Purification and NMR Studies of RNA Polymerase II C-Terminal Domain Phosphatase 1 Containing Ubiquitin Like Domain

Sunggeon Ko, Youngmin Lee, Jong-Bok Yoon, and Weontae Lee*

Department of Biochemistry, College of Life Science and Biotechnology, Yonsei University, Seoul 120-749, Korea
**E-mail: wlee@spin.yonsei.ac.kr*
Received December 24, 2008, Accepted March 11, 2009

RNA polymerase II C-terminal domain phosphatase 1 containing ubiquitin like domain (UBLCP1) has been identified as a regulatory molecule of RNA polymerase II. UBLCP1 consists of ubiquitin like domain (UBL) and phosphatase domain homologous with UDP and CTD phosphatase. UBLCP1 was cloned into the *E.coli* expression vectors, pET32a and pGEX 4T-1 with TEV protease cleavage site and purified using both affinity and gel-filtration chromatography. Domains of UBLCP1 protein were successfully purified as 7 mg/500 mL (UBLCP1, 36.78 KDa), 32 mg/500 mL (UBL, 9 KDa) and 8 mg/500 mL (phosphatase domain, 25 KDa) yielded in LB medium, respectively. Isotope-labeled samples including triple-labeled ($^{2}H/^{15}N/^{13}C$) UBLCP1 were also prepared for hetero-nuclear NMR experiments. ^{15}N-^{1}H 2D-HSQC spectra of UBLCP1 suggest that both UBL and phosphatase domain are properly folded and structurally independent each other. These data will promise us further structural investigation of UBLCP1 by NMR spectroscopy and/or X-ray crystallography.

Key Words: RNA Polymerase II, CTD-Phosphatase, NMR, Cloning, Purification

Introduction

RNA polymerase II (RNAPII) consists of folded domain and mobile c-terminal repeat domain (CTD). Folded domain mainly involves in mRNA biogenesis and mobile CTD controls the RNA polymerase II activity.[1] A long and highly flexible CTD consists of 7 repeats of hepta-peptide (YSPTSPS) which contains phosphorylated and dephosphorylated serine residues by CTD kinase and CTD phosphatase,[2] respectively. The degree of phosphorylation on CTD during the transcription process determines its binding partners required for mRNA splicing, capping and polyadenylation.[3-5] In the initiation stage of transcription, CTD of RNAPII was dephosphorylated[8] and it was hyperphosphorylated[9] in the elongation stage of transcription. The crystal structures of CTD phosphatases, FCP1 phosphatase and small c-terminal phosphatase (SCP1) have been reported recently, showing that they interact with CTD of RNAPII.[10,11] Recently, the unique CTD phosphatase containing ubiquitin like domain (UBLCP1) was identified and characterized.[12] It has been well-known that ubiquitin molecule is an important messenger for proteasome mediated proteolysis.[13,14] Since ubiquitin like domain of UDP showed both sequence and structure homology with ubiquitin, UDPs might be interact with 26S proteasome.[15] For instance, UBL domains of Rad 23 and Parkin interact with subunit of 26S proteasome.[16,17] In this report, we present structural and molecular information derived from molecular cloning and NMR spectroscopy, which will serve as an important clue for detailed structural study of UBLCP1.

