Structure and Carboxyl-terminal Domain (CTD) Binding of the Set2 SRI Domain That Couples Histone H3 Lys^{36} Methylation to Transcription

Erika Vojnic, Bernd Simon, Brian D. Strahl, Michael Sattler and Patrick Cramer

J. Biol. Chem. 2006, 281:13-15.
doi: 10.1074/jbc.C500423200 originally published online November 14, 2005

ACCELERATED PUBLICATION

This paper is available online at www.jbc.org

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 1, pp. 13–15, January 6, 2006

Structure and Carboxyl-terminal Domain (CTD) Binding of the Set2 SRI Domain That Couples Histone H3 Lys[36] Methylation to Transcription*[S]

Received for publication, October 25, 2005, and in revised form, November 10, 2005
Published, JBC Papers in Press, November 14, 2005, DOI 10.1074/jbc.C500423200

Erika Vojnic[‡], Bernd Simon[§], Brian D. Strahl[¶1], Michael Sattler[§2], and Patrick Cramer[‡3]

From the [‡]Gene Center, Department of Chemistry and Biochemistry, Ludwig-Maximilians-University of Munich, Feodor-Lynen-Strasse 25, 81377 Munich, Germany, the [§]EMBL, Meyerhofstrasse 1, 69120 Heidelberg, Germany, and the [¶]Department of Biochemistry and Biophysics, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599-7260

During mRNA elongation, the SRI domain of the histone H3 methyltransferase Set2 binds to the phosphorylated carboxyl-terminal domain (CTD) of RNA polymerase II. The solution structure of the yeast Set2 SRI domain reveals a novel CTD-binding fold consisting of a left-handed three-helix bundle. NMR titration shows that the SRI domain binds an Ser[2]/Ser[5]-phosphorylated CTD peptide comprising two heptapeptide repeats and three flanking NH_2-terminal residues, whereas a single CTD repeat is insufficient for binding. Residues that show strong chemical shift perturbations upon CTD binding cluster in two regions. Both CTD tyrosine side chains contact the SRI domain. One of the tyrosines binds in the region with the strongest chemical shift perturbations, formed by the two NH_2-terminal helices. Unexpectedly, the SRI domain fold resembles the structure of an RNA polymerase-interacting domain in bacterial σ factors (domain σ_2 in σ^{70}).

Gene transcription by RNA polymerase II (Pol II) is physically and functionally coupled to other nuclear events, most notably mRNA processing (1–7). Transcription-coupled events generally depend on the carboxyl-terminal repeat domain (CTD)[4] of the largest Pol II subunit, which binds many nuclear factors during transcription elongation. The CTD forms a mobile extension from the structural core of Pol II (8) and consists of heptapeptide repeats of the consensus sequence Tyr[1]-Ser[2]-Pro[3]-Thr[4]-Ser[5]-Pro[6]-Ser[7], which can be phosphorylated at residues Ser[2] and Ser[5]. The CTD phosphorylation pattern changes during the transcription cycle. Ser[5] phosphorylation occurs in promoter-proximal regions and leads to recruitment of the 5′-RNA capping enzyme (9–12). Ser[2] phosphorylation occurs in regions that are more distal from the promoter and triggers binding of the 3′-RNA processing machinery (10, 13).

Recently it emerged that transcription is also coupled to the alteration of chromatin structure. The histone methyltransferases Set1 and Set2, which catalyze methylation of histone H3 lysines Lys[4] and Lys[36], respectively, are associated with Pol II during elongation (reviewed in Refs. 14 and 15). Histone methylation apparently controls newly initiated Pol II, and two phases of histone H3 methylation can be distinguished after transcription initiation (16). Set1 association with Pol II is mediated by the Paf complex, which occurs in promoter regions, and depends on Ser[5] phosphorylation of the CTD (17, 18). In contrast, Set2 directly interacts with the phosphorylated CTD of Pol II and is observed throughout the coding region of genes (17–20). Set2 recruitment to Pol II requires the CTD kinase CTDK-I that phosphorylates Ser[2] residues in the CTD (17, 18, 20, 21).

