Biochemistry

 Volume 37, Number 51 December 22, 1998

New Concepts in Biochemistry

Structure and Function of the Core Histone N-Termini: More Than Meets the Eye†

Jeffrey C. Hansen,*,‡ Christin Tse,‡ and Alan P. Wolffe§

Department of Biochemistry, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, Texas 78284-7760, and Laboratory of Molecular Embryology, National Institute of Child Health and Human Development, National Institutes of Health, Building 18T, Room 106, Bethesda, Maryland 20892-5431

Received October 9, 1998

ABSTRACT: For two decades, the core histone N-termini generally have been thought of as unstructured domains whose function is to bind to DNA and screen negative charge. New data indicates that both the molecular mechanisms of action and biological functions of the core histone N-termini in chromatin are considerably more complex. At the level of the chromatin fiber, multiple distinct functions of the N-termini are required to achieve higher order chromatin condensation, two of which apparently involve protein−protein rather than protein−DNA interactions. In addition, the N-termini have been documented to participate in specific interactions with many chromatin-associated regulatory proteins. Here, we discuss evidence supporting the new concepts that when functioning in their natural chromatin context, (1) the N-termini are engaged primarily in protein−protein interactions, (2) as a consequence of these interactions the N-termini adopt specific secondary structure, (3) posttranslational modifications such as acetylation disrupt the ability of the N-termini to form secondary structure, and (4) because the N-termini perform essential roles in both chromatin condensation and also bind specific chromatin-associated proteins, the global structure and function of any given region of the genome will be determined predominantly by the core histone N-termini and their specific interaction partners.

The core histones H2A, H2B, H3, and H4 (Figure 1A) each consist of a uniquely structured histone fold domain involved in both histone−histone and histone−DNA interactions in the nucleosome (*1*, *2*) and a highly basic N-terminus whose functions are less understood (reviewed in refs *3* and *4*). The core histone N-termini present a puzzle to those attempting to decipher their structure−function relationships. Despite possessing no apparent propensity to function as anything other than uncoiled peptides that bind DNA due to their high positive surface charge density, the core histone N-termini are nevertheless among the mostly highly conserved stretches of 25−40 amino acids yet to be discovered. In particular, the amino acid sequence of both the yeast and human H3 and H4 N-termini are identical, with the exception of one conservative substitution (*5*). This implies that the N-termini do not function simply by interacting nonspecifically with the DNA components of chromatin, since such a mechanism would not require such an extraordinary degree of sequence conservation. Therein lies the enigma.

Beginning with the pioneering studies of Allfrey et al. (*6*), who in 1964 showed that acetylation of the core histone N-termini is strongly correlated with enhanced transcription, there has been intense interest in the structure and function of these histone domains. This interest has never been stronger, both because of a decade of renewed focus on

† This work was supported by NIH grant GM45916 to J.C.H.

* Corresponding author. Phone: 210-567-6980. Fax: 210-567-6595. E-mail: hansen@bioc02.uthscsa.edu.

‡ The University of Texas Health Science Center of San Antonio.

§ National Institutes of Health.

