

Chromatin Condensing Functions of the Linker Histone C-Terminal Domain Are Mediated by Specific Amino Acid Composition and Intrinsic Protein Disorder[†]

Xu Lu,^{‡,§} Barbara Hamkalo,^{||} Missag H. Parseghian,[⊥] and Jeffrey C. Hansen^{*,‡}

Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado 80523-1870, Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, California 92697, and Department of Research and Development, Peregrine Pharmaceuticals, 14272 Franklin Avenue, Tustin, California 92780

Received August 28, 2008; Revised Manuscript Received November 17, 2008

ABSTRACT: Linker histones bind to the nucleosomes and linker DNA of chromatin fibers, causing changes in linker DNA structure and stabilization of higher order folded and oligomeric chromatin structures. Linker histones affect chromatin structure acting primarily through their ~100-residue C-terminal domain (CTD). We have previously shown that the ability of the linker histone H1^o to alter chromatin structure was localized to two discontinuous 24-/25-residue CTD regions (Lu, X., and Hansen, J. C. (2004) *J. Biol. Chem.* 279, 8701–8707). To determine the biochemical basis for these results, we have characterized chromatin model systems assembled with endogenous mouse somatic H1 isoforms or recombinant H1^o CTD mutants in which the primary sequence has been scrambled, the amino acid composition mutated, or the location of various CTD regions swapped. Our results indicate that specific amino acid composition plays a fundamental role in molecular recognition and function by the H1 CTD. Additionally, these experiments support a new molecular model for CTD function and provide a biochemical basis for the redundancy observed in H1 isoform knockout experiments in vivo.

Linker histones (e.g., H1 and H5) are chromatin architectural proteins found in all eukaryotes (1, 2). They are abundant, with a stoichiometry of ~0.8 total linker histones per nucleosome in most tissues (ref 3 and references therein). Linker histones are modularly structured proteins that have an ~35-residue unstructured N-terminal domain (NTD),¹ a central globular winged helix domain, and an ~100-residue unstructured C-terminal domain (CTD) (4). Linker histones bind to chromatin fibers through interaction of the globular domain with nucleosomal sites(s) (1, 2, 5) and the CTD with linker DNA (6, 7). Higher eukaryotes have at least six somatic linker histone isoforms, which differ primarily in their CTD primary sequences (1, 8). The H1 isoform CTDs do, however, share a very similar and characteristic amino acid composition (9). At the molecular level, little is known about the actions of the isoforms. Linker histones are multifunctional, with roles in chromatin condensation (1, 2, 10, 11), nucleosome spacing (12, 13), specific gene expression (ref 13 and references therein), DNA methylation (13), and other nuclear processes. In addition to chromatin, linker histones bind to many nuclear proteins, e.g., DFF40/CAD (14), BAF (15), and Pur α /YB1 (16). Thus, a key unanswered question related to linker histone function is how these small, simply constructed proteins can specifically

recognize and interact with so many different macromolecular partners.

The relationships between linker histones, nucleosomal arrays, and chromatin fiber structure are well documented (1, 2, 10). In vitro, nucleosomal arrays are in salt-dependent equilibrium between unfolded, folded, and oligomeric conformational states (10, 11). Binding of linker histones to nucleosomal arrays affects chromatin structure in at least three distinct ways: First, the linker DNA between nucleosomes assumes an apposed stem/loop motif (7, 17, 18). Second, the folded conformational states of linker histone-bound chromatin fibers are stabilized relative to the folding seen with nucleosomal arrays alone; i.e., linker histones shift the equilibrium in favor of the folded chromatin fiber structures. Finally, Mg²⁺-dependent chromatin fiber oligomerization is also stabilized. Linker histones that lack their long unstructured CTD can bind to chromatin fibers but cannot mediate formation of apposed linker DNA segments or stabilization of folded or oligomeric conformational states (7, 19). Thus, the H1 CTD is essential for maintaining the structure and stability of chromatin fibers. Recently, CTD truncation studies showed that the abilities of the CTD to alter linker DNA structure and affect chromatin fiber folding and oligomerization were localized to two discontinuous 24-/25-residue regions (H1^o residues 98–122 and 147–170) (7). While these results ruled out one obvious model in which the entire CTD is needed to bind to DNA and neutralize negative charge, the molecular basis for this observation remains unresolved. One possibility is that each of these two regions contains structured primary sequence motifs that mediate function, e.g., stable α -helices. However, an alterna-

[†] This work was supported by NIH Grant GM45916 to J.C.H.

* To whom correspondence should be addressed. Tel: 970-491-5440. Fax: 970-491-0494. E-mail: jeffrey.c.hansen@colostate.edu.

[‡] Colorado State University.

[§] Current address: Van Andel Institute, 333 Bostwick Ave. N.E., Grand Rapids, MI 49503.

^{||} University of California, Irvine.

[⊥] Peregrine Pharmaceuticals.

¹ Abbreviations: NTD, N-terminal domain; CTD, C-terminal domain.

tive possibility is that the action of the two functional CTD regions is linked to intrinsic protein disorder (9).

The concept of intrinsic protein disorder has been explored extensively in recent years (see ref 20 and references therein). Intrinsically disordered protein domains mostly or completely lack classical secondary structure but undergo disorder to order transitions concomitant with binding to their macromolecular targets (9, 21, 22). While much progress has been made predicting the prevalence of intrinsically disordered protein domains in eukaryotic proteomes (23), there is a relative paucity of experimental biochemical data that address the molecular mechanisms that underlie how intrinsically disordered regions function. Several groups have speculated that intrinsically disordered domains contain short primary sequence elements imbedded within the domain (24–26). In addition, interesting attributes related to amino acid composition have been observed, such as reduced levels of hydrophobic residues and increased levels of charged residues (21, 27). We have noted that the amino acid composition of the somatic linker histone CTDs is distinctive; it is very similar between isoforms but differs from that of other intrinsically disordered protein regions such as yeast prion domains (9). On the basis of these observations, we have hypothesized that the chromatin-condensing functions of the CTD may be mediated by its specific amino acid composition (9). A potential role for amino acid composition is intriguing because dogma holds that protein function is linked to primary amino acid sequence rather than composition. To examine the role of primary sequence and specific amino acid composition, and to better understand how linker histone isoforms affect chromatin structure *in vitro*, here we have used sedimentation velocity and differential centrifugation to characterize the linker DNA structure and folded and oligomeric states of defined chromatin model systems bound to endogenous mouse somatic H1 isoforms and specific recombinant H1 CTD mutants. Our results support a mechanism for H1 CTD function that is based on its amino acid composition and intrinsic protein disorder rather than classical modes of molecular recognition. Further, these data provide a molecular explanation for how the linker histone isoform CTDs can have redundant functions in chromatin condensation *in vivo*.