Set2 interacts with the Pol II CTD via a novel domain, the Set2 Rpb1-interacting (SRI) domain (22, 23). The SRI domain of *S. cerevisiae* comprises the COOH-terminal residues 619–718 of Set2 (22). *In vitro*, the yeast Set2 SRI domain binds specifically and with high affinity to the CTD doubly phosphorylated at Ser[2] and Ser[5] (22). *In vivo*, deletion of the Set2 SRI domain abolishes H3 Lys[36] methylation and impairs transcription elongation (22), suggesting that the SRI domain is responsible for coupling transcription to histone methylation by Set2.

Here we report the solution structure of the Set2 SRI domain from the yeast *S. cerevisiae* and present NMR binding experiments with phospho-CTD peptides. Our results elucidate the molecular determinants for Set2 CTD binding, which underlies coupling of transcription to Set2-directed chromatin modification.

EXPERIMENTAL PROCEDURES

Sample Preparation—The region of the gene of the *Saccharomyces cerevisiae* Set2 protein (Swiss Prot P46995) encoding for Set2 residues 620–719 was cloned into a modified pET9d vector with an NH_2-terminal hexahistidine tag. The protein was overexpressed in *Escherichia coli* pLysS cells at 18 °C for 16 h. For labeling of the protein with $^{15}N/^{13}C$ or ^{15}N, cells were grown in M9 minimal medium supplemented with [$^{13}C_6$]glucose and/or $^{15}NH_4Cl$. Cell lysates were subjected to affinity chromatography on a nickel-nitrilotriacetic acid column (Quiagen), followed by cleavage of the hexahistidine tag with tobacco etch virus protease and dialysis overnight. The tag and the His_6-tagged protease were removed on a second Ni-NTA column. DNA was removed by cation exchange chromatography (Mono S, Amersham Biosciences). After gel filtration the sample was dissolved in 20 mM sodium phosphate, pH 6.5, 200 mM NaCl, 0.2 mM dithiothreitol. Edman sequencing of the protein confirmed the presence of four additional residues (GAMG) at the NH_2 terminus, which result from the cloning strategy. NMR samples were prepared in H_2O or 100% D_2O at 0.4–1 mM concentration of protein.

NMR Structure Determination—NMR spectra were acquired at 292 K on Bruker DRX500, DRX600, or DRX900 spectrometers with cryogenic triple resonance probes. Spectra were processed with NMRPipe (24) and analyzed using NMRVIEW (25). The ^1H, ^{13}C, and ^{15}N chemical shifts were assigned by standard methods (26). Distance restraints were derived from two-dimensional NOESY and ^{15}N- or ^{13}C-resolved three-dimensional NOESY. Restraints for the backbone angles ϕ and ψ were derived from TALOS (27). Slowly exchanging amide protons were identified from $^1H,^{15}N$ correlation experiments after dissolving of lyophilized protein in D_2O. ^{15}N relaxation (T1, T2) and heteronuclear (^1H)-^{15}N NOE was measured on a ^{15}N-labeled protein sample at 292 K as described (28) (supplemental Fig. S1). The experimentally determined distance and dihedral restraints (supplemental Table S1 and Fig. 1*C*) were applied in a simulated-annealing protocol using ARIA (29) and CNS (30). NOEs were manually assigned and distance calibrations were performed by ARIA. The final ensemble of NMR structures was refined in a shell of water molecules (31). Structural quality was analyzed with PROCHECK (32).

Phosphopeptide Interaction Studies—The phospho-CTD peptides used for binding experiments were chemically synthesized (one-repeat peptide, YpSPTpSPS; two-repeat peptide, SPS-YpSPTpSPS-YpSPTpSPS, pS = phosphoserine). For NMR titration, increasing amounts of the CTD peptide were added to a 0.4 mM solution of $^{15}N,^{13}C$-labeled SRI domain up to a 1.25-fold molar excess. Chemical shifts were monitored in two-dimensional $^1H,^{15}N$ HSQC experiments.

