Journal of the Korean Magnetic Resonance Society 2002, 6, 54-68

Structural Studies on RUNX of *Caenorhabditis elegans* by Spectroscopic Methods

Woo-Sung Son, Jong-Wan Kim, Hee-Chul Ahn, Sung-Jean Park, Suk-Chul Bae[1] and Bong-Jin Lee*

Research Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, San 56-1, Shillim-dong, Kwanak-gu, Seoul 151-742, Republic of Korea
[1]Department of Biochemistry, School of Medicine, and Medical Research Institute, Chungbuk National University, Cheongju 361-763, Republic of Korea
Received May 20, 2002

Abstract : PEBP2/CBF (Polyomavirus Enhancer-core Binding Protein 2/Core Binding Factor), represents a new family of heterodimeric transcription factor. Those members play important roles in hematopoiesis and osteogenesis in mouse and human. PEBP2/CBF is a sequence-specific DNA binding protein. Each member of the PEBP2/CBF family of transcription factors is composed of two subunits, α and β. The evolutionarily conserved 128 amino acid region in α subunit has been called the Runt domain, which harbors two different activities, the ability to bind DNA and interact with the β subunit. Recently, cDNA clones encoding the *C. elegans* Runt domain were isolated by screening a cDNA library. This gene was referred to *run* (Runt homologous gene). In this study, the basic experiments for the structural characterization of RUN protein were performed using spectroscopic methods. We have identified the structural properties of RUN using bioinformatics, CD and NMR. The limit temperature of the structural stability was up to 60℃ with irreversible thermal process, and the structure of RUN seems to adopt α helices and one or more β sheet or turn. The degree of NMR peak dispersion and intensity was increased by addition of glycine. Therefore, glycine could be used to alleviate the aggregation property of RUN in NMR experiment.

INTRODUCTION

The Runt domain is a highly conserved domain of 128 amino acid residues which serves as a scaffold for DNA binding and protein-protein interactions in a newly characterized family of transcription factors[1], PEBP2/CBF (Polyomavirus Enhancer-core Binding Protein 2/Core Binding Factor), whose members play important roles in hematopoiesis and osteogenesis in mouse and human[2-4]. PEBP2/CBF is a sequence-specific DNA binding protein. Each member of the PEBP2/CBF family of transcription factors is

* To whom : lbj@nmr.snu.ac.kr

composed of two subunits, α and β[5-9]. The α subunit binds to DNA and the β subunit increases the binding affinity of the α protein for DNA. Three closely related mammalian PEBP2α genes, termed *Runx1/PEBP2αB/CBFA2*, *Runx2/PEBP2αA/CBFA1* and *Runx3/PEBP2αC/CBFA3* have been reported previously[8, 10-13]. Each of the three PEBP2α/CBFA gene products is highly homologous to *Drosophila runt*[14] and *Lozenge*[15, 16] within a 128 amino acid region which harbors two different activities, the ability to bind DNA and the ability to interact with the β subunit. The evolutionarily conserved 128 amino acid region has been called the Runt domain[1]. Runt was initially characterized because of its role as one of the pair-rule genes during segmentation[17] and was subsequently found to have roles in two other developmental processes, fly-sex determination and neurogenesis[18, 19]. *Lozenge* is involved in prepatterning photoreceptor precursors in the developing fruit-fly eye[16]. The Runt domain is evolutionarily conserved and proteins containing this domain play important biological roles in the fruit fly, mouse and human. *Runx1*, also called *AML1*, has been identified at the breakpoint of chromosome translocations responsible for human leukemia[2-4]. Targeted disruption of the PEBPαB/AML1 gene resulted in lack of definitive hematopoiesis of fetal liver[20-21]. Analyses of the expression and function of Xaml, the Xenopus homolog of PEBP2αB/AML1, also suggested that PEBP2αB is required for the development of all blood lineages and that PEBP2αB may function in the early formation of hematopoietic stem cells[22]. It has been demonstrated that PEBP2αA/CBFA1, another PEBP2α gene, plays an essential role in osteogenesis. Mice with a homozygous mutation in PEBP2αA died just after birth due to breathing inability caused by a complete lack of ossification[23, 24]. The PEBP2αA gene is also responsible for the human disease, cleidocranial dysplasia (CCD), an autosomal dominant disorder[25, 26]. More recently, it has been demonstrated that *Runx3/PEBP2αC* is a major growth regulator of gastric epithelial cells and RUNX3 function is causally related to the genesis and progression of human gastric cancer[27].

Amino acid sequence homology searches in the *Caenohabditis elegans* database revealed that *C. elegans* contains a Runt domain-encoding gene in its genome. Recently, cDNA clones encoding the *C. elegans* Runt domain was isolated, by screening a cDNA library[28]. This gene was referred to as *run* (*Runx* homologous gene). More than 30% of the predicted genes from the genomic sequence of *C. elegans* have a human homolog, many of which have been implicated in human disease[29, 30]. This observation, along with increasing evidence that biological processes and molecular pathways are highly conserved, has made the nematode *C. elegans* a model organism for the study of developmental regulation. Also, It has been reported that the *run* gene is expressed exclusively in the nuclei of hypodermal seam cells and intestinal cells[31].

In this paper, the preliminary experiments for the determination of three-dimensional structure of RUN protein (129 AA) were performed. The purification protocol for this protein was optimized and the buffer condition was found which improves the temperature stability. These developments make it possible to obtain samples of the RUN protein that

suitable for structural studies by NMR.