EXPERIMENTAL PROCEDURES

Materials. 208-12 DNA (28–30) and chicken erythrocyte core histone octamers (31) were purified as described. All mouse linker histone H1° mutants were created based on the wild-type H1° plasmid, pET-H1°-11d (32), using standard methods (33). Plasmids were sequenced to confirm the desired mutations were present. Wild-type and mutant H1° proteins were expressed and purified in *E. coli* as described (7). Mouse endogenous linker histone variants were prepared as described (14). The following procedure was used to select the sequences of the two H1° mutants containing randomized sequences in their subdomain 1. Ten randomized sequences were first generated at Web site, http://www.cellbiol.com/scripts/randomizer/sequence_randomizer.html, using wild-type subdomain 1 as input. Program Excel (Microsoft, Seattle, WA) was then used to generate two random numbers between 1 and 10, and the two sequences corresponding to

Table 1: Primary Sequences of Linker Histone H1° CTD Regions 1–4 and Specific CTD Mutants^a

region	residues	sequence
4	171–194	SKPKKAKTVKPKAKSSAKRASKKK
3	147–170	KKKPAATPKKAKKPKVVKVVKPKA
2	123–146	AAKPKKAASKAPSKKPKATPVKKA
1	98–122	GDEPKRSVAFKKT KKEV KKVATPKK
RanS1#1		KTRKVDKPEPKKFKGEAVSKKAVTK
RanS1#2		SVKGAVEKVKAEKPKDKPTKTKFKRKK
S1TD		GDEPKRS N AFK KPKKEN KK N ATPKK

^a The amino acid differences between CTD region 1 and the S1TD mutant are shown in bold underline.

these two numbers were selected and named H1°RanS1#1 and H1°RanS1#2 (see Table 1 for sequences).

Reconstitution of Nucleosomal Arrays and Linker Histone-Containing Chromatin Fibers. Nucleosomal arrays in which approximately 50% of the sample was saturated with nucleosomes were assembled from 208-12 DNA and chicken histone octamers as described previously (7). Briefly, histone octamers and DNA were mixed in 10 mM Tris and 0.25 mM EDTA, pH 7.8 (TE), containing 2 M NaCl at a molar ratio of ~1.1. The mixture was subsequently dialyzed against TE buffer containing 1 M, 0.75 M, and 2.5 mM NaCl, respectively, for at least 6 h each. The mixture was then dialyzed against fresh TE buffer containing 2.5 mM NaCl (TEN) overnight. To assemble linker histone-containing chromatin model systems, the parent nucleosomal arrays were first brought to 50 mM NaCl by addition of appropriate amounts of 0.5 M NaCl. Linker histone was added to the solution at a ratio of ~1.3 mol linker histone/mol of 208 bp repeat, and the solution was dialyzed against TEN overnight (7, 18). The integrity of the reconstituted nucleosomal arrays and linker histone-containing chromatin fibers was confirmed by sedimentation velocity in TEN.

Sedimentation Velocity. Sedimentation velocity experiments were performed in a Beckman XLA or XLI ultracentrifuge equipped with scanner optics as described (7, 18, 34). Scans were analyzed by the modified method of van Holde and Weischet (35, 36) using Ultrascan data analysis software (35). This method yields the integral distribution of sedimentation coefficients of the chromatin sample, $G(s)$, plotted as boundary fraction versus the diffusion-corrected sedimentation coefficient (35, 36).

Differential Centrifugation. The assay for chromatin fiber oligomerization was performed as described (7, 18). Chromatin model systems (A_{260} 1.2–1.4) were mixed with an equal volume of $2 \times \text{MgCl}_2$ solution to achieve the desired final salt concentration. After incubation for 5 min at room temperature, samples were centrifuged for 5 min at 16000g in a microcentrifuge at room temperature. The supernatant was then removed and the A_{260} measured. The percentage of the sample remaining in the supernatant was plotted against MgCl_2 to yield the oligomerization curve (7, 18). Each data point in the plot reflected the mean of two to five experiments. The oligomerization curve was used to obtain the Mg^{50} , defined as the MgCl_2 concentration at which 50% of the sample had oligomerized and pelleted during centrifugation.

RESULTS

When saturated with 12 nucleosomes per DNA, the model nucleosomal arrays used in our experiments sediment at

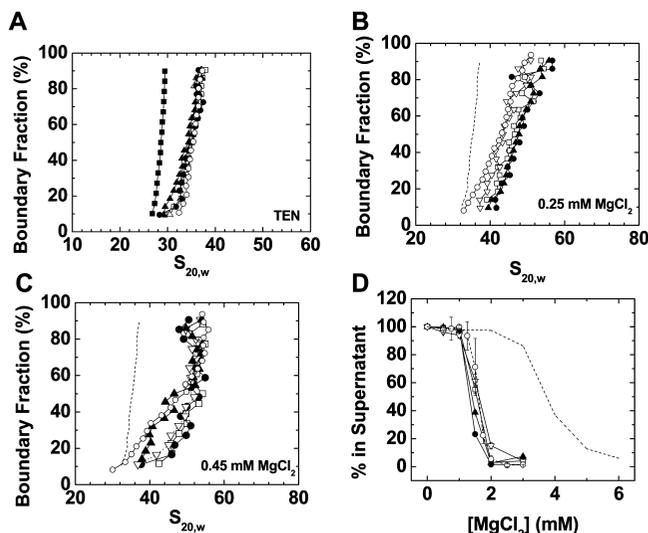


FIGURE 1: Characterization of model chromatin fibers bound to endogenous mouse somatic linker histone isoforms. (A) Sedimentation coefficient distributions in low salt TEN buffer of 208-12 nucleosomal arrays (dashed line) assembled with endogenous mouse H1^o (○), H1^{S-1} (●), H1^{S-2} (□), H1^{S-3} (▲), and H1^{S-4} (▽). (B) $G(s)$ plots of 208-12 chromatin fibers bound to wild-type H1^o (○), H1^{S-1} (●), H1^{S-2} (□), H1^{S-3} (▲), or H1^{S-4} (▽) in 0.25 mM MgCl₂. The $G(s)$ plot of H1^o-bound chromatin fibers in low salt (dashed line; data from panel B) is shown for reference. (C) $G(s)$ plots of 208-12 fibers bound to wild-type H1^o (○), H1^{S-1} (●), H1^{S-2} (□), H1^{S-3} (▲), or H1^{S-4} (▽) in 0.45 mM MgCl₂. (D) Oligomerization as a function of MgCl₂. Shown are data for 208-12 nucleosomal arrays (dashed line) and 208-12 chromatin fibers bound to wild-type H1^o (○), H1^{S-1} (●), H1^{S-2} (□), H1^{S-3} (▲), or H1^{S-4} (▽). All plots are representative of results seen with three to four different chromatin preparations.

