DOI 10.5012/bkcs.2011.32.5.1640

Solution Structure of Water-soluble Mutant of Crambin and Implication for Protein Solubility

Su-Jin Kang,† Jongsoo Lim, Bong-Jin Lee,† and Hee-Chul Ahn*

*College of Pharmacy, Dongguk University, Seoul 410-820, Korea. *E-mail: hcahn@dongguk.edu*
†Research Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, Seoul 151-742, Korea
Received January 24, 2011, Accepted March 29, 2011

Water-soluble mutant of intrinsically insoluble protein, crambin, was produced by mutagenesis based on the sequence analysis with homologous proteins. Thr1, Phe13, and Lys33 of crambin were substituted for Lys, Tyr, and Lys, respectively. The resultant mutant was soluble in aqueous buffer as well as in dodecylphosphocholine (DPC) micelle solution. The ^{1}H-^{15}N spectrum of the mutant crambin showed spectral similarity to that of the wild-type protein except for local regions proximal to the sites of mutation. Solution structure of water-soluble mutant crambin was determined in aqueous buffer by NMR spectroscopy. The structure was almost identical to the wild-type structure determined in non-aqueous solvent. Subtle difference in structure was very local and related to the change of the intra- and inter-protein hydrophobic interaction of crambin. The structural details for the enhanced solubility of crambin in aqueous solvent by the mutation were provided and discussed.

Key Words : Crambin, NMR, Solution structure, Protein solubility

Introduction

Solubility of proteins in aqueous buffer depends on the distribution of hydrophilic and hydrophobic amino acids on the surface of proteins. Hydrophobic residues predominantly occur in the core of globular protein. Proteins that have high hydrophobic amino acid content on the surface have low solubility in an aqueous solvent. Charged and polar surface residues interact with ionic groups in the solvent and increase the solubility of proteins.

If solubility of a protein could be controlled by change of the distribution of hydrophobic and hydrophilic amino acids without disruption of enzymatic activity, it would be very helpful to develop novel versions of enzyme in biological sciences as well as in biotechnology. For example, high concentration and long-term stability of a protein in aqueous solvent are required in varieties of applications in pharmaceuticals,[1] and biochemical studies,[2] and structural biology.[3,4]

Many strategies have been conducted to improve the yield of soluble proteins, such as optimization of the protocols for expression, purification, and solubilization.[3-7] Proteins expressed as inclusion bodies were denatured, purified, and refolded to yield soluble forms.[8,9] Several proteins showed enhanced solubility and long-term stability by simultaneous addition of charged amino acids L-Arg and L-Glu at 50 mM to the buffer, which increased the maximal concentration up to 8.7 times.[10]

Here, we present solubility-enhancement of a water-insoluble protein by mutagenesis. As a model system, we used crambin, a highly hydrophobic and water-insoluble plant protein isolated from the seeds of the plant *Crambe abyssinica*.[11] Crambin has been widely used in methodological development for the determination of highly accurate structure by X-ray crystallography[12,13] and for structure calculation by NMR spectroscopy.[14-19] Because of insolubility of crambin in aqueous solvent, those structures were determined from the crystals grown in 50% ethanol or in 75% acetone/25% water mixture. Recently, crambin was successfully produced in *Escherichia coli* with good yield and the structures in membrane-environments were determined by NMR spectroscopy.[20]

We produced the water-soluble version of crambin mutant based on the sequence analysis with homologous proteins, thionin family.[21] Three-dimensional structure of the mutant crambin were determined by NMR spectroscopy. The structural difference between the water-soluble mutant and insoluble wild-type crambin were compared and the effect of mutation for protein solubility was discussed based on the determined structure.

Experimental Section

Mutagenesis. Sites of mutation were selected based on the sequence alignment of thionin family and the selected residues of crambin were subjected to site-directed mutagenesis. Thr1, Phe13, and Lys33 were substituted for Lys, Tyr, and Lys, respectively. The final constructed plasmid was confirmed by DNA sequencing.

