Journal of Biomolecular NMR (2005) 32: 251–256
DOI 10.1007/s10858-005-8206-3

NMR structure note

Solution structure of an antifreeze protein CfAFP-501 from *Choristoneura fumiferana**

Congmin Li[a,b], Xianrong Guo[a,b], Zongchao Jia[d], Bin Xia[a,b,c] & Changwen Jin[a,b,c,**]
[a]*Beijing Nuclear Magnetic Resonance Center,* [b]*College of Chemistry and Molecular Engineering,* [c]*College of Life Sciences, Peking University, Beijing 100871, China,* [d]*Department of Biochemistry, Queen's University, Kingston Ontario K7L 3N6, Canada*

Received 30 March 2005; Accepted 25 May 2005

Key words: antifreeze protein, constrains, dynamics, Nuclear Magnetic Resonance, solution structure

Abstract

Antifreeze proteins (AFPs) are widely employed by various organisms as part of their overwintering survival strategy. AFPs have the unique ability to suppress the freezing point of aqueous solution and inhibit ice recrystallization through binding to the ice seed crystals and restricting their growth. The solution structure of CfAFP-501 from spruce budworm has been determined by NMR spectroscopy. Our result demonstrates that CfAFP-501 retains its rigid and highly regular structure in solution. Overall, the solution structure is similar to the crystal structure except the N- and C-terminal regions. NMR spin-relaxation experiments further indicate the overall rigidity of the protein and identify a collection of residues with greater flexibilities. Furthermore, Pro91 shows a *cis* conformation in solution instead of the *trans* conformation determined in the crystal structure.

Biological context

Antifreeze proteins (AFPs) are widely employed by various organisms as part of their overwintering survival strategy, especially when the temperatures are below the freezing point of their bodily fluids (Jia and Davies, 2002). AFPs have the unique ability to suppress the freezing point of aqueous solution (melting point is unaffected) and inhibit ice recrystallization through binding to the ice seed crystals and restricting their growth, known as the adsorption-inhibition mechanism (Fletcher et al., 2001; Jia and Davies, 2002). Numerous theoretical and experimental studies have been carried out in order to understand the interaction between AFPs and ice. However, the molecular mechanism of ice-binding for AFPs still remains poorly understood.

AFPs have been identified from various organisms, such as fish, plants, bacteria, fungi and insects (Fletcher et al., 2001; Graether and Sykes, 2004). The insect spruce budworm (*Choristoneura fumiferana* (Cf)) produces a number of isoforms of its highly active AFP (CfAFP) (Doucet et al., 2000). The 121-residue isoform 501, known as CfAFP-501, is one of the most active antifreeze proteins discovered thus far. A recombinant version of CfAFP-501 can depress the freezing point up to −4.7 °C at 0.05 mM protein concentration. The crystal structure of CfAFP-501 has been reported and revealed a typical β–helical structure similar to that of CfAFP-337 (Leinala et al., 2002a). The 31-amino acid insertion in CfAFP-501

* Structural data have been deposited at PDB (1Z2F) and Chemical shift data at BMRB (bmrb 6111).
** To whom correspondence should be addressed. E-mail: changwen@pku.edu.cn

folds into two additional loops with the conserved T-X-T motif, which extend the ice-binding surface and lead to 2–3 times activity than the CfAFP-337.

Solution NMR is a well-established technique to investigate the three-dimensional (3D) structure and dynamics of biomolecules at atomic level. In order to obtain further insights into the biochemical features of this protein, particularly its solution behavior, we have used high-resolution solution NMR to determine the 3D structure and characterize the dynamics of CfAFP-501.

Methods and results

The gene cloning, protein expression, refolding of inclusion bodies and purification of CfAFP-501 have been described previously (Leinala et al., 2002b; Li and Jin, 2004). Briefly, uniformly ^{15}N or $^{13}C/^{15}N$-labeled CfAFP-501 samples were prepared by growing the *E. coli* strain BL21(DE3)/pLysS cells containing the expression plasmid in the M9 minimal media. The refolding and purification procedures of CfAFP-501 protein followed that of CfAFP-337 (Gauthier et al., 1998). The isotope-labeled samples for NMR experiments were prepared in 50 mM sodium phosphate buffer at pH 5.7 with 50 mM sodium chloride, 90% H_2O/10% D_2O and DSS was added as the internal chemical shift reference. The final protein concentration was about 0.2 mM.