Materials and Methods

Cloning and expression of UBLCP1. The cDNA encoding UBLCP1 was obtained from Hela cell. UBLCP1 protein was cloned into pET32a (Novagen) and pGEX 4T-1 (Amersharm) vector. PCRs from cDNA were executed using primers containing BamH I/EcoR I for pET32a (for both UBLCP1 and phosphatase domain) or BamH I/Xho I for pGEX 4T-1 (for UBL domain). We designed sense primers encoded tobacco etch virus (TEV) protease recognition sequence (ENLYFQG) for tag cleavage from fusion proteins. Sense and antisense primer for UBLCP1 were 5'-cgc gga tcc gaa aac ctg tat ttt cag ggc atg gct ctc cct att att-3' and 5'-ccg gaa ttg tca tca ctg tcc ttg ctt ctt tga gag ata tct ttc cca -3', respectively. For PCR of UBLCP1 mutants, antisense primer of △1-80 (UBL domain) was 5'-ccg ctc gag tca tca ctc ctc acg agt tcc cat cat cat-3' and sense primers of △110-318 and △133-318 (phosphatase domain) were 5'-cgc gga tcc gaa aac ctg tat ttt cag ggc ccc agg gaa ggg aaa aag-3' and 5'-cgc gga tcc gaa aac ctg tat ttt cag ggc gat att gaa gat gaa gta-3', respectively. The cloning results were further validated by DNA sequencing (COSMO Co.) and plasmid containing UBLCP1 was transformed into *E.coli* over-expression strain BL21 (DE3) and selected by Luria-Bertani (LB) plate containing ampicillin (0.1 mg/mL). Induced cells were sonicated in lysis buffer, 50 mM $NaPO_4$, 300 mM NaCl, pH 6.5, protease inhibition cocktail (Roche) and soluble fraction was further identified using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Isotope (^{15}N, $^{13}C/^{15}N$ and $^{2}H/^{13}C/^{15}N$) labeling of UBLCP1. For hetero-nuclear NMR experiments, we prepared isotope-labeled UBLCP1 proteins. Cells were cultured in M9 minimal medium containing ^{15}N labeled NH_4Cl (^{15}N, 99 %, Cambridge Isotope Laboratories, Inc.) or ^{13}C labeled d-glucose (U-$^{13}C_6$, 99 %, Cambridge Isotope Laboratories, Inc.) together with ^{15}N labeled NH_4Cl and induced by 0.1 mM IPTG at 25 °C. We discarded M9 medium using centrifugation and washed the pellets using 1 X PBS three times. Cell pellets were stored in -80 °C. Before cell lysis, three times of freeze and thaw procedure were executed.

Purification of UBLCP1 domains. For TRX-His_6 tagged UBLCP1 domains (UBLCP1, phosphatase domain) from pET32a, cells were sonicated in UBLCP1 buffer A, 50 mM $NaPO_4$, 300 mM NaCl, 5 mM imidazole, pH 7.0, 5 mM β-mercaptoethanol. Cell lysate was centrifuged and performed Ni-NTA affinity chromatography. For TRX-His_6 tag cleavage, elutes were dialyzed to TEV buffer and executed enzyme reaction with 0.5 X TEV protease for 6 hr at 25 °C. The result of TEV protease reaction was confirmed by SDS-PAGE and Ni-NTA affinity chromatography was performed to remove remaining TEV protease and TRX-His_6 tags. To ensure protein purity, size exclusion chromatography using Superdex^{TM} 75 was finally used (GE Healthcare). Purified UBLCP1 proteins were finally concentrated to ~0.5 mM for NMR experiments. For GST tagged UBLCP1 (UBL domain) from pGEX 4T-1, UBLCP1 buffer B, 50 mM $NaPO_4$, 300 mM NaCl, pH 6.5, was used for cell lysis using sonication procedure. Supernatants were loaded on glutathione sepharose resin (Amersham). After elution using 10 mM reduced glutathione containing UBLCP1 buffer B, elutes were dialyzed to TEV reaction buffer, 25 mM $NaPO_4$, 300 mM NaCl, 5 mM β-mercaptoethanol, 5 mM imidazole, pH 6.5, and executed protease rea ction using 1 X TEV protease for 18 hr at 25 °C. To remove GST tag and TEV protease protein, both GST and Ni-NTA affinity chromatography were performed. Purified proteins were further identified by SDS-PAGE and size exclusion chromatography.

