.302	0.333	0.328	0.323	0.334	0.332	0.333

Together, the *R*-factor statistics, the low-resolution single isomorphous replacement with anomalous scattering (SIRAS) electron-density map and the filtered and superimposed SA refined MR solutions gave us confidence in the approximate positioning of the two AFP molecules in the unit cell, the backbone curvature of one AFP molecule and the direction of both AFP molecules. The precise register of each AFP molecule was still in doubt since the rotation of the molecules about their helical axes were not sampled in the search routine, the register was not evident in the SIRAS electron-density map and

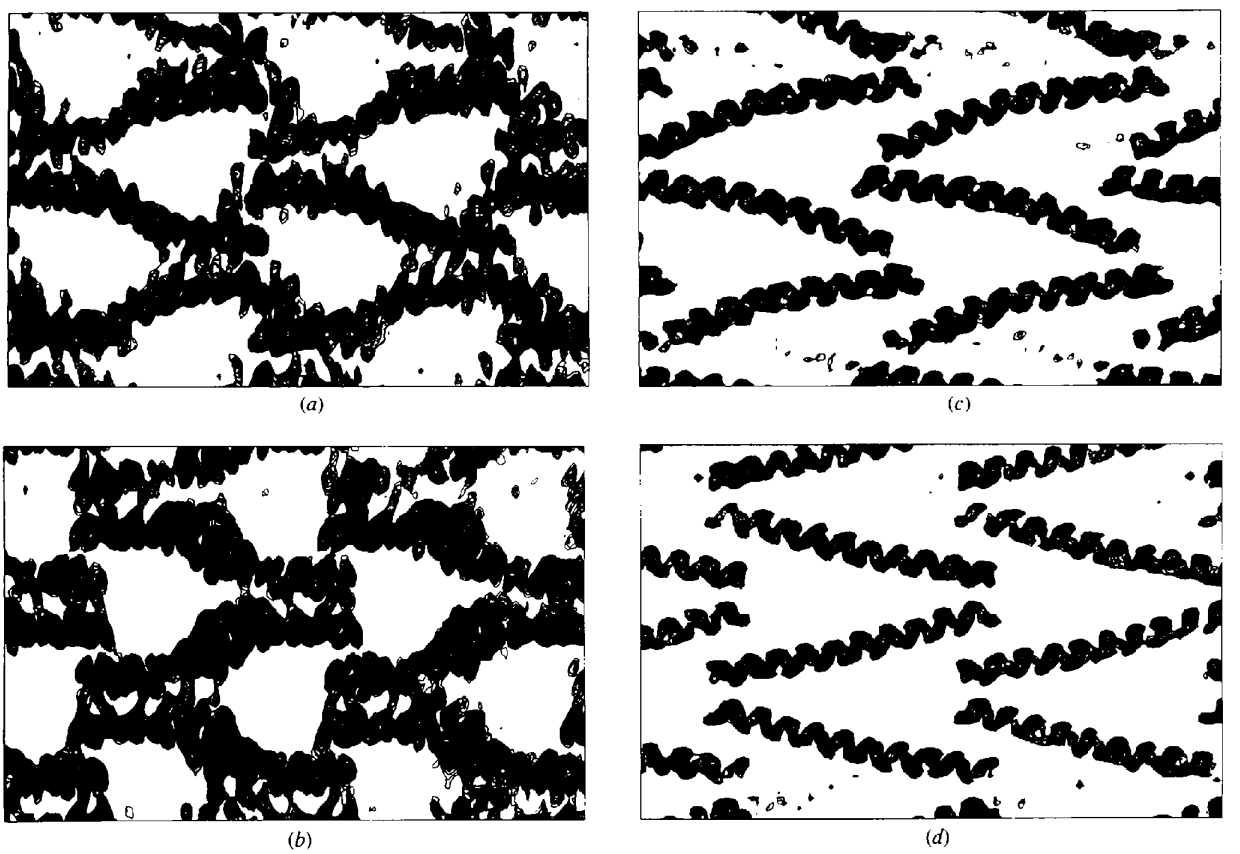

Fig. 5. Cross sections through a SIRAS (FOM $\times F_o$) electron-density map showing AFP molecule one (*a*) and two (*b*). Similarly, cross sections through an electron-density map (F_{calc}) of the top MR solution after SA refinement (*c*) and (*d*). The unique crystal axis *y* is vertical and the diagonal axis $x = -z$ lies horizontal. The origins of the SIRAS and MR maps were correlated for direct comparisons. Distortions in the SIRAS map arise from difficulties in the automatic boundary determination of the solvent-flattening routine. Use of the top MR solution for boundary calculation produced much improved maps. Resolution for all maps is 3–15 Å.