MATERIALS and METHODS

Materials

All materials were purchased from certified vendors in molecular biology grade for the reliability of results. Host (*E. coli* BL21 (DE3) Codon Plus-RIL™) and vector system (pET-15b) were purchased from Novagen Inc. (Darmstadt, Germany). All reagents except isotopes were from Sigma Chemical Company (St.Louis, U.S.A.). Isotopes such as ^{15}N sources ($(^{15}NH_4)_2SO_4$, $^{15}NH_4Cl$) were purchased from ISOTEC Inc. (Ohio, U.S.A.). Ni^{2+}-agarose column (His bind® Resin) was purchased from Novagen Inc. (Darmstadt, Germany). Chelating Sepharose Fast flow® from Amersham Pharmacia Biotech Inc. (Uppsala, Sweden) were used for the purification of His-tagged protein.

Sample preparation

The clone harboring the recombinant plasmid, which carries the gene for RUN was supplied from Bae and his cowokers engaged in Department of Biochemistry, School of Medicine, Chungbuk National University, Republic of Korea. More detailed information about this gene is deposited in the 'GenBank' (NIH), with Access ID AF153275. The *run* gene is composed of 11 exons and a 7.2kb long intron is located between exons 3 and 4. The vector system for the construction of the recombinant plasmid carrying RUN gene is pET-15b from Novagen Inc. The plasmid was cleaved at *NdeI* and *BamH1* cleavage sites, and the PCR fragment of RUN gene was inserted into the cleaved plasmid. The recombinant plasmid was transformed to competent cell, *E. coli* BL21 Codon Plus (DE3)-RIL™ . To evaluate whether the construction is correct, the overexpression test with IPTG was carried out, following SDS-PAGE test and the analysis of amino acid sequences. *E. coli*, containing the plasmid encoding RUN as described above, which had been stored at –80℃ was inoculated to [U-^{15}N]-M9 media containing ampicillin (50μg/ml). All steps for culture were processed at 37℃ and all steps for the purification of RUN were processed at 4℃. In the beginning, seed strain was inoculated to [U-^{15}N]-M9 media (5ml) and was scaled up to [U-^{15}N]-M9 media (100ml) at OD_{600}=0.7. That was once more scaled up to [U-^{15}N]-M9 media (2 liter). When OD_{600} was reached to 0.6, IPTG was added to culture media so that the final concentration of IPTG was 50μg/ml. After the additional culture for 6 hrs, the culture media was centrifuged at 9,000 rpm for 12 min and was decanted for the storage at –80℃.

The purification of RUN was started with lysis by sonic oscillation (sonication) of pellet. The pellets was suspended with the binding buffer, and sonicated until the solution was clear. The lysate was centrifuged twice at 9,000 rpm for 12 min. The solid material (Inclusion bodies) remained after the centrifugation was collected and stored at –80℃. Inclusion bodies containing RUN were resuspended and solubilized in 6M Urea. After

removal of insoluble materials by centrifugation, the crude protein mix loaded onto a His bind Resin. The washing process after the binding process was performed with the ten volumes of binding Buffer. To elute the target protein from His Bind Resin, we performed the elution with the imidazole gradient (0~2M). After nearly pure fractions was pooled, refolding by dialysis against 500mM NaCl and 20mM potassium phosphate buffer at pH 7.2 was performed. The concentration and buffer exchange with final buffer (20mM Sodium Phosphate buffer pH 6.0, 500mM NaCl and 1mM NaN_3) were processed by CentriPrep® and CentriCon®. As a result, the highly purified RUN was acquired.

Circular dichroism (CD) spectroscopy

Circular dichroism (CD) spectra were obtained on a JASCO J-715 spectropolarimeter (Victoria, British Columbia) to investigate the secondary structure of RUN. For the estimation of the secondary structure of proteins by CD spectrum (180~260 nm wavelength range), variable concentration samples (0.02mM~0.2mM, pH 4.0, 5.0, 6.0, 7.0 at 20℃) were prepared. Sample was dissolved in 20mM sodium phosphate buffer (pH 7.2), containing 500mM NaCl. All the processes were processed by the application program, 'JASCO Standard Analysis'. The secondary structure also was estimated with built-in software. To characterize the thermal stability of RUN, the temperature-dependent CD spectroscopy was performed. The region of 180~260 nm wavelength was monitored at 15℃~90℃ temperature range. To determine the effect of $MgSO_4$ on RUN, the CD spectroscopy with 200mM $MgSO_4$ was performed at pH 6.0. In all CD experiments, three consecutive scans per sample were performed in a 2 mm cell at 20℃. Three scans were added and averaged, followed by subtraction of the CD signal of the solvent. The helicity of the peptides was estimated from the mean residue ellipticity at 222 nm[32].

Nuclear magnetic resonance (NMR) spectroscopy

NMR tubes were purchased from Shigemi Inc. (Tokyo, Japan). All the measurements were conducted with Bruker DRX 500. To find the best condition for NMR measurements of RUN, 2D-1H,15N-HSQC spectrum were acquired at various conditions. NMR sample contains 20mM sodium phosphate buffer containing pH 6.0, 500mM NaCl and 1mM NaN_3. Additives are used by diluting each stock solution (1M Glycine, 0.1M CHAPS, 2M $MgSO_4$, 1M DTT). All NMR spectra were processed by using NMRPipe[33] software package and were analyzed with the NMRView[34] program on Indy®, Indigo2® and O2® from Silicon Graphics Inc. (California, U.S.A.).