RESULTS AND DISCUSSION

The Set2 SRI Domain Forms a Conserved Three-helix Bundle—The solution structure of the yeast Set2 SRI domain was determined by multidimensional NMR (supplemental Table S1; also see "Experimental Procedures"). The structure revealed three α-helices arranged in a left-handed bundle (Fig. 1). The NH_2-terminal helix $\alpha1$ is slightly kinked at residues Phe[639] and Val[640], and the linker between helices $\alpha1$ and $\alpha2$ includes a short 3_{10}-helical turn at residues Ser[650]–

* This work was supported in part by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Table S1 and Figs. S1–S3.

The atomic coordinates and structure factors (code 2C5Z) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] Supported by the National Institutes of Health and is a Pew Scholar in the Biomedical Sciences.

[2] To whom correspondence may be addressed. E-mail: sattler@embl-heidelberg.de.

[3] To whom correspondence may be addressed. Tel.: 49-89-2180-76951; Fax: 49-89-2180-76999; E-mail: cramer@LMB.uni-muenchen.de.

[4] The abbreviations used are: CTD, carboxyl-terminal domain; Pol, RNA polymerase II; SRI, Set2 Rpb1-interacting; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser and exchange spectroscopy; PDB, Protein Data Bank.

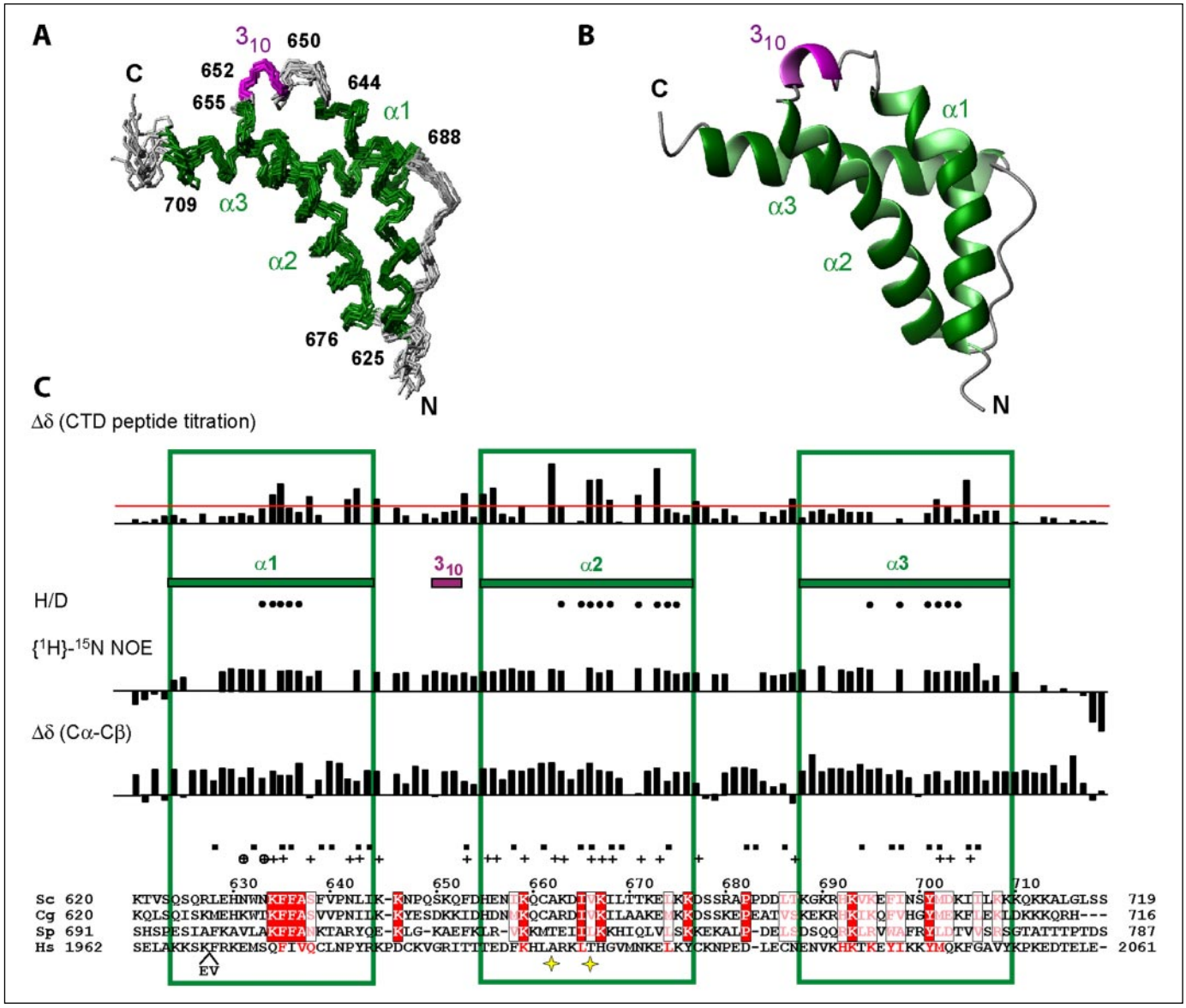

FIGURE 1. **Structure and CTD binding of the yeast Set2 SRI domain.** *A*, ensemble of final NMR structures. The three α-helices are shown in *green*, and a short 3_{10}-helix is shown in *pink*. *B*, ribbon diagram of the lowest energy structure in *A*. *C*, alignment of SRI domain sequences and NMR structure determination and CTD binding data. The secondary structure is shown above the sequence. Solvent-protected amide protons that show slow H/D exchange are indicated by *filled circles*. Secondary chemical shifts Δδ(Cα-Cβ) are indicated by *black bars*. Residues that experience large chemical shift perturbations upon addition of the CTD two-repeat phosphopeptide SPS-YpSPTpSPS-YpSPTpSPS (pS = phosphoserine) are indicated above the alignment with *crosses* and *circled crosses* for backbone and side chain amides, respectively. *Yellow stars* indicate residues Ala[662] and Val[666] that are implicated in binding of a CTD tyrosine side chain. Residues that are identical and conserved in fungal Set2 homologues are on *red background* and in *red*, respectively. Hydrophobic core residues are marked with a *black square*.