different registers were apparent in the superimposed MR solutions. Since threonine side chains are known to strongly favour a single conformation on α-helical structures (Piela, Nemethy & Scheraga, 1987; McGregor, Islam & Sternberg, 1987), and four are present in the flounder sequence (see Fig. 9 legend), we attempted to use this to resolve the remaining ambiguity. Threonine side chains were added at seven different registers to each AFP molecule in the top molecular-replacement solution. The seven registers tested corresponded to shifts from the top MR solution by up to three residues in either the amino- or carboxy-terminal directions. After four cycles of SA refinement at a higher resolution range (2.0–10 Å, 3000–300 K), the correct threonine placement on each AFP molecule was identified by a small margin in R factor (Table 4). Confirming the register selection, most of the remaining side chains were clearly apparent in $2|F_o| - |F_c|$ electron-density maps (Fig. 7).

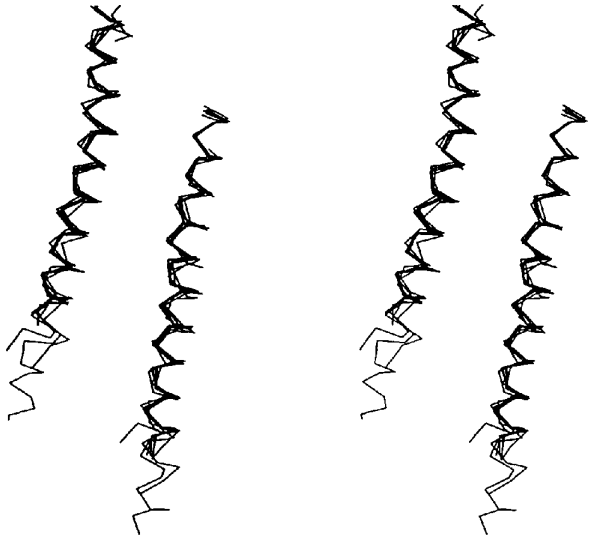

Fig. 6. Stereo plot of the origin correlated MR solutions.

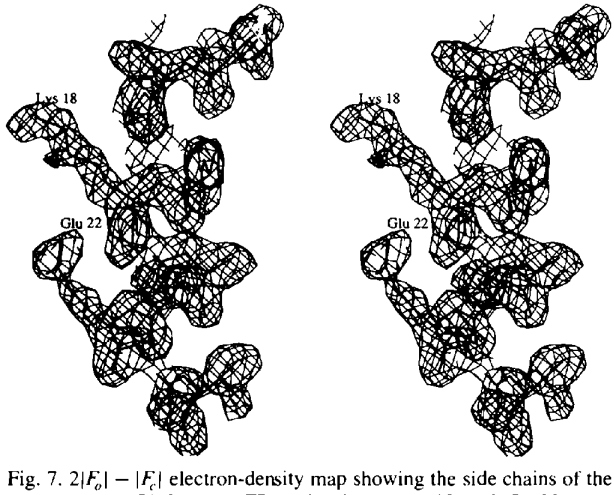

Fig. 7. $2|F_o| - |F_c|$ electron-density map showing the side chains of the Lys18 and Glu22 on AFP molecule 1. (Lys18 and Glu22 were omitted from the phase calculation for this map.) Such unbiased side-chain features confirmed the backbone register of each AFP molecule.

Table 4. *Identification of AFP backbone registers by crystallographic refinement of systematically placed threonine residues*

Register	R-factor
	Molecule 1
−3	30.90
−2	30.81
−1	30.46
0	30.10
+1	30.24
+2	30.43
+3	30.60
	Molecule 2
−3	30.37
−2	30.60
−1	30.43
0	29.41
+1	30.59
+2	31.43
+3	31.02

Notes: the side chains of threonine 2, 13, 24, and 35 were added to the backbone structures of the top SA-refined MR solution in seven different registers per AFP molecule. The seven registers correspond to shifts of −3 to +3 residues towards the amino- and carboxy-termini, respectively. In the +3, +2 and −2, −3 registers, the side chains of Thr35 and Thr2, respectively, were not included. Four cycles of SA refinement (3000–300 K) were performed on the 14 model structures using a resolution range of 2–10 Å. The same crystallographic scale factor was employed for each case.

