

# The Refined Crystal Structure of an Eel Pout Type III Antifreeze Protein RD1 at 0.62-Å Resolution Reveals Structural Microheterogeneity of Protein and Solvation

Tzu-Ping Ko,\* Howard Robinson,<sup>†</sup> Yi-Gui Gao,<sup>†</sup> Chi-Hing C. Cheng,<sup>‡</sup> Arthur L. DeVries,<sup>‡</sup> and Andrew H.-J. Wang\*<sup>†</sup>

\*Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan, and <sup>†</sup>Department of Biochemistry and <sup>‡</sup>Department of Animal Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801 USA

**ABSTRACT** RD1 is a 7-kDa globular protein from the Antarctic eel pout *Lycodichthys dearborni*. It belongs to type III of the four types of antifreeze proteins (AFPs) found in marine fishes living at subzero temperatures. For type III AFP, a potential ice-binding flat surface has been identified and is imbedded with side chains capable of making hydrogen bonds with a specific lattice plane on ice. So far, all crystallographic studies on type III AFPs were carried out using the Atlantic ocean pout *Macrozoarces americanus* as the source organism. Here we present the crystal structure of a type III AFP from a different zoarcid fish, and at an ultra-high resolution of 0.62 Å. The protein fold of RD1 comprises a compact globular domain with two internal tandem motifs arranged about a pseudo-dyad symmetry. Each motif of the “pretzel fold” includes four short  $\beta$ -strands and a  $3_{10}$  helix. There is a novel internal cavity of 45 Å<sup>3</sup> surrounded by eight conserved nonpolar residues. The model contains several residues with alternate conformations, and a number of split water molecules, probably caused by alternate interactions with the protein molecule. After extensive refinement that includes hydrogen atoms, significant residual electron densities associated with the electrons of peptides and many other bonds could be visualized.

## INTRODUCTION

RD1 is a small (7-kDa) globular protein from the Antarctic eel pout *Lycodichthys dearborni* (formerly *Rhigophila dearborni*). It belongs to the type III antifreeze proteins (AFPs) found in a number of related zoarcid fishes including Atlantic ocean pout and wolffishes (Hew et al., 1988; Scott et al., 1988), and Antarctic eel pouts (Cheng and DeVries, 1989; Wang et al., 1995). The first three-dimensional structure of type III AFP was determined using NMR (Sonnichsen et al., 1993) and x-ray crystallographic methods (Jia et al., 1996). The antifreeze property of AFPs derives from their ability to depress the freezing point of water by binding to ice crystals and inhibiting further ice growth. On type III AFPs, a potential flat ice-binding surface (IBS), imbedded with side chains capable of making hydrogen bonds with a specific lattice plane on ice, has been identified (Jia et al., 1996; Yang et al., 1998). In the following years, a number of crystal structures of type III AFP, including several mutants, were determined (Yang et al., 1998; Graether et al., 1999; Antson et al., 2001), in conjunction with physical chemical studies. These results agree well with the original proposal of the IBS, although there are multiple possibilities about its interactions with ice crystal, especially regarding what water in the ice lattice a specific polar side chain may bind to.

All type III AFPs are highly homologous as indicated by the extensive amino acid sequence identity. As shown in

Fig. 1, ~50% of the amino acid residues are identical. Recently, the human genome sequencing project identified a domain in sialic acid synthase (SAS) to be homologous to type III AFPs (International Human Genome Sequencing Consortium, 2001). This is the C-terminal domain (Baardsnes and Davies, 2001), and its deletion results in loss of the SAS activity (Dr. Chun-Hung Lin, IBC, personal communication). All of the crystallographic studies on type III AFPs carried out so far employed the Atlantic ocean pout *Macrozoarces americanus* as the source organism. Most of these structures were determined using crystals that belonged to the orthorhombic space group  $P2_12_12_1$  and had similar cell dimensions. In the study by Yang et al. (1998) a new monoclinic crystal form of  $P2_1$  was also reported. In this paper we present an orthorhombic crystal structure of type III AFP, RD1 from the Antarctic eel pout, *Lycodichthys dearborni*, at a resolution of 0.62 Å. This is the highest resolution of all type III AFP crystal structures known, and it permitted the identification of many interesting fine-scale structural features of the AFP that could be associated with function.

## MATERIALS AND METHODS

The protein was purified according to the procedures of Wang et al. (1995) and the crystal was grown by vapor diffusion of a solution containing 40 mg/ml protein, 4 mM Tris buffer (pH 7.5), and 20% ammonium sulfate against a reservoir of 50% ammonium sulfate at room temperature. High-resolution x-ray diffraction data were collected at 110 K using synchrotron at the Advanced Photon Source, Argonne National Laboratory (site SBC/ID19 with APS1 CCD detector). Because of the setup at APS ID19 station, several data collection protocols had to be tested. To minimize possible detector response variances toward different wavelengths (energies) a single wavelength was used to collect data of both low- and high-resolution ranges. Another difficulty was the severe intensity saturation problem for the low-resolution data. Very short exposure time and a highly attenuated beam had to be used.

Submitted June 10, 2002, and accepted for publication September 25, 2002.

Address reprint requests to Andrew H.-J. Wang, Institute of Biological Chemistry, Academia Sinica, Nankang, Taipei 11529, Taiwan. Tel.: 8-862-2788-1981; Fax: 8-862-2788-2043; E-mail: ahjwang@gate.sinica.edu.tw.

© 2003 by the Biophysical Society

0006-3495/03/02/1228/10 \$2.00

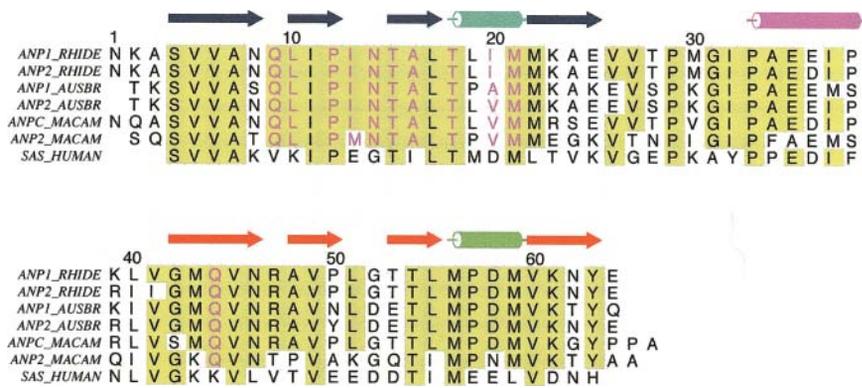


FIGURE 1 Sequence alignment of six type III antifreeze proteins and sialic acid synthase. The amino acid sequences of AFPs were obtained from the Swiss Protein Database, including four from the Antarctic eel pouts *Rhigophila dearborni* (RD1: anp1\_rhide and RD2: anp2\_rhide) and *Austrolycichthys brachycephalus* (AB1: anp1\_ausbr and AB2: anp2\_ausbr) and two from the North-Atlantic ocean pout *Macrozoarces americanus* (HPLC-12: anpc\_macam and HPLC-3: anp2\_macam). The latter two sequences correspond to the models of PDB entries 1MSI, 1HG7 (both HPLC-12) and 1OPS (HPLC-3) used in structural comparison with minor variations. The sequence of human sialic acid synthase was obtained from GeneBank

(AAF75261: sas\_human) and the aligned sequence corresponds to the C-terminal AFP-like domain of amino acid residues 294–353. Residues having five or more identities in the seven sequences are shaded in yellow. The eleven residues supposed to form the ice-binding surface (IBS) are highlighted in magenta. The two dyad-related  $\beta$ - $\beta$ - $\beta$ - $\beta$  motifs of the pretzel fold are also arranged in juxtaposition, with the secondary structural elements drawn on the top. These are colored blue and cyan for the N-terminal motif, red and green for the C-terminal motif, and magenta for the connecting  $3_{10}$ -helix.