10.1021/bi982409v CCC: $15.00
Published on Web 12/02/1998

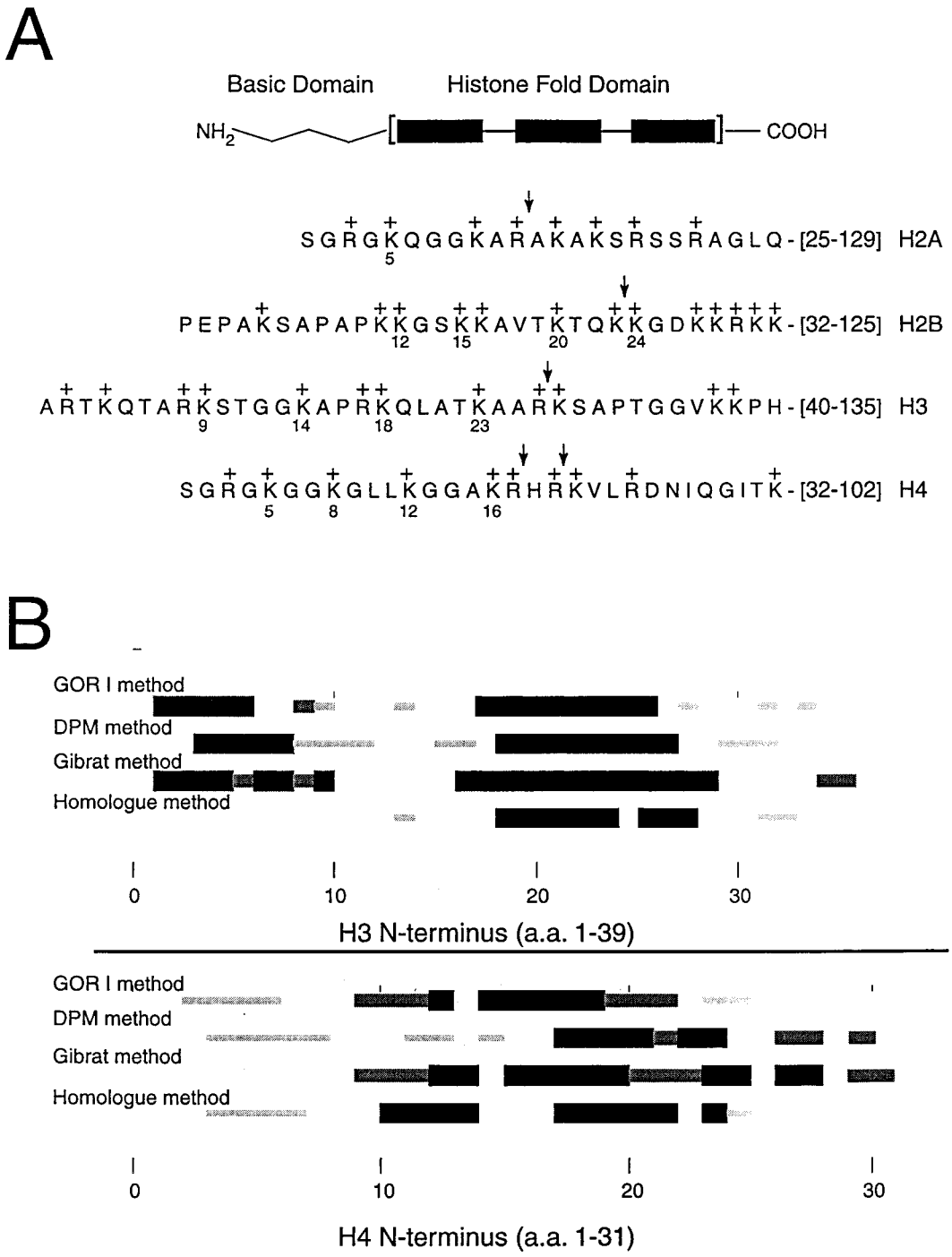

FIGURE 1: (A) Schematic illustration of the domain structure of the four core histones. Residue numbers in the N-termini indicate the sites of acetylation. The sites of proteolytic cleavage by trypsin are indicated by the arrows. Brackets indicate the residues that comprise the histone fold domain. Two each of the core histones comprise the histone octamer (1). A nucleosome core particle contains 146 bp of DNA wrapped ~1.75 times around the histone octamer (2). Histone octamers spaced at ~200 bp intervals along a DNA molecule form nucleosomal arrays (*11*). Nucleosomal arrays complexed with linker histones and other nonhistone proteins comprise chromatin (*3, 4*). (B) Prediction of secondary structure within the H3 and H4 N-termini. GORI (*55*), DPM (*56*), Gibrat (*57*), and Homologue (*58*) methods were employed as indicated, and have a 60−75% prediction accuracy based on analyses of proteins of known structure. Regions predicted to form α-helices, *β*-sheets, or *β*-turns are indicated as black rectangles, gray rectangles, or light gray rectangles, respectively. Gaps between regions of predicted secondary structure represent predicted stretches of random coils. The same analyses performed on the H2A and H2B N-termini indicated as a general trend that residues 9−14 of H2A and 10−21 of H2B also were predicted to form α-helices.