28–29 S in Mg²⁺-free TEN buffer (10, 18). As a control, the sample nucleosomal arrays used in Figures 1–5 were characterized by sedimentation velocity in TEN buffer and the boundaries analyzed to yield the diffusion-corrected integral distribution of sedimentation coefficients, $G(s)$ (35, 36). The data shown in Figure 1A indicated that roughly half of the sample sedimented at 28–29 S (boundary fraction 50–90%) and hence was saturated with 12 nucleosomes per DNA. The remainder of the sample sedimented at 26–27 S and thus contained 10–11 nucleosomes per DNA; i.e., they were slightly “subsaturated”. This is typical of array preparations used in previous studies (7, 18, 29, 31).

Effects of Endogenous Mouse Linker Histone Isoforms on Chromatin Fiber Structure. The CTDs of the somatic linker histone isoforms have divergent primary sequences (1, 8) but a very similar amino acid composition (9). Hence, to test the relative importance of primary sequence versus composition, we first characterized model systems assembled with five purified endogenous mouse H1 isoforms (H1^o, H1^{S-1}, H1^{S-2}, H1^{S-3}, H1^{S-4}; see Supporting Information Figure 1 for correlation of isoform nomenclatures). The experiments described below in each case assay three H1 CTD-mediated functions related to chromatin condensation: apposition of linker DNA structure, stabilization of the intrinsic folding of nucleosomal arrays (4, 7, 18), and stabilization of nucleosomal array oligomerization (7, 18).

Previous studies have shown that saturated 12-mer nucleosomal arrays bound to one linker histone per nucleosome sediment at ~36 S in TEN (7, 18). This increase in sedimentation coefficient from 29 to ~36 S is due to both

an increase in mass from linker histone binding to the nucleosomal arrays and the formation of the apposed stem–loop motif, the latter of which shortens the linker DNA and thus decreases the frictional coefficient of the fibers (7, 18). When mixed with the nucleosomal arrays at a ratio of 1.3 mol of linker histone/mol of nucleosome, the sedimentation coefficient distributions of all isoform sample fibers increased to 30–38 S. The $G(s)$ profiles in Figure 1B indicate that approximately 40% of each isoform (boundary fraction 50–90%) sedimented at 36–38 S and hence consisted of chromatin fibers saturated with both one nucleosome per 208 bp repeat and one H1 per nucleosome (7, 18). The remainder of each sample was subsaturated and lacked stoichiometric amounts of core histones, linker histones, or both. The data in Figure 1A demonstrate that all isoforms could bind to linker DNA and are consistent with each isoform equally inducing formation of the apposed linker DNA stem–loop motif.

Our observation of similar relative binding affinities of the somatic isoforms differs from that reported previously by several investigators (37, 38). However, these differences are likely due to the experimental conditions chosen to study binding. Talasz et al. (38) used gel electrophoretic mobility shift assays and nanomolar protein and nucleosome concentrations to conclude that there were ~8-fold differences (2–16 nM) in binding affinities of the somatic isoforms for 210 bp mononucleosomes. Suau and colleagues (37) found 20-fold differences in relative binding affinities. In contrast to these studies, binding to nucleosomal arrays in our work occurred at micromolar concentrations of linker histones and nucleosomal arrays and in low salt TEN buffer, i.e., conditions designed to saturate the model nucleosomal arrays with stably bound H1. Under these conditions we would not expect to see relatively small differences in binding affinities in the nanomolar range.

Chromatin folding is assayed by sedimentation velocity in the presence of MgCl₂ and is evidenced by an increase in sedimentation coefficient beyond 36 S (7, 10, 18, 39). Folding of 12-mer chromatin fibers is complete when a stable 55 S structure is achieved (7, 10, 18). To examine the effects of H1 proteins on stabilization of chromatin fiber folding, experiments were performed in 0.25 mM MgCl₂, which induces only partial folding, and 0.45 mM MgCl₂, which induces nearly complete folding of saturated chromatin fibers that contain stoichiometric amounts of bound linker histone. We purposely did not work at higher MgCl₂ concentrations to avoid induction of oligomerization. In 0.25 mM MgCl₂, small differences in the extent of folding were observed, with H1^{S-1} = H1^{S-2} = H1^{S-3} > H1^o = H1^{S-4} (Figure 1B). However, by 0.45 mM MgCl₂, all isoforms were capable of stabilizing the model chromatin fibers in 50–55 S structures, as indicated by the data between boundary fractions 50–90% of the $G(s)$ plots (Figure 1C). These data indicate that all H1 isoforms can stabilize chromatin fibers in extensively folded states. The differential effects of the isoforms in condensing saturated fibers in 0.25 mM MgCl₂ suggest that small isoform-dependent differences in the stability of the folded structures may exist.

Mg²⁺-dependent oligomerization is assayed by a differential centrifugation protocol that measures reversible, cooperative self-association of individual fibers (10, 40). Figure 1D shows the characteristic sigmoidal curves that

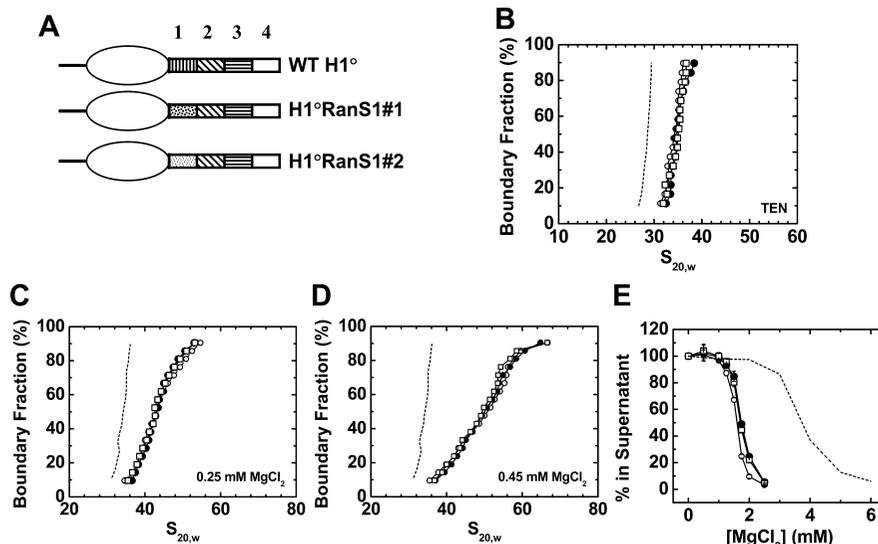


FIGURE 2: Randomization of the primary sequence of CTD region 1. (A) Schematic diagrams of wild-type H1^o, H1^oRanS1#1, and H1^oRanS1#2. (B) Sedimentation coefficient distributions in TEN buffer of 208-12 nucleosomal arrays (dashed line) assembled with wild-type H1^o (○), H1^oRanS1#1 (●), or H1^oRanS1#2 (□). (C) $G(s)$ plots of 208-12 chromatin fibers bound to wild-type H1^o (○), H1^oRanS1#1 (●), or H1^oRanS1#2 (□) in 0.25 mM MgCl₂. The $G(s)$ plot of H1^o-bound chromatin fibers in low salt (dashed line; data from Figure 1A) is shown for reference. (D) $G(s)$ plots of 208-12 fibers bound to wild-type H1^o (○), H1^oRanS1#1 (●), or H1^oRanS1#2 (□) in 0.45 mM MgCl₂. The $G(s)$ plot of H1^o-bound chromatin fibers in low salt (dashed line; data from Figure 1A) is shown for reference. (E) Oligomerization as a function of MgCl₂. Shown are data for 208-12 nucleosomal arrays (dashed line) and 208-12 chromatin fibers bound to wild-type H1^o (○), H1^oRanS1#1 (●), or H1^oRanS1#2 (□). All plots are representative of results seen with three to four different chromatin preparations.

were obtained for model chromatin fibers bound to each H1 isoform. Although very subtle differences in Mg⁵⁰ were observed, all isoforms led to essentially the same pronounced decrease in Mg⁵⁰ relative to the parent nucleosomal arrays.