2.4. *Refinement*

Ten percent of the structure-factor amplitudes were than partitioned for the analysis of R_{free} (Brünger, 1992*b*). After the addition of the remaining side chains, cycles of SA refinement and model rebuilding using *O* (Jones, Zou, Cowan & Kjeldgaard, 1991) were performed. SA omit maps were used to confirm all regions of the structure and all but the side chain of Asn27 in one AFP molecule appeared ordered. Resolution was then increased from 2.0 to 1.7 Å using cycles of SA refinement. Solvent was picked using the automated water-selection program *ASIR* (Tong, Berghuis, Chen & Luo, 1994) modified for use with *X-PLOR* and *O*. Only water molecules forming greater than one hydrogen bond were selected and a B-factor cutoff of 55 Å^2 was employed. 39 water molecules were included before R_{free} started to increase. A restrained isotropic B-factor refinement was performed on the final structure. The protein structure at 277 K without solvent was then used as the starting model for the

93 K structure. After cycles of rigid-body and SA refinement, SA omit maps were used to correct problem regions of the structure. Resolution was extended to 1.5 Å and solvent was picked as described for the 277 K structure. 130 water molecules were included before R_{free} increased. Well defined solvent molecules other than water were not apparent. Both AFP molecules in the 93 K crystal structure were completely ordered. A summary of the refinement statistics are given in Table 2. *B*-factor and Ramachandran plots for the 277 K and 93 K structures are shown in Fig. 8. The side-chain disorder of Asn27 in molecule one at 277 K is evident in the *B*-factor plot. All four AFP molecules are completely α-helical with the exception of the last peptide unit in each which adopts a 3_{10} conformation. The structure of the four AFP molecules at 277 K and 93 K are shown in Fig. 9.

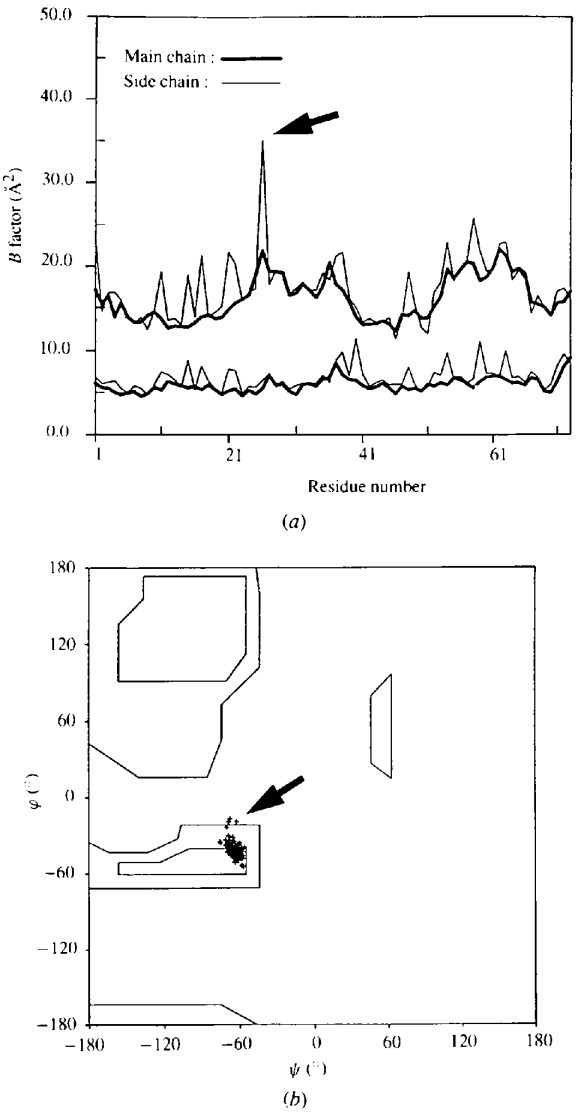

Fig. 8. (*a*) Residue averaged *B*-factor plot for the 277 K (top) and 93 K (bottom) winter flounder AFP crystal structures. Molecule one and two correspond to residues 1–37 and 38–72, respectively. Arrow shows disordered side chain of Asn27. Average solvent *B* factors for the 277 K and 93 K structures are 43.0 and 33.0 Å^2, respectively. (*b*) Ramachandran plot for the four AFP molecules of the 277 K and 93 K crystal structures. The four outlying points shown by arrow correspond to the 3_{10} conformations which terminate each AFP molecule.

Table 5. *Intermolecular hydrogen-bonding interactions of the 277 K crystal structure*

Residues 1–37 and 101–137 correspond to AFP molecules one and two of the asymmetric unit, respectively. Similar interactions are observed in the 93 K crystal structure.

Residue	Atom	Residue	Atom	Distance (Å)
Asp1	N	Glu122	OE2	3.0
Asp1	OD2	Lys118	NZ	3.1
Ala3	O	Arg37	NT1*	2.9
Asp5	OD1	Arg37	NH1	2.9
Asn16	OD1	Arg137	NE	3.0
Lys18	NZ	Thr135	OG1	3.3
Glu22	OE1	Asp101	N	2.7
Arg37	NE	Asn116	OD1	3.1
Asp101	OD1	Arg137	OT2*	3.1
Asp105	OD2	Arg137	NH1	2.9

* Non-standard nomenclature refers to the amidated carboxy terminus.

