BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

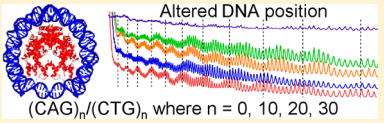
# CAG/CTG Repeats Alter the Affinity for the Histone Core and the Positioning of DNA in the Nucleosome

Catherine B. Volle[†] and Sarah Delaney*[,‡]

[†]Department of Molecular Biology, Cell Biology and Biochemistry and [‡]Department of Chemistry, Brown University, Providence, Rhode Island 02912, United States

Supporting Information

**ABSTRACT:** Trinucleotide repeats (TNRs) occur throughout the genome, and their expansion has been linked to several neurodegenerative disorders, including Huntington's disease. TNRs have been studied using both oligonucleotides and plasmids; however, less is know about how repetitive DNA responds to genomic packaging. Here, we investigate the behavior of CAG/CTG repeats incorporated into nucleosome core particles, the most basic unit of chromatin packaging. To assess the general interaction between CAG/CTG repeats and the histone core, we determined the efficiency with which various TNR-containing DNA substrates form nucleosomes, revealing that even short CAG/CTG tracts are robust incorporators. However, the presence of the Huntington gene flanking sequence (*htt*) decreases the rate of incorporation. Enzymatic and chemical probing revealed repositioning of the DNA in the nucleosome as the number of CAG/CTG repeats increased, regardless of the flanking sequence. Notably, the periodicity of the repeat tract remained unchanged as a function of length and is consistently 10.7 bp per helical turn. In contrast, the periodicity of the nonrepetitive flanking sequence varies and is smaller than the repeat tract at ∼10.0−10.5 bp per turn. Furthermore, while the CAG/CTG repeats remain as a canonical duplex in the nucleosome, nucleosome formation causes kinking in a secondary repeat tract in the *htt* gene, comprised of CCG/CGG repeats. This work highlights the innate ability of CAG/CTG repeats to incorporate and to position in nucleosomes and how that behavior is modulated by the *htt* flanking sequence. In addition, it illuminates the differences in packaging of healthy and diseased length repeat tracts within the genome.




Trinucleotide repeats (TNRs) are genetically unstable.[1−5] They occur throughout the genome, and their expansion or contraction, which is proposed to be mediated by the ability of repeats to form noncanonical secondary structures during DNA replication, transcription, and/or repair, can affect varied cellular processes such as gene expression, mRNA processing, and protein folding.[1−5] Furthermore, TNR expansion has been linked to a class of neurodegenerative disorders, including Huntington's disease (HD).[6] In the genome, the length of the repeat tract falls into one of three categories: a short, genetically stable repeat length found in healthy individuals, an intermediate, unstable repeat length that is prone to expansion, and a threshold repeat length that, once reached, leads to the manifestation of the disease state.[1−5] This threshold number of repeats varies by disease. The onset of HD occurs when there are more than 40 CAG/CTG repeats in the *htt* gene, whereas healthy individuals have between 5 and 36 repeats.[6]

In vitro and in vivo studies have shown that TNRs are found in both tightly packed, transcriptionally silent heterochromatin and the more loosely packed, transcriptionally active euchromatin.[7,8] Interestingly, HD repeats are found in association with histones containing euchromatin markers regardless of whether they are healthy or disease length.[9]

The most basic unit of packaging in chromatin is the nucleosome, which is comprised of ∼146 bp of DNA wrapped around a histone octamer.[10−14] Previous studies have shown that CAG/CTG repeats are preferential locations for nucleosome formation[15] and that as the length of the repeat tract increases, the ability of the DNA to incorporate into a nucleosome improves.[16] Notably, previous work has focused mainly on the incorporation of disease length TNRs into nucleosomes, and much less information about how repeat tracts found in healthy individuals are incorporated is available. However, to better understand the mechanism by which expansion occurs, it is important to consider the behavior of both healthy and disease length repeat tracts in a nucleosome. Furthermore, the DNA used in previous nucleosome incorporation studies was often constructed from genomic clones containing human repeat tracts; these DNA substrates often vary in their overall length, providing opportunities for multiple translational positions with respect to the histone core, and/or are removed from their genetic flanking sequence. For these reasons, we synthesized two series of 146 bp sequences containing CAG/CTG repeat tracts with lengths corresponding to those found in healthy individuals. In one series, a nucleosome positioning sequence flanks the repeats, while the other series uses the *htt* flanking sequence. We assessed the ability of these sequences to incorporate and to position in nucleosomes. These results describe how healthy length TNRs behave in chromatin and how this behavior compares to that of disease length repeat tracts. An understanding of how these sequences behave in nucleosomes defines how the functional organization of the *htt* gene, in particular the CAG/CTG

**Received:** October 17, 2012
**Revised:** November 14, 2012
**Published:** November 16, 2012

   dx.doi.org/10.1021/bi301416v | *Biochemistry* 2012, 51, 9814−9825

repeats, can influence the accessibility of the DNA, the progression of RNA and DNA polymerases, and the occurrence of genomic expansion.

## EXPERIMENTAL PROCEDURES

**Oligonucleotide Synthesis and Purification.** Oligonucleotides were synthesized using standard phosphoramidite chemistry on a BioAutomation DNA/RNA synthesizer (sequences can be found in Table S1 of the Supporting Information). The oligonucleotides were synthesized dimethoxytrityl (DMT)-on and were deprotected in $NH_4OH$ at 55 °C for at least 36 h. One round of high-performance liquid chromatography (HPLC) purification was conducted at 90 °C using a Polymer Laboratories Reverse phase column (4.6 mm × 250 mm) using 100 mM triethylammonium acetate (TEAA) in a 99% acetonitrile/1% water mixture (solvent A) and 100 mM TEAA in a 1% acetonitrile/99% water mixture (solvent B) as the mobile phases (for the gradient, the level of solvent A was increased from 5 to 25% over 25 min; 1 mL/min). Following removal of the 5′-DMT group by incubation in 80% glacial acetic acid for 12 min at room temperature, the oligonucleotides were dried, resuspended in formamide, and purified by 8% denaturing polyacrylamide gel electrophoresis (PAGE) (0.4 mm thick, 45 cm long). After electrophoresis at 65 W for 4 h, DNA was visualized by UV shadowing. Bands containing the 146-mer were excised from the gel, crushed, and soaked overnight at 37 °C in 10 mM Tris and 1 mM EDTA (pH 7.5), and the supernatant was collected. The DNA was then lyophilized, resuspended in 1 mL of water, and desalted using Sephadex G-50. The desalted DNA was then subjected to a second round of HPLC purification (for the gradient, the level of solvent A was increased from 0 to 25% over 40 min; 1 mL/min). The pUC19 fragment was amplified via polymerase chain reaction from plasmid pUC19 using the primers forward (5′-CCATTCGCCATTCAGGCTGCGCAA-3′) and reverse (5′-GCGGATAACAATTTCACACAGGAA-3′).

**Competitive Nucleosome Incorporation Assay.** Oligonucleotides were 5′-$^{32}$P end-labeled using T4 polynucleotide kinase (New England Biolabs) following the manufacturer's protocol. Radiolabeled DNA, along with a 1.5-fold excess of its complement, was resuspended to a final volume of 50 μL in Tris buffer [10 mM Tris-HCl, 100 mM NaCl, and 1 mM EDTA (pH 7.5)], heated at 95 °C for 5 min, and slowly cooled to 25 °C at a rate of 1 °C/min to form the duplex.

Nucleosomes were isolated from chicken erythrocytes following a previously published protocol.[17] Briefly, erythrocytes were collected by centrifugation and washed copiously with solution X [15 mM Tris, 60 mM KCl, 15 mM NaCl, 0.15 mM spermine, 0.5 mM spermidine, 0.34 M sucrose, 2 mM EDTA, 0.5 mM EGTA, 0.2 mM PMSF, 1 mM benzamidine, and 15 mM 2-mercaptoethanol (pH 7.5); the last three components were added immediately prior to use]. The cellular membranes were lysed by addition of 0.1% Nonidet p40 to solution X and gentle resuspension. The nuclei were then collected by centrifugation. The nuclei were again washed with solution X to remove hemoglobin. A final wash was performed by resuspending the nuclei in solution N [15 mM Tris, 60 mM KCl, 15 mM NaCl, 0.15 mM spermine, 0.5 mM spermidine, 0.34 M sucrose, 0.2 mM PMSF, and 15 mM 2-mercaptoethanol (pH 7.5); the last two components were added immediately prior to use].

The chromatin was digested by addition of $CaCl_2$ to a final concentration of 1 mM and incubated with 200 units/mL micrococcal nuclease (New England Biolabs). The digestion proceeded for 7 min at 37 °C before being quenched with EDTA (final concentration of 2 mM). Nuclei were then collected by centrifugation and lysed by addition of solution R [10 mM Tris, 0.2 mM EDTA, and 0.2 mM PMSF (pH 8.0); PMSF was added immediately prior to use]. The chromatin was then allowed to be shaken overnight at 4 °C, bringing the digested chromatin into solution.