When taken together, a major conclusion from the data in Figure 1 is that all three chromatin-condensing functions mediated by the H1 CTD are comparable for the different isoforms, despite divergent isoform CTD primary sequences. It follows from these results that amino acid composition may be a more important determinant of CTD action than the specific order of the amino acids.

Randomization of the Primary Sequence of Residues 98–122 (CTD Region 1) Does Not Abolish Function. To directly test the role of amino acid composition in CTD function, in the remainder of the studies we focused attention on the “functional” and “nonfunctional” CTD regions identified in previous CTD truncation experiments (7) (see Figure 2A). Residues 98–122 (CTD region 1) have been shown to be required for formation of the stem-loop linker DNA motif and stabilization of fiber folding and oligomerization. Residues 147–170 (CTD region 3) were found to assist in stabilizing folded chromatin fiber structures (7). Regions 2 (residues 123–146) and 4 (171–194) were nonfunctional; i.e., they could be deleted without any noticeable effects on H1 function (7). In the experiments described below, recombinant technology was used to scramble the primary sequence, mutate the amino acid composition, and switch the position of the functional and nonfunctional CTD regions of recombinant H1^o.

Figure 2A schematically illustrates two different mutants that were constructed in which the primary sequence of region 1 was randomized while the amino acid composition was kept constant. Table 1 shows that the two mutants have no primary sequence homology with wild-type region 1. Sedimentation velocity experiments in TEN indicated that both randomized region 1 mutants behaved the same at

binding to linker DNA and inducing stem-loop formation (Figure 2B). We also observed no differences in stabilization of folded chromatin fibers in 0.25 mM (Figure 2C) and 0.45 mM MgCl₂ (Figure 2D), and of Mg²⁺-dependent oligomerization (Figure 2E). The similar behavior of the two randomized mutants in Figure 2 also supports a mechanism of CTD function that does not involve primary sequence but is dependent on specific amino acid composition.

Mutation of the Amino Acid Composition of H1^o Residues 98–122 and 147–170 (CTD Regions 1 and 3) Affects Chromatin Folding. Given that scrambling the primary sequence did not abolish function, we attempted to disrupt H1 effects on chromatin structure by altering the specific amino acid composition of either CTD region 1 alone (H1^oS1TD) or CTD regions 1 and 3 together (H1^oS1S3TD) (Figure 3A). Three Val residues were replaced with Asn residues, which normally are absent from the linker histone CTD (9), and a Pro was inserted in place of a Thr (Table 1). Val is the only nonpolar amino acid present in significant quantities in the linker histone CTDs (9). The lysine content was kept constant to keep the high positive charge density of the mutants the same. Results indicated that both mutants behaved very similarly to wild type in terms of formation of the 36 S structure in low salt (Figure 3B) and salt-dependent oligomerization (Figure 3E). However, the H1^oS1TD mutant stabilized folded chromatin fiber structures to a lesser degree in both 0.25 and 0.45 mM MgCl₂, and the H1^oS1S3TD mutant was even less effective in both salts (Figure 3C,D). Specifically, in 0.45 mM MgCl₂, the mutant fibers folded roughly halfway between that of the wild-type fibers and H1-bound fibers completely lacking their C-terminal domains (Figure 3D). Thus, at least for the case of chromatin fiber folding, altering the amino acid content of CTD regions 1 and 3 led to reproducible disruption of CTD function. This partial disruption could be due to alteration of the amino acid composition per se or to point mutation(s)

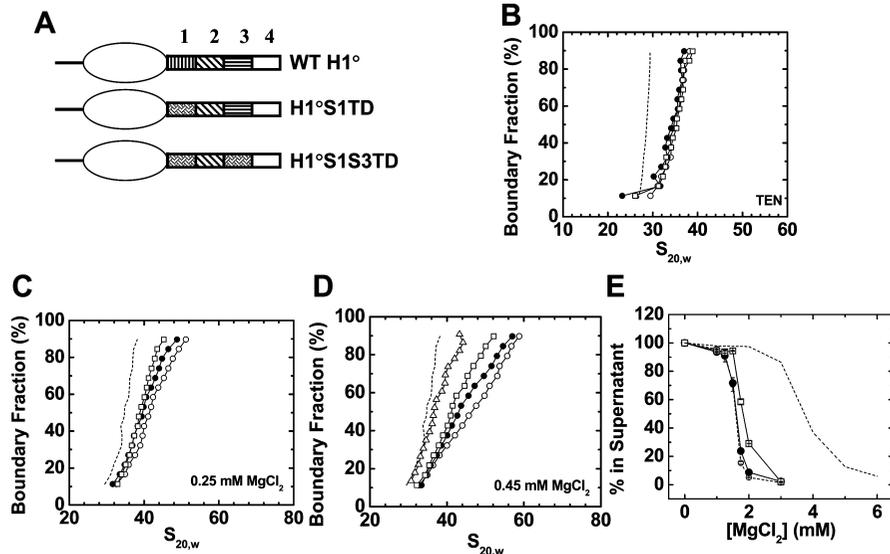


FIGURE 3: Mutagenesis of the amino acid composition of CTD region 1. (A) Schematic diagrams of wild-type H1^o, H1^oS1TD, and H1^oS1S3TD. (B) Sedimentation coefficient distributions in TEN buffer of 208-12 nucleosomal arrays (dashed line) assembled with wild-type H1^o (○), H1^oS1TD (●), or H1^oS1S3TD (□). (C) $G(s)$ plots of 208-12 chromatin fibers bound to wild-type H1^o (○), H1^oS1TD (●), or H1^oS1S3TD (□) in 0.25 mM MgCl₂. The $G(s)$ plot of H1^o-bound chromatin fibers in low salt (dashed line; data from Figure 1A) is shown for reference. (D) $G(s)$ plots of 208-12 fibers bound to wild-type H1^o (○), H1 CTD “tailless” (△), H1^oS1TD (●), or H1^oS1S3TD (□) in 0.45 mM MgCl₂. The $G(s)$ plot of H1^o-bound chromatin fibers in low salt (dashed line; data from Figure 1A) is shown for reference. (E) Oligomerization as a function of MgCl₂. Shown are data for 208-12 nucleosomal arrays (dashed line) and 208-12 chromatin fibers bound to wild-type H1^o (○), H1^oS1TD (●), or H1^oS1S3TD (□). All plots are representative of results seen with three to four different chromatin preparations.