chromatin per se, as well as because of the recent explosion of papers indicating that numerous proteins heretofore considered to be transcriptional activators, co-activators, or repressors are actually enzymes that specifically acetylate or deacetylate the N-termini, respectively (*7, 8*). However, despite over three decades of investigation and the availability of the crystal structures of both the histone octamer (*1*) and the nucleosome (*2*), the molecular mechanisms through which the core histone N-termini link chromatin structure and function remain to be established.

Only when these domains have been studied extensively in their biologically relevant context, i.e., nucleosomal arrays and chromatin, has a picture of N-termini function started to emerge that is consistent with their degree of evolutionary

conservation. In particular, it is now clear that the core histone N-termini mediate interaction of many nonhistone chromosomal proteins with the chromatin fiber and at the same time perform essential functions required for higher order chromatin folding. Furthermore, in each instance, the N-termini do not appear to work through a mechanism in which these domains behave as unstructured coils that bind to negatively charged macromolecular surfaces. These observations in turn suggest several intriguing new concepts about the structure of the core histone N-termini and how they function.

N-Termini Function in Chromatin: Multiple Roles for Protein−Protein Interactions. Approximately 30−40% of each of the core histone N-termini consist of lysines and arginines (Figure 1A). This high degree of positive surface charge density, together with the fact that these domains lack detectable secondary structure when studied as free histones or isolated peptides (reviewed in refs *9* and *10*), has led to the widespread perception that the N-termini function in chromatin primarily by binding as unstructured coils to the DNA component of nucleosomes (reviewed in refs *4*, *10*, and *11*). In support of this view, interactions of the N-termini with the DNA of isolated nucleosome core particles have been well documented in vitro (*3, 4, 9, 10*). However, it has recently been established that under physiological ionic conditions the core histone N-termini and analogous C-terminal domain of histone H2A are *not* stably bound to nucleosomal DNA in either native chromatin, reconstituted nucleosomal arrays, or mononucleosomes containing extra-nucleosomal DNA (*12*−*14*). These studies indicate that the mobile N-termini are forced to rearrange to a nucleosomal DNA location when intact chromatin is destroyed to create nucleosome core particles (*10, 11*). Consequently, the large body of results obtained from studies nucleosome core particles in retrospect have provided only glimpses of the complex functions of the core histone N-termini in their natural chromatin environment.

In this regard, recent incontrovertible evidence indicates that one of the primary functions of the N-termini in chromatin is to specifically interact with nonhistone regulatory proteins and multiprotein complexes. The first such proteins identified were SIR3, SIR4, and TUP1 from yeast (reviewed in refs *15*−*17*). The former two proteins are involved in establishing stable repression of the silent mating loci and telomeric chromatin, while the latter functions in α2-dependent gene repression. In both cases, in vivo analyses of the chromatin of yeast strains containing site-specific core histone mutants provided the first indication that the N-termini were functioning through protein−protein interactions (*15*−*19*). Subsequent in vitro experiments have confirmed direct interactions between the H3 and H4 N-termini and SIR3/SIR4 (*20, 21*) and TUP1 (22). Several chromatin remodeling complexes also have been shown to bind nucleosomes and nucleosomal arrays in part through interactions with the core histone N-termini, including the *Drosophila* NURF and yeast SWI/SNF and RSC complexes (*23, 24*). The yeast nucleosome assembly protein yNAP-1 requires the N-termini both to form complexes with core histones and to assemble regularly spaced nucleosomal arrays (*25*). Finally, it has recently been demonstrated that the HMG-14 domain responsible for chromatin unfolding specifically interacts with the N-terminus of histone H3 (*26*). The significance of these observations will be discussed further in the last section of the article.