The digested chromatin was collected and concentrated to a volume of no more than 4 mL using Amicon concentrators with a 3 kDa molecular mass cutoff, and the NaCl concentration was adjusted to 0.65 M by dropwise addition of 4 M NaCl. This solution was applied to a Sepharose 4B column (1.5 cm × 75 cm) equilibrated with solution S [20 mM Tris, 0.2 mM EDTA, 0.63 M NaCl, and 0.2 mM PMSF (pH 7.5); PMSF was added immediately prior to use]. Using a peristaltic pump, digested chromatin was eluted by flowing solution S at a rate of ∼0.5 mL/min. Twenty-five fractions containing 5 mL each were collected and assessed by SDS−PAGE to determine where the nucleosomes eluted, using the known molecular masses of the individual histone proteins.[18] Samples containing only the core histone proteins were pooled, concentrated, subjected to a second digestion with micrococcal nuclease to ensure nucleosomes were present as mononucleosomes, and repurified using the same procedure. Nucleosomes were dialyzed against solution D [10 mM Tris, 0.2 mM EDTA, and 0.2 mM PMSF (pH 7.5); PMSF was added immediately prior to use] and concentrated to ∼1 mL; the concentration was determined by the Bradford assay, and the final concentration was adjusted to 1 mg/mL with solution D.

Four incorporation reaction mixtures were prepared per DNA substrate; three reaction mixtures contained nucleosomes, while one reaction mixture lacked nucleosomes.[19] The reaction mixture lacking nucleosomes serves as a control for migration of the DNA duplex during native PAGE. Each 50 μL incorporation reaction mixture consisted of 100 ng of radiolabeled 146 bp duplex, 2.5 μg of calf thymus DNA, 25 μg of isolated nucleosomes, 1 M NaCl, and 10× exchange buffer (EB) [1× EB consists of 20 mM Tris, 1 mM EDTA, and 1 mM 2-mercaptoethanol (pH 8.0)]. For reaction mixtures that lacked nucleosomes, an equal volume of solution D was substituted. The initial incorporation reaction mixture was incubated for 1 h at room temperature. Next, 17 μL of 1× EB was added to each reaction mixture and each incubated for 1 h at room temperature. The addition of 16.5 μL of 1× EB followed by incubation for 1 h at room temperature was repeated twice, doubling the reaction volume. Then 100 μL of 1× EB was added to the reaction mixture and the mixture incubated for 15 min at room temperature; this process was repeated twice more. After a final addition of 100 μL of 1× EB, bringing the final reaction volume to 500 μL, the reaction mixture was allowed to incubate for 1 h at 37 °C. Each reaction mixture was then transferred to an Amicon microconcentrator with a 3 kDa molecular mass cutoff and centrifuged at 4 °C for 30 min at 14000*g*, reducing the final volume to ∼30 μL. These incorporation reactions were completed a total of three times for each DNA sequence for a total of nine incorporations per DNA.

**Native Polyacrylamide Gel Electrophoresis.** Samples (30 μL) from the competitive nucleosome incorporation assay were mixed with 6 μL of nondenaturing dye [15% ficoll, 0.25% bromophenol blue (w:v), and 0.25% xylene cyanol (w:v)], and 10 μL was applied to a 5% native polyacrylamide gel (4 mm

thick, 45 cm long). Samples were electrophoresed at 100 V for 12 h, and the products were visualized by phosphorimagery, using a Bio-Rad FX or Bio-Rad Pharos scanner and the accompanying Quantity One software. These conditions provided clear resolution among single-stranded, duplex, and nucleosome samples. The ratio of incorporated DNA (DNA in a nucleosome) to free DNA was determined by comparing the number of counts in each band.

**DNase I, Exo III, and S1 Nuclease Reactions.** For enzymatic probing, DNA was exchanged into the nucleosomes in a manner similar to that described for the competitive nucleosome incorporation assays, with a few alterations. First, no competitor DNA was added to the reaction mixture, and the amount of radiolabeled duplex was increased to 10 pmol. Second, the exchange buffer was altered [1× EB; 5 mM Tris (pH 7.5)]. Finally, the incubation time was shortened; while the first 1 h incubation remained, the subsequent 1 h incubations were changed to 45 min, and the 15 min incubations changed to 10 min. This procedure produced samples in which ∼90−95% of the DNA incorporated into the nucleosome.

After concentration of the exchanged nucleosomes, enzymatic probing reaction mixtures were prepared, combining 0.4 unit of DNase I (New England Biolabs), 50 units of Exo III (New England Biolabs), or 42.5 units of S1 nuclease (Promega) with 5 μL of the appropriate buffer for the enzyme to be used: DNase I buffer [100 mM Tris, 25 mM $MgCl_2$, and 5 mM $CaCl_2$ (pH 7.6)], NEB1 buffer [100 mM Bis Tris propane-HCl, 100 mM $MgCl_2$, and 10 mM DTT (pH 7.0)], or S1 nuclease buffer [500 mM sodium acetate, 2.8 M NaCl, and 45 mM $ZnSO_4$ (pH 4.5)]. The reaction volume was adjusted to 50 μL with exchanged, concentrated nucleosomes. The samples were incubated at 37 °C for 5 min and then the reactions quenched by the addition of 6 μL of 0.1 M EDTA and 6 μL of 50% glycerol. Samples were then loaded onto a 5% native polyacrylamide gel (8 mm thick, 16 cm long) and electrophoresed at 4 °C and 150 V for 3.5 h. The gel was used to expose a phosphorplate for approximately 25 min. The plate was imaged using a Bio-Rad Pharos scanner, and the image was used to identify the position of the DNA incorporated into nucleosomes. These bands were excised, crushed, and soaked overnight at 37 °C in 10 mM Tris and 1 mM EDTA (pH 7.5). After centrifugation to pellet the remaining gel, the supernatant was concentrated in a 15 mL volume Amicon concentrator with a 3 kDa molecular mass cutoff. Centrifugation occurred for 50 min at 4 °C and 3220*g*, leaving ∼0.5 mL of recovered nucleosomes. SDS was added to a final concentration of 0.01%, and samples were heated at 90 °C for 10 min. To preserve the duplex structure, this step was omitted for S1 nuclease reactions conducted after the incorporated DNA had been separated from the histone core. Samples were then extracted twice with 400 μL of a phenol/chloroform/isoamyl alcohol mixture and precipitated with ethanol, and the resulting pellet was dried under vacuum.

Reactions were also performed with the free duplex. In the Exo III and S1 nuclease reactions, the amount of enzyme added was the same as that for the nucleosome reactions. However, the enzyme concentration of DNase I optimized for the nucleosome reactions resulted in overcleavage of the free duplex, so a 1:10 dilution was used. The duplex was annealed as described previously, and 5 pmol was used to adjust the reaction volume to 50 μL. In all enzymatic reactions of the free duplex, the incubation time was shortened to 2 min and the reactions were quenched by addition of 6 μL of 0.1 M EDTA and 6 μL of 1 mg/mL calf thymus DNA. The DNA was immediately precipitated with ethanol and dried under vacuum. A standard Maxam−Gilbert A/G reaction performed on either strand of S1 [S1a or S1b (Table S1 of the Supporting Information)] was used to create a marker.

After the DNA pellets had dried, the number of counts was determined in a Perkin-Elmer scintillation counter and diluted with formamide to 5000 counts/μL. Three microliters of these samples were applied to an 8% denaturing polyacrylamide gel (0.4 mm thick, 65 cm long) and electrophoresed at 65 W for 2.5 h. The gel was dried and used to expose a phosphorplate for 96 h.

**Hydroxyl Radical Footprinting.** Nucleosome preparation was the same for the hydroxyl radical reactions as it was for the enzymatic probing. To generate hydroxyl radicals, 7.5 μL each of a 1 mM $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$/2 mM EDTA mixture, 10 mM sodium ascorbate, and 0.6% $H_2O_2$ was added to 52.5 μL of exchanged nucleosomes, bringing the total reaction volume to 75 μL. The solution was incubated for 15 min at room temperature. The reactions were quenched by addition of 8 μL of 50% glycerol and 8 μL of 0.1 M EDTA, and samples were immediately applied to a native polyacrylamide gel as described above. The sample preparation process was the same as that for the enzymatic probing experiments from this point forward. Free duplex reactions were conducted under the same reaction conditions as the nucleosome samples except they were incubated for 5 min at room temperature, quenched with 8 μL of 50% glycerol and 8 μL of 1 mg/mL calf thymus DNA, and precipitated with ethanol. From this point forward, the sample preparation followed the same procedure that was used for the enzymatic probing samples. Each sample was electrophoresed in three separate gels, one of which was run in the same manner as described above. However, to better visualize reactivity at the 5′ or 3′ end of the DNA, samples were also electrophoresed for 1.75 h (to visualize the 5′ end) or 3.5 h (to visualize the 3′ end).

**Image Processing.** Phosphorplates were imaged using a Bio-Rad Pharos scanner and the associated Quantity One software. Histograms were generated using Quantity One, but all other image processing was done using Bio-Rad's Image Lab. First, the gel image was aligned, and the lanes were identified. The lane width and position were adjusted, and the position of each band was manually determined. Using the accompanying Maxam−Gilbert sequencing reactions, each band was assigned by nucleotide and the corresponding intensity determined. Intensity was graphed as a function of nucleotide and smoothed by applying a moving average with a window of 5, and those data were imported into Origin 8.2 (OriginLab Corp.). The data were used to create a graph, which was baseline subtracted and used for curve fitting experiments.