RESULTS and DISCUSSION

The overexpression and purification of RUN

RUN is a protein with 171 amino acids (RUN expression region containing 148

amino acids and other amino acids including hexa-his tag). Its molecular weight is 19343Da The level of overexpression was checked by SDS-PAGE (Fig. 1). The overexpressed band was major one among the produced bands (so called, 'main band' or 'target band'). The target band was thickened remarkably while the others were not increased by the IPTG induction. Because the efficiency of the overexpression of RUN was very high, it was not so difficult to purify the large quantity of protein. To acquire the best induction point, the growth curve of cells in the presence or absence of IPTG was checked. Through various experiment to find the best condition of induction, the induction time, the final concentration of IPTG (50μg/ml) and culture temperature (37℃) were determined respectively. After the lysis of cells, the protein was not nearly found in the soluble fraction. Because the protein is very insoluble, the refolding step is required. In refolding process, large amount of protein was precipitated. For this reason, the purification of RUN was very difficult. After many buffer conditions were tested, the clue of solving this problem was found. When 0.5M NaCl was added in every step, RUN did not precipitate in refolding step.

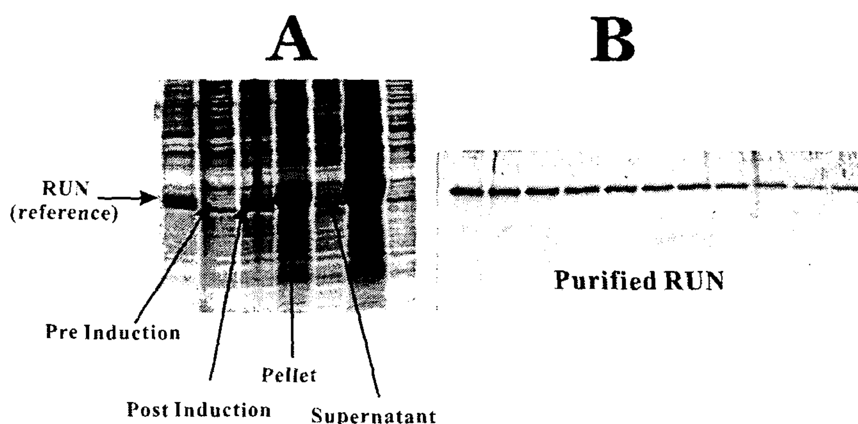

Fig. 1. Overexpression of RUN in *E. coli* and the purification of RUN. *Panel A:* The cells were disrupted by sonication, and samples were analyzed on a 15% SDS-tricine gel. Uninduced cells (pre induction), induced cells (post induction), supernatant and pellet are shown. *Panel B:* Electrophoresis of RUN purified by His-bind column. 18μl of aliquots for the each fraction were separated through a 15% SDS-tricine gel.

The estimation of structural properties using bioinformatics

The hydropathy plot[35, 36] of a certain protein shows the clue in knowing the secondary structure of the target protein, although it means the result of simulation with given sequences. If a series of residues are estimated to hydrophobic regions, the block may be buried in the molecular structure and may form a hydrophobic core. In RUN, the hydrophobic block and hydrophilic block were appeared alternatively with random tendency (so called, 'zigzag' pattern). Region of hexa-his tag was highly hydrophilic (Fig. 2). The

secondary structure of RUN was estimated on the web by PSA database server[37-40] and GOR database server[41-43]. From the Fig. 3 and Fig. 4, this protein is likely to adopt 3 helices and several strands in structure. The estimated component of RUN is 15.20% alpha helix (26AA / 171AA), 28.07% extended strand (48AA / 171AA) and 56.73% random coil (97AA / 171AA).

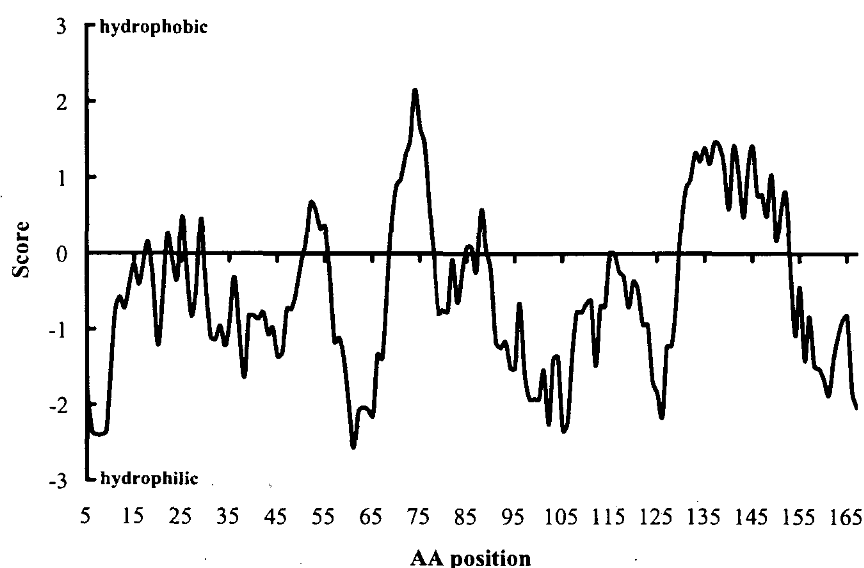

Fig. 2. The hydropathy plot of RUN.