Preparation of Crambin^{KYK}. The expression and purification of T1K, F13Y, and I33K mutant crambin (referred as to crambin^{KYK}) were performed by the previously described procedures for the preparation of the wild-type crambin sample.[20] ^{13}C and ^{15}N labeled crambin^{KYK} was prepared from *Escherichia coli* culture in M9 minimal medium containing $^{15}NH_4Cl$ (1 g/L) and/or ^{13}C glucose (2 g/L).

NMR Spectroscopy. Crambin^{KYK} was dissolved in 20 mM potassium phosphate buffer, pH 6.0, containing 10% D_2O. The final concentration of NMR sample was 0.5 mM.

NMR experiments were performed at 25 °C on Varian VNMRS 900 spectrometer equipped with a triple-resonance ColdProbe and a pulsed field gradient unit. Backbone resonance assignments were accomplished from ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC), 3D-HNCO, and HNCACB spectra. Resonances of side-chains were assigned from 3D-HCCH-COSY spectrum. Distance constraints for crambin^{KYK} were obtained by analyses of NOE cross-peaks from the 3D ^{15}N- and ^{13}C-edited NOESY-HSQC spectra with 80 ms mixing times, respectively. 150 mM dodecyl phosphocholine (DPC) was added to NMR sample to measure the ^{1}H-^{15}N HSQC spectrum of crambin^{KYK} in membrane environment. The ^{1}H-^{15}N resonance assignments for crambin^{KYK} in DPC micelles were accomplished by analysis of H_N-H_N NOE cross-peaks from the 3D ^{15}N-edited NOESY-HSQC spectrum of the micelle sample. All spectra were processed with nmrPipe[22] and analyzed with NMRView.[23]

Structure Determination. The structure of crambin^{KYK} in the aqueous solution was calculated from 797 distance constraints, 62 torsion angle constraints (31 ϕ and 31 ψ), and 30 hydrogen-bond constraints. The TALOS software[24] was used to determine torsion angle constraints from the assigned chemical shift values. The NOE cross-peaks from the 3D ^{15}N- and ^{13}C-edited NOESY-HSQC spectra of crambin^{KYK} were assigned semi-automatically by using the program ARIA 2.0.[25] NOE-based distance restraints were derived based on the peak volumes. The structure was calculated by using the program CNS 1.2[26] and ARIA 2.0. Hydrogen bond restraints were incorporated into the final structure calculation based on the secondary structure. In the final ARIA run, 100 structures were calculated and among them the 20 energy-lowest structures were selected and water-refined under the OPLS force field.[27] The water-refined structures were used for statistical analysis by using PROCHECK-NMR 3.4.[28]

Results and Discussion

Mutagenesis. Sequence alignment of 37 proteins in thionin family showed the consensus sequence (Fig. 1(a)).[29] Based on the consensus sequence, three residues, Thr1, Phe13, and Ile33, were selected and substituted for Lys, Tyr, and Lys, respectively. The net charge of the wild-type crambin was zero, since crambin contains one arginine and one glutamate. However, the net charge of the mutated crambin resulted in +2, since two lysine residues were introduced. The mutant crambin, crambin^{KYK}, was soluble in aqueous buffer, whereas the wild-type crambin was completely insoluble. The molecular weight of crambin^{KYK} was 4789.0 Da, which was determined by MALDI-TOF mass spectrometry and consistent with the theoretical value, 4788.5 Da.