The key features that are apparent in the AFP crystal structure include a Lys18–Glu22 intrahelical salt bridge, elaborate N- and C-terminal cap structures and a flat, rigidly constrained ice-binding surface. A detailed description of these features and their relevance to AFP function have been presented elsewhere (Sicheri & Yang, 1995).

3. Results and discussion

3.1. *Crystal packing arrangement*

Crystal packing can be visualized as the repeat of two distinct sheets of AFP molecules which lay perpendicular to the *xz* plane and contain the ⟨101⟩ vector. Each sheet is composed of one of the AFP molecules in the asymmetric unit (see Fig. 5, packing family *A*). Within a sheet, an AFP molecule makes head-to-tail contacts with its symmetry-related partners while across the sheets, each AFP molecule makes contacts with four non-symmetry-related molecules. Head-to-tail contacts involving electrostatic and hydrogen-bonding interactions are listed in Table 5. AFP molecules on adjacent sheets form wide crossing angles that vary from 26 to 32° due to differences in the local backbone curvature in each AFP molecule. Intermolecular contacts between AFP molecules on different sheets involve electrostatic, hydrogen bonding and hydrophobic interactions. Key electrostatic and hydrogen-bonding interactions are listed in Table 5 while the four relatively hydrophobic crossing sites between adjacent helices are shown in Fig. 10.

All of the AFP's side-chain structure is solvent exposed and can thus be potentially affected by crystal-packing interactions. Since significant non-isomorphism is observed between the 277 K and 93 K crystal structures, four AFP molecules are available to discern biologically relevant structure from crystal imposed artifacts. Regarding the protein's helix-stabilizing structure, most of the N- and C-terminal cap structure is highly conserved. However, crystal-packing interactions involving the side chain of Asp1 in three out of four AFP molecules, gives rise to a slightly different N-cap structure shown in Fig. 11. Additionally, in three out of four molecules, the Lys18 side chains are involved in crystal-packing contacts which prevent intra-molecular interactions with the Glu22 side chains. With respect to the proteins ice-binding structure, in five out of eight instances (two leucine residues per AFP molecule) crystal-packing interactions involving the leucine side chains prevent intramolecular interactions with adjacent asparagine side chains. Furthermore, the conformation of Thr35 in two molecules and Asn27 in one are distorted by crystal-packing interactions.

The packing arrangement of the AFP molecules with relatively small contact patches and wide crossing angles between any two AFP molecules does not suggest an intermolecular association that may be relevant in solution or once the AFP is bound to the surface of ice. This is in agreement with solution studies which shows that the AFP has little tendency to aggregate (Chakrabartty, Ananthanarayanan & Hew, 1989) and with estimates of AFP surface densities on ice based on ellipsometry measurements (Wilson, Beaglehole & DeVries, 1993).

3.2. *Comparison to previous X-ray crystal structure*

In the previous X-ray crystallographic analysis of the winter flounder AFP structure (Yang *et al.*, 1988), Patterson map features were visually interpreted to identify the tilt of the protein molecules and the relative positioning between the two AFP molecules in the asymmetric unit. In order to avoid a bias towards the previous structural model which failed to refine, the visual interpretations were not taken into account in the present structure determination. Our crystal structure now confirms that the previous interpretations of Patterson map features were in fact valid. It appears that the previous model failed to refine due to the difference in curvature between the search model and one of the AFP molecules in the asymmetric unit. This difference was not surmountable by the positional refinement routines previously employed but were resolved by the *X-PLOR* SA-refinement routine employed in the present study. Overall, the packing arrangement of the AFP molecules in the previous