Secondary structure components by circular dichroism (CD) spectroscopy

CD spectra of RUN were different from typical spectra of other proteins comprised of all α helices. From the analysis of the spectrum with a built-in program, it was revealed that this protein adopted 35.0% β sheet structure at pH 6 (Fig. 5). To find the best conditions for NMR measurements, the temperature-dependent stability of RUN protein was checked. The structure of RUN was corrupted above 60℃ (Fig. 6-A). CD spectrum acquired before heating is quite different from that acquired after heating at the same temperature (Fig. 6-B). Therefore, RUN is irreversible in thermal denature process. In order to investigate the pH dependence of RUN, CD experiments were performed on the samples between pH 4 and pH 7. CD spectrum acquired at pH 6 had more typical α helix pattern than that acquired at any other pH (Fig. 5). As a result, the best pH condition for NMR experiment was determined as pH 6. To identify effect of $MgSO_4$ on RUN, 200mM $MgSO_4$ was added to sample for CD experiments. From the comparison of CD spectra, it was found that RUN had more regular structure with $MgSO_4$ (Fig. 7).

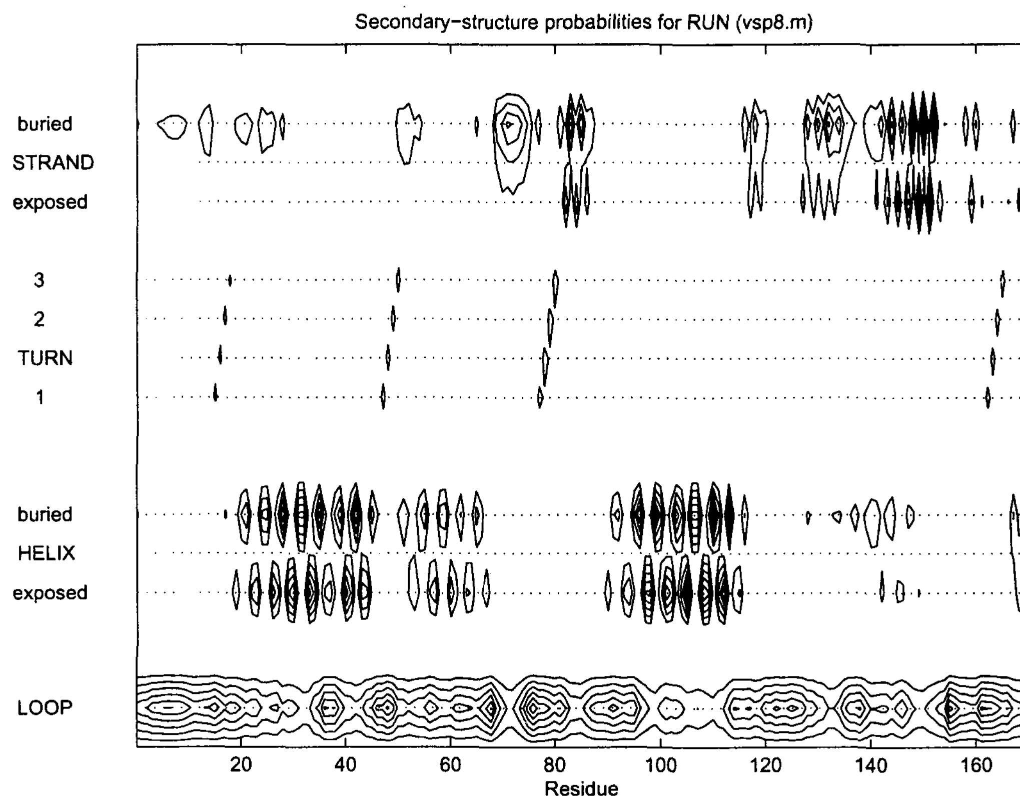

Fig. 3. The estimated secondary structure of RUN using bioinformatics (contour line form)