To determine the periodicity of the DNA around the histone core, a sine wave was fit to the hydroxyl radical footprinting data, the phase of which is equal to the best-fit period of the DNA.[13] Using the nonlinear curve fitting function in Origin, a best-guess sine wave, with a phase of ∼10, was fit to the data. From the initial guess, the software refined the fit until $\chi^2$ was optimized. The sine wave was applied to the entire length of the DNA or to certain sections, such as the repeat tract or flanking sequences.

Origin was also used to determine the location of the maximal and minimal reactivity to the hydroxyl radical, which was used to construct translational positioning maps, revealing

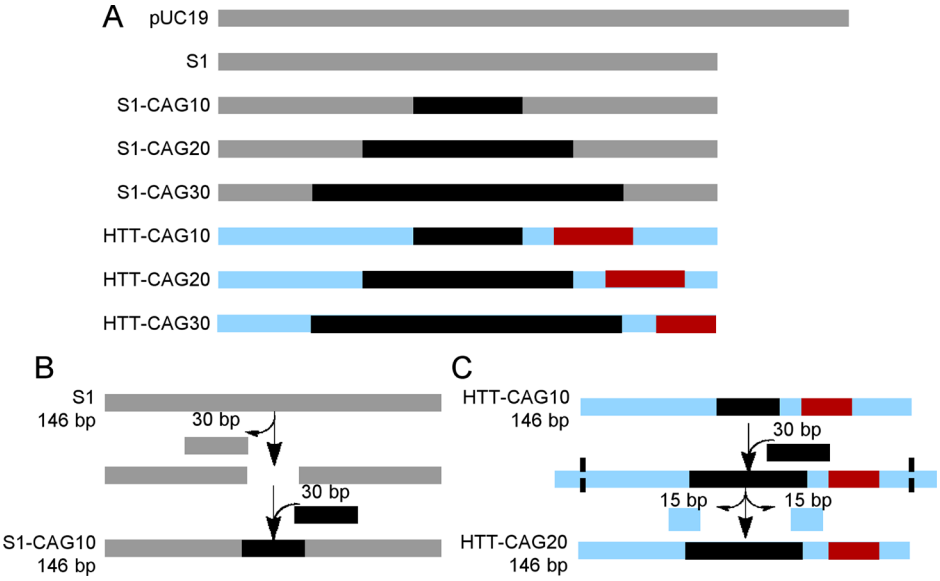



**Figure 1.** (A) Schematic representations of the DNA used in this study and its design. S1 and pUC19 serve as controls and contain no repeat tracts. S1-CAG10, S1-CAG20, and S1-CAG30 contain 10, 20, and 30 CAG/CTG repeats (black), respectively, within the S1 flanking sequence (gray). HTT-CAG10, HTT-CAG20, and HTT-CAG30 also contain 10, 20, and 30 CAG/CTG repeats (black), respectively; however, these substrates contain the *htt* flanking sequence (NG_009378.1, blue). The CCG/CGG repeats of the *htt* flanking sequence are colored dark red. (B) Schematic design of the S1 series of DNA in which the CAG/CTG repeats replace the center of the 146 bp S1 control. (C) Schematic design of the HTT series of DNA in which the size of the repeat tract was increased, and the flanking sequence was pushed outward and extra sequence trimmed from the ends to maintain a 146 bp length. Notably, all the 146 bp DNA sequences were prepared directly by phosphoramidite chemistry, and the schematics in panels B and C represent our design strategy rather than our experimental preparation. Please refer to the Supporting Information for the DNA sequences.

the location of the dyad axis.[20] The nucleotide position of each maximum was determined, and the number of nucleotides between two consecutive maxima ($y = \max_n - \max_{n-1}$) was plotted against the average value of the maxima [$x = (\max_n + \max_{n-1})/2$]. The same process was applied to the minima. The points determined for both the maxima and minima were smoothed with a moving average window of 2 or 3 and graphed.

## RESULTS AND DISCUSSION

**CAG/CTG Repeats Increase the Extent of Nucleosome Formation, but the Effect Is Offset by the *htt* Flanking Sequence.** To understand the behavior of CAG/CTG repeats in chromatin, we first performed competitive nucleosome incorporation assays. In the first series of DNA substrates, we started with 146 bp control duplex S1, a strong nucleosome positioning sequence with a single rotational and translational setting, which is derived from the pD89 sequence (Figure 1A).[20,21] To determine the innate ability of healthy length CAG/CTG repeats to incorporate into nucleosomes, we needed substrates that allowed us to systematically vary the number of repeats while keeping the primary sequence of the flanking region consistent between all the substrates in the series, ensuring that any observed changes are due solely to the increasing repeat number and not to differences between the primary sequence of the flanking DNA. Thus, the repeat-containing DNA was designed by replacing the center of the S1 duplex with 10, 20, and 30 CAG/CTG repeats to generate S1-CAG10, S1-CAG20, and S1-CAG30, respectively, while maintaining the overall 146 bp length (Figure 1B).

While the S1 sequence has been used in previous nucleosome investigations and therefore serves as a good control, CAG/CTG repeats are not flanked by the S1 sequence in the genome. Thus, we designed a second series of substrates containing the same CAG/CTG repeat tract lengths but changed the flanking sequence to the *htt* gene, creating HTT-CAG10, HTT-CAG20, and HTT-CAG30. These substrates were designed to mimic an expansion event; the TNR region was expanded, and the *htt* flanking sequence was pushed outward, allowing us to determine what effect the *htt* flanking sequence has on healthy length CAG/CTG repeats. The ends of the duplex were trimmed to maintain a 146 bp length (Figure 1C).

The extent to which a particular duplex incorporates into a nucleosome was determined using a competitive nucleosome incorporation assay.[19] In this assay, the DNA of interest was radiolabeled and exchanged onto the histone core of nucleosomes isolated from chicken erythrocytes. Unlabeled competitor DNA, in this case calf thymus DNA, acts as an internal control and, as it was in gross excess over the radiolabeled DNA, approximates an equilibrium distribution of the DNA incorporated into the nucleosome. Following the incorporation assay, native PAGE was used to separate radiolabeled duplex incorporated into nucleosomes from free, unincorporated duplex. As a mixed-sequence control, we also used a 261 bp fragment of pUC19 and examined its incorporation into a nucleosome.

We find that the S1 duplex incorporates into a nucleosome with the same efficiency as the pUC19 duplex, each with an incorporation ratio of ~2, indicating that there are two radiolabeled DNA duplexes incorporated into a nucleosome for every one radiolabeled duplex that remains as free DNA (Figure 2A,D). These results indicate that, while the S1 duplex was chosen because of its positioning ability, it incorporates with the same efficiency as the mixed-sequence pUC19 DNA. Notably, because some radiolabeled DNA remains free and is


dx.doi.org/10.1021/bi301416v | *Biochemistry* 2012, 51, 9814−9825

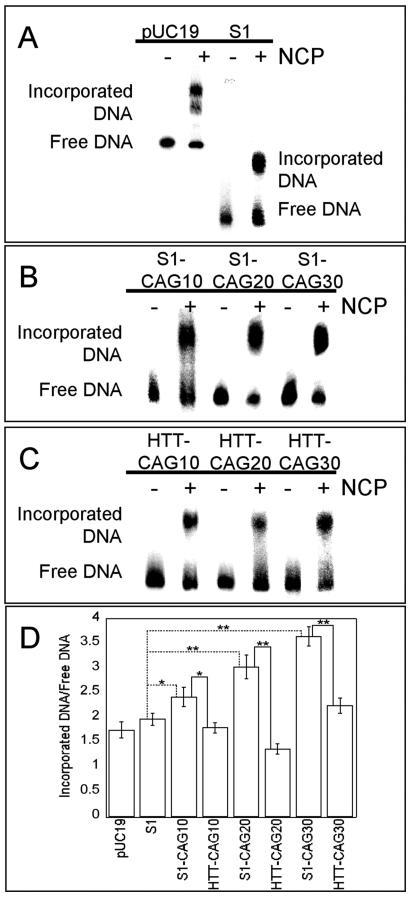



**Figure 2.** Representative competitive nucleosome incorporation reactions. To analyze the ability of each substrate to incorporate into a nucleosome, competitive nucleosome incorporation assays were performed under equilibrium conditions and the products of the reaction were resolved by native PAGE. An incorporation performed in the absence of chicken nucleosomes (NCP) (−NCP) and an incorporation performed in the presence of NCP (+NCP) are shown. The incorporation assays with pUC19 and S1 controls are shown in panel A; the incorporation assays with S1 DNA containing 10, 20, or 30 CAG/CTG repeats are shown in panel B, and the incorporation assays with HTT DNA are shown in panel C. By quantifying the amount of radioactivity in both the incorporated DNA and free DNA bands, we calculated the ratio of incorporated DNA to free DNA for each substrate (D). The error bars represent the standard error for each substrate. Data represent a total of three biological replicates per substrate, each consisting of three technical replicates, for nine total values. $*p \leq 0.05$ and $**p \leq 0.005$ by a Student's $t$ test.

not incorporated into a nucleosome, the S1 duplex provides a suitable baseline from which there is the potential to observe either an increased or a decreased level of incorporation upon introduction of a CAG/CTG sequence.

Interestingly, by native PAGE, we observe two distinct bands for the pUC19 duplex incorporated into nucleosomes, indicating that the pUC19 fragment forms at least two species of nucleosomes with different translational positions, in which the translational position is defined as the position of the histone core relative to the ends of the DNA duplex (Figure 2A). Although in previous studies[22−24] multiple, distinct species are not observed for nucleosomes formed from this pUC19 fragment, it is possible that our PAGE conditions allowed for resolution of multiple species.