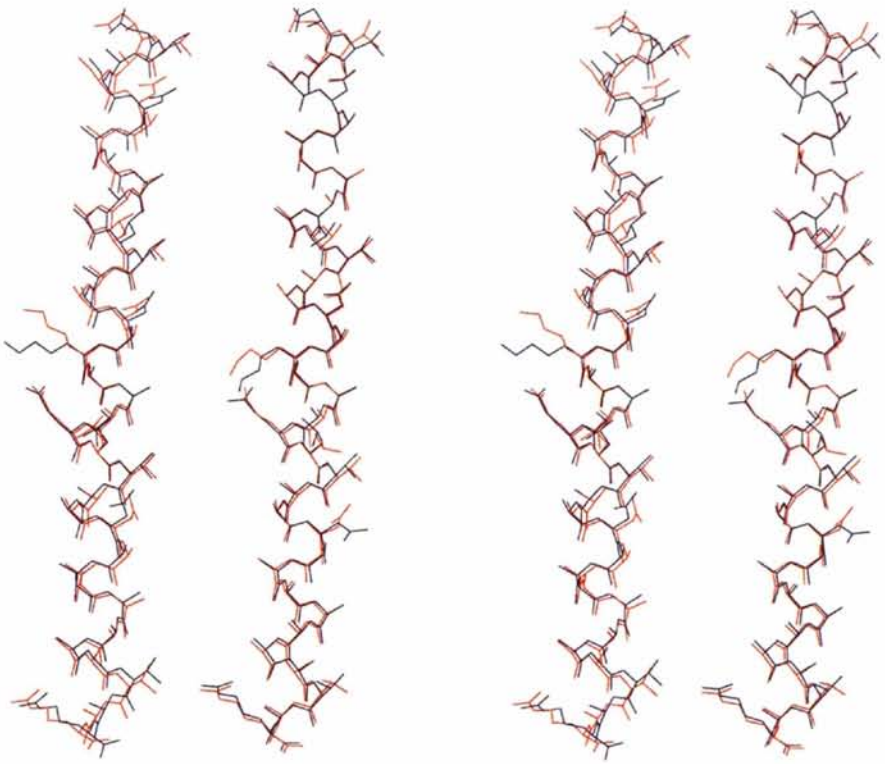

Fig. 9. Stereo plot of the winter flounder AFP molecules. Red and blue colouring denote the 277 K and 93 K structures, respectively. AFP molecules one are shown on the left and all molecules are aligned with amino termini pointing up. Average r.m.s. difference between the equivalent molecules at 277 K and 93 K is 0.26 Å. Average r.m.s. difference between non-equivalent molecules at 277 K and 93 K = 1.52 Å. The AFP amino-acid sequence is DTASD-AAAAAALTAANAKAAAELT-AANAAAAAAATAR-NH2 (Davies, Roach & Hew, 1982).

determination is similar to that of the present structure (Fig. 12). However, the backbone curvature of one AFP molecule, the direction of the other and the backbone register of both molecules differed significantly. Reflecting these differences, the previous X-ray model was unable to locate the $LaNO_3$ or $TbNO_3$ positions in difference Fourier maps.

3.3. *Comparisons to modeling studies*

The low-resolution image of the AFP provided by the first crystallographic analysis (Yang *et al.*, 1988) spawned numerous modeling studies to elucidate side-chain conformations that gave the AFP its ice binding ability and exceptional α-helical stability. These included simpler studies of the AFP in vacuum or in solution (Chou, 1992; McDonald, Brady & Clancy, 1993; Jorgenson *et al.*, 1993) and more complex studies of the AFP on the surface of ice (Wen & Laursen, 1992; Lal, Clark, Lips, Ruddock & White, 1993; Madura *et al.*, 1994). Since these studies have had a great influence on the current thinking of how AFP's bind to ice, comparisons with the high-resolution AFP crystal structure are warranted. What follows is a brief look at the salient features of the simpler and then the more complex theoretical studies.

With regard to the AFP's helix-stabilizing features, the salt-bridge interaction between the side chains of Lys18 and Glu22 was reproduced in two of the three studies (Jorgenson *et al.*, 1993; McDonald *et al.*, 1993). These two studies were also able to reproduce part of the protein's N-terminal cap structure. All three studies failed to reproduce the C-terminal cap structure and this may be attributed in part to how the carboxy terminus of the AFP was modeled. Jorgenson *et al.* (1993) analyzed an immature form of the AFP which contained an additional