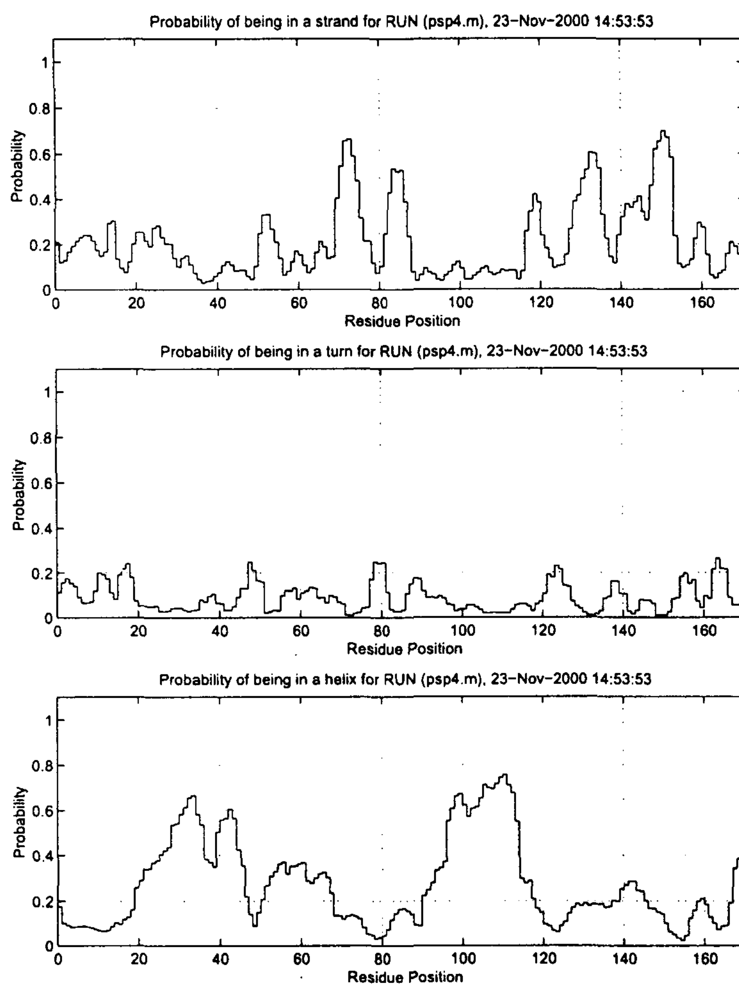

Fig. 4. The estimated secondary structure of RUN using bioinformatics (diagram form).

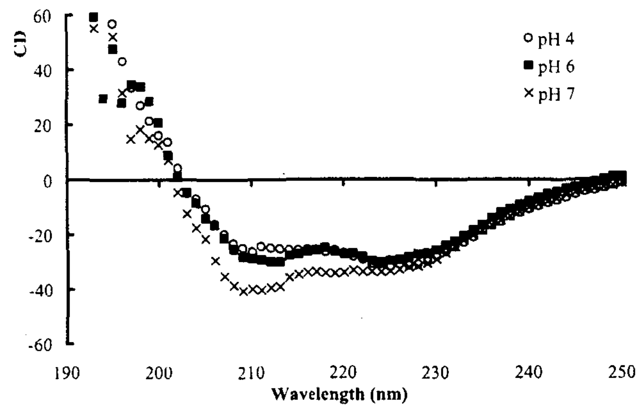

Fig. 5. The change in the components of secondary structures by pH change. CD spectra were monitored at pH 4 (○), pH 6 (■), pH 7 (×).

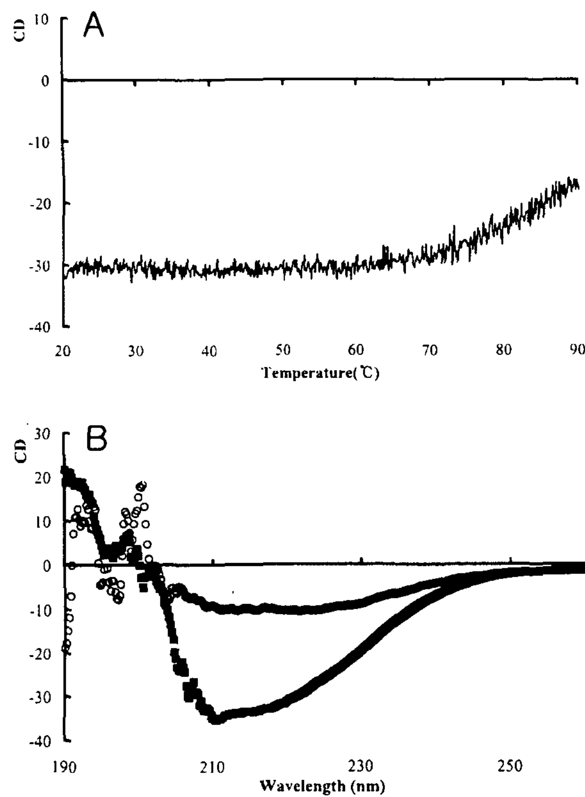

Fig. 6. The temperature-dependent stability of RUN. *Panel A:* The CD spectrum of temperature scan from 20℃ to 90℃. *Panel B:* The CD spectrum acquired before heating (■) and acquired after heating (○) at the same temperature (20℃).

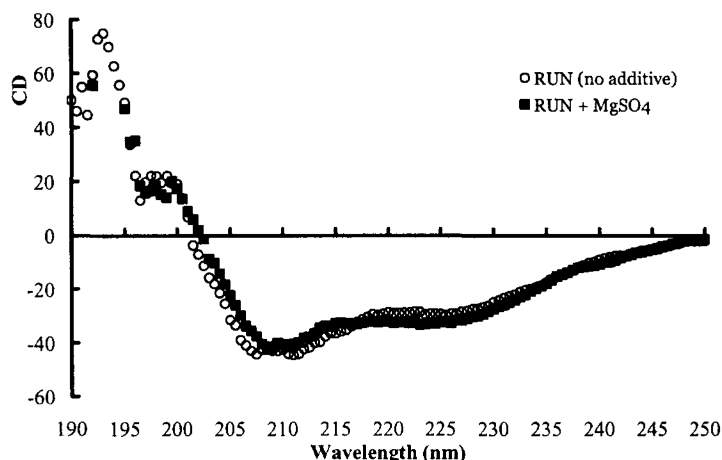

Fig. 7. The effect of $MgSO_4$ on RUN. CD spectra were acquired with no $MgSO_4$ (■) and with 200mM $MgSO_4$ added condition (○).