When the central base pairs of the S1 duplex are replaced by CAG/CTG repeats, there is a significant increase in the ratio of DNA found in a nucleosome relative to free DNA (Figure 2B,D). Even the addition of only 10 CAG/CTG repeats (S1-CAG10) provides a small but significant increase in the level of nucleosome incorporation compared to the S1 control; the ratio increases from 2 to 2.4. Addition of an additional 10 CAG/CTG repeats (S1-CAG20) further increases the incorporation ratio to 3. The largest CAG/CTG repeat tract studied here, within S1-CAG30, has the highest incorporation ratio of the investigated substrates (3.7). Because the primary sequence of the flanking DNA is maintained throughout the S1 series, the increase in the level of incorporation can be attributed to the increase in the number of CAG/CTG repeats.

The increased incorporation ratio for DNA containing CAG/CTG repeats relative to mixed-sequence controls has traditionally been ascribed to the flexibility induced by the AG/CT dinucleotide, which lowers the free energy of duplex bending.[24−27] Furthermore, the CAG/CTG duplex has been shown to be an inherently flexible molecule based on electrophoretic mobility and cyclization kinetics.[28] It is thought that more flexible DNA, with its lower free energy of bending, is better able to bend around the histone core, leading to higher incorporation ratios.[29]

While it is important to understand the innate influence of CAG/CTG repeats on nucleosome incorporation, it is equally important to consider the effect of the *htt* flanking sequence. To this end, we repeated the competitive nucleosome incorporation using the HTT series of DNA. The HTT substrates all had significantly lower incorporation ratios than their corresponding S1 counterparts (Figure 2C,D). The ratios for HTT-CAG10, HTT-CAG20, and HTT-CAG30 are 1.8, 1.2, and 2.2, respectively. The significant decrease observed for each substrate in the HTT series indicates that the *htt* flanking sequence has an inherent negative effect on nucleosome incorporation relative to the corresponding S1 flanking sequence, offsetting the benefit from the CAG/CTG repeats.

It is important to note that the HTT series contains not only the primary CAG/CTG repeats but also a secondary CCG/CGG repeat in the flanking sequence (Figure 1A), a TNR that has altered nucleosome incorporation in various contexts.[22−24,30,31] In contrast to the inherent flexibility of the CAG/CTG repeats, CCG/CGG repeats are quite stiff.[24−27] The lack of sequential A/T base pairs within the relatively G/C rich *htt* flanking sequence, which are known to be important in facilitating bending around the histone core, also likely contributes to the observed decrease in the level of incorporation.[32]

The trends we observe in the S1 series are consistent with previously reported results for longer, disease length CAG/CTG repeats, in which the presence of the repeats was found to facilitate nucleosome formation. By direct observation of nucleosome formation by electron microscopy (EM), repeat tracts containing up to 250 CAG/CTG repeats have been shown to be preferential sites for nucleosome formation.[15] Furthermore, repeat tracts containing either 75 or 130 CAG/CTG repeats have been shown to significantly increase the ratio of incorporated DNA to free DNA.[16] Importantly, because of the large number of repeats, the flanking sequence does not contribute to the incorporation in these experiments, where nucleosome formation occurs almost exclusively within the repeat tract.

Interestingly, it has been reported that there is a negligible difference in the incorporation efficiency between DNA containing 10 CAG/CTG repeats and DNA containing either 55 or 62 CAG/CTG repeats.[33] However, the 55 CAG/CTG repeat sequence forms nucleosomes with several translational positions, many of which contain either the 3′ or 5′ flanking sequence at the same position as the nucleosomes formed with 10 CAG/CTG repeats, indicating that the flanking sequence may be dominating the incorporation process. Here, we observe the ability of the *htt* flanking sequence to decrease the level of incorporation of CAG/CTG repeats, also masking the inherent favorability associated with CAG/CTG repeats.

**DNase I and Exo III Reveal the Position of DNA in the Nucleosome.** DNase I is an endonuclease that creates single-strand nicks.[34−37] In separate experiments in which either the CAG-containing (Figure 3) or CTG-containing (Figure S1 of the Supporting Information) strand is radiolabeled, both free duplex and nucleosome samples were subjected to digestion by DNase I, and the products were resolved by denaturing PAGE. While DNase I does not possess strict sequence dependency, it does cleave preferentially at some sequences, mainly depending on minor groove width[38,39] and DNA flexibility.[40,41] Indeed, we observe unequal reactivity along all of the free duplex samples.

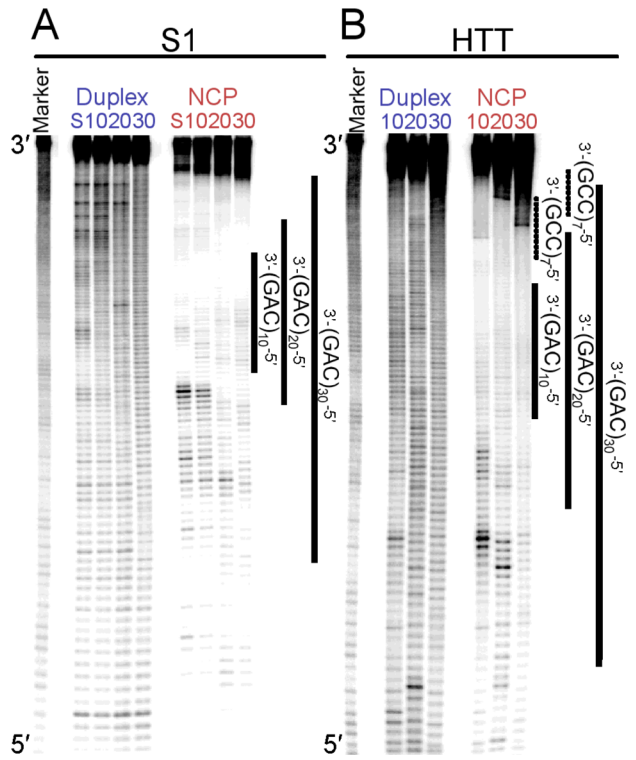



**Figure 3.** DNase I digestion reveals successful incorporation of the DNA around the histone core, as evidenced by the altered reaction pattern in the nucleosome samples as compared to that of the duplex. DNase I cleavage of the CAG-containing strands of the S1 substrates is shown in panel A, while cleavage of the CAG-containing strands of the HTT substrates is shown in panel B. Reactions were performed on both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. The number of repeats is indicated above each lane, with S indicating the S1 control. The marker lane consists of the Maxam−Gilbert A/G sequencing reaction performed on S1a. The location of the repeat tracts is indicated on the right-hand side of the gel.

When the DNA is incorporated into a nucleosome, its cleavage by DNase I changes, revealing a pattern of regions with high reactivity toward DNase I followed by regions with little or no reactivity toward DNase I. We observe this reactivity for both the S1 and HTT series incorporated into nucleosomes (Figure 3 and Figure S1 of the Supporting Information). Where the DNA backbone faces out into solution, DNase I can cleave, but DNA facing into the histone core will be inaccessible to the enzyme.[34−37] The DNA at the end of the duplex is more readily cleaved than the DNA in the center, producing dark bands near the top of each lane, because of weakened interactions between the extremes of the DNA duplex and the histone core.[42]

Exo III is a 3′ to 5′ exonuclease, which digests duplex DNA.[43−45] When the enzyme encounters DNA that is interacting with the histone core it pauses, which results in intense strand cleavage. These locations of intense strand cleavage map the translational position of the DNA. As with the DNase I reactions, in separate experiments in which either the CAG-containing (Figure 4) or CTG-containing (Figure S2 of the Supporting Information) strand was radiolabeled, both free duplex and nucleosome substrates were subjected to partial digestion by Exo III and the products were resolved by denaturing PAGE. Like DNase I, Exo III has some sequence specificity,[46] indicated by the uneven reaction along the free duplex samples for both the S1 and HTT series.