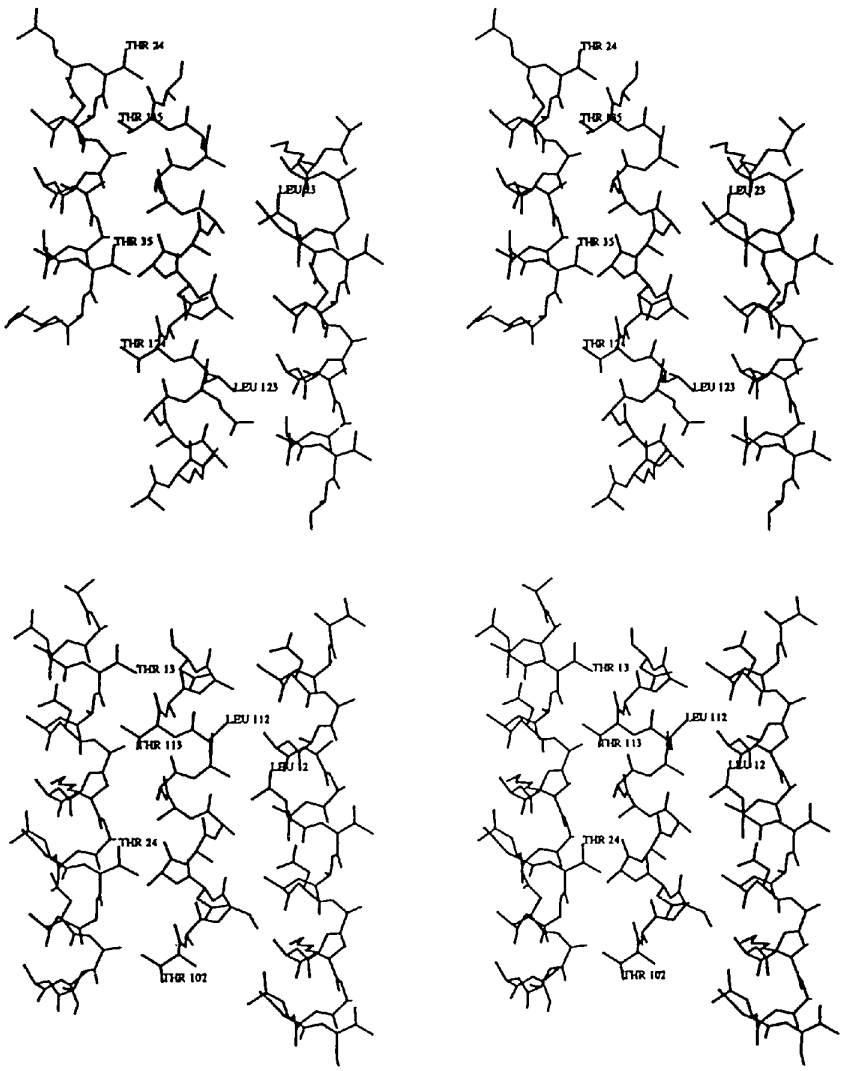

Fig. 10. Stereo plot of the hydrophobic crossing sites between non-symmetry-related AFP molecules. 621, 777, 941 and 796 Å^2 of surface area are buried in the four contact points. Intermolecular hydrogen-bonding contacts are listed in Table 4.

glycine residue rather than the amidated carboxy terminus of the mature AFP (Hew *et al.*, 1986). In the studies of Chou (1991) and McDonald *et al.* (1993), the amide modification of the carboxy terminus was neglected. In the crystal structure, the amidated carboxy terminus appears to play a critical role in defining the C-cap structure. With regard to the AFP's ice-binding structure, all three studies accurately reproduce the conformation of the threonine side chains but in most instances asparagine and aspartate side chains were found to adopt unrestrained extended conformations which differed from the contracted and restrained conformations observed in the crystal structure. The study of McDonald *et al.* (1993) was unique in observing that the AFP's backbone structure had a predisposition towards bending. The curvature of AFP molecule 1 in the 277 K and 93 K crystal structures shows that the AFP molecule is in fact flexible but evidence for a predisposition towards a particular bend is not apparent.

Since the simulation studies of AFP molecules in vacuum or in solution had limited success in reproducing the side-chain features apparent in the crystal structure, we question the accuracy of results derived from more complex systems which explicitly treated both AFP and ice. In three separate studies, which were able to account for the AFP's orientational binding preferences on the {20$\bar{2}$1} ice plane (Knight, Cheng & DeVries, 1991), the resultant protein structures and the physical basis for the AFP's binding preferences differed greatly. In Wen & Laursen's (1992) study, binding specificity was attributed to optimal hydrogen-bonding interactions arising from a spatial match between ice-binding groups on the