Gln[652]. A hydrophobic core is formed by numerous residues located at the interface between the three helices, including four residues in the two regions linking the helices (Fig. 1*C*). Consistently the heteronuclear {^{1}H}-^{15}N NOE measurements demonstrate that the polypeptide backbone in all three helices and the connecting linker regions is rigid (Fig. 1*C* and supplemental Fig. S1). The hydrophobic core residues are generally conserved across species (Fig. 1*C*), demonstrating that our structure is a good model for SRI domains in Set2 of other species.

The SRI Domain Defines a Novel CTD-binding Fold—Comparison with the five known structures of CTD-binding domains reveals that the SRI domain defines a novel CTD-binding fold. Other CTD-binding domains include FF domains, CTD-interacting domains, WW domains, BRCT domains, and a domain in the Cgt1 subunit of the 5′-capping enzyme (reviewed in Ref. 7). Of these, FF and CTD-interacting domains also form helical bundles (33, 34), but, in contrast to the SRI domain, the superhelical arrangement in these two domains is right-handed (supplemental Fig. S2). Thus the six CTD-binding domains that have been structurally characterized use different folds for specific CTD recognition.

The SRI Domain Binds a Two-repeat CTD Phosphopeptide—To characterize the CTD-binding determinants of the SRI domain, we performed NMR titration experiments with Ser[2]/Ser[5]-phosphorylated CTD peptides (Fig. 1*C*). A phosphopeptide consisting of a single CTD repeat (YpSPTpSPS, pS = phosphoserine; Fig. S3*A*) did not perturb chemical shifts in a two-dimensional ^{1}H,^{15}N HSQC spectrum, indicating that there is no significant binding (data not shown). However, titration with a peptide that comprised two CTD repeats and three flanking NH_2-terminal residues (SPS-YpSPTpSPS-YpSPTpSPS) resulted in many strong chemical shift perturbations (Fig. 1*C* and supplemental Fig. S3). From the titration data the dissociation constant is estimated to be in the low micromolar range, comparable with the reported approximate affinity of 6 μM for a CTD phosphopeptide comprising three repeats (22).