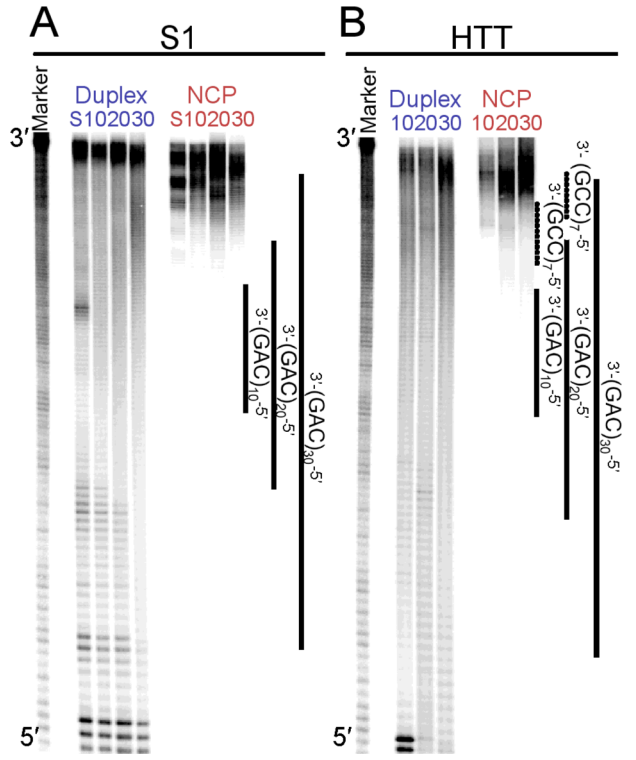



**Figure 4.** Exonuclease III digestion reveals weaker interactions between the ends of the DNA and the histone core. Exo III digestion of the CAG-containing strands of the S1 substrates is shown in panel A, while digestion of the CAG-containing strands of the HTT substrates is shown in panel B. Reactions were performed on both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. The number of repeats is indicated above each lane, with S indicating the S1 control. The marker lane consists of the Maxam−Gilbert A/G sequencing reaction performed on S1a. The location of the repeat tracts is indicated on the right-hand side of the gel.


dx.doi.org/10.1021/bi301416v | *Biochemistry* 2012, 51, 9814−9825

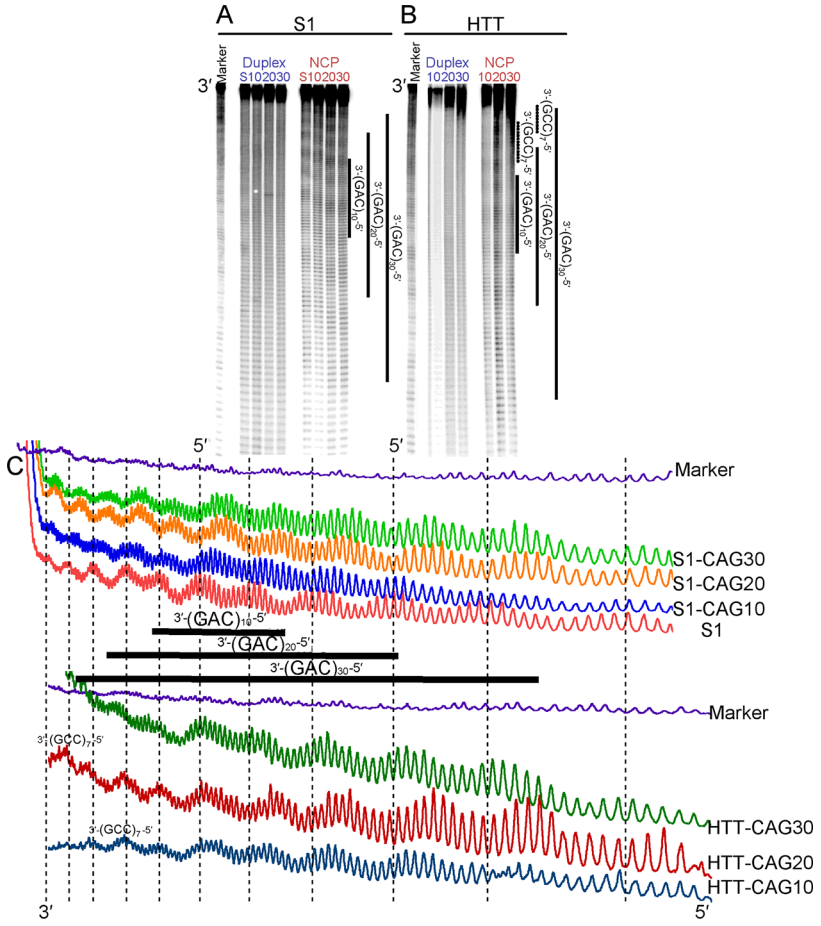



**Figure 5.** Hydroxyl radical footprinting reveals the periodicity of the DNA around the histone core. Radiolabeled samples, both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates, were exposed to hydroxyl radicals, revealing a characteristic pattern of oscillating high and low reactivity as the DNA wraps around the histone core. The reactivity toward the hydroxyl radical of the CAG-containing strands of the S1 substrates is shown in panel A, while the reactivity of the CAG-containing strands of the HTT substrates is shown in panel B. The number of repeats is indicated above each lane, with S indicating the S1 control. The marker lane consists of the Maxam−Gilbert A/G sequencing reaction performed on S1a. The location of the repeat tracts is indicated on the right-hand side of the gel. Panel C contains histograms generated from the gels presented in panels A and B. The dashed lines indicate the maxima of the S1 substrate.

There is a clear change in the Exo III digestion pattern when the duplex is incorporated into a nucleosome (Figure 4 and Figure S2 of the Supporting Information). The enzyme reactivity in the nucleosome samples is limited to the ends of the DNA. In the S1 series, there are three or four highly reactive sites observed regardless of the strand that is radiolabeled (Figure 4A and Figure S2A of the Supporting Information). While the presence of multiple reactive sites could indicate that the nucleosomal S1 series is forming multiple translational positions, it is more likely that, because of salt-induced shielding of the extremes of the DNA duplex, the interactions between the DNA and the histone core are weaker at the ends, allowing some Exo III to push the DNA off the histone core and continue its digestion.[42] Furthermore, the S1 control sequence has been shown previously to adopt a single translational position under similar conditions.[20,21]

There are only slight differences observed for Exo III digestion of the HTT substrates, indicating that the three HTT DNAs form nucleosomes with similar translational positions (Figure 4B and Figure S2B of the Supporting Information). There is only one main reactive site in each of the HTT substrates, indicating that the *htt* flanking sequence is in contact with the histone core.

**Hydroxyl Radical Footprinting Reveals That CAG/CTG Repeats Alter the Rotational Positioning of DNA in a Nucleosome.** While the enzymatic probing with DNase I and Exo III reveals the general position of DNA in the nucleosome, the results can be complicated by the sequence preferences associated with each enzyme. A more sensitive method of determining the rotational position of the DNA in a nucleosome is hydroxyl radical footprinting. The hydroxyl radical reacts preferentially with the C5′ hydrogen atom of the sugar−phosphate backbone, located in the minor groove, inducing cleavage at every nucleotide. Regions of the most intense strand cleavage indicate where the minor groove is solvent-exposed.[13,21,36,47,48]

To assess the position of each strand of the duplex with respect to the histone core, reaction mixtures were preformed in separate experiments in which either the CAG-containing (Figure 5) or CTG-containing (Figure S3 of the Supporting Information) strand was radiolabeled. The products of the reactions were resolved by denaturing PAGE, in which 100−120 nucleotides were observed by standard electrophoresis procedures. The PAGE images were used to create histograms for each sample (Figure 5C and Figure S3C of the Supporting


dx.doi.org/10.1021/bi301416v | *Biochemistry* 2012, 51, 9814−9825

Table 1. Numbers of Base Pairs per Helical Turn for S1 and HTT Nucleosomal Substrates

| substrate | total | repeat | flanking (5′ end) | flanking (3′ end) |
|---|---|---|---|---|
| S1 | 10.35 ± 0.02 | N/A[a] | N/A[a] | N/A[a] |
| S1-CAG10 | 10.44 ± 0.04 | 10.71 ± 0.02 | 10.34 ± 0.04 | 10.30 ± 0.04 |
| S1-CAG20 | 10.61 ± 0.02 | 10.70 ± 0.02 | 10.34 ± 0.04 | 10.30 ± 0.07 |
| S1-CAG30 | 10.62 ± 0.03 | 10.66 ± 0.03 | 10.40 ± 0.11 | 10.30 ± 0.06 |
| HTT-CAG10 | 10.55 ± 0.04 | 10.69 ± 0.03 | 10.54 ± 0.08 | 10.01 ± 0.01 |
| HTT-CAG20 | 10.52 ± 0.02 | 10.67 ± 0.03 | 10.52 ± 0.09 | 10.51 ± 0.03 |
| HTT-CAG30 | 10.71 ± 0.05 | 10.70 ± 0.04 | 10.55 ± 0.26 | 10.50 ± 0.03 |

[a]Not applicable.

Information), which were processed for use in calculating the periodicity of the DNA (Table 1).

When the free duplex is exposed to hydroxyl radicals, the reaction produces a nearly even pattern of reactivity along the DNA, with a slight increase in reactivity at purines (Figure 5A,B and Figure S3A,B of the Supporting Information).[20,36,48,49] Once the DNA is incorporated into a nucleosome, we observe an oscillating pattern of high and low reactivity toward the hydroxyl radical. Areas of increased reactivity indicate regions where the minor groove faces away from the histone core and is accessible to the hydroxyl radical, whereas areas of reduced reactivity indicate regions where the minor groove faces in toward the histone core and is protected from cleavage by hydroxyl radicals.

Analysis of the hydroxyl radical footprinting reactions on the S1 series reveals the inherent positioning ability of the CAG/CTG repeat tract. The pattern of reactivity is clearly different among the S1 series in both the gels and the resulting histograms (Figure 5A,C and Figure S3A,C of the Supporting Information). The vertical, dashed lines in Figure 5C indicate the position of each maxima for the S1 substrate, from which the offset of the maxima and minima of other substrates is determined. Using the S1 control as a baseline, we observe that the maxima of S1-CAG20 and S1-CAG30 are shifted by ∼5 nucleotides; where a maximum occurs in S1, the S1-CAG20 and S1-CAG30 have minima. While the offset between S1 and the S1-CAG10 substrate is not as pronounced, the nucleotide with maximal reactivity is offset by ∼1−2 nucleotides in the repeat region (Figure 5C). However, the shift appears to be localized near the repeat tract, as the extremes of the flanking sequence share maxima and minima with the S1 substrate.