All NMR experiments were performed at 290 K at the Beijing Nuclear Magnetic Resonance Center (BNMRC) on Bruker Avance 500 MHz (with cryoprobe) and 800 MHz spectrometers, both equipped with triple-resonance probes and pulsed-field gradients. Data were processed with NMRPipe (Delaglio et al., 1995) and analyzed using NMRview (Johnson and Blevins, 1994). The chemical shift assignments have been reported previously (Li and Jin, 2004). Briefly, the 3D triple-resonance spectra of HNCA, HNCACB, CBCA(CO)NH, HBHA(CO)NH and HNCO were collected to obtain the backbone resonance assignments. The 3D spectra H(CCO)NH-TOCSY, (H)C(CO)NH-TOCSY, HCCH-TOCSY, HCCH-COSY and ^{15}N-edited TOCSY-HSQC, CCH-COSY were collected for the side chain resonance assignments. 3D ^{15}N-NOESY-HSQC (mixing times of 50 and 100 ms), and ^{13}C-NOESY HSQC (mixing time 80 ms) were recorded to confirm the assignments and generate distance constraints for structure calculations. The chemical shift assignments have been deposited in the BioMagResBank (http://www.wisc.edu) under accession number 6111. The heteronuclear $^{1}H-^{15}N$ cross relaxation (NOE) experiment was performed on an 800 MHz NMR spectrometer at 290 K. The NOE values were determined from the ratio of peak heights for experiments with and without ^{1}H saturation (Farrow et al., 1994).

The solution structure of CfAFP-501 has been determined using inter-proton nuclear Overhauser effect (NOE)-derived distance constraints in combination with the dihedral angles and hydrogen and disulfide bonds information. TALOS (Cornilescu et al., 1999) and CSI (Wishart and Sykes, 1994) programs were used to predict dihedral angles ϕ and ψ constraints. Hydrogen bond restraints were determined from hydrogen-deuterium exchange experiments and intermediate range NOEs in conjunction with secondary structural information. Five disulfide bonds, Cys4–Cys17, Cys25–Cys38, Cys56–Cys68, Cys93–Cys116 and Cys98–Cys111 were confirmed by NOE contacts and the C^{β} chemical shifts of the cysteine residues. A total of 1564 NOE distance constraints, 116 dihedral angle constraints and 70 hydrogen bond restraints were used for the final calculations of the structures. 200 structures were initially calculated by CYANA (Güntert et al., 1997). The 100 lowest-energy structures were used as input structures and refined using AMBER7 (Case et al., 2002). Note that only partial stereo-specific assignments could be obtained and chi-1 angle constraints could not be measured. Consequently, side chain positioning at the protein surface is the result of the combination of NOEs, intrinsic packing constraints, and rotamer preference of the AMBER7 force field. Finally, the 20 lowest-energy structures were selected. PROCHECK (Laskowski et al., 1996) was used to analyze the quality of structures, and MOLMOL (Koradi et al., 1996) was used to evaluate the RMSD values. The 3D coordinates of CfAFP-501 have been deposited in the Protein Data Bank (PDB) under the accession number 1Z2F. A superimposition of 20 representative structures, together with the ribbon diagram representations of the mean structure, are shown in Figure 1a, b and c, respectively. The structural statistics is shown in Table 1.

In the crystal structure (PDB entry 1M8N), CfAFP-501 structure is dimeric conformation