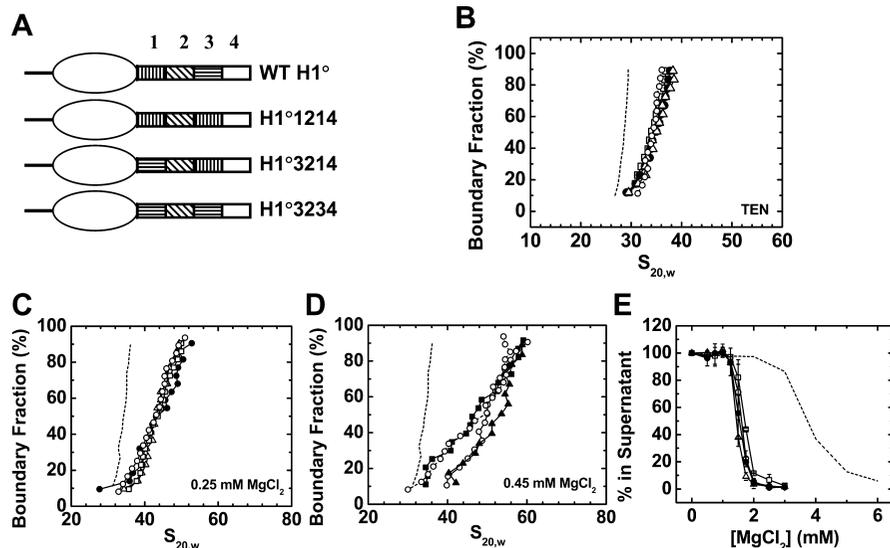


FIGURE 4: Exchanging and swapping H1^o CTD regions 1 and 3. (A) Schematic diagrams of wild-type H1^o, H1^o1214, H1^o3214, and H1^o3234. (B) Sedimentation coefficient distributions in TEN buffer of 208-12 nucleosomal arrays (■) and 208-12 chromatin fibers bound to wild-type H1^o (○), H1^o1214 (●), H1^o3214 (△), or H1^o3234 (□). (C) $G(s)$ plots of 208-12 fibers bound to wild-type H1^o (○), H1^o1214 (●), H1^o3214 (△), or H1^o3234 (□) in 0.25 mM MgCl₂. The $G(s)$ plot of H1^o-bound chromatin fibers in low salt (dashed line; data from Figure 1A) is shown for reference. (D) $G(s)$ plots of 208-12 fibers bound to wild-type H1^o (○), H1^o1214 (●), H1^o3214 (△), or H1^o3234 (□) in 0.45 mM MgCl₂. The $G(s)$ plot of H1^o-bound chromatin fibers in low salt (dashed line; data from Figure 1A) is shown for reference. (E) Oligomerization as a function of MgCl₂. Shown are data for 208-12 nucleosomal arrays (dashed line) assembled with wild-type H1^o (○), H1^o1214 (●), H1^o3214 (△), or H1^o3234 (□). All plots are representative of results seen with three to four different chromatin preparations.

that disrupt the ability of CTD region 1 to form a key secondary structure element upon binding to DNA (see Discussion).

All CTD Regions Are Functionally Interchangeable Given Proper Positioning Relative to the Globular Domain. In the final set of experiments, a series of mouse H1^o CTD mutants were constructed in which either the order or position relative to the globular domain of the various functional and nonfunctional CTD regions were altered (Figures 4A and

5A). The primary sequences of the four CTD regions are listed in Table 1 and show that there is little sequence conservation between the regions. Table 2 shows the amino acid composition of the four regions, which is much more similar. Relative to wild type, H1 proteins were constructed that either exchanged the positions of regions 1 and 3 (H1^o3214), replaced region 3 with a second copy of region 1 (H1^o1214), or replaced region 1 with a second copy of region 3 (H1^o3234) (Figure 4A). Because the positions of

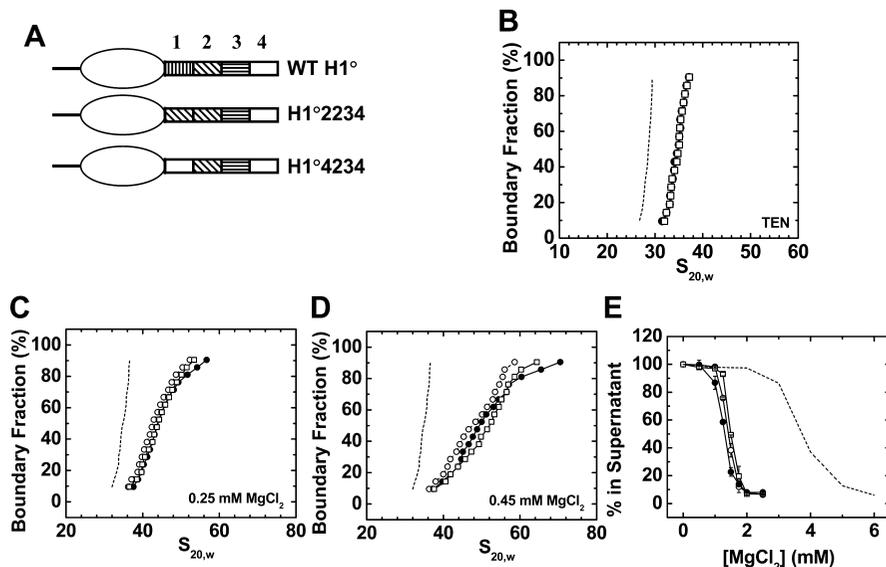


FIGURE 5: Replacement of H1° CTD region 1 with regions 2 or 4. (A) Schematic diagrams of wild-type H1°, H1°2234, and H1°4234. (B) Sedimentation coefficient distributions in low salt TEN buffer of 208-12 nucleosomal arrays assembled with wild-type H1° (○), H1°2234 (●), or H1°4234 (□). (C) Sedimentation velocity analysis. $G(s)$ plots of 208-12 chromatin fibers bound to wild-type H1° (○), H1°2234 (●), or H1°4234 (□) in 0.25 mM MgCl₂. The $G(s)$ plot of H1°-bound chromatin fibers in low salt (dashed line; data from Figure 1A) is shown for reference. (D) $G(s)$ plots of 208-12 fibers bound to wild-type H1° (○), H1°2234 (●), or H1°4234 (□) in 0.45 mM MgCl₂. The $G(s)$ plot of H1°-bound chromatin fibers in low salt (dashed line; data from Figure 1A) is shown for reference. (E) Oligomerization as a function of MgCl₂. Shown are data for 208-12 nucleosomal arrays (dashed line) and 208-12 chromatin fibers bound to wild-type H1° (○), H1°2234 (●), or H1°4234 (□). All plots are representative of results seen with three to four different chromatin preparations.