By overlaying the footprinting patterns obtained for the complementary strands, we are able to identify the minor groove, allowing analysis of the rotational position of the DNA using both strands (Figure S4 of the Supporting Information).[20] Consistent with the maxima and minima observed in the histograms, the rotational positions of the S1 and S1-CAG10 substrates are quite similar but differ from the rotational positions of the S1-CAG20 and S1-CAG30 substrates. This result indicates that for CAG/CTG repeat tracts longer than 20, the repeats override the S1 flanking sequence to define the rotational position (Figure S4A−D of the Supporting Information).

While the HTT series possesses some small changes in the maxima and minima present in the gel and resulting histograms, they are much more uniformly aligned than those within the S1 series (Figure 5B,C and Figure S3 of the Supporting Information). Upon closer examination of the footprinting patterns for the complementary strands, HTT-CAG10 and HTT-CAG20 share a rotational position, which is different than the rotational position occupied by HTT-CAG30 (Figure S4E−G of the Supporting Information). In these constructs, it is likely that the secondary CCG/CGG repeat, which is reduced to only two of the seven repeats in HTT-CAG30, is playing some role in determining the rotational position of the HTT substrates. S1-CAG30 and HTT-CAG30 have the same rotational position, indicating that the repeats and not the flanking sequence are determining the rotational position in these substrates. However, unlike the S1 substrates, 20 CAG/CTG repeats is not enough to overcome the influence of the *htt* flanking sequence; it takes 30 CAG/CTG repeats.

**CAG/CTG Repeats Possess a Periodicity of ∼10.7 bp per Turn, Regardless of Flanking Sequence.** Upon fitting a sine wave to the hydroxyl radical footprinting data generated for the S1 series of substrates, we observe a general increase in the average periodicity for the entire DNA as the repeat tract expands. The S1 substrate has a periodicity of 10.35 ± 0.02 bp/turn (Table 1), consistent with previously reported values for this substrate, as well as similar DNA incorporated into nucleosomes.[10,13,20] For S1-CAG10, the periodicity increases to 10.44 ± 0.04 bp/turn, a change of ∼0.1 bp/turn, which represents a global underwinding of the duplex by ∼1 bp. The periodicity of the S1-CAG20 DNA further increases to 10.61 ± 0.02 bp/turn, or a global underwinding of ∼2 bp relative to the S1 substrate. However, this appears to be the maximal change in periodicity associated with the addition of CAG/CTG repeats because S1-CAG30 has a periodicity comparable to that of S1-CAG20, consistent with our proposal that the CAG/CTG repeats have fully dominated the influence of the S1 flanking sequence in the S1-CAG20 and S1-CAG30 substrates.

The periodicity of the HTT substrates is different from that of the S1 series. HTT-CAG10 and HTT-CAG20 have a periodicity of ∼10.55 bp/turn, but the periodicity of the HTT-CAG30 substrate is larger at 10.71 ± 0.05 bp/turn. This translates to a global underwinding of ∼2 bp for HTT-CAG10 and HTT-CAG20 and ∼3 bp for HTT-CAG30, compared to S1. The relative periodicities are consistent with the rotational position determined for each substrate; the substrates where the CAG/CTG repeat is the determinant of rotational position (S1-CAG20, S1-CAG30, and HTT-CAG30) all have a similar periodicity.

When a sine wave is fit to the hydroxyl radical footprinting data corresponding to specific regions of the DNA substrates, an interesting trend is observed. The periodicity of all the CAG/CTG repeat regions is ∼10.7 bp/turn (Table 1 and Table S2 of Supporting Information). On the basis of both chemical and enzymatic probing, each CAG/CTG repeat tract resides at the dyad axis, which is known to have a naturally larger periodicity than other regions of DNA in the nucleosome,[10,13,20] but as the repeat tract continues to expand off the dyad axis, the periodicity associated with the repeat tract remains constant at ∼10.7. Taken together, the shared


dx.doi.org/10.1021/bi301416v | *Biochemistry* 2012, 51, 9814−9825

periodicity of the repeat regions from both the S1 and HTT substrates indicates that the preferred periodicity of the CAG/CTG repeat tract in a nucleosome, under these conditions, is ~10.7 bp/turn. Furthermore, the increased periodicity of the CAG/CTG repeat tracts relative to the flanking sequence provides additional insight into why CAG/CTG repeats are such good incorporators and prefer to associate with the dyad axis; their inherent flexibility along with a preferred rotational setting congruent with that of the dyad axis allows them to facilitate nucleosome formation at the dyad axis.

**CAG/CTG Repeats Alter the Translational Positioning of DNA in the Nucleosome.** We also examined the periodicity associated with each maximum and minimum in the hydroxyl radical footprinting data. By graphing these values, we generated the translational positioning curves shown in Figure 6, allowing us to identify extremes in the periodicity with respect to the DNA backbone.[20] The position of the maximum in the curve defines the dyad axis, which is given in Table 2. The dyad axis of each repeat-containing substrate is located well within the repeat tract, consistent with previous observations that CAG/CTG repeats of various lengths preferentially assemble at the dyad axis.[15,16,33] Upon addition of 10 CAG/CTG repeats to the S1 substrate, there is a shift in the position of the dyad axis of 2 bp. An additional 10 and 20 repeats shift the dyad another 2 and 3 bp, respectively. These shifts cannot be due solely to changes in rotational positioning; as stated above, the difference between the S1 and S1-CAG10 or S1-CAG10 and S1-CAG20 substrates only amounts to 1−2 bp over the entire duplex, and the periodicity at the dyad axis is the same. The observed shift in the position of the dyad axis must be at least partially due to an altered translational positioning. However, when the repeats are placed within the *htt* flanking sequence, the position of the dyad axis remains constant at position 71, a further example of the ability of the *htt* flanking sequence to modulate the innate behavior of the CAG/CTG repeat.

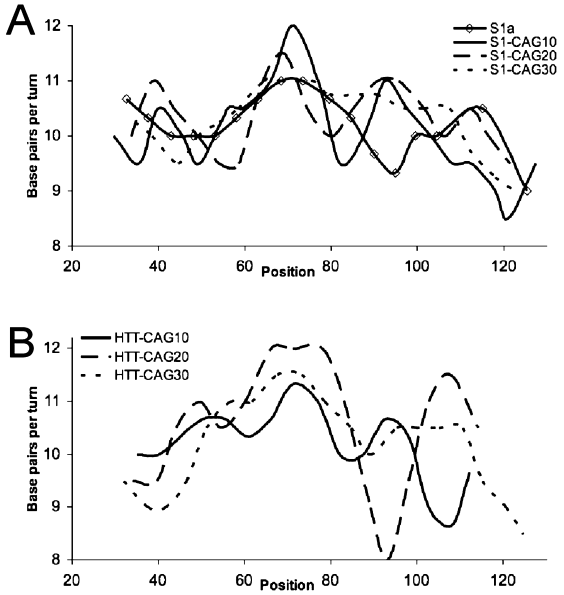



**Figure 6.** Translational positioning curves reveal a change in the position of the dyad axis for both the S1 and HTT series. Translational positioning curves were created by determining the location of and distance between each consecutive maximum and minimum in the hydroxyl radical footprinting data. The locations of the maxima and minima are then graphed with respect to their periodicity. The curve reveals the variation in periodicity around the histone core, and the highest peak indicates the location of the dyad axis. In the S1 series (A), the translational positioning curves reveal 2−5 bp offsets in the location of the dyad axis. In comparison, the HTT series (B) all have the same dyad axis placement.

**Table 2. Positions of the Dyad and Relative Shifts**

| substrate | dyad position | shift (no. of nucleotides)[a] |
|---|---|---|
| S1 | 73 | − |
| S1-CAG10 | 71 | −2 |
| S1-CAG20 | 69 | −4 |
| S1-CAG30 | 68 | −5 |
| HTT-CAG10 | 71 | −2 |
| HTT-CAG20 | 71 | −2 |
| HTT-CAG30 | 71 | −2 |

[a]With respect to the S1 control.

It is important to consider that the length of the DNA substrates used here is limited to the length of DNA associated with a single nucleosome. This length minimizes the possibility that the DNA would form a heterogeneous population of nucleosomes with multiple translational positions, which would confound our analysis. Within the cell, the histone core will have access to a longer DNA substrate with which to form a nucleosome, and while a nucleosome will likely form at the CAG/CTG repeat, given the affinity of the CAG/CTG repeat tract for the histone core, the nucleosomes formed with longer DNA substrates may have multiple DNA positions. However, it is likely that even with longer DNA substrates, CAG/CTG repeats will still alter DNA positioning, though the alterations may be more complex.