NMR Spectra of Crambin^{KYK} in Aqueous and Micelle Solution. Most of all ^{1}H, ^{15}N, and ^{13}C resonances of crambin^{KYK} were unambiguously assigned (data not shown). Fig. 1B shows the ^{1}H-^{15}N HSQC spectrum of ^{13}C and ^{15}N-labeled crambin^{KYK}. The ^{1}H-^{15}N correlations from each residues

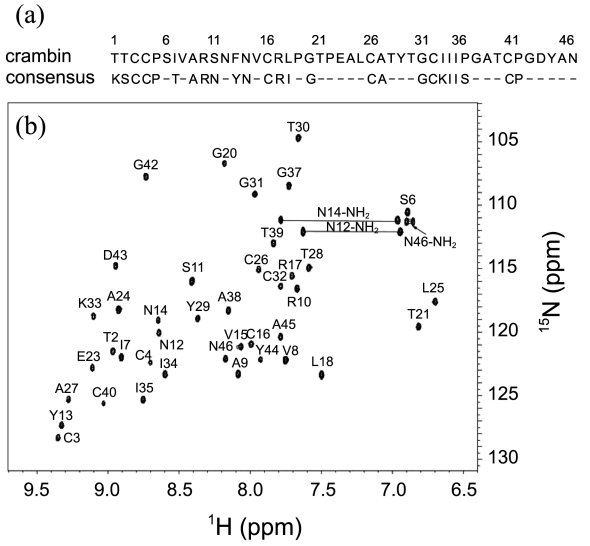

Figure 1. Amino acid sequence of crambin and the consensus sequence of thionin family (a). ^{1}H-^{15}N HSQC spectrum of ^{13}C and ^{15}N-labeled crambin^{KYK} in phosphate buffer, pH 6.0 (b). The resonances were labeled with one-letter amino acid codes with respective sequence.

were well-dispersed, suggesting crambin^{KYK} is folded in aqueous solution. Since the wild-type crambin is not soluble, the ^{1}H-^{15}N HSQC spectrum of crambin in aqueous solution is not available. The spectra of crambin in acetone/water mixture and DPC micelles were reported previously.[20] Thus, we added 150 mM DPC to crambin^{KYK} sample and recorded the ^{1}H-^{15}N HSQC spectrum (Fig. 2) to compare with that of the wild-type protein in DPC micelles.

Crambin^{KYK} was also soluble in DPC micelles, suggesting that the mutant crambin is soluble in aqueous solvent as well as membrane-mimetic environment. The ^{1}H-^{15}N HSQC spectra of the wild-type and mutant crambin in DPC micelles were

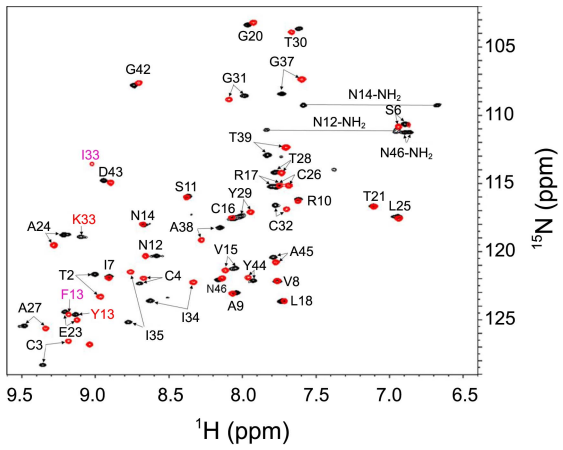

Figure 2. Overlaid ^{1}H-^{15}N HSQC spectra of crambin^{KYK} (black) and wild-type crambin (red). The corresponding residues are indicated. The original amino acids are colored in pink (F13 and I33 for crambin) and the mutated ones in red (Y13 and K33 for crambin^{KYK}). Since N-terminal first residue is usually not detectable in ^{1}H-^{15}N HSQC spectrum, T1 and K1 for the wild-type and mutant crambin, respectively, were not shown.