Two data sets were collected in the end, with resolution ranges of 50–0.96 Å and 1.2–0.62 Å. They were merged before use in the structural analysis. The wavelength was 0.6668 Å and the crystal-to-detector distance was 117 mm for both data sets, but the detector was placed at  $2\theta = 0$  and  $30^\circ$ , respectively. A scalar value of 474.65 was computed between 11,483 matched reflection pairs in the overlapped resolution range between 1.2 Å and 0.96 Å in the two data sets with an  $R$ -merge value of 18.5%, then 18,840 reflections in the low-resolution data set were added to the high-resolution data set by applying the scalar factor. The programs DENZO and SCALEPACK (Otwinowski and Minor, 1997) were used for data processing. The integration procedures might not be fully optimized due to the ultra-high resolution nature of the data. Some errors might be introduced into the resultant structure factors. Relevant statistical numbers are listed in Table 1.

Initial structure was determined by molecular replacement methods using X-PLOR (Brunger, 1993), with a search model from NMR experiments and a 1.3 Å resolution data set collected earlier (unpublished results). It turned out to be similar to other type III AFPs. The programs CNS (Brunger et al.,

1998), SHELX-97 (Sheldrick and Schneider, 1997), and O (Jones et al., 1991) were used in computational least-square refinement and in manual modifications of the model according to the electron density maps. The model after preliminary refinement yielded an  $R$ -value of 0.17 at 1.0 Å resolution. It contained all 64 amino acid residues and 92 water molecules. All temperature factors were treated as isotropic and all occupancies were fixed as unity. Further refinement at a higher resolution (e.g., 0.95 Å) did not give satisfactory results using the “normal” protocols for protein crystals.

The difference Fourier maps calculated at this stage began to show alternate conformations of several residues, and the model was modified accordingly (vide infra). Water molecules were included with a criterion of density level higher than  $1.5\sigma$  in the  $2F_o - F_c$  map. Several cases of split water were also observed. By employing SHELX-97, all hydrogen atoms were explicitly attached to the protein model with idealized geometry. Water molecules, however, were treated as oxygen atoms because the electron density maps did not allow precise identification of their associated hydrogen atoms. Six anisotropic temperature factors were allowed to vary for each nonhydrogen atom. Inasmuch as the protein atoms still assume full occupancies, water molecules were now allowed to refine with fractional occupancies. Using these additional constructs, usually performed for “small molecules” only, the  $R$ -values of high-resolution shells were improved significantly, and the refinement could subsequently be extended to 0.62 Å. Most individual nonhydrogen atoms were now seen as discrete spheres in the electron density maps, as shown in Fig. 2.

The refinement continued with alternating routines of manual model adjustments and computational parameter variations. Most of the modifications at these final stages were deletion of weak solvent molecules and addition of new ones, as well as minor positional shifts of some stronger waters. Conjugate gradient algorithm was employed in beginning cycles of SHELX refinement, which were followed by several additional full-matrix least squares cycles. Throughout the refinement, the protein model was restrained with parameters of Engh and Huber (1991). The thermal parameters of bonded atoms and atoms close in space were also restrained. For water molecules, the six anisotropic components were restrained to approximate isotropic behavior. The total number of variables is 6818 for the final model, in which the degree of freedom is small as compared with the 118,101 observed data used in refinement. However, a complete release of the restraints did not yield convergent results. The geometry of the protein model became worse, with broader distributions of bond lengths and angles and significant nonplanarity for some planar groups (e.g., peptide bonds and side-chain carboxyl and amide groups), whereas higher  $R$ -values were obtained. Therefore, the refinement concluded with a restrained model. Statistics are shown in Table 1 also.

For model analysis and comparison, the programs CCP4 (Collaborative Computational Project Number 4, 1994), PROCHECK (Laskowski et al.,

TABLE 1 Crystallographic data of the antifreeze protein RD1

Data collection	Normal setup	High-angle setup
Unit cell (Å)	32.56, 39.52, 44.67	32.50, 39.50, 44.64
Resolution (Å)	50–0.96 (0.99–0.96)	1.2–0.62 (0.64–0.62)
No. of observations	129867 (647)	256420 (18801)
Unique reflections	30323 (545)	108526 (10198)
Average $I/\sigma(I)$	15.7 (0.8)	9.9 (1.2)
Completeness (%)	84.0 (15.4)	97.0 (91.8)
$R_{\text{merge}}$ (%)	6.8 (59.9)	7.3 (65.3)
Refinement	CNS-1.0	SHELX-97
Resolution range (Å)	20–1.0	50–0.62
No. of reflections	31130	118101
$R_{\text{crist}}$ for 95% data	0.168	0.133
$R_{\text{free}}$ for 5% data	0.195	0.155
RMSD bond lengths (Å)	0.025	0.012
RMSD bond angles ( $^\circ$ )	1.8	2.1
Average B of non-H atoms	8.05/568	8.91/730
Protein backbone atoms	5.15/256	4.59/258
Protein side-chain atoms	8.12/220	7.06/232
Water molecules	15.68/92	15.33/240
Most favored $\phi$ , $\psi$ angles (%)	96.2	96.2
Other allowed $\phi$ , $\psi$ angles (%)	3.8	3.8

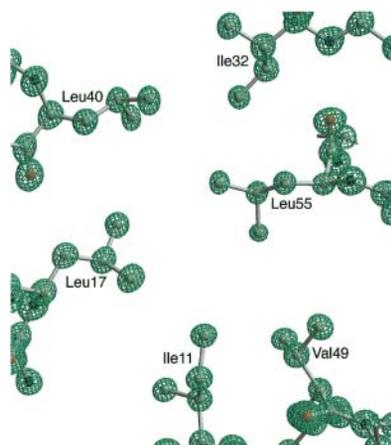


FIGURE 2 Representative ultra-high-resolution electron density map. The final model of RD1 was superimposed on a  $2F_o-F_c$  map for a central hydrophobic core region. The map was calculated using all data at 0.62-Å resolution and contoured at  $3\sigma$  level.

1993), GRASP (Nicholls et al., 1991), and O were used. Three models from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank were used for comparison, including two orthorhombic structures and one monoclinic crystal structure at 1.25 Å, 1.15 Å, and 2.0 Å resolution, with access codes 1MSI, 1HG7, and 1OPS, respectively. The programs ALSRIPT (Barton, 1993), BOBSCRIPT (Esnouf, 1997), MOLSCRIPT (Kraulis, 1991), Raster3D (Merrit and Murphy, 1994), and GRASP were employed in producing figures.

## RESULTS AND DISCUSSION

### Overall structure of the protein

The protein fold of RD1 is identical to other type III AFPs. It comprises a compact, globular, single domain with two internal tandem motifs arranged about a pseudo-dyad symmetry, as shown in Fig. 3. Each motif of the “pretzel fold” (Yang et al., 1998) includes four short  $\beta$ -strands and a one-turn  $3_{10}$  helix. The motifs are connected by a 15-residue loop that contains a two-turn  $3_{10}$ -helix. The strands are intertwined into three small antiparallel  $\beta$ -sheets; two are three-stranded and one two-stranded. Two proline residues, 12 and 50, which are related by the internal dyad axis, can be clearly seen attached to the pair of  $\beta$ -strands on top of Fig. 3 A.