Regions in the SRI Domain That Interact with the CTD—Residues that show strong chemical shift perturbations of their backbone NH groups cluster in two regions on the SRI domain structure (Fig. 2*A*). The first region includes residues Lys[634], Phe[635] in α1, and Ala[662], Val[666], Lys[667], Thr[670], Thr[671], and Glu[673] in α2, whereas the second region includes residues Phe[653], His[655], Glu[656] in the α1-α2 linker, and residue Ile[705] in α3 (Figs. 1*C* and 2*A* and supplemental Fig. S3). With the exception of Ile[705], the strongest perturbations upon peptide binding were observed in region 1 (Phe[635], Ala[662], Val[666], Lys[667], and Glu[673]). In this region, the side chain NH_2 groups of residues Asn[631] and

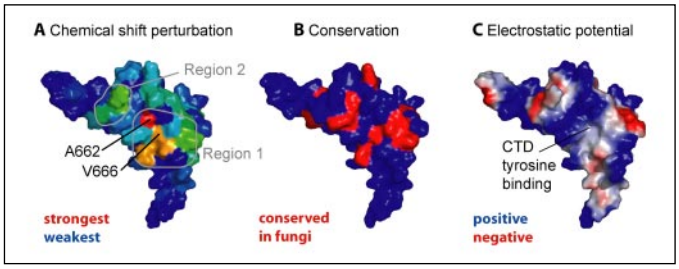

FIGURE 2. **Surface analysis of the Set2 SRI domain.** Surface representation of the SRI domain color-coded according to chemical shift perturbation of backbone NH and side chain NH_2 groups upon binding of the CTD peptide (supplemental Fig. S3*A*), colored from *red* to *blue* for strong to weak perturbations, respectively (*A*); amino acid conservation among fungal homologues in the alignment of Fig. 1*C* (*B*); and electrostatic surface potential (*blue* and *red* for positive and negative charges, respectively) (*C*).

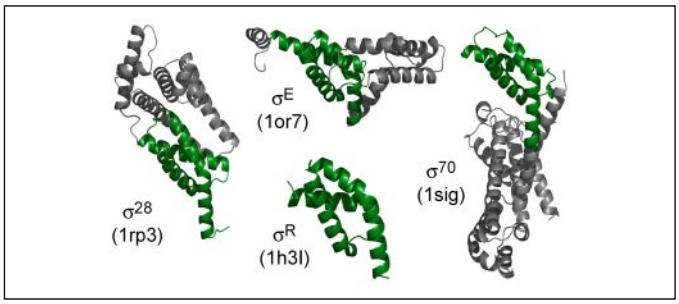

FIGURE 3. **The Set2 SRI domain resembles a domain in bacterial σ factors.** The domain in the σ factor that resembles the SRI fold is highlighted in *green*. The PDB codes of the structures are given in *parentheses*.

Asn^{633} also show significant chemical shift perturbations (supplemental Fig. S3*B*). Both regions are conserved among fungal Set2 homologues (Fig. 2*B*), befitting the conserved function of the *Saccharomyces pombe* and *Neurospora crassa* Set2 homologues (35, 36). The observation of two putative CTD-binding regions, and the finding that two CTD repeats are required for SRI domain binding, indicate that the phospho-CTD extends over a long distance along helices α1 and α2 and the connecting linker.