In the genome, CAG/CTG repeats have been found in heterochromatin and euchromatin.[7−9,50−52] Both healthy and disease length repeats associated with HD are associated with the more open, transcriptionally active euchromatin. In contrast, while shorter CAG/CTG tracts associated with myotonic dystrophy (DM1) are present in regions of euchromatin, the expanded, disease length repeats are associated with several heterochromeric markers. Notably, disease length DM1 repeat tracts can contain up to 1800 CAG/CTG repeats, which would span ~27 nucleosomes, compared to the single nucleosome that healthy length repeats would occupy. The innate incorporation ability of CAG/CTG repeats, along with a defined periodicity compatible with the underwinding associated with the dyad axis, is consistent with the shift from euchromatin to heterochromatin associated with repeat expansion in DM1, though it is likely that other cellular factors play a role as well.[8] Conversely, the CAG/CTG repeats associated with HD expand over a much shorter range, with the longest expanded repeat tracts containing ~100 repeats, which would span fewer than two nucleosomes.[6] Thus, while a nucleosome is more likely to position within the disease length *htt* repeat tract, the gene remains as euchromatin and transcription can proceed normally. However, the same gene containing a healthy length repeat tract must balance the competing incorporation and positioning influences of the repeat tract and flanking sequence, likely leading to a difference in functional organization, and thus different levels of DNA accessibility, which could affect genomic expansion.

**The CCG/CGG Repeat in the *htt* Flanking Sequence "Kinks" When Incorporated into the Nucleosome.** S1 nuclease, which recognizes single-stranded regions of DNA or

RNA, has been used previously to characterize the noncanonical secondary structures formed by CAG or CTG repeat oligonucleotides.[53,54] To assess the ability of CAG/CTG repeats to form noncanonical structures in the nucleosome, we incubated both free duplex and nucleosomal substrates from the HTT series with S1 nuclease (Figure 7A,B). We found no

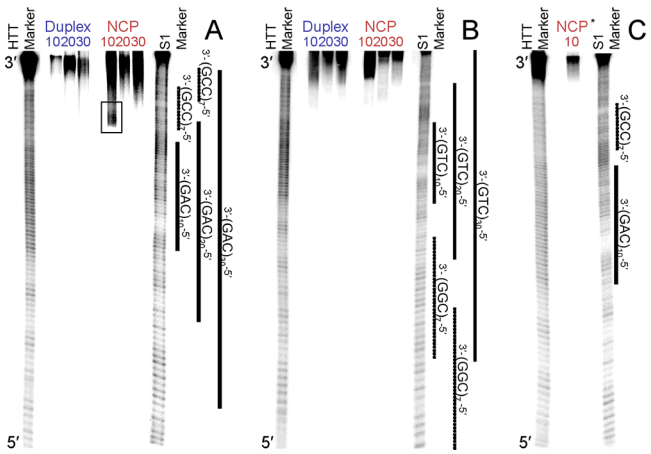



**Figure 7.** S1 nuclease probing reveals that kinking occurs in the secondary CCG/CGG repeat of the HTT-CAG10 DNA when incorporated into a nucleosome. The HTT series of substrates, with either the CAG-containing strand radiolabeled (A) or the CTG-containing strand radiolabeled (B), was incubated with S1 nuclease, a probe for single-stranded DNA. Reactions were performed with both free duplex (labeled in blue) and nucleosome (NCP, labeled in red) substrates. A region of reactivity (A, boxed region) is observed for HTT-CAG10 incorporated into a nucleosome (A), but once that DNA is removed from a nucleosomal context, the reactivity is no longer observed (C; NCP* indicates that the DNA was incorporated but was removed from the NCP before probing). The S1 nuclease reaction mixtures were electrophoresed along with an S1 marker and an HTT marker, consisting of the Maxam–Gilbert A/G sequencing reaction performed on S1a or the CAG-containing strand of HTT-CAG10.

evidence of noncanonical structure formation in any of the CAG/CTG repeat regions, as indicated by a lack of strand cleavage. However, there is reactivity in the secondary CCG repeat tract of the incorporated HTT-CAG10 substrate (Figure 7A, boxed region), indicating that there is a region of unpaired DNA in the nucleosome. We refer to this region as "kinked". While we expect that the complementary region in the CGG repeat is also kinked, no cleavage by S1 nuclease was observed (Figure 7B). However, on the basis of the hydroxyl radical footprinting, the CGG repeats complementary to the kinked region are facing in toward the histone core, making them inaccessible to the nuclease. Thus, the complementary strand may be kinked as well but cannot be detected by S1 nuclease.

As a control, we incorporated the HTT-CAG10 substrate into a nucleosome, removed the proteins by performing a phenol/choloroform/isoamyl alcohol extraction, and treated the freed duplex DNA with S1 nuclease. Upon removal of the DNA from the constraints of the histone core, no cleavage of DNA by nuclease was observed, indicating that kinking is a reversible process (Figure 7C).

The region in which this kink occurs is distinguished by its overwinding. The CCG/CGG repeat in the HTT-CAG10 duplex is located in an area of the nucleosome that is locally overwound; the center of the $(CCG)_7$ repeat falls in a trough on the translational positioning curve for the HTT-CAG10 nucleosome (Figure 6B). Furthermore, the entire flanking region possesses a global overwinding; the periodicity of the CCG repeat-containing flanking sequence is much lower than the periodicities determined for any of the other flanking sequences (Table 1). It has previously been shown that superhelical locations (SHLs) 1.5 and 4.5 have a distinct bending of the DNA, but the kink does not form at these locations.[10] The formation of the kink is likely due in combination to the stiffness of the CCG/CGG repeats and the stress of bending around the histone core.[24−27] Kinks have been observed in previous experiments with superhelically stressed DNA and play a part in relieving some of this stress.[55−57] We observe that once the CCG/CGG repeat is moved to a different location in the nucleosome, kinks no longer form. In HTT-CAG20, the CCG/CGG repeats are shifted close to the end of the duplex such that it is likely more favorable to simply dissociate from the histone core than to kink the DNA. This is reflected in the Exo III footprint of HTT-CAG20 nucleosomes; Exo III shows strong digestion through the CCG/CGG repeat tract (Figure 4B). In HTT-CAG30, part of the CCG/CGG repeat is removed and no cleavage by S1 nuclease is observed. It is unknown whether cellular conditions will allow for kinking, yet this process of relieving superhelical stress could have interesting consequences within the cell.

## CONCLUSIONS

In conclusion, while CAG/CTG repeats are naturally strong nucleosome incorporation and positioning sequences, the effect of healthy length repeats is moderated by the *htt* flanking sequence. When placed within the S1 sequence, the CAG/CTG repeats are able to significantly increase the level of incorporation with each addition of 10 repeats. However, when the repeats are flanked by the *htt* gene sequence, each substrate showed a significant decrease in the level of nucleosome incorporation compared to the corresponding S1 substrate. Twenty CAG/CTG repeats are sufficient to override the preferred positioning of the S1 flanking sequence and adopt the orientation relative to the histone core most favorable for the CAG/CTG repeat tract. However, in the *htt* gene sequence, 30 CAG/CTG repeats are required to override the positioning effect of the *htt* flanking sequence and adopt the most favorable position for the repeat tract. Interestingly, the only feature of the CAG/CTG repeats that is not influenced by the flanking sequence is their periodicity in the nucleosome; the periodicity is ∼10.7 bp/turn in each substrate. Thus, while disease length CAG/CTG repeats will likely interact with the histone core in a manner dictated solely by the CAG/CTG repeat, CAG/CTG repeat tracts found in healthy individuals must balance the incorporation and positioning preferences of both the repeat and the flanking sequence, leading to different levels of nucleosome occupancy and forms of positioning between healthy and disease length repeat tracts, the consequences of which could alter the propensity for genomic expansion and manifestation of the disease state within the packaged genome.

## ASSOCIATED CONTENT

### Supporting Information

List of DNA sequences and reactions on and periodicities of the CTG-containing strands. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION


### Corresponding Author

*Brown University, 324 Brook St., Providence, RI 02912. E-mail: sarah_delaney@brown.edu. Telephone: (401) 863-3590. Fax: (401) 863-9368.



### Funding

This work was supported by the National Institute of Environmental Health Sciences (R01ES019296) and was conducted with government support under and awarded by the Department of Defense, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a, to C.B.V.

### Notes

The authors declare no competing financial interest.


## ACKNOWLEDGMENTS


We thank Prof. Tom Tullius and Shakti Ingle (Boston University, Boston, MA) for advice pertaining to hydroxyl radical footprinting experiments and Kelly Schermerhorn and Ji Huang for helpful discussions.


## ABBREVIATIONS

TNRs, trinucleotide repeats; HD, Huntington's disease; DM1, myotonic dystrophy; DRPLA, dentatorubral-pallidoluysian atrophy; NCP, nucleosome core particle; DMT, dimethoxytrityl; TEAA, triethylammonium acetate; EDTA, ethylenediaminetetraacetic acid; PMSF, phenylmethanesulfonyl fluoride; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate.

## REFERENCES


(1) Gatchel, J. R., and Zoghbi, H. Y. (2005) Diseases of unstable repeat expansion: Mechanisms and common principles. *Nat. Rev. Genet. 6*, 743−755.

(2) McMurray, C. T. (2010) Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet. 11*, 786−799.

(3) López Castel, A., Cleary, J. D., and Pearson, C. E. (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol. 11*, 165−170.

(4) Mirkin, S. M. (2007) Expandable DNA repeats and human disease. *Nature 447*, 932−940.

(5) Liu, Y., and Wilson, S. H. (2012) DNA base excision repair: A mechanism of trinucleotide repeat expansion. *Trends Biochem. Sci. 37*, 162−172.

(6) The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell 72*, 971−983.

(7) Dion, V., and Wilson, J. H. (2009) Instability and chromatin structure of expanded trinucleotide repeats. *Trends Genet. 25*, 288−297.