Table 2: Amino Acid Composition of the H1° Regions

region	Lys ^a	Ala	Pro	Gly	Arg	Glu	Asp	Asn	Gln	Ser	Thr	Cys	Met	His	Val	Ile	Leu	Phe	Trp	Tyr
1	36	8	8	4	4	8	4	0	0	4	8	0	0	0	12	0	0	4	0	0
2	38	29	17	0	0	0	0	0	0	8	4	0	0	0	4	0	0	0	0	0
3	46	17	17	0	0	0	0	0	0	0	4	0	0	0	17	0	0	0	0	0
4	46	17	8	0	4	0	0	0	0	17	4	0	0	0	4	0	0	0	0	0

^a Each amino acid residue is expressed as percent of the total composition.

key functional regions had been moved or replaced with a different region, one might expect the mutants to be nonfunctional. However, given the results from Figures 1–3, and the shared similarity in amino acid composition of the two functional regions, our prediction was that the mutants would behave the same as wild-type H1°.

Figure 4B shows that the $G(s)$ profiles in TEN of 208-12 model systems bound to wild-type H1°, and all three CTD region 1/3 swap mutants were identical within error. When the same samples were examined for folding in 0.25 mM (Figure 3C) and 0.45 mM (Figure 3D) MgCl₂, there was no appreciable difference in the $G(s)$ plots of the saturated fraction (boundary fraction = 50–90%) at both salt concentrations. Hence the stabilization of folding also was unaffected by the mutations. The differences in the subsaturated portion of the samples were most likely due to differences in the amount of bound H1. Figure 4E shows that all four samples exhibited nearly identical oligomerization behavior and Mg⁵⁰ values. Collectively, the data in Figure 1 indicate that subdomains 1 and 3 were interchangeable for nucleosomal array binding, linker DNA stem-loop motif formation, and stabilization of condensed chromatin fiber structures, despite having different primary sequences. An even more telling experiment examined mutants in which CTD region 1 was replaced with either nonfunctional CTD region 2 (H1°2234) or region 4 (H1°4234) (Figure 5A). These mutants bound to nucleosomal arrays and formed the ~36 S fiber structure the same as wild type (Figure 5B).

Similarly, the extent of folding in 0.25 mM (Figure 5C) and 0.45 mM MgCl₂ (Figure 5D) was nearly identical to wild type (if anything the mutants were slightly more folded at the higher salt concentration). As shown in Figure 5E, the oligomerization curves of both mutants also were essentially indistinguishable from wild type. Thus, both of the CTD regions identified as nonfunctional in truncation experiments could replace native CTD region 1 if moved to the appropriate position along the H1° polypeptide chain.

DISCUSSION

The H1 CTD Is an Intrinsically Disordered Domain. Intrinsically disordered regions of proteins lack secondary structure but nevertheless are functional (see refs 9, 20, 22, and 41 and references therein). In particular, they often are involved in binding and macromolecular recognition. Further, intrinsically disordered regions often undergo a disorder to order transition that is coupled to recognition of their binding partners (9, 22, 41). Early experiments showed that the H1 CTD is unstructured in solution (4, 42, 43). Despite lacking native structure, the H1 CTD can bind DNA (43–45), nucleosomal arrays (7, 18, 19), and specific proteins (14–16). When H1 CTD peptides bind to DNA, they assume α -helical structure (43, 44). By these criteria, the H1 CTD meets the definition of an intrinsically disordered protein domain.

Amino Acid Composition as a Determinant of Molecular Recognition and H1 CTD Function. In a previous study we characterized H1° mutants in which 24- or 25-residue

segments were successively deleted from the C-terminus of the CTD (7). We found that deletion of regions 1 (residues 98–122) and 3 (residues 147–170) together led to loss of apposed linker DNA structure and destabilization of condensed chromatin fibers, while regions 2 (residues 123–146) and 4 (residues 171–194) could be deleted without effect (7). At the time we termed regions 1–4 “subdomains”. A straightforward explanation for these results is that there are stable structured primary sequence element(s) in CTD regions 1 and 3 that mediate macromolecular recognition and function. In this case, scrambling the primary sequence should disrupt H1 function. However, in both mutants tested we observed that the primary sequence of CTD region 1 could be scrambled without influencing the ability of this CTD region to bind to linker DNA and stabilize condensed chromatin fibers (Figure 2). Thus, the chromatin-condensing functions of the H1 CTD do not appear to be dependent on CTD primary sequence.

One factor that was held constant in the scrambling experiments was amino acid composition. The amino acid composition of the H1 CTDs is distinctive and almost constant between isoforms. Approximately 40% of the residues are Lys and 20–30% Ala. The other residues are mainly Ser/Thr, Pro, Val, and some Gly (9). For all practical purposes, 13 of the common amino acids are absent from the isoform CTDs, including the bulky hydrophobic residues. In addition to the sequence scrambling experiments, three additional results support a role for specific amino acid composition as a mechanistic determinant of H1 CTD action. First, all somatic linker histone isoforms functioned nearly identically. The isoforms have divergent primary sequences (1, 8) but have maintained a very similar amino acid composition (9). Second, replacement of three Val residues and one Thr residue in CTD region 1 with Asn and Pro, which led to an altered amino acid composition depleted of nonpolar residues, partially disrupted H1 effects on chromatin fiber folding (Figure 3C,D). Finally, CTD regions 2, 3, and 4 all could replace functional region 1 if they were placed in the proper position along the polypeptide chain (Figures 4 and 5). None of the four regions has primary sequence homology (Table 1), but their amino acid composition is more similar (Table 2). Collectively, these results are consistent with a mechanism in which specific amino acid composition mediates H1 CTD function in linker DNA binding and chromatin fiber condensation.

Amino acid composition has recently been shown to determine the function of the prion domains of the yeast proteins, Ura1p and Sup35p (9, 46, 47). Interestingly, the amino acid composition of the prion domains is Asn and Gln rich and differs fundamentally from that of the linker histone CTDs. The prion domains are intrinsically disordered but form parallel in-register β -sheet structures concomitant with molecular recognition and amyloid fiber assembly (48). The prion domain results and the H1 CTD work reported here indicate that different specific amino acid compositions can dictate formation of different secondary structures when intrinsically disordered domains interact with their targets. While little is known about the protein chemistry that underlies the connections between amino acid composition, intrinsic protein disorder, and molecular recognition, the widespread prevalence of intrinsically disordered domains in eukaryotic proteomes (23, 49) suggests that mechanisms

of macromolecular recognition based on amino acid composition may be common.