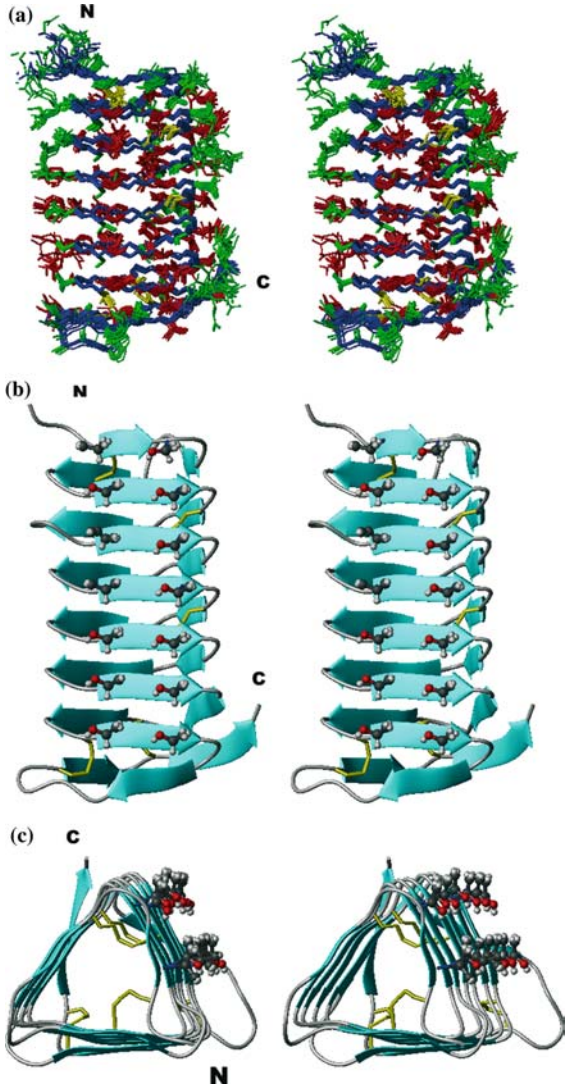

Figure 1. The solution structure of CfAFP-501. (a) A stereo view of the 20 lowest-energy conformers with superposition for best fit for the backbone heavy atoms of residues 10–101. The backbone, side-chains of regular secondary structures, and the remaining side-chains heavy atoms are colored in blue, red and green, respectively. The disulfide bonds are colored in yellow. (b) Ribbon diagram of the energy-minimized mean structure of CfAFP-501 with the ice-binding surface toward the viewer. The oxygen atoms of the threonine side-chains on the ice-binding surface are shown in red spheres. (c) Ribbon diagram of the mean structure with the triangle cross-section towards the viewer. The graphics were generated using MOLMOL (Koradi et al., 1996).

(Leinala et al., 2002b). In order to characterize the oligomerization of CfAFP-501 in solution, Gel-filtration and ultracentrifugation experiments were performed under the same conditions used in NMR experiments at the temperature of 290 K.

Table 1. Structural statistics

Distance constraints	
Intra-residue	543
Sequential	428
Medium-range	97
Long-range	496
Total	1564
Dihedral angle constraints (ϕ and ψ)	116
Hydrogen bond constraints	70
S–S Bond constraints	5
Structure statistics (20 structures)	
CYANA target function value	0.68 ± 0.11
Violation statistics	
NOE violation (>0.3 Å)	0
Torsion angle violation ($>5°$)	0
AMBER Energies (kcal/mol)	
Mean AMBER energy	-5815.95 ± 6.64
NOE distance constraints violation energy	10.79 ± 2.17
Torsion angle constraints violation energy	4.28 ± 1.90
Ramachandran statistics	
Residues in most favored regions (c/o)	84.5
Residues in additional allowed regions (c/o)	14.7
Residues in generously allowed regions (c/o)	0.8
Residues in disallowed regions (c/o)	0.0
RMSD to the mean structure (residue 10–121)	
Backbone heavy atoms	
All residues (Å)	0.55 ± 0.15
Regular secondary structure (Å)	0.36 ± 0.08
All heavy atoms	
All residues (Å)	1.23 ± 0.20
Regular secondary structure (Å)	1.15 ± 0.21

The result indicated that CfAFP-501 exists mostly in the monomeric state. This is in general agreement with the observation of the shorter isoform, CfAFP-337 (Graether et al., 2003).