(8) Wang, Y.-H. (2007) Chromatin structure of repeating CTG/CAG and CGG/CCG sequences in human disease. *Front. Biosci. 12*, 4731−4741.

(9) Goula, A.-V., Berquist, B. R., Wilson, D. M., Wheeler, V. C., Trottier, Y., and Merienne, K. (2009) Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. *PLoS Genet. 5*, e1000749.

(10) Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature 389*, 251−260.

(11) Kornberg, R. D. (1974) Chromatin structure: A repeating unit of histones and DNA. *Science 184*, 868−871.

(12) Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., and Klug, A. (1984) Structure of the nucleosome core particle at 7 Å resolution. *Nature 311*, 532−537.

(13) Hayes, J. J., Tullius, T. D., and Wolffe, A. P. (1990) The structure of DNA in a nucleosome. *Proc. Natl. Acad. Sci. U.S.A. 87*, 7405−7409.

(14) Hayes, J. J., Clark, D. J., and Wolffe, A. P. (1991) Histone contributions to the structure of DNA in the nucleosome. *Proc. Natl. Acad. Sci. U.S.A. 88*, 6829−6833.

(15) Wang, Y. H., Amirhaeri, S., Kang, S., Wells, R. D., and Griffith, J. D. (1994) Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. *Science 265*, 669−671.

(16) Wang, Y. H., and Griffith, J. (1995) Expanded CTG triplet blocks from the myotonic dystrophy gene create the strongest known natural nucleosome positioning elements. *Genomics 25*, 570−573.

(17) Hayes, J. J., and Lee, K. M. (1997) In vitro reconstitution and analysis of mononucleosomes containing defined DNAs and proteins. *Methods 12*, 2−9.

(18) Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M., and Richmond, T. J. (1997) Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol. 272*, 301−311.

(19) Shrader, T. E., and Crothers, D. M. (1989) Artificial nucleosome positioning sequences. *Proc. Natl. Acad. Sci. U.S.A. 86*, 7418−7422.

(20) Ober, M., and Lippard, S. J. (2007) Cisplatin damage overrides the predefined rotational setting of positioned nucleosomes. *J. Am. Chem. Soc. 129*, 6278−6286.

(21) Studitsky, V. M., Clark, D. J., and Felsenfeld, G. (1994) A histone octamer can step around a transcribing polymerase without leaving the template. *Cell 76*, 371−382.

(22) Wang, Y. H., Gellibolian, R., Shimizu, M., Wells, R. D., and Griffith, J. (1996) Long CCG triplet repeat blocks exclude nucleosomes: A possible mechanism for the nature of fragile sites in chromosomes. *J. Mol. Biol. 263*, 511−516.

(23) Wang, Y. H., and Griffith, J. (1996) Methylation of expanded CCG triplet repeat DNA from fragile X syndrome patients enhances nucleosome exclusion. *J. Biol. Chem. 271*, 22937−22940.

(24) Mulvihill, D. J., Nichol Edamura, K., Hagerman, K. A., Pearson, C. E., and Wang, Y.-H. (2005) Effect of CAT or AGG interruptions and CpG methylation on nucleosome assembly upon trinucleotide repeats on spinocerebellar ataxia, type 1 and fragile X syndrome. *J. Biol. Chem. 280*, 4498−4503.

(25) Metzenberg, S. (1996) On the formation of nucleosomes within the FMR1 trinucleotide repeat. *Am. J. Hum. Genet. 59*, 252−253.

(26) Satchwell, S. C., and Travers, A. A. (1989) Asymmetry and polarity of nucleosomes in chicken erythrocyte chromatin. *EMBO J. 8*, 229−238.

(27) Sivolob, A. V., and Khrapunov, S. N. (1995) Translational positioning of nucleosomes on DNA: The role of sequence-dependent isotropic DNA bending stiffness. *J. Mol. Biol. 247*, 918−931.

(28) Chastain, P. D., and Sinden, R. R. (1998) CTG repeats associated with human genetic disease are inherently flexible. *J. Mol. Biol. 275*, 405−411.

(29) Lowary, P. T., and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol. 276*, 19−42.

(30) Godde, J. S., Kass, S. U., Hirst, M. C., and Wolffe, A. P. (1996) Nucleosome assembly on methylated CGG triplet repeats in the fragile X mental retardation gene 1 promoter. *J. Biol. Chem. 271*, 24325−24328.

(31) Hsu, Y. Y., and Wang, Y.-H. (2002) Human fragile site FRA16B DNA excludes nucleosomes in the presence of distamycin. *J. Biol. Chem. 277*, 17315−17319.

(32) Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I. K., Wang, J.-P. Z., and Widom, J. (2006) A genomic code for nucleosome positioning. *Nature 442*, 772−778.

(33) Godde, J. S., and Wolffe, A. P. (1996) Nucleosome assembly on CTG triplet repeats. *J. Biol. Chem. 271*, 15222−15229.

(34) Lutter, L. C. (1979) Precise location of DNase I cutting sites in the nucleosome core determined by high resolution gel electrophoresis. *Nucleic Acids Res. 6*, 41−56.
(35) Fox, K. R. (1998) DNase I Footprinting. *Methods Mol. Biol. 90*, 1−22.
(36) Vitolo, J. M., Thiriet, C., and Hayes, J. J. (2001) DNase I and hydroxyl radical characterization of chromatin complexes. *Current Protocols in Molecular Biology*, Chapter 21, Unit 21.4, Wiley, New York.
(37) Lutter, L. C. (1989) Digestion of nucleosomes with deoxyribonucleases I and II. *Methods Enzymol. 170*, 264−269.
(38) Drew, H. R., and Travers, A. A. (1984) DNA structural variations in the *E. coli* tyrT promoter. *Cell 37*, 491−502.
(39) Drew, H. R. (1984) Structural specificities of five commonly used DNA nucleases. *J. Mol. Biol. 176*, 535−557.
(40) Hogan, M. E., Roberson, M. W., and Austin, R. H. (1989) DNA flexibility variation may dominate DNase I cleavage. *Proc. Natl. Acad. Sci. U.S.A. 86*, 9273−9277.
(41) Brukner, I., Jurukovski, V., and Savic, A. (1990) Sequence-dependent structural variations of DNA revealed by DNase I. *Nucleic Acids Res. 18*, 891−894.
(42) Davis, W. B., Bjorklund, C. C., and Deline, M. (2012) Probing the Effects of DNA-Protein Interactions on DNA Hole Transport: The N-Terminal Histone Tails Modulate the Distribution of Oxidative Damage and Chemical Lesions in the Nucleosome Core Particle. *Biochemistry 51*, 3129−3142.
(43) Riley, D., and Weintraub, H. (1978) Nucleosomal DNA is digested to repeats of 10 bases by exonuclease III. *Cell 13*, 281−293.
(44) Prunell, A. (1983) Periodicity of exonuclease III digestion of chromatin and the pitch of deoxyribonucleic acid on the nucleosome. *Biochemistry 22*, 4887−4894.
(45) Archer, T., and Ricci, A. (1999) Exonuclease III as a probe of chromatin structure in vivo. *Methods Enzymol. 304*, 584−599.
(46) Linxweiler, W., and Hörz, W. (1982) Sequence specificity of exonuclease III from E. coli. *Nucleic Acids Res. 10*, 4845−4859.
(47) Jain, S. S., and Tullius, T. D. (2008) Footprinting protein-DNA complexes using the hydroxyl radical. *Nat. Protoc. 3*, 1092−1100.
(48) Tullius, T. D., and Dombroski, B. A. (1986) Hydroxyl radical "footprinting": High-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. *Proc. Natl. Acad. Sci. U.S.A. 83*, 5469−5473.
(49) Burrows, C. J., and Muller, J. G. (1998) Oxidative Nucleobase Modifications Leading to Strand Scission. *Chem. Rev. 98*, 1109−1152.
(50) Otten, A. D., and Tapscott, S. J. (1995) Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc. Natl. Acad. Sci. U.S.A. 92*, 5465−5469.
(51) Klesert, T. R., Otten, A. D., Bird, T. D., and Tapscott, S. J. (1997) Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. *Nat. Genet. 16*, 402−406.
(52) Thornton, C. A., Wymer, J. P., Simmons, Z., McClain, C., and Moxley, R. T. (1997) Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. *Nat. Genet. 16*, 407−409.
(53) Sobczak, K., and Krzyzosiak, W. J. (2005) CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts. *J. Biol. Chem. 280*, 3898−3910.
(54) Amrane, S., Saccà, B., Mills, M., Chauhan, M., Klump, H. H., and Mergny, J.-L. (2005) Length-dependent energetics of $(CTG)_n$ and $(CAG)_n$ trinucleotide repeats. *Nucleic Acids Res. 33*, 4065−4077.
(55) Allemand, J. F., Bensimon, D., Lavery, R., and Croquette, V. (1998) Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proc. Natl. Acad. Sci. U.S.A. 95*, 14152−14157.
(56) Randall, G. L., Zechiedrich, L., and Pettitt, B. M. (2009) In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form. *Nucleic Acids Res. 37*, 5568−5577.
(57) Lionberger, T. A., Demurtas, D., Witz, G., Dorier, J., Lillian, T., Meyhöfer, E., and Stasiak, A. (2011) Cooperative kinking at distant sites in mechanically stressed DNA. *Nucleic Acids Res. 39*, 9820−9832.