A New Working Model for H1 CTD Action. Many events are involved in H1 regulation of chromatin structure in vitro, including (but not limited to) binding to chromatin fiber targets, formation of apposed linker DNA, stabilization of fiber folding, and stabilization of fiber oligomerization. Because of its importance in regulating genome structure and integrity, the molecular mechanism of H1 function has been studied for nearly 30 years. Based on its uniform high positive charge density (~ 4 Lys per 10 residues along all somatic H1 CTDs; see ref 50), the H1 CTD initially was proposed to function through an electrostatic mechanism in which the CTD was bound to linker DNA and screened negative charge, thereby facilitating close approach of nucleosomes (50, 51). Inherent in this proposal was that the entire CTD was engaged with linker DNA. However, as described extensively above, only two discontinuous regions (CTD regions 1 and 3 of 25 and 24 residues in length, respectively) are involved in mediating linker DNA apposition and stabilization of chromatin fiber condensation (7).

How does one explain this result given the functional interchangeability of the four CTD subdomains identified in the present studies (Figures 4 and 5)? In other words, if all regions can substitute for one another, why are some regions “functional” while others are “nonfunctional” in the intact protein? We hypothesize that this reflects the importance of the position of the CTD regions relative to another structural and functional reference point in the protein, in this case the centrally located globular domain. The globular domain binds to nucleosomes, although the details of this interaction remain unresolved (5). In our working model, binding of the globular domain helps to position CTD regions 1 and 3 to interact with linker DNA and possibly other chromatin components as well. The initial interaction can be purely electrostatic (7), but the specific amino acid composition of the CTD allows formation of one or more short α -helices in these regions upon DNA binding that mediate CTD function in chromatin. We note that there is one Pro per seven residues on average in the H1 CTDs and that Pro residues often are found at the beginning of α -helices (52), as well as being helix breakers. The finding that the amino acid composition mutants only affected chromatin folding (Figure 3) indicates that the three CTD-mediated functions may involve different specific disorder to order transitions. In this regard, several different α -helices may form in any given stretch of 24 CTD residues.

If only two discontinuous 24-residue CTD regions and a 73 residue long CTD are needed to stabilize condensed chromatin fibers (7), why are all the H1 isoform CTDs nearly 100 residues in length and maintain a consistent amino acid composition throughout the entire domain? The answer, at least in part, is that the H1 CTD can mediate interactions with specific proteins as well as chromatin. In the case of the apoptotic nuclease, DFF40/CAD, it has been shown that the mouse somatic H1 isoforms bound to and activated the enzyme equally well. Further, any CTD peptide greater than 47 residues in length was capable of activating the enzyme (14). However, for the H1-DFF40/CAD interaction, CTD regions 2, 3, and 4 (but not CTD region 1) were needed to mediate the protein–protein interaction in truncation experiments. On the basis of these results, we hypothesize that

H1-DFF40/CAD binding also is mediated by the specific amino acid composition of the H1 CTD but that a different specific pattern of α -helices form when the 72 C-terminal most CTD residues interact with the DFF40/CAD surface than when the CTD binds to chromatin. By acting through a mechanism involving intrinsic disorder and amino acid composition, the linker histone CTD is likely to be involved in competitive interactions with both the chromatin fiber and many other nuclear proteins. This “molecular plasticity” in binding may explain much of the multifunctionality of the H1 proteins.

Redundant Functions of the Linker Histone Isoforms in Chromatin Condensation. Skoultchi and colleagues have investigated linker histone isoform function in vivo by knocking out the H1 isoform genes singly (12), in pairs (12), and in triple combinations (13). They found that isoforms could be knocked out singly and in pairs without noticeable effects on chromatin structure or biological phenotype and that expression of the other isoforms compensated for the knocked out isoforms to maintain total wild-type H1 stoichiometry (3, 12). Their data indicated that the functions of the isoforms were redundant in vivo at the level of maintaining large-scale genome structure and organization. The results of our in vitro experiments in high Mg^{2+} showing similar condensation abilities of five of the mouse H1 isoforms (Figure 1D,E) support the in vivo data. We hypothesize that the characteristic amino acid composition of the H1 isoform CTDs leads to redundant functions of the H1 isoforms in chromatin condensation when the isoforms are stably bound to chromatin fibers and chromosomal domains, acting through the mechanisms described above. Simultaneously, isoform-specific differences in other molecular properties, e.g., chromatin binding affinity, protein–protein interactions, and posttranslational modifications, can lead to isoform-specific effects on gene expression and other biological processes involving linker histones.

ACKNOWLEDGMENT

We thank Drs. Steven McBryant, Heather Szerlong, and Robert Woody for critically reading the manuscript.

SUPPORTING INFORMATION AVAILABLE

One figure showing the correlations between three isoform nomenclatures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

- van Holde, K. E. (1988) *Chromatin*, Springer-Verlag, New York.
- Wolffe, A. (1998) *Chromatin: structure and function*, 3rd ed., Academic Press, San Diego.
- Woodcock, C. L., Skoultchi, A. I., and Fan, Y. (2006) Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res.* 14, 17–25.
- Allan, J., Hartman, P. G., Crane-Robinson, C., and Aviles, F. X. (1980) The structure of histone H1 and its location in chromatin. *Nature* 288, 675–679.
- Thomas, J. O. (1999) Histone H1: location and role. *Curr. Opin. Cell Biol.* 11, 312–317.
- Henzel, M. J., Lever, M. A., Crawford, E., and Th'ng, J. P. (2004) The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. *J. Biol. Chem.* 279, 20028–20034.
- Lu, X., and Hansen, J. C. (2004) Identification of specific functional subdomains within the linker histone H10 C-terminal domain. *J. Biol. Chem.* 279, 8701–8707.
- Ponte, I., Vidal-Taboada, J. M., and Suau, P. (1998) Evolution of the vertebrate H1 histone class: evidence for the functional differentiation of the subtypes. *Mol. Biol. Evol.* 15, 702–708.
- Hansen, J. C., Lu, X., Ross, E. D., and Woody, R. W. (2006) Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J. Biol. Chem.* 281, 1853–1856.
- Hansen, J. C. (2002) Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. *Annu. Rev. Biophys. Biomol. Struct.* 31, 361–392.
- Widom, J. (1998) Structure, dynamics, and function of chromatin in vitro. *Annu. Rev. Biophys. Biomol. Struct.* 27, 285–327.
- Fan, Y., Nikitina, T., Morin-Kensicki, E. M., Zhao, J., Magnuson, T. R., Woodcock, C. L., and Skoultchi, A. I. (2003) H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. *Mol. Cell. Biol.* 23, 4559–4572.
- Fan, Y., Nikitina, T., Zhao, J., Fleury, T. J., Bhattacharyya, R., Bouhassira, E. E., Stein, A., Woodcock, C. L., and Skoultchi, A. I. (2005) Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* 123, 1199–1212.
- Widlak, P., Kalinowska, M., Parseghian, M. H., Lu, X., Hansen, J. C., and Garrard, W. T. (2005) The histone H1 C-terminal domain binds to the apoptotic nuclease, DNA fragmentation factor (DFF40/CAD) and stimulates DNA cleavage. *Biochemistry* 44, 7871–7878.
- Montes de Oca, R., Lee, K. K., and Wilson, K. L. (2005) Binding of barrier to autointegration factor (BAF) to histone H3 and selected linker histones including H1.1. *J. Biol. Chem.* 280, 42252–42262.
- Kim, K., Choi, J., Heo, K., Kim, H., Levens, D., Kohno, K., Johnson, E. M., Brock, H. W., and An, W. (2008) Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription. *J. Biol. Chem.* 283, 9113–9126.
- Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., and Woodcock, C. L. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14173–14178.
- Carruthers, L. M., Bednar, J., Woodcock, C. L., and Hansen, J. C. (1998) Linker histones stabilize the intrinsic salt-dependent folding of nucleosomal arrays: mechanistic ramifications for higher-order chromatin folding. *Biochemistry* 37, 14776–14787.
- Allan, J., Mitchell, T., Harborne, N., Bohm, L., and Crane-Robinson, C. (1986) Roles of H1 domains in determining higher order chromatin structure and H1 location. *J. Mol. Biol.* 187, 591–601.
- Tomba, P., and Fuxreiter, M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* 33, 2–8.
- Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function. *Biochemistry* 41, 6573–6582.
- Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) Intrinsically disordered protein. *J. Mol. Graphics Modell.* 19, 26–59.
- Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* 337, 635–645.
- Mohan, A., Oldfield, C. J., Radivojac, P., Vacic, V., Cortese, M. S., Dunker, A. K., and Uversky, V. N. (2006) Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.* 362, 1043–1059.
- Vacic, V., Oldfield, C. J., Mohan, A., Radivojac, P., Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. *J. Proteome Res.* 6, 2351–2366.
- Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* 338, 1015–1026.
- Vucetic, S., Brown, C. J., Dunker, A. K., and Obradovic, Z. (2003) Flavors of protein disorder. *Proteins* 52, 573–584.
- Georgel, P., Demeler, B., Terpening, C., Paule, M. R., and van Holde, K. E. (1993) Binding of the RNA polymerase I transcription complex to its promoter can modify positioning of downstream nucleosomes assembled in vitro. *J. Biol. Chem.* 268, 1947–1954.
- Gordon, F., Luger, K., and Hansen, J. C. (2005) The core histone N-terminal tail domains function independently and additively