NMR spectroscopy. UBLCP1 proteins were dissolved in NMR buffer containing 25 mM HEPES, 100 mM NaCl, 1 mM DTT at pH 7.0. Both ^{1}H-^{15}N transverse relaxation optimized spectroscopy based heteronuclear single quantum coherence (TROSY-HSQC)[18] and ^{1}H-^{15}N HSQC[19] were performed for UBLCP1, phosphatase domain and UBL domain in Bruker DRX 900 MHz and 500 MHz equipped with Cryoprobe^{TM}. For sequential assignments, TROSY type triple resonance experiments[20] such as TROSY-HNCACB, TROSY-CBCA-CONH and TROSY-HNCA, were performed using 70% $^{2}H/^{13}C/^{15}N$ labeled UBLCP1 in Bruker DRX 500 MHz equipped with Cryoprobe^{TM}. All spectra were processed using NMRpipe program[21] and analyzed using Sparky Ver. 3.113 in Linux workstation.

Results and Discussion

Plasmid constructs of UBLCP1s. We performed sequence alignment and secondary structure prediction using PSI-PRED program (protein structure prediction server - *http://bioinf.cs.ucl.ac.uk/psipred/*) to identify the structural domains of UBLCP1. Based on domain analysis results, UBLCP1, UBL domain and phosphatase domain were cloned into the *E.coli* over-expression vectors. Especially, we designed two phosphatase domains ($UBLCP1_{110\text{-}318}$ and $UBPCP1_{133\text{-}318}$) to confirm domain prediction (Fig. 1A) and cloned them. UBLCP1s were expressed as TRX/His_6/thrombin protease recognition sequence and TEV protease recognition sequence (UBLCP1) from pET32a and GST/thrombin protease recognition sequence with TEV protease recognition sequence from pGEX 4T-1(Fig. 1B and 1C).

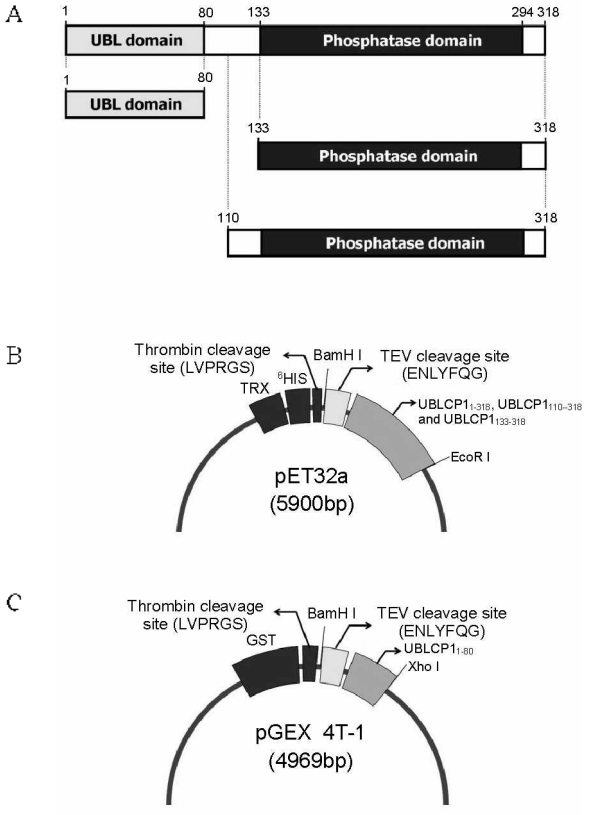

Figure 1. Domain construction and vector maps of UBLCP1. Domains are determined by sequence analysis and secondary structure prediction (A). The maps of *E.coli* expression vectors containing UBLCP1s were presented (B and C).

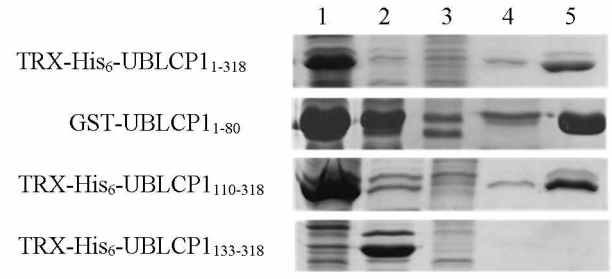

Figure 2. The SDS pages showing solubility of UBLCP1 domains. UBLCP1 fusion proteins were identified by 15% SDS-PAGE and the molecular weight was measured using molecular size marker, Mark 12 (Invitrogen Co.). Lane 1 and 2 indicate supernatant and precipitation of the cell lysate. Lane 3, 4 and 5 indicate the elution profile of affinity chromatography, flow-through, washing and elution, respectively.