In the refined model, the peptide bond between Thr28 and Pro29 has *cis*-configuration, as observed in other crystal structures of type III AFPs. Pro29 is located on a surface loop (Fig. 3 A) and no obvious reason can be addressed for its special *cis*-configuration, as the oxygen atoms of Thr28 and Pro29 both interact with the bulk solvent. The first peptide bond between Asn1 and Lys2 has alternate conformations, differing by  $\sim 55^\circ$  rotation about the  $C_\alpha-C_\alpha$  axis. The side chains of four amino acid residues also show alternate conformations: Glu25, Met30, Met43, and Met56. Examples are presented in Fig. 4. There are three alternate nonhydrogen atoms in Glu25 (CD, OE1, and OE2) and Met56 (CG, SD, and CE), whereas Met30 and Met43 have a single methyl group in alternate conformations. All but two residues

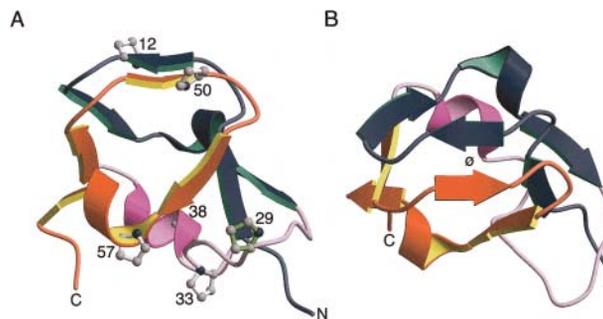


FIGURE 3 Overall structure of the RD1 molecule. The two similar motifs of the pretzel fold are shown in blue (N-terminal) and red (C-terminal). These are connected by a  $3_{10}$ -helix, shown in magenta. In (A) all of the six proline residues in the RD1 molecule are labeled, with the *cis*-proline 29 shown in green. In (B) the molecule is turned  $\sim 90^\circ$  to view along the pseudo-dyad axis.

(Asn14 and Asn62) have their dihedral angles within the most favored regions as defined by PROCHECK. Both asparagines have the  $\alpha_L$ -like conformation ( $[\phi, \psi] = [72^\circ, 8^\circ]$  and  $[65^\circ, 15^\circ]$ , respectively), which is normally found for a glycine residue in the third position of a type II turn (Richardson, 1981).

There are 43 direct intramolecular hydrogen bonds in the RD1 molecule, of which 31 are between peptide backbone atoms. Most of the hydrogen bond distances between the NH and CO groups fall within the standard range with an average value of  $2.83 \pm 0.02$  Å. These are bonds between  $\beta$ -strands and in the  $3_{10}$  helices and  $\beta$ -turns. For example, both of the pseudo-dyad-related residues Pro12 and Pro50 (Fig. 3 A) are in the first position of a type II  $\beta$ -turn. The oxygen atoms of Pro12 and Pro50 are hydrogen bonded to the backbone nitrogens of Thr15 and Thr53, respectively. In addition, the backbone O and N atoms of preceding residues Ile11 and Val49 also make two dyad-related hydrogen bonds. All proline residues except Pro29 are involved in backbone hydrogen bonds and presumably play structural roles. The other 12 bonds include three neutral bonds and three salt bridges between side chains, plus six bonds between side-chain and backbone atoms. Interestingly, there are two

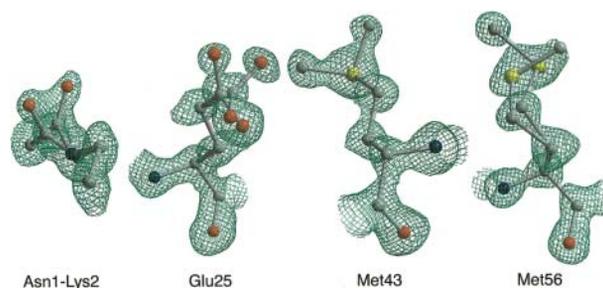


FIGURE 4 Alternate conformations in the RD1 crystal structure. Four examples of alternative conformers, including the peptide bond between Asn1 and Lys2, the carboxyl group of Glu25, the methyl group of Met43, and most of the side chain of Met56, are shown in ball-and-stick representations, and are superimposed on the final  $2F_o-F_c$  map, contoured at  $1.5\sigma$  level.

intraresidue hydrogen bonds with well-defined geometry and corresponding density in the Fourier maps: the side chains of Glu35 and Asp58 turn back with the OE2 and OD1 atoms, respectively, hydrogen bonded to the backbone N atoms of the same residue.

The RD1 model has root mean square deviations (RMSD) in bond lengths and angles of 0.012 Å and 2.1° from ideal values. These are typical of well-refined protein structures, despite the ultra-high resolution data used. Deviations of the parameters of the polypeptide chain are also within the normal range from ideal values. However, three residues show unusual geometry. First, the side chain of Leu10 has torsion angles of  $\chi_1 = 180^\circ$  and  $\chi_2 = 300^\circ$ , which is not favored. It is in contact with a nonpolar patch of Leu19, Ile37, and Pro38 on a symmetry-related molecule. Presumably, crystal-packing interactions of the side chains force Leu10 into such a special conformation (vide infra). Second, the peptide torsion angles of Met21–Met22 ( $\omega = -165^\circ$ ) and Gln44–Val45 ( $\omega = 168^\circ$ ) deviate by more than 10° from the ideal planar value. They are probably caused by preferred hydrogen bonds of the carbonyl oxygen atoms of Met21 and Gln44 with the backbone nitrogen atoms of Asn8 and Lys61, respectively. The optimized O–N distances of 2.80 and 2.81 Å as well as exact alignment of the hydrogen bonds would be altered if the peptides were more planar.

For distribution analysis, the temperature factors of all atoms are converted to isotropic values (Table 1). Of the 490 nonhydrogen atoms in the protein model, 484 temperature factors are less than 20 Å<sup>2</sup>, and only six are between 20 and 30 Å<sup>2</sup>. These include the side-chain atoms CE and NZ of Lys2, OE1 of Glu35, OE2 of Glu36, NE2 of Gln44, and OE2 of Glu64. They are all on the protein surface and facing solvent. The average temperature factor of 530 hydrogen atoms in the model is 7.95 Å<sup>2</sup>. Seven of them are between 20 and 30 Å<sup>2</sup>, and three are between 30 and 40 Å<sup>2</sup>. These correspond to the hydrogen atoms attached to CE and NZ of Lys2, CE of the second conformer of Met30, and NE2 of Gln44. The water molecules have higher temperature factors, inasmuch as most (214) of them are between 5 and 25 Å<sup>2</sup>, and only six are between 30 and 40 Å<sup>2</sup>.

### Residual electron densities and cavity

After extensive refinement, the *R*-value did not reduce to less than 10%, as usually required for “small molecule” crystals. This may be due in part to scale problems with the disordered bulk solvent, which occupies about a quarter of the solvent space (vide infra). If the data were divided according to resolution, the linear scalar values between the observed and calculated structure factors showed a concave curve with a minimum of 0.959 located in the shell of 1.0–0.94 Å, whereas at both ends the scalar values were 1.168 and 1.146 for 50–2.5 Å and 0.64–0.62 Å shells, respectively. Although a bulk solvent model was employed in SHELX refinement, with the parameters *g* and *U* of 0.629 and 6.09, respectively,

it seemed insufficient to overcome the scaling problem. On the other hand, although the fully unrestrained refinement was attempted, it did not yield a convergent and geometrically sensible model with lower *R*-values. It is possible that the full matrix least squares conjugate gradient method is not optimal for this particular case. At this stage of the refinement, it is probably not justified to proceed with the charge-density refinement procedure as yet.