CTD Tyrosine Side Chains Contribute to SRI Domain Binding—The peptide titration experiments also revealed that the two-repeat CTD peptide (supplemental Fig. S3*A*) binds to the SRI domain via its tyrosine residues. Intermolecular NOEs between both CTD tyrosine side chains and the SRI domain were detected (data not shown). Preliminary assignments indicate that one of the tyrosine side chains is in proximity of residues Ala^{662} and Val^{666} in region 1 (Figs. 1*C* and 2*B*). These two residues are part of a hydrophobic patch between helices α1 and α2 and flanked by positively charged surfaces (Fig. 2*C*), as expected for interaction with the negatively charged phospho-CTD. Interestingly, the tyrosine-proximal residue Ala^{662} is identical in human Set2, as are Phe^{635}, Glu^{656}, and Glu^{673} in the putative CTD-binding regions (Fig. 1*C*). In the three known CTD-protein complex structures, the Y1 side chain is also involved in hydrophobic contacts (34, 37, 38), suggesting that Y1 binding is a general feature of CTD recognition. Previous studies revealed that the CTD can adopt different conformations (reviewed in Ref. 7), and this structurally versatile nature of the CTD discourages any detailed model building.

The SRI Domain Resembles a Polymerase-interacting Domain in Bacterial σ Factors—Comparison of our structure with known folds in the data base (DALI (39)) strikingly shows that the SRI domain resembles a region in bacterial σ factors (Fig. 3). The four highest hits were the σ factors σ^{28} (PDB code 1rp3), σ^{E} (PDB-code 1or7), σ^{R} (PDB code 1h3l), and σ^{70} (PDB code 1sig), which show DALI scores of 5.6, 5.4, 5.1, and 4.9, respectively, and root mean square deviations between 3.3 and 3.7 Å. The region in σ^{70} that is structurally related to the SRI domain is domain 2 (σ_2), which interacts with the clamp region of the core RNA polymerase upon formation of the holoenzyme (40). The σ_2 domain is involved in binding the −10 element of promoter DNA and contributes to DNA melting during initiation (reviewed in Ref. 41). In the eukaryotic initiation complex, promoter DNA around position −10 lies near the NH_2-terminal domain of the initiation factor TFIIEα (42), which shows weak sequence homology (43) and structural similarity (44) to the bacterial σ_2 domain. We speculate that the eukaryotic TFIIEα NH_2-terminal domain, which may contact promoter DNA, and the Set2 SRI domain, which binds the negatively charged phospho-CTD, both evolved from the bacterial σ_2 domain.

Acknowledgments—We thank C. Buchen, L. Lariviere, and other members of the Cramer laboratory (Gene Center Munich) for help. We thank G. Stier and A. Lingel (EMBL, Heidelberg, Germany) and K. Kizer (University of North Carolina) for help.