Expression and purification of UBLCP1. UBLCP1 domains were successfully expressed in 5 mL LB medium using 0.1 mM IPTG at 25 °C. After Ni-NTA or GST affinity chromatography, UBLCP1, UBL domain and phosphatase domain ($UBLCP1_{110\text{-}318}$) showed an excellent solubility and the final yields are ~7 mg, ~32 mg and ~8 mg for 500 mL culture, respectively. However, $UBLCP1_{133\text{-}318}$ showed low induction rate which was about 30 % of $UBLCP1_{110\text{-}318}$ (data not shown) and it becomes insoluble form in lysis buffer (Fig. 2). According to the expression and solubility data, it is plausible that

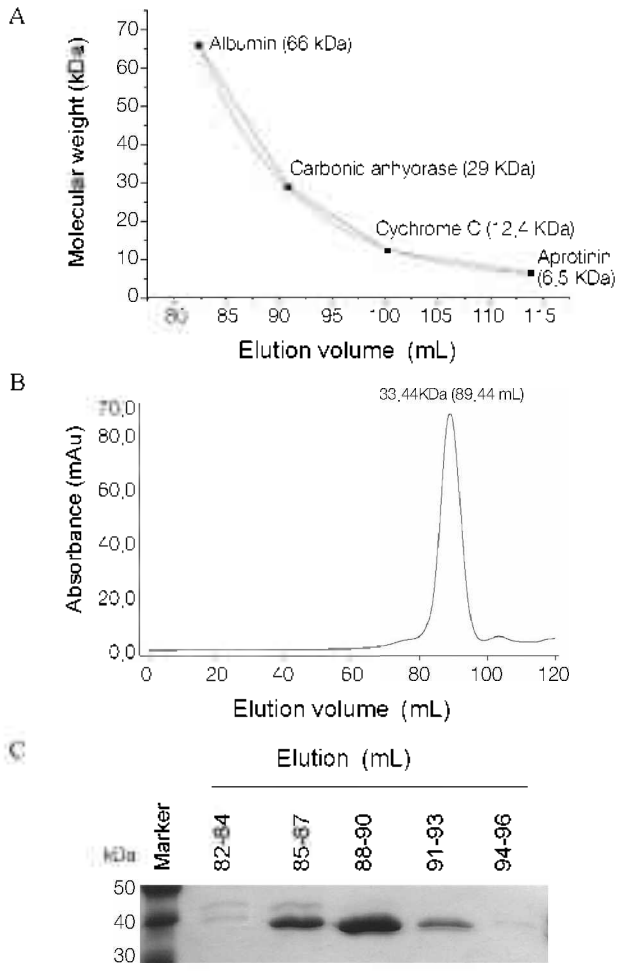

Figure 3. Elution profile of the size exclusion chromatography. The standard marker proteins were loaded and analyzed by Origin 7.01 (Origin Lab.) (A). UBLCP1 domains were eluted from 82 mL to 96 mL and the highest elute fraction was detected in 89.44 mL (B). These were further examined by 15% SDS-PAGE (C).