Nevertheless, when the refined model is superimposed on the final  $F_O - F_C$  difference Fourier map, some interesting features can be seen. These are probably a consequence of the atomic model used in refinement, which only adopts a spherical or ellipsoidal approach for calculating structure factors and does not take valence electrons and lone pairs into account. In Fig. 5 A, the map for a typical peptide bond is shown. Residual electron densities are seen, particularly about the peptide carbonyl oxygen which is more electronegative than other atoms. In Fig. 5 B, all of the ordered peptide bonds are superimposed and significant features about the oxygen atom come out clearly. In addition to the bond electrons of the carbonyl group, some possible densities for lone pairs of the oxygen atom also become visible. Bond electrons in the other three bonds between  $C_\alpha$ –C, C–N, and N– $C_\alpha$  can also be visualized with certainty. Beyond the  $C_\alpha$  atom, electron densities for other bonds, including both peptide and side-chain bonds, are concentrated in two regions, corresponding to two favored orientations of the adjacent peptide units.

In a recent work on crambin by Jelsch et al. (2000), the crystal structure was analyzed at 0.54-Å resolution by refinement using a multipolar charged atom model. The deformation density maps, which were calculated for the differences between the actual electron density of the molecule and the “normal” model of spherical atoms, clearly showed the distribution of bond electrons and the lone pair electrons of the peptide bonds. Although we have not carried out similar refinement procedures of charge density refinement for RD1, the current model of spherical atoms turned out to be also effective in visualization of the valence electrons. In Fig. 5 C, residual electron densities are seen in the middle of most bonds. Besides, significant volumes of densities are present about the sulfur atoms in the side chains of two methionines, 21 and 22. There are 229 side-chain bonds in an RD1 molecule. If the  $F_O - F_C$  map is contoured at 2- $\sigma$  level, residual densities are seen for 144 bonds; if it is at 3- $\sigma$  level, 78 have corresponding densities; if it is at 4- $\sigma$  level, densities for only 20 bonds are seen. Of these 20 bonds, 11 belong to methionine and five belong to serine and threonine residues.

In the RD1 molecule there is a cavity that is not occupied by protein or solvent atoms and does not have residual density. It has a volume of 45 Å<sup>3</sup> as calculated using GRASP and is surrounded by eight conserved nonpolar residues: Val5, Ala7, Ile11, Leu17, Met21, Met22, Val49, and Leu55, shown in Fig. 6. Such a cavity has not been reported for type III AFP before (Yang et al., 1998; Antson et al., 2001). However, in

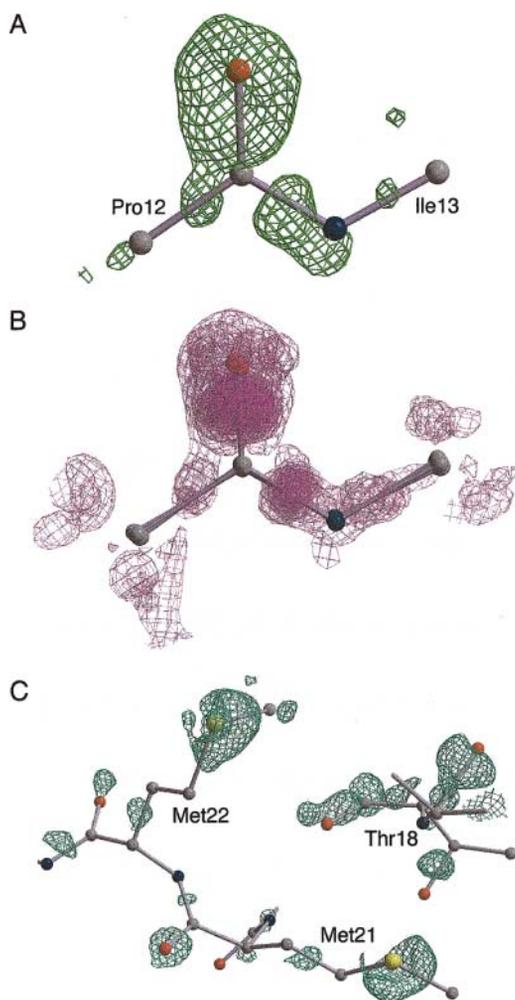


FIGURE 5 Residual densities in the RD1 crystal. The refined model is superimposed on the final  $F_o - F_c$  difference Fourier map. In (A) the map is contoured at  $3\text{-}\sigma$  level for the peptide bond between Pro12 and Ile13. In (B) all of the 62 peptide bonds, excluding the first one that has alternate conformations, are superimposed and the densities are contoured at  $4\text{-}\sigma$  level. For the *cis*-peptide between Thr28 and Pro29, the bond of N-C $\delta$  is used instead of N-C $\alpha$ . In (C) the map is contoured at  $3\text{-}\sigma$  level for a region about Met21 and Met22.

the monoclinic crystal structure (RCSB code 1OPS) a similar cavity of  $22 \text{ \AA}^3$  can also be identified. The absence of a detectable cavity in both high-resolution orthorhombic structures with RCSB codes 1MSI and 1HG7 is probably due to the side chain of Leu55, which has alternate conformations. Apparently the cavity offers Leu55 more space and rotational freedom. Although the precise function of the cavity has not been determined, a possibility is to provide flexibilities of core hydrophobic side chains in the process of protein folding. We are currently making mutants in this region to investigate the effects of various side chains on the stability of RD1.

### Crystal packing and structure comparison

The specific volume of the RD1 crystal is  $2.08 \text{ \AA}^3/\text{Da}$ , which

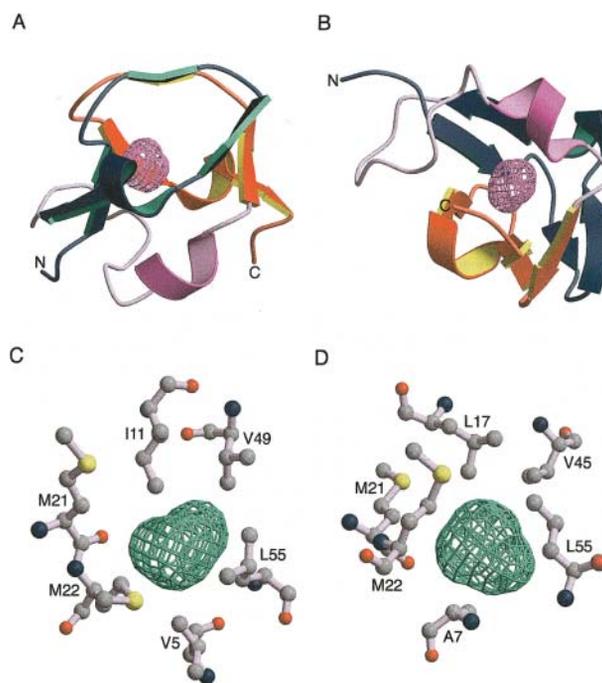


FIGURE 6 Cavity in the RD1 molecule. In (A) and (B) the RD1 molecule is shown in two views by ribbon representations as in Fig. 3. The cavity is shown as mesh and colored in magenta. In (C) and (D) are two orthogonal closeup views of the cavity, colored cyan, and its nine surrounding residues, which are labeled. The cavity was constructed using GRASP.