CfAFP-501 is a cysteine, threonine and serine enriched protein. The protein was stabilized by hydrophobic side chain stacking inside the core of the protein, five disulfide bridges and the inter-loop hydrogen bond networks. The disulfide bonds are important for the protein folding and stability. After adding dithiothreitol (DTT) to break the disulfide bonds, the protein became unfolded gradually. In addition, the extensive amount of hydrogen bonds helps rigidify the entire protein. Figure 2a shows the $\{^1H\}-^{15}N$ heteronu-

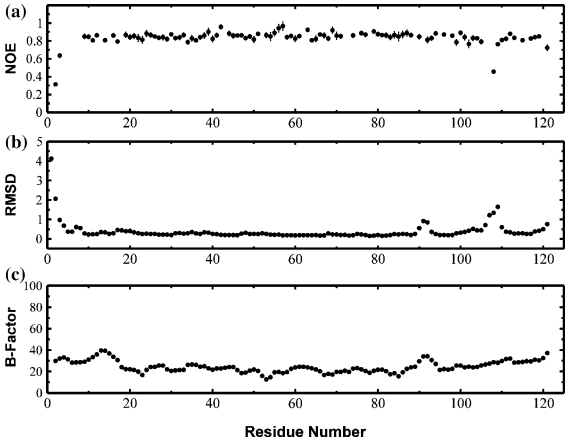

Figure 2. Analysis of flexibility of the CfAFP-501 protein. (a) The plot of the $\{^1H\}-^{15}N$ heteronuclear NOE values *versus* residue numbers for CfAFP-501. The experiments were performed on a Bruker Avance 800 MHz spectrometer at 290 K. (b) The RMSD (in Å) of the backbone nitrogen atoms *versus* residue numbers. (c) The B-factors of the backbone nitrogen atoms in the crystal structure of CfAFP501.

clear NOE values *versus* the protein sequence. It indicates that CfAFP-501 is rigid overall except the N-terminal residues and residues 106–110 near the C-terminus. The RMSD of backbone nitrogen atoms of the solution structure, together with that of the B-factor values of the crystal structure are shown in Figure 2b and 2c, respectively. It demonstrates a good correlation between the RMSD and the $\{^1H\}-^{15}N$ heteronuclear NOE values.

An overlay of the C^{α} trace of the solution structure (mean structure) of CfAFP-501 (PDB entry 1Z2F) with that of the crystal structure (PDB entry 1M8N) is shown in Figure 3. The RMSD value between the solution and the crystal structures is 0.38 Å for residues in the regular secondary structure. The structures are similar overall except the N-terminal region, and two loop regions including residues 90–93 and 106–110 near the C-terminal region. Both loop regions contain proline residues Pro91 and Pro107 exhibit larger RMSD values. In the crystal structure, both X-prolyl bonds are in a *trans* conformation. Interestingly, in the solution structure the Gly90-Pro91 bond is a *cis* conformation. This was identified by NOE contacts and the chemical shifts (Wüthrich, 1986). NOE contacts between proline residue (i) and previous residue ($i-1$) can be used to identify the conformation of X-prolyl bond. For Pro91, only contacts between $H^{\alpha(i-1)}$ and $H^{\alpha(i)}$, but not between $H^{\alpha(i-1)}$ and $H^{\delta(i)}$, were found. This indicates

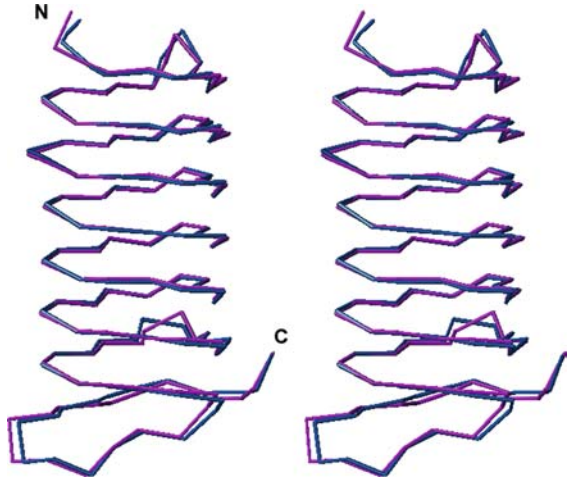

Figure 3. An overlay of the C^{α} trace of solution structure (blue, PDB entry 1Z2F) with the crystal structure (magenta, PDB entry 1M8N) of CfAFP-501.