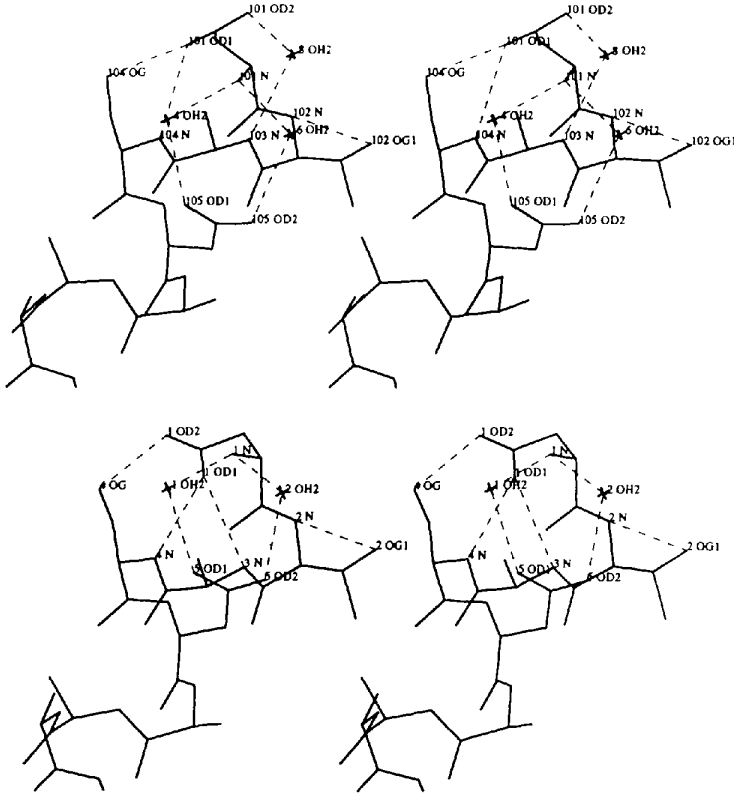

Fig. 11. Stereoview of N-terminal cap structures. The top structure, observed in three out of four AFP molecules, differs slightly from the bottom structure due to crystal-packing interactions which distort the conformation of the Asp1 side chain. Water molecules are denoted as OH2.

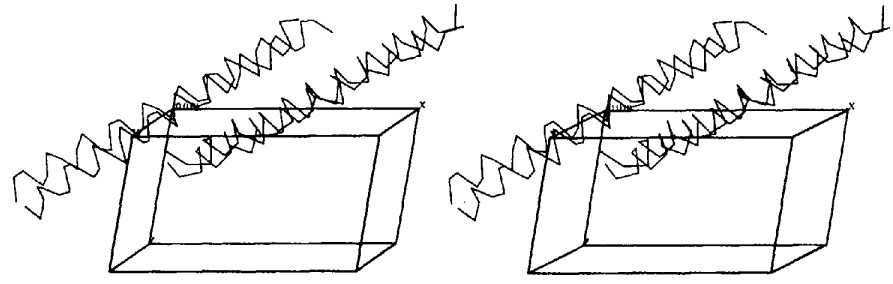

Fig. 12. Superimposed Cα traces of the previous (red) and present (blue) AFP X-ray crystal structures. Due to significant errors in the previous structure, the correlation of origins could only be performed visually.

AFP and the water molecules on the ice surface. In the study of Lal *et al.* (1993) the binding specificity of the AFP was attributed to a van der Waals complementarity between the AFP's ice-binding surface and a ridge and valley topology of the ice plane. In the study of Madura *et al.* (1993), aspartate and asparagine ice-binding side chains were found to integrate directly into the ice lattice. Considering the complexity of the simulation systems, it is not surprising that different results were obtained. The AFP molecule has many degrees of conformational freedom which are unlikely to be fully sampled in a simulation study and hence simulation results will be highly dependent on the starting conformation of the AFP and ice models employed.

The high-resolution crystal structure of the winter flounder AFP should facilitate future simulation studies by providing a more accurate starting model. By reducing the complexity of the protein structure, it should be possible to improve the accuracy of simulation results further. We suggest that side chains involved solely in helix stabilization be removed and treated implicitly with α-helical backbone constraints. This may prove more effective in maintaining the AFP's α-helical conformation over longer simulation periods. Problems with fraying helix termini and backbone distortions were observed in a number of simulation studies and these may adversely effect the study of the AFP's interaction with ice.

3.4. *Implications of molecular-replacement methodology*

The simplicity of predominantly α-helical structures such as the winter flounder AFP, the peptitergent PD_1 (Schafmeister, Miercke & Stroud, 1993) and the model four helix bundle protein α1 (Hill, Anderson, Wesson, DeGrado & Eisenberg, 1990), would seem to make them well suited for structure determination using molecular-replacement techniques with idealized α-helices as search models. However, a survey of the literature, shows that such an approach is not common. To date, only the peptitergent PD_1 structure has been correctly determined using this approach and a successful outcome can be attributed in great part to the crystal system's convenient packing arrangement with one molecule in the asymmetric unit. Our results clearly demonstrate why α-helical structures are likely to pose problems for solution by the aforementioned technique. One problem arises from the symmetry of the α-helical structure itself. This makes distinguishing correct helix directions and rotations about the helix axes difficult since a sufficient number of the side-chain conformations which can break the rotational symmetry are not usually known beforehand. Furthermore, in the structure determination reported here, the knowledge of the threonine side-chain conformations did not help to break rotational ambiguity until after the backbone structure had been extensively refined using SA. Determining the correct translation solutions can also present a problem depending on the tilt of the α-helical molecules with respect to the target symmetry axis. The translation of a molecule along the direction of its helical axis is intrinsically difficult to determine due to its periodic nature.

Difficulties encountered in rotation and translation searches are obviously compounded by having a greater number of molecules in the asymmetric unit and the deviation of the search model from the actual structure presents an even greater problem. Both of these problems were encountered in the AFP structure determination and were surmounted by a multiparameter search and refinement scheme. While an R-factor comparison after the SA-refinement stage was sufficient to distinguish the correct solution from a large number of potential solutions, the use of a heavy-atom derivative to filter and correlate the origin of a large number of solutions (90 in our case) provided a powerful and complementary means to this end.