corresponds to  $\sim 41\%$  solvent content (Matthews, 1968). In the orthorhombic crystal, each RD1 molecule makes lattice contacts with six symmetry-related neighbors, shown in Fig. 7. There are three types of crystal contacts. Direct interactions are listed in Table 2. Not taking water molecules into account, the first type (A) involves 10 and 11 residues in molecules related by  $(x, y, z)$  and  $(x-1/2, 1/2-y, -z)$ , respectively. They make four direct hydrogen bonds, including two salt bridges, and bury  $351$  and  $393 \text{ \AA}^2$  surface areas, respectively. The second type (B) involves eight and seven residues in molecules related by  $(x, y, z)$  and  $(1-x, y-1/2, 1/2-z)$ , respectively. There is only one direct hydrogen bond, but there are also a number of nonpolar interactions at this interface. The surface areas buried are  $246$  and  $292 \text{ \AA}^2$ , respectively. The third type (C) involves five and four residues in molecules related by  $(x, y, z)$  and  $(1/2-x, -y, z-1/2)$ , respectively. They make two hydrogen bonds and the buried areas are  $171$  and  $159 \text{ \AA}^2$ , respectively. The total crystal-contact surface area is  $1612 \text{ \AA}^2$  per RD1 molecule, encompassing  $43\%$  of the molecular surface area of  $3783 \text{ \AA}^2$ .

Among the residues at crystal contacts, seven have more than  $50 \text{ \AA}^2$  surface area buried by the interface: Asp58, Thr28\*, and Leu51\* (asterisks designate residues of the counter molecule) in interface A, Leu19, Leu10\*, and Ile13\* in interface B, and Asn1 in interface C. Asp58 forms a salt bridge with Lys23\*, and the hydrogen bond distance of  $2.46 \text{ \AA}$  to Glu25\* is short. This is probably due to the disordered

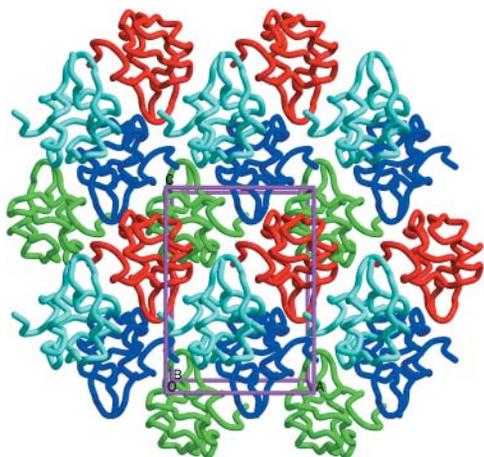


FIGURE 7 Packing of RD1 molecules in the crystal. The four symmetry-related protein molecules in a unit cell are colored red, green, blue, and cyan, and molecules related by unit lattice translations are shown in same colors.

side chain of Glu25\*, which also forms a salt bridge with Lys23\*. Thr28\* is hydrogen bonded to Asn62, and it is also in contact with Lys39 through interface *C*. Leu51\* interacts with a nonpolar patch of Pro29 and Met56. Both conformers of Met56 are well accommodated. Leu19 is also located in a nonpolar patch and interacts with Leu10\*, Ala48\*, and Pro50\*. Leu10\* is facing Leu19, Ile37, and Pro38, and Ile13\* facing Ile20, presumably also involved in hydrophobic interactions. The amino group of Asn1 is hydrogen bonded to two backbone oxygen atoms of Lys39\* and Val41\*, whereas the side chain is surrounded by Lys39\* and Met43\*, which has two side-chain conformers.

As mentioned above, all type III AFPs are highly

TABLE 2 Specific bonds at lattice contact interfaces of the orthorhombic RD1 crystal

Resid-1	Atom-1	Resid-2	Atom-2	d (Å)	Type of interaction
A. Molecules 1 and 2 are related by $(x, y, z)$ and $(x-1/2, 1/2-y, -z)$					
Pro 29	CG	Leu 51	CD2	3.43	van der Waals contact
Arg 47	NH1	Glu 25	OE1	3.03	Salt bridge
Met 56	CE	Leu 51	CG	3.28	van der Waals contact
Asp 58	OD2	Lys 23	NZ	3.44	Salt bridge
Asp 58	OD2	Glu 25	OE2	2.46	Hydrogen bond
Asn 62	N	Thr 28	OG1	2.95	Hydrogen bond
B. Molecules 1 and 2 are related by $(x, y, z)$ and $(1-x, y-1/2, 1/2-z)$					
Asn 8	OD1	Asn 14	ND2	3.04	Hydrogen bond
Leu 19	CB	Ala 48	CB	3.66	van der Waals contact
Ile 20	CG1	Ile 13	CD1	3.82	van der Waals contact
Ala 24	CB	Pro 12	CG	3.96	van der Waals contact
Ile 37	CD1	Pro 12	CG	3.64	van der Waals contact
Ile 37	CG2	Leu 10	CD1	3.85	van der Waals contact
Pro 38	CD	Leu 10	CG	3.88	van der Waals contact
Pro 38	CG	Leu 10	CB	3.89	van der Waals contact
C. Molecules 1 and 2 are related by $(x, y, z)$ and $(1/2-x, -y, z-1/2)$					
Lys 39	O	Asn 1	N	2.68	Hydrogen bond
Val 41	O	Asn 1	N	2.92	Hydrogen bond
Lys 39	CG	Thr 28	CG2	3.83	van der Waals contact

homologous and have identical protein folds. If the RD1 molecule is superimposed on other type III AFPs, the difference in coordinates would be minimal. Upon comparison using the LSQ procedure of O with 1.0-Å matching criterion, the RMSD for RD1 and the 1HG7 models is 0.253 Å between 61 pairs of  $C\alpha$  atoms. The RMSD for RD1 and 1MSI models is 0.279 Å between 63 pairs of  $C\alpha$  atoms, and that for RD1 and 1OPS is 0.453 Å between 59 atom pairs. The model of 1OPS has larger deviations from other models, probably because it is in a different crystal form and the protein molecules are involved in different lattice contacts (vide infra). It has RMSD from the 1HG7 and 1MSI models of 0.445 and 0.385 Å for 62 and 60 atom pairs, respectively, whereas the RMSD between the latter two models is only 0.241 Å for 64 atom pairs. In addition to the N- and C-terminal ends, the regions of largest deviations are located about residues 27–28 and 38–39, with  $C\alpha$  coordinate differences of  $\sim 0.8$  and 1.2 Å, respectively. These two loop regions are located at the margin of the two homologous motifs (Fig. 1), and are presumably less structurally constrained. In fact, both are involved in lattice contact interactions.

In the monoclinic crystal form, each AFP molecule makes lattice contacts with eight symmetry-related neighbors. Direct interactions are listed in Table 3. The first type of interface (*A*) involves four residues from each molecule, related by unit translation along the *a*-axis. It buries 143 and 145 Å<sup>2</sup> surface areas on the two apposing molecules. The second type (*B1*) involves nine and ten residues and buries 385 and 352 Å<sup>2</sup> on molecules related by  $(x, y, z)$  and  $(x, 1/2+y, -z)$ ; the third type (*B2*) involves four and two residues and buries 101 and 110 Å<sup>2</sup> on molecules related by  $(x, y, z)$  and  $(-x, 1/2+y, 1-z)$ , respectively. The fourth type (*C*) involves five and four residues on molecules related by unit translation along the *c*-axis, and buries 176 and 208 Å<sup>2</sup> surface areas, respectively. The total surface area buried by crystal contact is 1620 Å<sup>2</sup> per molecule, which covers  $\sim 44\%$  of the molecular surface. As seen in Table 3, the side chains of Thr27, Asn28, and Gln39 all participate in specific hydrogen bonds with neighboring molecules.