In addition to interacting with specific chromatin associated proteins, the core histone N-termini also are essential for condensation of nucleosomal arrays and chromatin. While it has been known for many years that the core histone N-termini contribute to the compaction of the chromatin fiber (ref *27*; reviewed in refs *4* and *9*−*11*), only recently has the mechanistic complexity of N-termini involvement in chromatin condensation been documented. With it has come the realization that the molecular mechanisms through which the N-termini mediate chromatin condensation are fundamentally different than the DNA charge neutralization-based scheme that was first proposed (*27*). Progress in this area has emerged as a direct consequence of the availability of defined nucleosomal array and chromatin model systems that are assembled in vitro from pure histone and DNA components (*28*). Initial work with these systems revealed that nucleosomal arrays reconstituted with intact histone octamers exist in equilibrium between unfolded, moderately folded, and extensively folded conformational states (*29, 30*). The moderately folded conformation is an irregular open helical structure (*31*) that results from close approach of adjacent nucleosomes (*32*). Longer range internucleosomal interactions in cis lead to formation of the extensively folded conformation, which is as compact as a canonical 30 nm fiber (*30*). Importantly, linker histones stabilize the moderately and extensively folded conformational states of nucleosomal arrays, but do not otherwise alter the intrinsic folding pathway observed when linker histones are absent (*33*). Nucleosomal arrays condense further by forming oligomeric species through a process that is reversible and cooperative (*11, 34*).

Evidence that the core histone N-termini are required for all aspects of nucleosomal array condensation comes from studies showing that arrays assembled from trypsinized histone octamers lacking their N-terminal domains can neither fold (*13, 35*) nor oligomerize (*34*) in physiological salt. Characterization of nucleosomal arrays assembled from histone octamers that specifically lacked only the H3/H4 or H2A/H2B N-termini (*36, 37*) or were differentially acetylated (*38*) have revealed the mechanistic complexity of N-termini involvement in array condensation and also support the concept that N-termini function in large part through protein−protein interactions. The selective proteolysis experiments showed that under physiological ionic conditions only the H3/H4 N-termini are required to mediate formation of the moderately folded conformation and that high concentrations of divalent cations could replace this function. In contrast, although both the H2A/H2B and H3/H4 N-termini are required to form extensively folded nucleosomal arrays while oligomerization can occur in the presence of either the H3/H4 or H2A/H2B N-termini alone, in the absence of the N-termini, neither of these transitions can be induced by high salt concentrations (*36*). These results indicate that different molecular mechanisms are involved in the processes through which the N-termini mediate both short and long-range nucleosome−nucleosome interactions in cis, as well as the interfiber interactions, leading to oligomerization.

Importantly, the inability of high salt to induce higher order folding and oligomerization of trypsinized nucleosomal

arrays also demonstrates that the N-termini facilitate array condensation through mechanisms other than simple protein−DNA interactions (*36*). This conclusion has been confirmed by studies of differentially acetylated nucleosomal arrays, which showed that an unexpectedly low threshold level of acetylation (equivalent to neutralization of only ∼10% of the total positive charges in the N-termini of each nucleosome) was able to destabilize higher order folding to the same extent as proteolytic removal of all the N-termini (*38*). Given their known ability to bind chromatin-associated proteins, the simplest explanation for the nonelectrostatic functions of the N-termini in chromatin condensation is that they also involve protein−protein interactions. Likely candidates include internucleosomal N-termini interactions (*36, 38*) and (or) interaction of one the N-termini with exposed histone fold domains of other nucleosomes (*2*).

Are the Core Histone N-Termini Structured When Participating in Macromolecular Interactions? Considering that the N-termini are capable of participating in multiple types of protein−protein interactions in chromatin, it is natural to assume that these domains possess specific secondary structure. Grunstein and colleagues (*39*) have showed using helical wheel alignment that a key functional region of the H4 N-terminus in principle can form an amphipathic α-helix. In addition, prediction algorithms uniformly confirm that each of the N-termini possesses the propensity to form one or more significant stretches of secondary structure (Figure 1B). Thus, while the N-termini are unstructured coils in solution when studied as free histones and isolated peptides (*9, 10*), there is now compelling evidence that the N-termini adopt specific secondary structure when participating in macromolecular interactions.