REFERENCES

1. Dahmus, M. E. (1995) *Biochim. Biophys. Acta* **1261,** 171–182
2. Palancade, B., and Bensaude, O. (2003) *Eur. J. Biochem.* **270,** 3859–3870
3. Sims, R. J., III, Mandal, S. S., and Reinberg, D. (2004) *Curr. Opin. Cell Biol.* **16,** 263–271
4. Zorio, D. A., and Bentley, D. L. (2004) *Exp. Cell Res.* **296,** 91–97
5. Proudfoot, N. (2004) *Curr. Opin. Cell Biol.* **16,** 272–278
6. Maniatis, T., and Reed, R. (2002) *Nature* **416,** 499–506
7. Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. (2005) *Genes Dev.* **19,** 1401–1415
8. Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001) *Science* **292,** 1863–1876
9. Ho, C. K., Sriskanda, V., McCracken, S., Bentley, D., Schwer, B., and Shuman, S. (1998) *J. Biol. Chem.* **273,** 9577–9585
10. Komarnitsky, P., Cho, E. J., and Buratowski, S. (2000) *Genes Dev.* **14,** 2452–2460
11. McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D. L. (1997) *Genes Dev.* **11,** 3306–3318
12. Cho, E. J., Takagi, T., Moore, C. R., and Buratowski, S. (1997) *Genes Dev.* **11,** 3319–3326
13. Cho, E. J., Kobor, M. S., Kim, M., Greenblatt, J., and Buratowski, S. (2001) *Genes. Dev.* **15,** 3319–3329
14. Gerber, M., and Shilatifard, A. (2003) *J. Biol. Chem.* **278,** 26303–26306
15. Hampsey, M., and Reinberg, D. (2003) *Cell* **113,** 429–432
16. Morillon, A., Karabetsou, N., Nair, A., and Mellor, J. (2005) *Mol. Cell* **18,** 723–734
17. Krogan, N. J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D. P., Beattie, B. K., Emili, A., Boone, C., Shilatifard, A., Buratowski, S., and Greenblatt, J. (2003) *Mol. Cell Biol.* **23,** 4207–4218
18. Ng, H. H., Robert, F., Young, R. A., and Struhl, K. (2003) *Mol. Cell* **11,** 709–719
19. Strahl, B. D., Grant, P. A., Briggs, S. D., Sun, Z. W., Bone, J. R., Caldwell, J. A., Mollah, S., Cook, R. G., Shabanowitz, J., Hunt, D. F., and Allis, C. D. (2002) *Mol. Cell. Biol.* **22,** 1298–1306
20. Xiao, T., Hall, H., Kizer, K. O., Shibata, Y., Hall, M. C., Borchers, C. H., and Strahl, B. D. (2003) *Genes Dev.* **17,** 654–663
21. Li, B., Howe, L., Anderson, S., Yates, J. R., III, and Workman, J. L. (2003) *J. Biol. Chem.* **278,** 8897–8903
22. Kizer, K. O., Phatnani, H. P., Shibata, Y., Hall, H., Greenleaf, A. L., and Strahl, B. D. (2005) *Mol. Cell Biol.* **25,** 3305–3316
23. Phatnani, H. P., Jones, J. C., and Greenleaf, A. L. (2004) *Biochemistry* **43,** 15702–15719
24. Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6,** 277–293
25. Johnson, B. A., and Blevins, R. A. (1994) *J. Biomol. NMR* **4,** 603–614
26. Sattler, M., Schleucher, J., and Griesinger, C. (1999) *Prog. NMR Spectrosc.* **34,** 93–158
27. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR* **13,** 289–302
28. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* **33,** 5984–6003
29. Linge, J. P., O'Donoghue, S. I., and Nilges, M. (2001) *Methods Enzymol.* **339,** 71–90
30. Brünger, 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., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
31. Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M., and Nilges, M. (2003) *Proteins* **50,** 496–506
32. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR* **8,** 477–486
33. Allen, M., Friedler, A., Schon, O., and Bycroft, M. (2002) *J. Mol. Biol.* **323,** 411–416
34. Meinhart, A., and Cramer, P. (2004) *Nature* **430,** 223–226
35. Morris, S. A., Shibata, Y., Noma, K., Tsukamoto, Y., Warren, E., Temple, B., Grewal, S. I., and Strahl, B. D. (2005) *Eukaryot. Cell* **4,** 1446–1454
36. Adhvaryu, K. K., Morris, S. A., Strahl, B. D., and Selker, E. U. (2005) *Eukaryot. Cell* **4,** 1455–1464
37. Fabrega, C., Shen, V., Shuman, S., and Lima, C. D. (2003) *Mol. Cell* **11,** 1549–1561
38. Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T., and Noel, J. P. (2000) *Nat. Struct. Biol* **7,** 639–643
39. Holm, L., and Sander, C. (1995) *Trends Biochem. Sci.* **20,** 478–480
40. Murakami, K. S., Masuda, S., and Darst, S. A. (2002) *Science* **296,** 1280–1284
41. Gross, C. A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998) *Cold Spring Harbor Symp. Quant. Biol.* 63, 141–155
42. Forget, D., Langelier, M.-F., Therien, C., Trinh, V., and Coulombe, B. (2004) *Mol. Cell. Biol.* **24,** 1122–1131
43. Okhuma, Y., Sumimoto, H., Hoffmann, A., Shimasaki, S., Horikoshhi, M., and Roeder, R. (1991) *Nature* **354,** 398–401
44. Meinhart, A., Blobel, J., and Cramer, P. (2003) *J. Biol. Chem.* **278,** 48267–48274