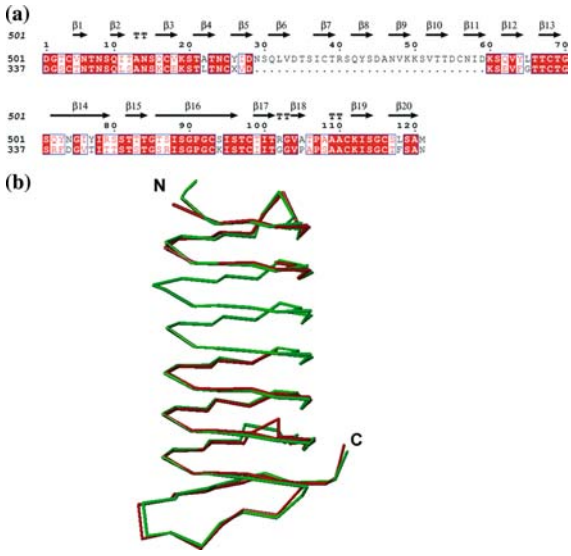

Figure 4. (a) A comparison of amino acid sequences of CfAFP-501 and CfAFP-337. (b) A comparison of the crystal structure of CfAFP-337 (red) and solution structure of CfAFP-501 (green) by mapping the N-terminal and C-terminal C^{α} trace of CfAFP-337 on the C^{α} trace of CfAFP-501, respectively.

that the Gly90-Pro91 bond is in a *cis* conformation. In contrast, for the Pro107 only contacts between $H^{\alpha(i-1)}$ and $H^{\delta(i)}$, but not between $H^{\alpha(i-1)}$ and $H^{\alpha(i)}$, were found, demonstrating a *trans* conformation for the Thr106–Pro107 bond. The chemical shift difference of C^{β} and C^{γ} ($\Delta\beta\gamma \equiv \delta[C^{\beta}]-\delta[C^{\gamma}]$) of proline residue can also be used to identify the *cis*/*trans* conformation of X-prolyl bonds (Schubert et al., 2002). The chemical shift differences $\Delta\beta\gamma$ are

9.69 ppm for Pro91and 3.49 ppm for Pro107, which are typical values corresponding to the *cis* and *trans* conformation, respectively.

Furthermore, we examined *cis*/*trans* conformation of the prolines in the isoform of CfAFP-337 (Leinala et al., 2002a). The comparisons of the amino acid sequences and an overlay of the solution structure of CfAFP-501 with the crystal structure of CfAFP-337 (PDB entry 1L0S) are shown in Figure 4a and b, respectively. Excluding the extra two loop of CfAFP-501 (CfAFP-501-Δ-2-loop), the sequence identity between CfAFP-501-Δ-2-loop and CfAFP-337 is 68%. Pro91 and Pro107 of CfAFP-501 correspond to Pro60 and Pro76 of CfAFP-337 in the protein sequences. Pro91 of CfAFP-501 and Pro60 of CfAFP-337 are in a *cis* conformation, while Pro107 of CfAFP-501 and Pro76 of CfAFP-337 are in a *trans* conformation. The solution structures of CfAFP-337 (PDB entry 1EWW and 1N4I) were determined at temperatures of 303 and 278 K (Graether et al., 2000, 2003). The structures showed a *trans* conformation of Pro60 of CfAFP-337 at both temperatures, which is different from the crystal structure. By checking the chemical shift differences of Pro60 in CfAFP-337, we found that the chemical shift difference between C^{β} and C^{γ} ($\Delta\beta\gamma = \delta[C^{\beta}]\text{-}\delta[C^{\gamma}]$) of Pro60 is 2.82 ppm (BMRB accession number 5572) at 303 K which corresponds to a possible *trans* conformation. Meanwhile, the chemical shift difference $\Delta\beta\gamma$ of Pro60 is 9.33 ppm at 278 K (BMRB accession number 5573), which is a possible indication of *cis* conformation. However, further information on NOE contacts is required to confirm the *cis*/*trans* conformation.

To investigate the possible *cis*/*trans* conformational switch in Pro91 at different temperatures, we have also assigned the ^{13}C chemical shifts of Pro91 side-chain at temperatures of 303 and 278 K using 3D triple resonance experiments HNCA and CCH-COSY. The chemical shift difference $\Delta\beta\gamma$ is 9.07 ppm at 303 K, 9.69 ppm at 290 K and 9.10 ppm at 278 K. The chemical shift differences at temperature from 278 to 303 K are typical values for the *cis* conformation.