Using circular dichroism, Barnéres et al. (*40*) have demonstrated that the H3 and H4 N-termini are ∼50% α-helical when bound to the nucleosomal DNA of nucleosome core particles. Due to the phenomenon of N-termini rearrangement during nucleosome core particle isolation, the relevance of these observations to N-termini function in chromatin remains to be established (see above). Nevertheless, these results provide direct biochemical evidence for interaction-induced secondary structure formation of the N-termini in vitro. Studies of specific yeast N-termini mutants have provided strong indirect evidence that the N-termini are structured while interacting with chromatin associated proteins in vivo and in vitro. Amino acids 1−20 of the H4 N-terminus are required both for repression of the silent mating loci and telomeric silencing in *S. cerevisiae* in vivo (*18, 41, 42*) and for binding of the H4 N-termini to SIR3 and SIR4 in vitro (*20*), and this same stretch of amino acids is predicted to have the ability to form an α-helix (Figure 1B). Similarly, amino acids 21−28 of the H3 N-terminus are required for H3-TUP1 interactions in vitro (*22*) and also have significant propensity to form an α-helix (Figure 1B). The observation that the histone acetyltransferase Gcn5p triacetylates the H3 N-terminus of free histones but only monoacetylates the H3 N-termini of nucleosome core particles and nucleosomal arrays (*43*) further suggests that the N-termini are structured when complexed with other macromolecules. Finally, Roth and co-workers have recently concluded that the biological effects of Gcn5p-dependent acetylation in vivo are mediated through disruption of structured domains in the N-termini rather than through neutralization of positive charge (*44*). Thus, not only do the N-termini appear to be structured when participating in macromolecular interactions, but recent evidence suggests that the function of acetylation may be to disrupt these secondary structure motifs (*22, 38, 44*). It should be noted that both interaction-induced secondary structure formation (see refs *45*−*49* and references therein) and acetylation-dependent disruption of secondary structure (*50, 51*) have been demonstrated previously in other systems.

Modulation of Global Chromosomal Domain Structure and Function by the Core Histone N-Termini. If alteration of the stability of the nucleosome per se is considered to be the primary mechanism for regulating the function of chromatin, as is commonly proposed throughout both the chromatin and gene expression literature (reviewed in refs *4* and *17*), it is not apparent why the N-termini should have evolved dual essential functions relating to higher order chromatin folding and interaction with chromatin-associated regulatory proteins. An alternative concept is that a primary mechanism for regulating chromatin function involves modulation of the higher order structure of discrete chromosomal domains consisting of ≥2−3 nucleosomes as a consequence N-termini interactions with nonhistone proteins.

The bulk of the chromatin fibers in the nucleus are complexed with linker histones and exist as a highly condensed "30 nm" fiber (*52*). Given that linker histones do not interfere with essential structural roles of the core histone N-termini in chromatin folding (*33*), in principle, both the structure and function of any given specific region of the chromosomal fiber can easily be "remodeled" by chromatin associated proteins that specifically bind the N-termini. When in the vicinity of the chromatin fiber, such proteins will be in direct competition with the various internucleosomal N-termini interactions required for chromatin folding. Under conditions where the regulatory protein(s) bind to the chromatin fiber, a subset of the N-termini will be recruited away from their internucleosomal interactions and consequently will no longer be able to mediate chromatin condensation. This necessarily will cause local reorganization of the fiber, where the condensed state of the remodeled region will simultaneously be influenced by the properties of the specific chromatin associated protein(s) and the types and extent of remaining internucleosomal N-termini interactions. Depending on the specific regulatory proteins and core histone N-termini involved, the bulk 30 nm chromatin fiber could be molded into essentially an unlimited number of specific structural and functional states. This includes everything from decondensed fibers with greatly enhanced biological activity (*38*) to extensively condensed heterochromatin encompassing genes that are permanently silenced. In this fashion, we envision that both the structure and function of any given region of a chromosomal fiber will be dictated largely by the combination of the core histone N-termini and their specific interaction partners in that region of the fiber. An additional attractive feature of this model is that because the chromatin fiber can in principle be converted into a decondensed active state through strategic interaction of regulatory protein(s) with the core histone N-termini, a priori, there is no need for widespread dissociation of linker histones as well (*53, 54*).