Table 1. Structural statistics for the final 20 structures of water-soluble crambin mutant

Number of experimental restraints	
NOE distance restraints	797
Hydrogen bond restraints	30
Torsion angle restraints	62
Number of violations	
NOE > 0.3 (Å)	0
dihedral angle > 0.5 (°)	0
Energies (kcal/mol)	
E_{total}	−1508.54 ± 32.71
E_{bond}	1.78 ± 0.15
E_{angle}	16.58 ± 0.97
$E_{improper}$	1.67 ± 0.45
E_{dihed}	207.77 ± 3.64
E_{vdw}	−361.39 ± 5.85
E_{elec}	−1374.96 ± 32.05
E_{NOE}	5.20 ± 2.16
RMSD from experimental restraints	
NOE (Å)	0.011 ± 0.002
Torsion angle (°)	0.024 ± 0.078
RMSD from the idealized geometry	
Bonds (Å)	0.002 ± 0.000
Bond angles (°)	0.299 ± 0.009
Improper angles (°)	0.180 ± 0.023
Ramachandran plot (%)[a]	
Most favored	92.1
Additionally allowed	2.9
Generously allowed	0
Disallowed	0
RMSD of well-ordered region (Å)[b]	
Backbone	0.36
Heavy atoms	0.74

[a]As determined y PROCHECK-NMR[28]. [b]RMSD was calculated for the well-ordered region of crambin^{KYK} (Thr2-Ile35)

overlaid in Figure 2. The spectrum of crambin^{KYK} in DPC micelles corresponded well to that of the wild-type crambin in DPC micelles. Most of the ^{1}H-^{15}N resonances of crambin^{KYK} were identical to or slightly shifted from the respective resonances of the wild-type crambin, except for about one fourth of crambin residues close to the mutated ones. The most dramatic change in the chemical shift occurred at the residues, Thr2, Cys3, Gly31, Ile34, Ile35, Gly37, and Thr39, which are close to the mutated residues, two lysines at 1st and 33rd position in the sequence or in tertiary structure. Other significant spectral change happened on the residues of the second α-helix, spanning from Glu23 to Tyr29, which might be affected by the mutation of phenylalanine to tyrosine at the position 13 (discussed below).

Three-dimensional Structure of Crambin^{KYK}. To evaluate the effect of mutation on the structure of crambin, we determined the solution structure of crambin^{KYK} in aqueous buffer and compared it with the previously reported ones. The solution structure of crambin^{KYK} was calculated with the 797 NOE-based distance restraints and 62 dihedral angle restraints from TALOS analysis (Table 1). The energy-lowest ensemble structures of crambin^{KYK} were well converged (Fig. 3(a)) and did not show any experimental violations (Table 1). Crambin^{KYK} had the identical structure to the previous structures of crambin, even though the structures of wild-type protein were determined in acetone/water mixture,[14,20] or in DPC micelles,[20] or from the crystal grown in ethanol/water mixture.[13] The RMSD of C_α traces of the wild-type (PDB:1EJG)[13] and mutant (this work) crambin was 0.82 Å (Fig. 3(c)).

In Figure 3(c), the ribbon presentation of crambin^{KYK} structure was shown with the side-chain configurations of the mutated residues, Lys1, Tyr13, and Lys33, which coincided well with the wild-type configurations. The side-chains of Lys1 and Lys33 were protruded from the center of the structure, while that of Tyr13 was situated at the center of two α-helices. Thus, the effects of T1K and I33K mutation were limited to the proximal residues such as Thr2, Cys3, Gly31, Ile34, Ile35, Gly37, and Thr39 (Fig. 2). However, the presence of hydroxyl group in tyrosine replacing phenylalanine might bring about more significant chemical shift changes of the whole residues on Helix II rather than the