- during salt-dependent oligomerization of nucleosomal arrays. *J. Biol. Chem.* 280, 33701–33706.
30. Simpson, R. T., Thoma, F., and Brubaker, J. M. (1985) Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: a model system for study of higher order structure. *Cell* 42, 799–808.
 31. Hansen, J. C., Ausio, J., Stanik, V. H., and van Holde, K. E. (1989) Homogeneous reconstituted oligonucleosomes, evidence for salt-dependent folding in the absence of histone H1. *Biochemistry* 28, 9129–9136.
 32. Wellman, S. E., Song, Y., Su, D., and Mamoon, N. M. (1997) Purification of mouse H1 histones expressed in *Escherichia coli*. *Biotechnol. Appl. Biochem.* 26 (Part 2), 117–123.
 33. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
 34. Carruthers, L. M., Tse, C., Walker, K. P., III., and Hansen, J. C. (1999) Assembly of defined nucleosomal and chromatin arrays from pure components. *Methods Enzymol.* 304, 19–35.
 35. Demeler, B., and van Holde, K. E. (2004) Sedimentation velocity analysis of highly heterogeneous systems. *Anal. Biochem.* 335, 279–288.
 36. van Holde, K. E., and Weischet, W. O. (1978) Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. *Biopolymers* 17, 1387–1403.
 37. Orrego, M., Ponte, I., Roque, A., Buschati, N., Mora, X., and Suau, P. (2007) Differential affinity of mammalian histone H1 somatic subtypes for DNA and chromatin. *BMC Biol.* 5, 22.
 38. Talasz, H., Sapojnikova, N., Helliger, W., Lindner, H., and Puschendorf, B. (1998) In vitro binding of H1 histone subtypes to nucleosomal organized mouse mammary tumor virus long terminal repeat promotor. *J. Biol. Chem.* 273, 32236–32243.
 39. Tse, C., and Hansen, J. C. (1997) Hybrid trypsinized nucleosomal arrays: identification of multiple functional roles of the H2A/H2B and H3/H4 N-termini in chromatin fiber compaction. *Biochemistry* 36, 11381–11388.
 40. Schwarz, P. M., Felthouser, A., Fletcher, T. M., and Hansen, J. C. (1996) Reversible oligonucleosome self-association: dependence on divalent cations and core histone tail domains. *Biochemistry* 35, 4009–4015.
 41. Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11, 739–756.
 42. Hartman, P. G., Chapman, G. E., Moss, T., and Bradbury, E. M. (1977) Studies on the role and mode of operation of the very-lysine-rich histone H1 in eukaryote chromatin. The three structural regions of the histone H1 molecule. *Eur. J. Biochem.* 77, 45–51.
 43. Vila, R., Ponte, I., Collado, M., Arrondo, J. L., and Suau, P. (2001) Induction of secondary structure in a COOH-terminal peptide of histone H1 by interaction with the DNA: an infrared spectroscopy study. *J. Biol. Chem.* 276, 30898–30903.
 44. Clark, D. J., Hill, C. S., Martin, S. R., and Thomas, J. O. (1988) Alpha-helix in the carboxy-terminal domains of histones H1 and H5. *EMBO J.* 7, 69–75.
 45. Mamoon, N. M., Song, Y., and Wellman, S. E. (2005) Binding of histone H1 to DNA is described by an allosteric model. *Biopolymers* 77, 9–17.
 46. Ross, E. D., Baxa, U., and Wickner, R. B. (2004) Scrambled prion domains form prions and amyloid. *Mol. Cell. Biol.* 24, 7206–7213.
 47. Ross, E. D., Edskes, H. K., Terry, M. J., and Wickner, R. B. (2005) Primary sequence independence for prion formation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12825–12830.
 48. Shewmaker, F., Ross, E. D., Tycko, R., and Wickner, R. B. (2008) Amyloids of shuffled prion domains that form prions have a parallel in-register beta-sheet structure. *Biochemistry* 47, 4000–4007.
 49. Dosztanyi, Z., Chen, J., Dunker, A. K., Simon, I., and Tompa, P. (2006) Disorder and sequence repeats in hub proteins and their implications for network evolution. *J. Proteome Res.* 5, 2985–2995.
 50. Subirana, J. A. (1990) Analysis of the charge distribution in the C-terminal region of histone H1 as related to its interaction with DNA. *Biopolymers* 29, 1351–1357.
 51. Clark, D. J., and Kimura, T. (1990) Electrostatic mechanism of chromatin folding. *J. Mol. Biol.* 211, 883–896.
 52. Richardson, J. S., and Richardson, D. C. (1988) Amino acid preferences for specific locations at the ends of alpha helices. *Science (New York)* 240, 1648–1652.

BI801636Y