The monoclinic AFP crystals of 1OPS diffracted x rays to only 2.0 Å resolution, whereas the RD1 crystal showed an effective resolution of 0.62 Å. Apparently, the difference is not correlated to the total surface area of crystal contact, because the areas are about the same in the two crystal forms. Neither can it be explained by solvent contents, because the orthorhombic crystals have slightly larger  $V_m$  values. It is likely that crystal quality is determined by the overall arrangements of protein molecules. Thus, the monoclinic crystals are less ordered than the orthorhombic crystals, consistent with the fact that the high-resolution structures of 1MSI and 1HG7, as well as RD1, are all from the orthorhombic crystals.

All orthorhombic crystal structures of type III AFPs have identical space group ( $P2_12_12_1$ ) and very similar unit-cell

**TABLE 3** Specific bonds at lattice contact interfaces of the monoclinic 1OPS crystal

Resid-1	Atom-1	Resid-2	Atom-2	d (Å)	Type of interaction
A. Molecules 1 and 2 are related by $(x, y, z)$ and $(x+1/2, y, z)$					
Ala 20	CB	Thr 47	CG2	3.18	van der Waals contact
B1. Molecules 1 and 2 are related by $(x, y, z)$ and $(-x, y+1/2, -z)$					
Pro 33	CG	Lys 51	CD	3.64	van der Waals contact
Glu 36	OE2	Lys 25	NZ	2.66	Salt bridge
Gln 39	OE1	Asn 28	N	2.93	Hydrogen bond
Gln 39	NE2	Asn 28	O	3.05	Hydrogen bond
Tyr 63	O	Lys 25	NZ	3.06	Hydrogen bond
Ala 65	OXT	Val 26	N	2.94	Hydrogen bond
B2. Molecules 1 and 2 are related by $(x, y, z)$ and $(-x, y+1/2, 1-z)$					
Lys 43	CE	Pro 12	CG	3.95	van der Waals contact
C. Molecules 1 and 2 are related by $(x, y, z)$ and $(x, y, z+1)$					
Met 13	SD	Asn 28	CG	3.81	van der Waals contact
Asn 12	ND2	Thr 27	O	2.92	Hydrogen bond

dimensions. In other words, they are quasi-isomorphous, irrespective of their different amino acid sequences. Nevertheless, the crystals of 1MSI and 1HG7 diffracted x rays to only 1.25 and 1.15 Å resolution, significantly lower as compared with RD1. The  $V_m$  and crystal contact areas are not very different and, again, these cannot explain the diffraction results. In fact, the residues involved in direct crystal contact interactions are mostly conserved (Table 2). In the 1HG7 structure, only three residues are different from RD1 in these regions: Arg23 (Lys), Ser24 (Ala), and Arg39 (Lys). Arg23 forms a conserved salt bridge with Asp58, and the CB atom of Ser24 also remains in contact with Pro12 whereas the OG atom is facing the bulk solvent. Arg39, however, makes quite different interactions. In 1HG7, there is a sulfate ion bound to the guanidinium group of Arg39, and also to Thr28, Asn46, and Gly62 of two other symmetry-related molecules. In contrast, no sulfate molecule was observed in the RD1 crystal. The side chain of Lys39 in RD1 makes direct contact with Thr28 of a symmetry-related molecule, whereas in 1HG7 there is no such contact, and the side chain of Thr28 has alternate conformations. At interface *C* of 1HG7, the only direct interaction is between the terminal amino group and the carbonyl oxygen atoms of Arg39 and Val41. In addition, at interface *B*, the side chain of Leu10 in 1HG7 is not constrained into an unfavored conformation by lattice contact. Similar situations are seen in the 1MSI structure, but there is no sulfate ion and no alternate conformer for Thr28.

As mentioned above, the polypeptide termini of RD1 and the regions near residues 27–28 and 38–39 show largest RMSD in  $C\alpha$  coordinates upon superposition with other models. It is not only because of the intrinsic peptide flexibility for lack of structural role, but also caused by tight crystal packing of the molecule. The more involved interactions at interfaces *B* and *C* may account for the superior quality of the orthorhombic RD1 crystal, which diffracts x rays very well and allows data collection at ultra-high resolution.

## Solvent model and ice-binding surface

There are 240 water molecules in the refined model, but only 47 are well defined in the electron density maps. These include structural waters that each of them binds to at least one protein atom. An example is shown in Fig. 8 *A*, where one of the water molecules forms four hydrogen bonds with protein. This water molecule belongs to the first “good” category. The other waters in Fig. 8 *A* are not as well defined, judged by the lower fractional occupancies, but they still have corresponding densities in the map. These belong to the second “not-so-good” category that includes 122 of the total 240 waters. In Fig. 8 *B*, some water molecules at crystal contact are shown (and so is Leu10 discussed above). The three RD1 molecules are colored orange, green, and blue. The nature of this interface is nonpolar but five of the water molecules form a pentagon inside the hydrophobic pocket. Six of the eight waters shown here belong to the first category and two belong to the second.

The refined structure of RD1 also contains water molecules in dual locations. Protein or nucleic acid crystals are frequently analyzed to less than 1.0-Å resolution, in which no such “split” waters can be discerned easily. If an elongated density is observed, it is usually treated as single water with higher temperature factor. At 0.62-Å resolution, however, some of the disordered waters can be explicitly modeled. Fig. 8 *C* shows an example where two of the waters are seen as “split” molecules, i.e., which have dual locations and fractional occupancies. This is caused by alternate interactions with the protein and other solvent molecules. For example, the central water in Fig. 8 *C* can make a hydrogen bond with the carbonyl oxygen of Ala48, but only one alternative can make a proper bond with one of its neighbors. A weak water molecule making a non-ideal hydrogen bond of 2.55 Å is also seen. The water bound to the peptide O of Ala48 belongs to the first category, and the others are examples of, again, the second category of waters.

The remaining 71 of the 240 water molecules did not behave well upon refinement and belong to the third category. They were added according to the residual densities in the bulk solvent region because the presence of these waters somehow reduced both  $R$  and  $R_{\text{free}}$  values, but actually do not have corresponding density level of more than 1.5  $\sigma$  in the final 2F<sub>O</sub>-F<sub>C</sub> map. This could be explained by the use of ammonium sulfate in crystallization, which resulted in a dense solvent background. The interstitial volume of solvent corresponds to  $\sim 24,800$  Å<sup>3</sup> per unit cell. The 240 water molecules of the refined model occupy 18,000 Å<sup>3</sup> (4500 Å<sup>3</sup> per asymmetric unit) of the volume, leaving 6800 Å<sup>3</sup> (1700 Å<sup>3</sup> per asymmetric unit) void volume. Thus, the number of other disordered water molecules not included in the model is probably more than 60, as estimated using the existing solvent model.