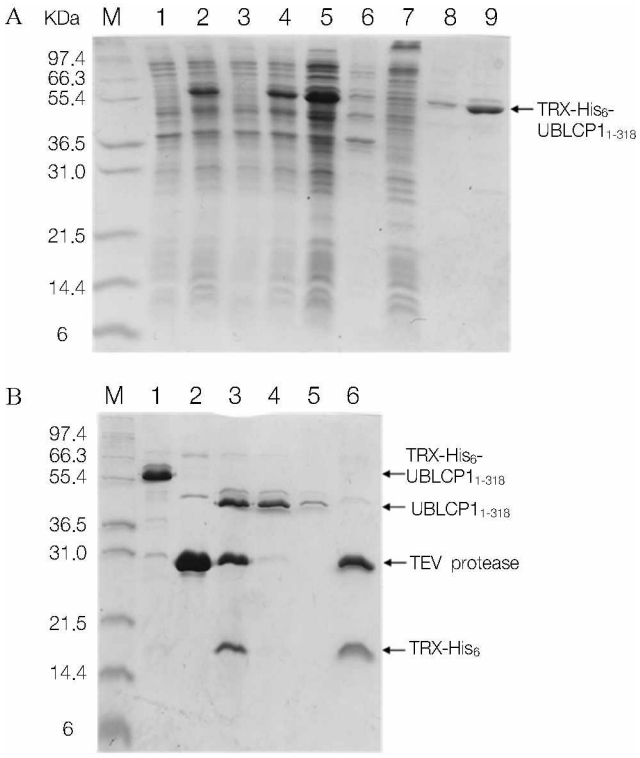

Figure 4. Purification profile of 70% $^{2}H/^{13}C/^{15}N$ labeled UBLCP1. The protein expression and Ni-NTA affinity chromatography were presented (A). Lane 1 and 2 indicate before and after induction profiles in 100% H_2O M9 medium and lane 3 and 4 presented before and after induction profiles in 100% deuterium M9 medium containing ^{13}C-d-glucose and $^{15}NH_4Cl$. Lane 5 and 6 indicate cell lysate (supernatant), precipitation, respectively. Lane 7, 8 and 9 show profiles of the flow-through, washing and elution data of Ni-NTA affinity chromatography, respectively. Data from TEV cleavage reaction are presented in B. Lane 1, 2, 3, 4, 5 and 6 indicate TRX-His_6-UBLCP1, TEV protease, product after TEV protease reaction, flow-through product of Ni-NTA affinity chromatography, proteins after washing procedure and final protein after elution procedure, respectively.

small α-helix in N-terminal (110-132) is necessary to stabilize the structure of phosphatase domain. Although we did not determine the complete three-dimensional structure, we hypothesize that the small α-helix in N-terminal of UBLCP1 could contribute the global structure of phosphatase domain, affecting the protein stability and solubility. Based on the successful determination of structural domains of UBLCP1 which UBL domain is residue 1-80 and phosphatase domain is residue 110-318, we successfully cloned and purified UBLCP1s. Final yields of UBLCP1, UBL domain, and phosphatase domain from 1 L LB medium culture were ~8.5 mg, ~24 mg and ~7.8 mg, respectively. UBLCP1, UBL domain and phosphatase domains were prepared to 0.5 mM (4.8 mg/300 uL), 1.5 mM (4 mg/300 uL) and 1 mM (7.5 mg/300 uL) concentration, respectively.

Size exclusion chromatography of UBLCP1. To determine the molecular weight of UBLCP1 domains, size exclusion chromatography was executed using Superdex^{TM} 75. Albumin (66 KDa), carbonic anhydrase (29 KDa), cytochrome C (12 KDa) and aprobinin (6.5 KDa) were used to calculate the actual molecular weight of each domain by program Origin 7.1 (Fig. 3A). UBLCP1 domains were eluted from 85 mL to 93ml volume and the highest absorbance was detected at 89.4ml position (Fig. 3B and 3C). Based on this analysis, UBLCP1 is determined as a monomer (~33.4 KDa) in our experimental condition.

Structural information of UBLCP1 domains. ^{1}H-^{15}N TROSY-HSQC spectrum shows that most of backbone resonances were broaden due to T2 relaxation effect. To overcome the T2 relaxation effect, we exchanged 70% of protons to deuterium in UBLCP1. Since deuterium labeled house keeping enzyme shows lower enzyme activity than wild one, *E.coli* grown cell from deuterium environment shows low viability and protein expression.[22] To ensure over-expression of UBLCP1 protein in deuterium environments, we carried out deuterium adaptation procedure using different contents of deuterium in M9 media and finally purified triple-labeled protein sample for NMR experiments (Fig. 4). Most of backbone amide proton resonances were observed in HSQC spectrum of UBLCP1, UBL and phosphatase domain, respectively (Fig. 5A, 5B and 5C). Whereas UBL domain shows well-dispersed resonances,