Nuclear magnetic resonance (NMR) spectroscopy and structural properties

From many 2D-^{1}H, ^{15}N-HSQC spectra, the best condition for NMR measurements was checked (Fig. 8, 9). Spectral resolution was monitored according to increasing the temperature. However, any significant change was not detected. The spectral resolution was slightly improved by addition of glycine but was not effected by addition of other additives. From all the data processed, it is likely that RUN protein have a tendency to aggregate in high concentration for NMR experiment. This aggregation was reduced by addition of glycine (Fig. 9). Although NMR spectra were not sufficient for the determination of three-dimensional structure of RUN, they could be used to extract some useful structural information. From the study of the temperature-dependent stability, it was decided that the optimal temperature for NMR studies is 30℃, though this protein was stable up to 60℃. In high concentration for NMR experiment, RUN is likely to aggregate. And this aggregation is general in RUNX. Not only RUN but other free form of RUNX proteins aggregate in high concentration[44-47]. There is no example of RUNX to be completely refolded in free form. From the fact that the structural studies of several RUNX are reported from modified sample or DNA complex sample[44-47], it is suggested that structure of free form of RUN be flexible. Therefore, this flexible RUN is likely to aggregate with each other in high concentration.

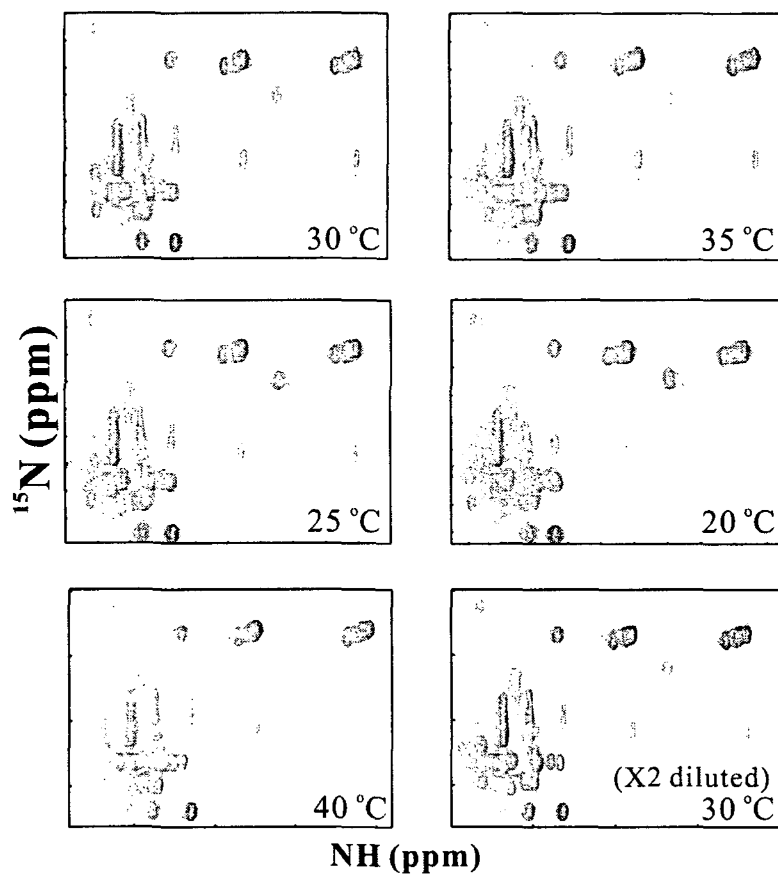

Fig. 8. The HSQC spectra of RUN monitored according to increasing the temperature

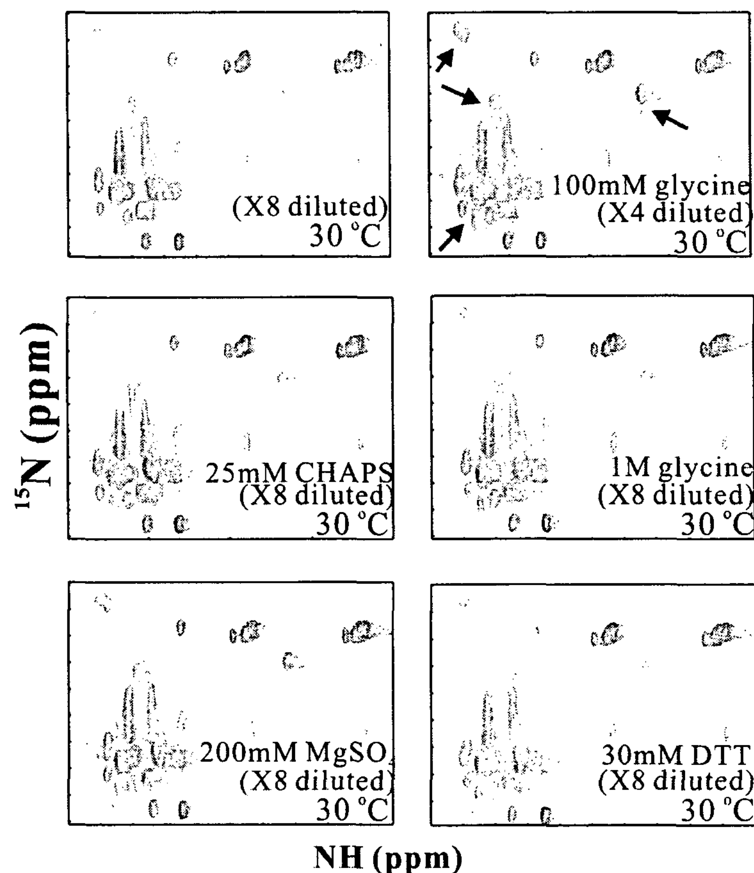

Fig. 9. The HSQC spectra of RUN monitored according to adding additives.