When the crystal structure of the type III AFP was first determined (Jia et al., 1996), it was proposed that the protein

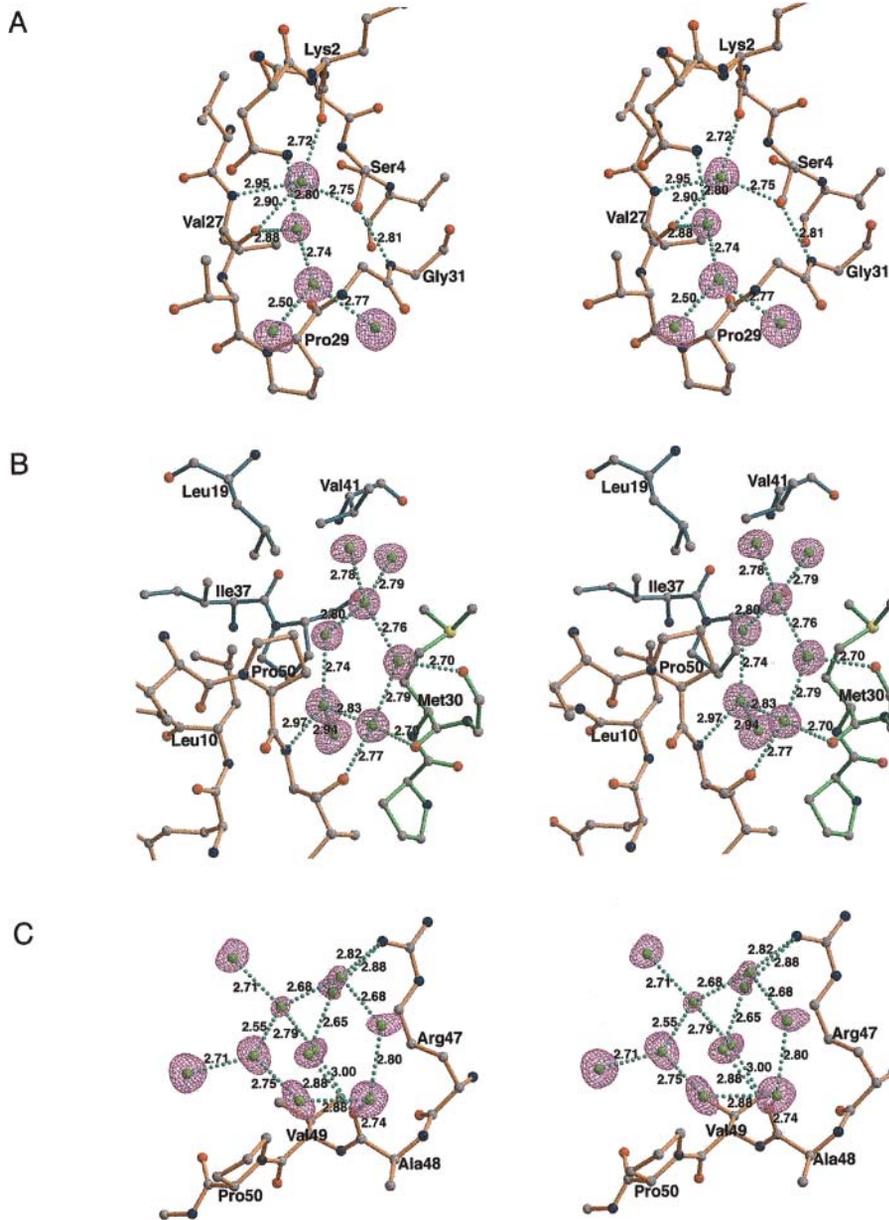


FIGURE 8 Examples of water molecules in the RD1 crystal. The protein and water molecules are shown as ball-and-stick representation and green spheres. These are superimposed on the  $2F_o - F_c$  map calculated using the final refined model and contoured at  $2\text{-}\sigma$  level. For clarity, densities are shown about the water molecules only. Relevant hydrogen bonds are shown as strings of small beads and also labeled with the interatomic distances. See text for further description.

binds to ice by formation of hydrogen bonds. Later research suggested that additional interactions with hydrophobic side chains might also play an important role in ice binding (Yang et al., 1998; Antson et al., 2001). In Fig. 9, the structures of the IBS residues are compared, and they show minimal deviations from one another. The flat and mostly neutral surface on RD1, as shown in Fig. 10, is supposed to bind to a particular lattice plane of ice, presumably with a similar mechanism proposed by Jia et al. (1996), Yang et al. (1998), and Antson et al. (2001). The ice-binding surface of RD1 is composed of 11 residues, which are shown in Fig. 11 *A*. In the crystal structure of RD1, there are 44 ordered water molecules near the IBS, shown in Fig. 11 *B*. These are all within 4.0-Å distance from the ice-binding residues, including Gln9, Leu10, Pro12, Ile13, Asn14, Thr15, Ala16,

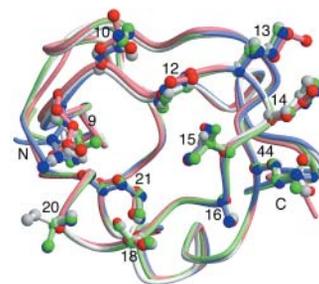


FIGURE 9 Structure comparison of the IBS residues. The model of RD1, colored white, is superimposed on those of other type III AFP models corresponding to PDB entries 1OPS, 1MSI, and 1HG7, colored red, green, and blue, respectively. Only minor structural variations are observed for the 11 ice-binding side chains.

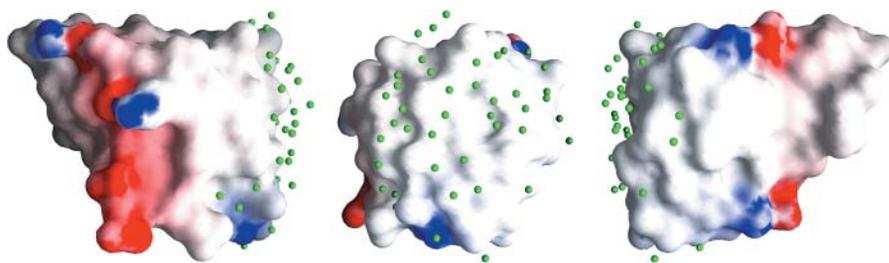


FIGURE 10 The ice-binding surface of RD1. Three orthogonal views of a surface charge potential representation of the RD1 molecule by using GRASP. The red and blue colors show negative and positive charges on the surface, respectively, with a range of  $-10$  to  $+10 k_B T$ . Neutral surface regions are colored white. Water molecules on the IBS (see text for details) are colored in green. The molecule is supposed to bind ice with the front surface and portion of the bottom surface.

Thr18, Ile20, Met21, and Gln44, and have corresponding densities of more than  $2\text{-}\sigma$  level in the final  $2F_o - F_c$  map. In cases of split waters, averaged coordinates are used. If they are superimposed on the 1HG7 model, 28 of the 44 waters have equivalents within  $1.0 \text{ \AA}$  distance. If the matching criterion is  $0.5 \text{ \AA}$ , 14 waters still have equivalents in both models. Among the 28 water molecules, 19 waters make 22 well-defined hydrogen bonds with the protein, including 17 bonds with backbone atoms and five bonds with the side-chain atoms of Gln9, Asn14, and Thr18. The remaining nine equivalent waters and the 16 unmatched waters are bound to other water molecules, which form a cage-like structure, as shown in Fig. 11 *A*. Similar results are obtained if the 1MSI model is used, with 22 pairs of water molecules matched by  $1.0 \text{ \AA}$  distance. However, if the 1OPS model is compared, only six equivalent waters are found. They bind to the backbone atoms of Leu10, Ala16, Leu17, Ile20, and Leu51, and the side chain of Gln9. The distribution of these water molecules appears to be sporadic and featureless, and probably unrelated with any ice structure.