Discussion and conclusion

The CfAFP-501 shows a regular left-handed β-helical structure containing seven left-handed loops with 15–17 residues per loop, and followed by one right-handed loop with 18 residues. Each left-handed turn contains three β-strands. The parallel β-strands form three β-sheet platforms with rectangular shape (Figure 1b). The cross-section perpendicular to the axis of the β-helix has a triangular shape (Figure 1c). One of the three platforms contains two ranks of aligned threonine residues known as the repetitive T-X-T motif, which forms the putative ice-binding surface (Graether et al., 2000; Liou et al., 2000). Compared with the CfAFP-337, the 31 amino acids insertion forms two extra loops within CfAFP-501 to extend the height of the prism and increase the size of the ice-binding surface (Marshall et al., 2004). On the other hand, CfAFP-501 has 2–3 times the thermal hysteresis activity of CfAFP-337. This fact demonstrates that the size of the ice-binding face is an important determinant of its antifreeze activity. The higher activity of CfAFP-501, as compared with the shorter isoform, CfAFP-337 is mainly due to the increased length of the β-helical platform and consequently the increased ice-binding area.

In the solution structure of CfAFP-501, the overall shape of a triangular prism is 34 Å in height with each side of the triangle being 16–18 Å in length. In loop1, loop3 and loop4, the motifs V-X-T, I-X-T, V-X-T substituted the TXT motif on the ice-binding surface, respectively. The average distance between hydroxyl groups of threonine within the T-X-T motif is 6.75 Å. The average distance between equivalent hydroxyls of T-X-T motifs in adjacent loops is 4.57 Å. As in accordance with the crystal structural studies, the two-dimensional array of threonine side chains makes an excellent match to the repeated spacing between oxygen atoms in the ice lattice on the basal plane along the *a*-axis (4.52 Å). This lattice matching is instrumental for ice-binding function of the AFP. Our work demonstrates that CfAFP-501 retains its rigid and highly regular structure in solution, which essentially needs no conformational change for ice recognition and binding.

Acknowledgements

All NMR experiments were carried out at the Beijing NMR Center, Peking University. We thank Weibin Gong and Dr. Yingang Feng for

useful discussions. We thank Ping Wei for assistance on the ultracentrifugation experiments. This work was supported by the National Natural Science Foundation of China (No. 30325010) to C.J.

References

Case, D.A. et al. (2002) *AMBER 7*, University of California, San Francisco, CA.
Cornilescu, G. et al. (1999) *J. Biomol. NMR*, **13**, 289–302.
Delaglio, F. et al. (1995) *J. Biomol. NMR*, **6**, 277–293.
Doucet, D. et al. (2000) *Eur. J. Biochem.*, **267**, 6082–6088.
Farrow, N.A. et al. (1994) *Biochemistry*, **33**, 5984–6003.
Fletcher, G.L. et al. (2001) *Annu. Rev. Physiol.*, **63**, 359–390.
Gauthier, S.Y. et al. (1998) *Eur. J. Biochem.*, **258**, 445–453.
Graether, S.P. et al. (2000) *Nature*, **406**, 325–328.
Graether, S.P. et al. (2003) *J. Mol. Biol.*, **327**, 1155–1168.
Graether, S.P. and Sykes, B.D. (2004) *Eur. J. Biochem.* **271**, 3285–3296.
Güntert, P. et al. (1997) *J. Mol. Biol.*, **273**, 283–298.
Jia, Z. and Davies, P.L. (2002) *Trends Biochem. Sci.* **27**, 101–106.
Johnson, B.A. and Blevins, R.A. (1994) *J. Biomol. NMR* **4**, 603–614.
Koradi, R. et al. (1996) *J. Mol. Graph.*, **14**, 51–55.
Laskowski, R.A. et al. (1996) *J. Biomol. NMR*, **8**, 477–486.
Leinala, E.K. et al. (2002a) *Structure*, **10**, 619–627.
Leinala, E.K. et al. (2002b) *J. Biol. Chem.*, **277**, 33349–33352.
Li, C. and Jin, C. (2004) *J. Biomol. NMR* **30**, 101–102.
Liou, Y.C. et al. (2000) *Nature*, **406**, 322–324.
Marshall, C.B. et al. (2004) *Biochemistry*, **43**, 11637–11646.
Schubert, M. et al. (2002) *J. Biomol. NMR*, **24**, 149–154.
Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR* **4**, 171–180.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* Wiley, New York, NY.