CONCLUSIONS

In this study, we have identified the structural properties of RUN using bioinformatics CD and NMR. The limit temperature of the structural stability was up to 60℃ with irreversible thermal porcess and the structure of RUN might adopt α helices and one or more β sheet or turn. In NMR experiment, the degree of peak dispersion and intensity was increased by addition of glycine. To get more significant structural information, optimization of refolding step and advanced NMR spectroscopy are required.

Acknowledgements

This work was supported by National Research Laboratory Program (M1-0203-00-0075) from the Korea Institute Science & Technology Evaluation and Planning, Republic of Korea. This work was also supported in part by 2002 BK21 project for Medicine, Dentistry, and Pharmacy.

REFERENCES

1. H. Kagoshima, K. Shigesada, M. Satake, Y. Ito, H. Miyoshi, M. Ohki, M. Pepling and P. Gergen, *Trends Genet.*, **9**, 338-341 (1993).
2. T. R. Golub, G. F. Barker, S. K. Bohlander, S. W. Hiebert, D. C. Ward, P. Bray-Ward, E. Morgan, S. C. Raimondi, J. D. Rowley and D. G. Gilliland, *Proc. Natl. Acad. Sci. USA.*, **92**, 4917-4921 (1995).
3. H. Miyoshi, T. Kozu, K. Shimizu, K. Enonoto, N. Maseki, Y. Kaneko, N. Kamada and M. Ohki, *EMBO J.*, **12**, 2715-2721 (1993).
4. S. P. Romana, M. Mauchauffe, M. Le Coniat, I. Chumakov, D. Le Paslier, R. Berger and O. A. Bernard, *Blood*, **85**, 3662-3670 (1995).
5. S. C. Bae, E. Ogawa, M. Maruyama, H. Oka, M.Satake, K. Shigesada, N. A. Jenkins, D. J. Gilbert, N. G. Copeland and Y. Ito, *Mol. Cell. Biol.*, **14**, 3242-3252 (1994).
6. S. C. Bae and Y. Ito, *Histol. Histopathol.*, **14**, 1213-1221 (1999).
7. E. Ogawa, M. Inuzuka, M. Maruyama, M. Satake, M. Naito-Fujimoto, Y. Ito and K. Shigesada, *Virology*, **194**, 314-331 (1993).
8. E. Ogawa, M. Maruyama, H. Kagoshima, M. Inuzuka, J. Lu, M. Satake, K. Shigesada and Y. Ito, *Proc. Natl. Acad. Sci. USA.*, **90**, 6859-6863 (1993).
9. S. Wang, Q. Wang, B. E. Crute, I. N. Melnikova, S. R. Keller and N. A. Speck, *Mol. Cell. Biol.*, **13**, 3324-3339 (1993).
10. S. C. Bae, Y. Yamaguchi-Iwai, E. Ogawa, M. Maruyama, M. Inuzuka, H. Kagoshima, K. Shigesada, M. Satake and Y. Ito, *Oncogene*, **8**, 809-814 (1993).
11. S. C. Bae, E. Takahashi, Y. W. Zhang, E. Ogawa, K. Shigesada, Y. Namba, M. Satake and Y. Ito, *Gene*, **159**, 245-248 (1995).
12. D. Levanon, V. Negreanu, Y. Bernstein, I. Bar-Am, L. Avivi and Y. Groner, *Genomics*, **23**, 425-432 (1994).
13. H. Miyoshi, K. Shimizu, T. Kozu, N. Maseki, Y. Kaneko and M. Ohki, *Proc. Natl. Acad. Sci. USA.*, **88**, 10431-10434 (1991).
14. M. A. Kania, A. S. Bonner, J. B. Duffy and J. P. Gergen, *Genes Dev.*, **4**, 1701-1713 (1990)
15. J. Canon and U. Banerjee, *Cell Dev. Biol.*, **11**, 327-336 (2000).
16. A. Daga, C. A. Karlovich, K. Dumstrei and U. Banerjee, *Genes Dev.*, **10**, 1194-1205 (1996).
17. J. P. Gergen and E. F. Wieschaus, *Dev. Biol.*, **109**, 321-335 (1985).
18. J. B. Duffy, M. A. Kania and J. P. Gergen, *Development*, **113**, 1223-1230 (1991).
19. J. B. Duffy and J. P. Gergen, *Genes Dev.*, **5**, 2176-2187 (1991).