## CONCLUSION

The structural refinement of RD1 at  $0.62\text{-}\text{\AA}$  resolution revealed several interesting features regarding the micro-heterogeneity of the protein and solvent structures. One disordered peptide bond and four side chains were observed and explicitly modeled with dual conformations. A number of water molecules were also split into pairs of very close locations. These were a result of hydrogen bond combina-

tions with a slightly different repertoire. The difference maps calculated using the refined model also allowed visualization of many bond electrons and possible lone pairs. The overall structure of RD1 is essentially identical to other members of this protein family, which shares a common ice-binding mechanism. Although all orthorhombic crystals of type III AFPs had very similar unit-cell dimensions, the RD1 molecules appeared most ordered in the crystal probably due to only a few minor differences in lattice contact interactions. In addition, a cavity in the RD1 molecule was identified. Site-directed mutations on the surrounding amino acid residues would lead to elucidation of the function of this unusual structural feature.

In summary, the structural refinement of biological macromolecules at such an ultra-high resolution is complex and challenging. Our results here illustrated that more work needs to be done to reach a satisfactory conclusion, not the least including: *i*) ample synchrotron beam time to collect many more data sets using newer generation of detectors and data processing protocols, *ii*) testing different refinement programs, and *iii*) treatment of imperfect models (bulk solvents, multiple conformers, etc.). This new work will be carried out in the future.

The authors are grateful to the Advanced Photon Source for generous beam time allocation, and to Andrzej Joachimiak and Ruslan Sanishvili for help with data collection.

This work was supported by grant NSF OPP 99-09841 to A.L.D. and C-H.C.C. from the National Science Foundation, United States, and grant NSC 90-2321-B-001-015 to A.H-J.W. from the National Science Council, Republic of China.

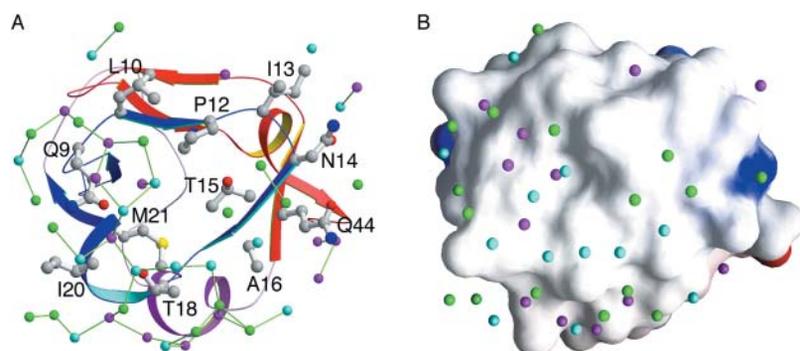


FIGURE 11 Ice-binding residues and water molecules on the IBS of RD1. Water molecules are colored in magenta, cyan, and green, respectively, according to the distance criteria of less than  $0.5 \text{ \AA}$ ,  $0.5\text{--}1.0 \text{ \AA}$  and more than  $1.0 \text{ \AA}$  from their equivalents in the 1HG7 structure. In (*A*) the water molecules with distances less than  $3.2 \text{ \AA}$  from each other are connected with thin bonds. In (*B*) the RD1 molecule is shown as surface charge potential representation. The view is similar to that in Fig. 10, with the molecule rotated counterclockwise by  $\sim 100^\circ$ .

## REFERENCES

- Antson, A. A., D. J. Smith, D. I. Roper, S. Lewis, L. S. D. Caves, C. S. Verma, S. L. Buckley, P. J. Lillford, and R. E. Hubbard. 2001. Understanding the mechanism of ice binding by type III antifreeze proteins. *J. Mol. Biol.* 305:875–889.
- Baardsnes, J., and P. L. Davies. 2001. Sialic acid synthase: the origin of fish type III antifreeze protein? *Trends Biochem. Sci.* 26:468–469.
- Barton, G. L. 1993. ALSSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* 6:37–40.
- Brunger, A. T. 1993. X-PLOR Version 3.1: A System for X-Ray Crystallography and NMR. Yale University Press, New Haven, Connecticut.
- Brunger, A. T., P. D. Adams, G. M. Clore, W. L. Delano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.* D54:905–921.
- Cheng, C.-H. C., and A. L. DeVries. 1989. Structures of antifreeze peptides from the Antarctic fish *Austrolycichthys brachycephalus*. *Biochim. Biophys. Acta.* 997:55–64.
- Collaborative Computational Project Number 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* D50:760–763.
- Engh, R. A., and R. Huber. 1991. Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr.* A47:392–400.
- Esnouf, R. M. 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* 15:132–134.
- Graether, S. P., C. I. DeLuca, J. Baardsnes, G. A. Hill, P. L. Davies, and Z. Jia. 1999. Quantitative and qualitative analysis of type III antifreeze protein structure and function. *J. Biol. Chem.* 274:11842–11847.
- Hew, C. L., N.-C. Wang, S. Joshi, G. L. Fletcher, G. K. Scott, P. H. Hayes, and B. Buettner. 1988. Multiple genes provide the basis for antifreeze protein diversity and dosage in the ocean pout, *Macrozoarces americanus*. *J. Biol. Chem.* 263:12049–12055.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature.* 409:860–921.
- Jelsch, C., M. M. Teeter, V. Lamzin, V. Pichon-Pesme, R. H. Blessing, and C. Lecomte. 2000. Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. *Proc. Natl. Acad. Sci. USA.* 97:3171–3176.
- Jia, Z., C. I. DeLuca, H. Chao, and P. L. Davies. 1996. Structural basis for the binding of a globular AFP to ice. *Nature.* 384:285–288.
- Jones, T. A., J. Y. Zou, S. W. Cowan, and M. Kjeldgaard. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* A47:110–119.
- Kraulis, P. J. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24:946–950.
- Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26:283–291.
- Matthews, B. W. 1968. Solvent content of protein crystals. *J. Mol. Biol.* 33:491–497.
- Merrit, E. A., and M. E. P. Murphy. 1994. Raster3D Version 2.0: a program for photorealistic molecular graphics. *Acta Crystallogr.* D50:869–873.
- Nicholls, A., K. A. Sharp, and B. Honig. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins.* 11:281–296.
- Otwinowski, Z., and W. Minor. 1997. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276:307–326.
- Richardson, J. S. 1981. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 34:167–339.
- Scott, G. K., P. H. Hayes, G. L. Fletcher, and P. L. Davies. 1988. Wolfish antifreeze protein genes are primarily organized as tandem repeats that each contain two genes in inverted orientation. *Mol. Cell. Biol.* 8:3670–3675.
- Sheldrick, G. M., and T. R. Schneider. 1997. SHELXL: high resolution refinement. *Methods Enzymol.* 277:319–343.
- Sonnichsen, F. D., B. D. Sykes, H. Chao, and P. L. Davies. 1993. The nonhelical structure of AFP type III. *Science.* 259:1154–1157.
- Wang, X., A. L. DeVries, and C.-H. C. Cheng. 1995. Antifreeze peptide heterogeneity in an Antarctic eel pout includes an unusually large major variant comprised of two 7 kDa type III AFPs linked in tandem. *Biochim. Biophys. Acta.* 1247:163–172.
- Yang, D. S. C., W.-C. Hon, S. Bubanko, Y. Xue, J. Seetharaman, C. L. Hew, and F. Sicheri. 1998. Identification of the ice-binding surface on a type III AFP with a “flatness function” algorithm. *Biophys. J.* 74: 2142–2151.