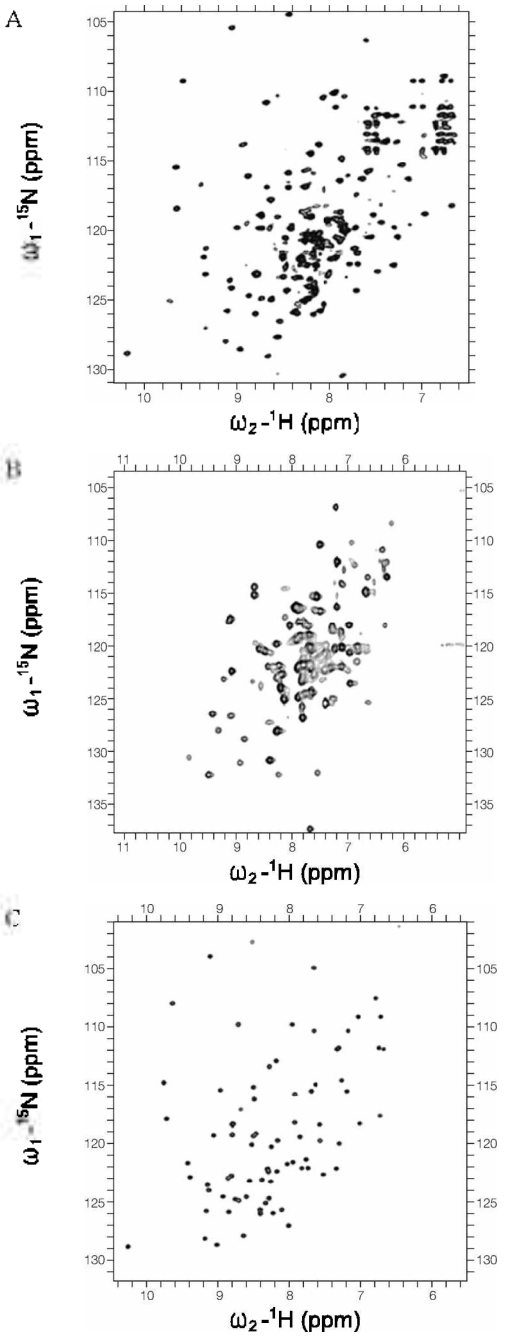

Figure 5. ^{1}H-^{15}N HSQC spectra of UBLCP1 domains. ^{1}H-^{15}N TROSY-HSQC spectra of UBLCP1 and phosphatase domain were presented in (A) and (B). ^{1}H-^{15}N HSQC spectra of UBL domain is displayed in (C). NMR spectra of UBLCP1 and phosphatase domain were performed in Bruker DRX 900 MHz with Cryoprobe™ at 298 K. NMR spectra of UBL domains were collected in Bruker DRX 500 MHz spectrometer equipped with Cryoprobe™.

a number of backbone resonances of the phosphatase domain are observed near 8 ppm, implying that phosphatase domain might possesses partially disordered structure (Fig. 5B and 5C). In addition, comparison of HSQC spectrum suggests that the phosphatase and UBL domains are structurally independent each other (Fig. 5A). The complete resonance assignment for structural elucidation of UBLCP1 using TROSY-based NMR data is currently in progress.

Conclusion

Human UBLCP1 domains are successfully cloned into the *E.coli* expression vectors, pET32a and pGEX 4T-1 with TEV protease cleavage site and purified using both affinity and gel-filtration chromatography. UBLCP1, UBL domain and phosphatase domain were purified and determined as a monomeric conformation. Isotope-labeled proteins including triple-labeled (^{2}H/^{15}N/^{13}C) UBLCP1 were also prepared for NMR experiments and ^{15}N-^{1}H 2D-HSQC spectra suggest that both UBL and phosphatase domain are correctly folded and they are structurally independent by a flexible linker. This data will be of use in performing detailed structural and functional study of UBLCP1.