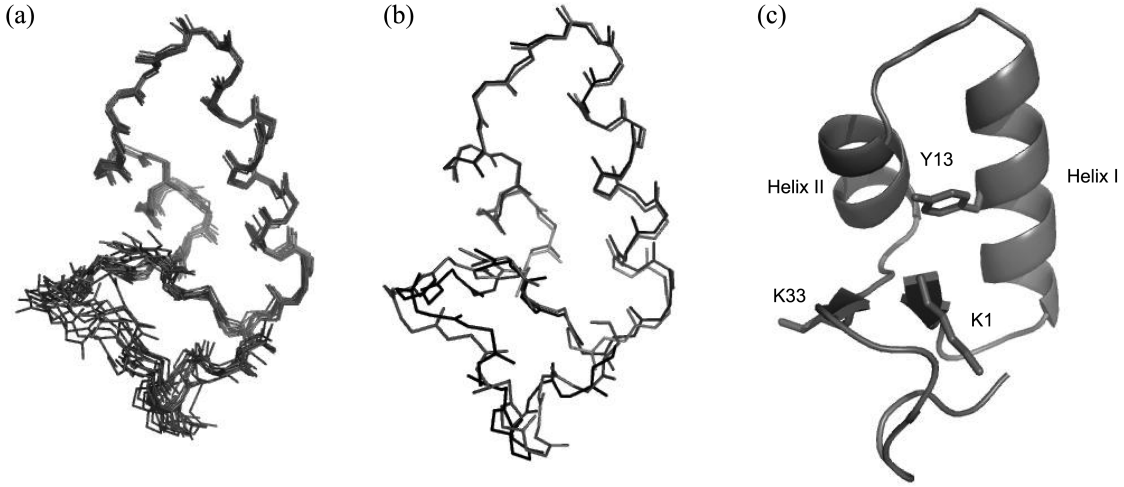

Figure 3. 20 energy-lowest ensemble structure of crambin^{KYK} (a). The overlaid backbone structure of the wild-type (gray) and mutant (black) crambin (b). The RMSD of Cα traces was 0.82 Å. Side-chain orientations of the mutated residues and the secondary structure of crambin^{KYK} (c).

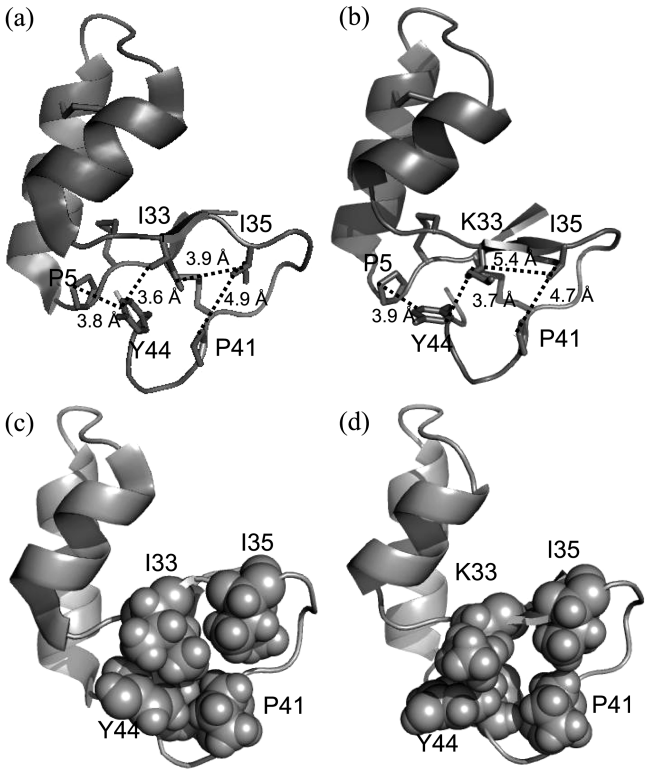

Figure 4. Structural details of the wild-type and mutant crambin. The hydrophobic interactions were presented by means of distances between side-chains in the wild-type (a) and mutant (b) crambin. The space-filling presentation of the hydrophobic residues in crambin (c) and crambin^{KYK} (d).

residues on Helix I (Fig. 2 and Fig. 3(c)).