REFERENCES

1. Arents, G., Burlingame, R. W., Wang, B. W., Love, W. E., and Moudrianakis, E. N. (1991) *Proc. Natl. Acad. Sc.i U.S.A. 88*, 10148−10152.
2. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) *Nature 389*, 251−260.
3. van Holde, K. E. (1988) *Chromatin*, Springer-Verlag, New York.
4. Wolffe, A. P. (1998) *Chromatin: Structure & Function*, 3rd ed., Academic Press, New York.
5. Smith, M. M., and Andreasson, O. S. (1983) *J. Mol. Biol 169*, 633−690.
6. Allfrey, V., Faulkner, R. M., and Mirsky, A. E. (1964) *Proc. Natl. Acad. Sci. U.S.A. 51*, 786−794.
7. Wade, P. A., and Wolffe, A. P. (1997) *Curr. Biol. 7*, R82−R84.
8. Pazin, M. J., and Kadonaga, J. T. (1997) *Cell 89*, 325−328.
9. Bohm, L., and Crane-Robinson, C. (1984) *Biosci. Rep. 4*, 365−386.
10. Hansen, J. C. (1997) *Chemtracts: Biochem. Mol. Biol. 11*, 56−69.
11. Fletcher, T. M., and Hansen, J. C. (1996) *Crit. Rev. Euk. Gene Exp. 6*, 149−188.
12. Usachenko, S. I., Bavykin, S. G., Gavin, I. M., and Bradbury, E. M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 6845−6849.
13. Fletcher, T. M., and Hansen, J. C. (1995) *J. Biol. Chem. 270*, 25359−25362.
14. Lee, K. M., and Hayes, J. J. (1998) *Biochemistry 37*, 8622−8628
15. Loidl, P. (1994) *Chromosoma 103*, 441−449.
16. Roth, S. (1995) *Curr. Opin. Genet. Dev. 5*, 168−173.
17. Kingston, R. E., Bunker, C. A., and Imbalzano, A. N. (1996) *Genes Dev. 10*, 905−920.
18. Johnson, L. M., Fisher-Adams, G., and Grunstein, M. (1992) *EMBO J. 11*, 2201−2209.
19. Roth, S. Y., Shimizu, M., Johnson, L., Grunstein, M., and Simpson, R. T. (1992) *Genes Dev. 6*, 411−425.
20. Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M., and Grunstein, M.(1995) *Cell 80*, 583−592.
21. Hecht, A., Strahl-Bolsinger, S., and Grunstein, M. (1996) *Nature 383*, 92−96.
22. Edmondson, D. G., Smith, M.M., and Roth, S. Y. (1996) *Genes Dev. 10*, 1247−1259.
23. Georgel, P. T., Tsukiyama, T., and Wu, C. (1997) *EMBO J. 16*, 4717−4726.
24. Logie, C., Tse, C., Hansen, J. C., and Peterson, C. (submitted for publication).
25. McQuibban, G. A., Commisso-Cappelli, C. N., and Lewis, P. N. (1998) *J. Biol. Chem. 273*, 6582−6590.
26. Treichmann, L., Martin, B., and Bustin, M. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 5468−5473.
27. Allan, J., Harborne, N., Rau, D. C., and Gould, H.(1982) *J. Cell Biol. 93*, 285−297.
28. Simpson, R. T., Thoma, F., and Brubaker, J. M. (1985) *Cell 42*, 799−808.
29. Hansen J. C., Ausio J., Stanik V. H., and van Holde K. E (1989) *Biochemistry 28*, 9129−9136.
30. Schwarz, P. M., and Hansen, J. C. (1994) *J. Biol. Chem. 269*, 16284−16289.
31. Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Hoster, A J., and Woodcock, C. W. (1998) *Proc. Natl. Acad. Sci. U.S.A.* (in press).
32. Fletcher, T. M., Serwer, P., and Hansen, J. C. (1994) *Biochemistry 33*, 10859−10863.
33. Carruthers, L. M., Bednar, J., Woodcock, C. W., and Hansen, J. C. (1998) *Biochemistry*, 37, 14776−14787.
34. Schwarz, P. M., Felthauser, A., Fletcher, T. M., and Hansen, J. C. (1996) *Biochemistry 35*, 4009−4015.
35. Garcia-Ramirez, M., Dong, F., and Ausio, J. (1992) *J. Biol. Chem. 267*, 19587−19595.
36. Tse, C., and Hansen, J. C. (1997) *Biochemistry 36*, 11381-11388.
37. Moore, S. C., and Ausio, J. (1997) *Biochem. Biophys. Res. Commun. 230*, 136−139.
38. Tse, C., Sera, T., Wolffe, A. P., and Hansen, J. C. (1998) *Mol. Cell. Biol. 18*, 4629−4638.
39. Johnson, L. M., Fisher-Adams, G., and Grunstein, M. (1992) *EMBO J. 11*, 2201−2209.
40. Baneres, J.-L., Martin, A., and Parello, J. (1997) *J. Mol. Biol 273*, 503−508.
41. Johnson, L. M., Kayne, P. S., Kahn, E. S., and Grunstein, M. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6286−6290.
42. Aparicio, O. M., Billington, B. L., and Gottschling, D. E. (1991) *Cell 66*, 1279−1287.
43. Tse, C., Georgieva, E. I., Ruiz-Garcia, A. B., Sendra, R., and Hansen, J. C. (1998) *J. Biol. Chem. 273*, 32388−32392.
44. Zhang, W., Bone, J. R., Edmondson, D. G., Turner, B. M., and Roth, S. Y. (1998) *EMBO J. 17*, 3155−3167.
45. Padmanabhan, S., Capp, M. W., Anderson, C. F., and Record, M. T., Jr. (1997) *Biochemistry 36*, 5193−5206.
46. Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) *Science 270*, 262−269.
47. Daughdrill, G. W., Chadsey, M. S., Karlinsey, J. E., Hughes, K. T., and Dahlquist F. W. (1997) *Nat. Struct. Biol. 4*, 285−291.
48. Fletcher, C. M., and Wagner, G. (1998) *Protein Sci. 7*, 1639−1642.
49. Karantza, V., Freire, E., and Moudrianakis, E, N. (1996) *Biochemistry 35*, 2037−2046.
50. Xu, X., Cooper, L. G., Dimario, P. J., and Nelson, J. W. (1994) *Biopolymers 35*, 93−102.
51. Batra, P. P., Roebuck, M. A., and Uetrecht, D. (1990) *J. Protein Chem. 9*, 37−44.
52. Woodcock, C. L., Frado, L. L., and Ratner, J. B. (1984) *J. Cell Biol. 99*, 42−52.
53. Ericsson, C., Grossbach, U., Björkroth, B., and Daneholt, B. (1990) *Cell 60*, 73−83.
54. Kamakaka, R. T., and Thomas, J. O. (1990) *EMBO J. 9*, 3997−4006.
55. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol. 120*, 97−120.
56. Deleage, G., and Roux, B. (1987) *Protein Eng. 1*, 289−294.
57. Gibrat, J. F., Garnier, J., and Robson, B. (1987) *J. Mol. Biol. 198*, 425−443.
58. Levin, J. M., Robson, B., and Garnier, J. (1986) *FEBS Lett. 205*, 303−308.

BI982409V