In conclusion, lone α-helical molecules can present a number of difficulties for structure determination by MR methods. However, through (1) a careful visual interpretation of Patterson map features, (2) a rational search of the unknown rotation and translation parameters, (3) SA refinement of the better solutions, and (4) a filtering and correlation of origins of the better solutions, it may be possible to solve the structure of other lone α-helical proteins and possibly more complex systems.

4. Summary

The high-resolution crystal structure of the winter flounder AFP has been determined. The molecular-replacement methods employed to solve this problematic structure may prove useful for solving other simple α-helical systems.

We acknowledge the Canadian Space Agency for funding and the National Cancer Institute for allocation of computing time and staff support at the Biomedical Supercomputing Center of the Frederick Cancer Research and Development Center.*

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1WFA, R1WFASF, 1WFB, R1WFBSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: AM0027).

References

Brünger, A. T. (1992*a*). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR.* New Haven, Yale University Press, CT, USA.

Brünger, A. T. (1992*b*). *Nature (London)*, **355**, 472–475.

Chakrabartty, A., Ananthanarayanan, V. A. & Hew, C. L. (1989). *J. Biol. Chem.* **264**, 11307–11312.

Chakrabartty, A. & Baldwin, R. L. (1996). *Adv. Prot. Chem.* In the press.

Chou, K. C. (1992). *J. Mol. Biol.* **223**, 509–517.

Davies, P. L., Roach, A. H. & Hew, C. L. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 335–339.

Fitzgerald, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.

Furey, W. & Swaminathan, S. (1990). *PHASES. A Program package for the processing and analysis of diffraction data from macromolecules.* Am. Crystallogr. Assoc. Meet. Abs. Ser. 2, Vol. 18, pp. 73. Abstract PA33.

Hew, C. L., Wang, N. C., Yan, S., Cai, H., Sclater, A. & Fletcher, G. L. (1986). *Eur. J. Biochem.* **160**, 267–272.

Hew, C. L. & Yip, C. (1976). *Biochem. Biophys. Res. Commun.* **71**, 845–850.

Hill, C. P., Anderson, D. H., Wesson, L., DeGrado, W. F. & Eisenberg, D. (1990). *Science*, **249**, 543–546.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Jorgensen, H., Mori, M., Matsui, H., Kanaoka, M., Yanagi, H., Yabusaki, Y. & Kikuzono, Y. (1993). *Protein Eng.* **6**, 19–27.

Knight, C. A., Cheng, C. C. & DeVries, A. L. (1991). *Biophys. J.* **59**, 409–418.

Lal, M., Clark, A. H., Lips, A., Ruddock, J. N. & White, D. N. J. (1993). *Faraday Discuss.* **95**, 299–306.

McDonald, S. M., Brady, J. W. & Clancy, P. (1993). *Biopolymers*, **33**, 1481–1503.

McGregor, M. J., Islam, S. A. & Sternberg, M. J. E. (1987). *J. Mol. Biol.* **198**, 295–310.

Madura, J. D., Wierzbicki, A., Harrington, J. P., Maughon, R. H., Raymond, J. A. & Sikes, C. S. (1994). *J. Am. Chem. Soc.* **116**, 417–418.

Piela, L., Nemethy, G. & Scheraga, H. A. (1987). *Biopolymers*, **26**, 1273–1286.

Rao, B. N. N. & Bush, C. A. (1987). *Biopolymers*, **26**, 1227–1244.

Schafmeister, C. E., Miercke, L. J. W. & Stroud, R. M. (1993). *Science*, **262**, 734–738.

Sicheri, F. S. & Yang, D. S. C. (1995). *Nature (London)*, **375**, 427–431.

Sonnichsen, F. S., Sykes, B. D., Chao, H. & Davies, P. L. (1993). *Science*, **259**, 1154–1157.

Terwilliger, T. C., Kim, S. H. & Eisenberg, D. (1987). *Acta Cryst.* **43**, 1–5.

Tong, H., Berghuis, A. M., Chen, J. & Luo, Y. (1994). *J. Appl. Cryst.* **27**, 421–426.

Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.

Wen, D. & Laursen, R. A. (1992). *Biophys. J.* **63**, 1659–1662.

Wilson, P. W., Beaglehole, D. & DeVries, A. L. (1993). *Biophys. J.* **64**, 1878–1884.

Yang, D. S. C., Chung, Y. J., Chen, P., Rose, J. P. & Hew, C. L. (1986). *J. Mol. Biol.* **189**, 725.

Yang, D. S. C., Sax, M., Chakrabartty, A. & Hew, C. L. (1988). *Nature (London)*, **333**, 232–237.