To compare the solution structures of the wild-type and mutant crambin, we used Protein Interactions Calculator (PIC, http://crick.mbu.iisc.ernet.in/~PIC)[30] to investigate various kinds of interactions within proteins, such as disulfide bonds, hydrophobic interactions, ionic interactions, hydrogen bonds, and so on. Most of the interactions categorized above were conserved in the mutant crambin, however, critical difference was found in the intra-protein interactions between the solvent exposed residues. Those were related to the mutation of isoleucine to lysine at position 33. The hydrophobic interactions of the solvent exposed residues of the mutant and wild-type crambin are shown in Figure 4(a) and (b). In the wild-type crambin (Fig. 4(b)), hydrophobic interactions between the second β-strand (I33-I35) and loop region (P41-Y44) are critical to maintain the crambin fold. In crambin^{KYK}, those hydrophobic interactions were conserved except for the interaction between the mutated Lys33 and Ile35 (Fig. 4(a)). Since the side-chain of Lys33 in crambin^{KYK} is positively charged and protruded from the center of the protein, it could not provide the hydrophobic interaction with the side-chain of Ile35, while the side-chains of Ile33 and Ile35 in the wild-type crambin obviously showed the hydrophobic interaction. Fig. 4(c) and (d) show the space-filling presentation of hydrophobic side-chains in the wild-type and mutant crambin, in which the solvent accessible hole was seen only in the mutant protein.

Implication of Water-solubility of Crambin^{KYK}. The analysis of the difference in the structures of the wild-type and mutant crambin suggested the possible explanation for the solubility enhancement of crambin^{KYK} in aqueous solution. First, the mutation in this study gave rise to positive net charge in crambin^{KYK}, since two lysine residues replaced non-charged threonine and isoleucine at 1st and 33rd positions of the wild-type protein. The net-charge of crambin^{KYK} was +2 and pI of the mutant was 8.32 (5.73 for crambin). Thus, the increase of the positively charged residues on protein surface might enhance the solubility of the mutant crambin in aqueous solution. In addition, the weakened hydrophobic interactions between the second β-strand and loop region should be also attributed to the enhancement of the solubility in aqueous solution. We believe that the surface exposed hydrophobic residues of crambin, especially, Ile33, Ile35, Pro41, and Tyr44, might provide the hydrophobic core for self-aggregation and be the main cause of the insolubility of crambin in aqueous solution. Indeed, in the structure of crambin in DPC micelles, the hydrophobic surfaces were shielded under the micelles in which the hydrophobic chains of DPC interact with the surfaces and mask them from the contact to water molecules.[20] However, in crambin^{KYK}, the hydrophobic interactions in the same region were weakened and the planarity made by four residues, Lys33, Ile35, Pro41, and Tyr44, was distorted since the side-chain of Lys33 was protruded outward against the corresponding plane in the wild-type crambin (Fig. 4(d)). Therefore, intermolecular hydrophobic interactions between crambin^{KYK} proteins would be also much weaker than those between the wild-types and would not accelerate the protein precipitation in aqueous solvent. Finally, water molecules would be able to permeate into the center of crambin^{KYK}, whereas water molecules would be expelled by the hydrophobic surface in the wild-type crambin.

In conclusion, we made water-soluble version of crambin mutant based on the sequence analysis. Three dimensional structure of crambin mutant was determined in aqueous solution by NMR spectroscopy and the structural detail provided the explanation for the water-solubility. This work showed that the careful design and mutation enhance the solubility of a protein without disruption of three-dimensional structure and would be expanded to produce soluble proteins for the structural and biological investigations.

Acknowledgments. We thank the staffs at the KIST 900 MHz NMR facility. This work was supported by a grant of the Korea Healthcare technology R&D Project, Ministry for Health & Welfare, Republic of Korea (A092006).