Acknowledgments. This work was supported by Korea Research Foundation Grant (KRF-2007-C00444 to W. Lee) and by 900 MHz NMR spectrometer at the Korea Basic Science Institute.

References

1. Cramer, P.; Armache, K. J.; Baumli, S.; Benkert, S.; Brueckner, F.; Buchen, C.; Damsma, G. E.; Dengl, S.; Geiger, S. R.; Jasiak, A. J.; Jawhari, A.; Jennebach, S.; Kamenski, T.; Kettenberger, H.; Kuhn, C. D.; Lehmann, E.; Leike, K.; Sydow, J. F.; Vannini, A. In *Annual Review of Biophysics and Biomolecular Structure*; Rees, D. C.; Sheetz, M. P.; Williamson, J. R., Eds.; 2008; Vol. 37, p 337-352.
2. Dahmus, M. E. *Journal of Biological Chemistry* **1996**, *271*, 19009-19012.
3. Dahmus, M. E. *Biochim. Biophys. Acta* **1995**, *1261*, 171-182.
4. Maniatis, T.; Reed, R. *Nature* **2002**, *416*, 499-506.
5. Palancade, B.; Bensaude, O. *Eur. J. Biochem.* **2003**, *270*, 38593870.
6. Proudfoot, N. J.; Furger, A.; Dye, M. J. *Cell* **2002**, *108*, 501-512.
7. Sims, R. J., 3rd; Mandal, S. S.; Reinberg, D. *Curr. Opin. Cell Biol.* **2004**, *16*, 263-271.
8. Layboum, P. J.; Dahmus, M. E. *J. Biol. Chem.* **1989**, *264*, 66936698.
9. O'Brien, T.; Hardin, S.; Greenleaf, A.; Lis, J. T. *Nature* **1994**, *370*, 75-77.
10. Collet, J. F.; Stroobant, V.; Pirard, M.; Delpierre, G.; Van Schaftingen, E. *J. Biol. Chem.* **1998**, *273*, 14107-14112.
11. Yeo, M.; Lin, P. S.; Dahmus, M. E.; Gill, G. N. *J. Biol. Chem.* **2003**, *278*, 26078-26085.
12. Zheng, H.; Ji, C.; Gu, S.; Shi, B.; Wang, J.; Xie, Y.; Mao, Y. *Biochem. Biophys. Res. Commun.* **2005**, *331*, 1401-1407.
13. Ciechanover, A. *EMBO Journal* **1998**, *17*, 7151-7160.
14. Hershko, A. *Cell Death and Differentiation* **2005**, *12*, 1191- 1197.
15. Jentsch, S.; Pyrowolakis, G. *Trends Cell Biol.* **2000**, *10*, 335- 342.
16. Elsasser, S.; Gali, R. R.; Schwickart, M.; Larsen, C. N.; Leggett, D. S.; Muller, B.; Feng, M. T.; Tubing, F.; Dittmar, G. A.; Finley, D. *Nat. Cell Biol.* **2002**, *4*, 725-730.
17. Sakata, E.; Yamaguchi, Y.; Kurimoto, E.; Kikuchi, J.; Yokoyama, S.; Yamada, S.; Kawahara, H.; Yokosawa, H.; Hattori, N.; Mizuno, Y.; Tanaka, K.; Kato, K. *EMBO Rep.* **2003**, *4*, 301-306.
18. Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366-12371.
19. Bodenhausen, G.; Ruben, D. J. *Chem. Phys. Lett.* **1980**, *69*, 185-189.
20. Salzmann, M.; Wider, G.; Pervushin, K.; Senn, H.; Wüthrich, K. *Journal of the American Chemical Society* **1999**, *121*, 844-848.
21. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277-293.
22. Hochuli, M.; Szyperski, T.; Wüthrich, K. *Journal of Biomolecular NMR* **2000**, *17*, 33-42.