20. T. Okuda, J. van Deursen, S. W. Hiebert, G. Grosveld and J. R. Downing, *Cell*, **84**, 321-330 (1996).
21. Q. Wang, T. Stacy, J. D. Miller, A. F. Lewis, T. L. Gu, X. Huang, J. H. Bushweller, J. C. Bories, F. W. Alt, G. Ryan, P. P. Liu, A. Wynshaw-Boris, M. Binder, M. Marin-Padilla, A. H. Sharpe and N. A. Speck, *Cell*, **87**, 697-708 (1996).
22. W. D. Jr. Tracey, M. E. Pepling, M. E. Horb, G. H. Thomsen and J. P. Gergen, *Development*, **125**, 1371-1380 (1998).
23. T. Komori, H. Yagi, S. Nomura, A. Yamaguchi, K. Sasaki, K. Deguchi, Y. Shimizu, R. T. Bronson, Y. H. Gao, M. Inada, M. Sato, R. Okamoto, Y. Kitamura, S. Yoshiki and T. Kishimoto, *Cell*, **89**, 755-764 (1997).
24. F. Otto, A. P. Thornell, T. Crompton. A. Denzei, K. C. Gilmour, I. R. Rosewell, G. W. Stamp. R. S. Beddington, S. Mundlos, B. R. Olsen, P. B. Selby and M. J. Owen, *Cell*, **89**, 765-771 (1997).
25. B. Lee, K. Thirunavukkarasu, L. Zhou, L. Pastore, A. Baldini, J. Hecht, V. Geoffroy, P. Ducy and G. Karsenty, *Nat. Genet.*, **16**, 307-310 (1997).
26. S. Mundlos, F. Otto, C. Mundlos, J. B. Mulliken, A. S. Aylsworth, S. Albright, D. Lindhout, W. G. Cole, W. Henn, J. H. Knoll, M. J. Owen, R. Mertelsmann, B. U. Zabel and B. R. Olsen, *Cell*, **89**, 773-779 (1997).
27. Q. L. Li, K. Ito, C. Sakakura, H. Fukamachi, K. Inoue, X. Z. Chi, K. Y. Lee, S. Nomura, C. W. Lee, S. B. Han, H. M. Kim, W. J. Kim, H. Yamamoto, N. Yamashita, T. Yano, T. Ikeda, S. Itohara, J. Inazawa, T. Abe, A. Hagiwara, H. Yamagishi, A Ooe, A. Kaneda, T. Sugimura, T. Ushijima, S. C. Bae and Y. Ito, *Cell*, **109**, 113-124 (2002).
28. S. C. Bae and J. Lee, *Gene*, **241**, 255-258 (2000).
29. A. G. Fraser, R. S. Kamath, P. Zipperlen, M. Martinez-Campos, M. Sohrmann and J. Ahringer, *Nature*, **408**, 325-330 (2000).
30. P. Gonczy, G. Echeverri, K. Oegema, A. Coulson, S. J. Jones, R. R. Copley, J. Duperon, J. Oegema, M. Brehm, E. Cassin, E. Hannak, M. Kirkham, S. Pichler, K. Flohrs, A. Goessen, S. Leidel, A. M. Alleaume, C. Martin, N. Ozlu, P. Bork and A. A. Hyman, *Nature*, **408**, 331-336 (2000).
31. S. H. Nam, Y. H. Jin, Q. L. Li, K. Y. Lee, G. B. Jeong, Y. Ito, J. H. Lee and S. C. Bae, *Mol. Cell. Biol.*, **22**, 547-554 (2002).
32. Y. H. Chen, J. T. Yang and H. M. Martinez, *Biochemistry*, **11**, 4120-4131 (1972).
33. F. Delaglio, S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer and A. Bax, *J. Biomol. NMR*, **6**, 277-293 (1995).
34. B. A. Johnson and R. A. Blevins, *J. Biomol. NMR*, **4**, 603-614 (1994).
35. J. Kite and R. F. Doolittle, *J. Mol. Biol.*, **157**, 105-132 (1982).
36. http://www.expasy.ch/cgi-bin/protscale.pl
37. C. M. Stultz, J. V. White and T. F. Smith, *Protein Science*, **2**, 305-314 (1993).
38. J. V. White, C. M. Stultz and T. F. Smith, *Mathematical Biosciences*, **119**, 35-75 (1994).
39. C. M. Stultz, R. Nambudripad, R. H. Lathrop and J. V. White, *Protein Structural Biology*

in Bio-Medical Research, JAI Press, Greenwich, pp. 447-506 (1997).
40. http://bmerc-www.bu.edu/psa/request.htm
41. J. Garnier, D. Osguthorpe and B. Robson, *J. Mol. Biol.*, **120**, 97-120 (1978).
42. J. Garnier, J. F. Gibrat and B. Robson, *Methods in Enzymology*, **266**, 540-553 (1996).
43. http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html
44. M. Wolf-Watz, X. Q. Xie, M. Holm, T. Grundstrom and T. Hard, *Eur. J. Biochem*, **261**, 251-260 (1999).
45. T. Nagata, V. Gupta, D. Sorce, W. Y. Kim, A. Sail, B. T. Chai, K, Shigesada, Y. Ito and M H. Werner, *Nat. Struct. Biol.*, **6**, 615-619 (1999).
46. M. Goger, V. Gupta, W. Y. Kim, K. Shigesada, Y. Ito and M. H. Werner, *Nat. Struct. Biol.* **6**, 620-623 (1999).
47. T. H. Tahirov, T. Inoue-Bungo, H. Morii, A. Fujikawa, M. Sasaki, K. Kimura, M. Shiina, K. Sato, T. Kumasaka, M. Yamamoto, S. Ishii and K. Ogata, *Cell*, **104**, 755-767 (2001).