References

1. Wang, W. *Int. J. Pharm.* **1999**, *185*, 129.
2. Bondos, S. E.; Bicknell, A. *Anal. Biochem.* **2003**, *316*, 223.
3. Bagby, S.; Tong, K. I.; Ikura, M. *Methods Enzymol., Pt. B* **2001**, *339*, 20.
4. Pedelacq, J. D.; Piltch, E.; Liong, E. C.; Berendzen, J.; Kim, C. Y.; Rho, B. S.; Park, M. S.; Terwilliger, T. C.; Waldo, G. S. *Nat.*

Biotechnol. **2002**, *20*, 927.
5. Zhou, P.; Lugovskoy, A. A.; Wagner, G. *J. Biomol. NMR* **2001**, *20*, 11.
6. Edwards, A. M.; Arrowsmith, C. H.; Christendat, D.; Dharamsi, A.; Friesen, J. D.; Greenblatt, J. F.; Vedadi, M. *Nat. Struct. Biol*. **2000**, *7*, 970.
7. Savchenko, A.; Yee, A.; Khachatryan, A.; Skarina, T.; Evdokimova, E.; Pavlova, M.; Semesi, A.; Northey, J.; Beasley, S.; Lan, N.; Das, R.; Gerstein, M.; Arrowmith, C. H.; Edwards, A. M. *Proteins* **2003**, *50*, 392.
8. Misawa, S.; Kumagai, I. *Biopolymers* **1999**, *51*, 297.
9. De Bernardez Clark, E. *Curr. Opin. Biotechnol*. **2001**, *12*, 202.
10. Golovanov, A. P.; Hautbergue, G. M.; Wilson, S. A.; Lian, L.-Y. *J. Am. Chem. Soc*. **2004**, *126*, 8933.
11. Van Etten, C. H.; Nielson, H. C.; Peters, J. E. *Phytochemistry* **1965**, *4*, 467.
12. Hendrickson, W. A.; Teeter, M. M. *Nature* **1981**, *290*, 107.
13. Jelsch, C.; Teeter, M. M.; Lamzin, V.; Pichon-Lesme, V.; Blessing, B.; Lecomte, C. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3171.
14. Bonvin, A. M.; Rullmann, J. A.; Lamerichs, R. M.; Boelens, R.; Kaptein, R. *Proteins* **1993**, *15*, 385.
15. Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett*. **1988**, *239*, 129.
16. Xu, Y.; Wu, J.; Gorenstein, D.; Braun, W. *J. Magn. Reson*. **1999**, *136*, 76.
17. Spronk, C. A.; Linge, J. P.; Hibers, C. W.; Vuister, G. W. *J. Biomol. NMR* **2002**, *22*, 281.
18. Linge, J. P.; Williams, M. A.; Spronk, C. A.; Bonvin, A. M.; Nilges, M. *Proteins* **2003**, *50*, 496.
19. Jee, J.; Ahn, H.-C. *Bull. Korean Chem. Soc*. **2009**, *30*, 1139.
20. Ahn, H.-C.; Juranic, N.; Macura, S.; Markley, J. L. *J. Am. Chem. Soc*. **2006**, *128*, 4398.
21. Stec, B. *Cell. Mol. Life Sci*. **2006**, *63*, 1370.
22. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277.
23. Johnson, B. *Methods Mol. Biol*. **2004**, *278*, 313.
24. Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 289.
25. Linge, J. P.; O'Donoghue, S. I.; Nilges, M. *Methods Enzymol*. **2001**, *339*, 71.
26. Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Cryst. D* **1998**, *54*, 905.
27. Linge, J. P.; Williams, M. A.; Spronk, C. A.; Bonvin, A. M.; Nilges, M. *Proteins* **2003**, *50*, 496.
28. Laskowski, R. A.; Rullmannn, M. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. *J. Biomol. NMR* **1996**, *8*, 477.
29. Finn, R. D.; Mistry, J.; Tate, J.; Coggill, J.; Heger, A.; Pollington, J. E.; Gavin, O. L.; Gunesekaran, P.; Ceric, G.; Forslund, K.; Holm, L.; Sonnhammer, E. L.; Eddy, S. R.; Bateman, A. *Nucleic Acids Res*. **2010**, *38*, D211.
30. Tina, K., G.; Bhadra, R.; Srinivasan, N. *Nucleic Acids. Res*. **2007**, *35*, W473.