Nucleosome Structural Features and Intrinsic Properties of the TATAAACGCC Repeat Sequence*

Hans R. Widlund‡§, Prasad N. Kuduvalli¶, Martin Bengtsson‡, Hui Cao‡, Thomas D. Tullius‖, and Mikael Kubista‡ **

From the ‡Department of Biochemistry and Biophysics, Lundberg Institute, Chalmers University of Technology, SE 413 90 Göteborg, Sweden, the ¶Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205-2185, and the ‖Department of Chemistry, Boston University, Boston, Massachusetts 02215

Nucleosomes, the fundamental building blocks of chromatin, play an architectural role in ensuring the integrity of the genome and act as a regulator of transcription. Intrinsic properties of the underlying DNA sequence, such as flexibility and intrinsic bending, direct the formation of nucleosomes. We have earlier identified genomic nucleosome-positioning sequences with increased *in vitro* ability for nucleosome formation. One group of sequences bearing a 10-base pair consensus repeat sequence of TATAAACGCC had the highest reported nucleosome affinity from genomic material. Here, we report the intrinsic physical properties of this sequence and the structural details of the nucleosome it forms, as analyzed by footprinting techniques. The minor groove is buried toward the histone octamer at the AA steps and facing outwards at the CC steps. By cyclization kinetics, the overall helical repeat of the free DNA sequence was found to be 10.5 base pairs/turn. Our experiments also showed that this sequence is highly flexible, having a *J*-factor 25-fold higher than that of random sequence DNA. In addition, the data suggest that twist flexibility is an important determinant for translational nucleosome positioning, particularly over the dyad region.

DNA packaging into nucleosomes, the basic repeating units of chromatin, involves the wrapping of 146 bp[1] of double-stranded DNA into almost two complete turns around the histone octamer. The histone proteins have been highly conserved through evolution and are designed to bind to virtually any DNA sequence within the nucleus. There are, however, several known sequences that show a considerably higher ability to bind the histone octamer compared with bulk DNA.

About 90% of the DNA in an eukaryotic cell is complexed with histones to form chromatin fibers. This represents a tremendous obstacle to transcription, replication, and repair machinery that requires access to these DNA regions (1). The location of a nucleosome on the DNA sequence is determined by several factors. At the primary level of compaction, the DNA sequence itself is responsible for determining whether or not a nucleosome is positioned due to inherent intrinsic mechanical properties. *In vivo*, secondary effects, such as the interaction of DNA with non-histone proteins and other ligands, and boundary effects can determine the basic and higher order positioning of nucleosomes in chromatin (2).

Several DNA sequence motifs have been studied in an effort to determine the organization of nucleosome-positioning signals at the level of primary DNA sequence. Travers and co-workers (3) investigated the sequence properties of the DNA in a library of nucleosomal DNA from chicken erythrocytes. They found that AA/TT dinucleotides were present where the minor groove was compressed and facing inward toward the histone octamer. Conversely, CG/CC dinucleotides were located where the minor groove was wider and facing outwards. These dinucleotides also showed a preferential distribution of 10–11-bp periodicity, indicating the importance of anisotropic DNA bendability in nucleosome positioning (4). Highly flexible poly(A-T) DNA has been shown to incorporate into nucleosomes more readily than bulk DNA (5). This and other work suggest that DNA in the nucleosome is under torsional stress, and consequently, more flexible sequences would be favored to position nucleosomes.

Based on these experimental data, an artificial nucleosome-positioning sequence, denoted TG, was constructed (6). This motif contains a $5'$-$(A/T)_3NN(G/C)_3NN$-$3'$ sequence repeated 10 times in the TG-5 sequence. TG-5 showed high affinity for binding the histone octamer *in vitro*, but failed to position a nucleosome *in vivo* (7, 8). These results suggest that anisotropic bendability that allows strong rotational positioning is not sufficient to position nucleosomes *in vivo*.

Widom and coworkers (9 and 10) advocate that nucleosome positioning is not a precise mechanism, but rather a thermodynamic equilibrium system. Since nucleosomes are dynamic entities formed under equilibrium conditions, there is always a statistical opportunity that every possible site is occupied at one time or another. However, some sites are preferred over the others as the system evolves toward a minimum in free energy. Occupancy of these sites would be expected to follow a Boltzmann distribution. In an *in vitro* system, it is possible to create conditions in which no other factors are present that are capable of affecting the free energy of the system. In cells, of course, there are a number of such factors that either passively (DNA

§ Supported by the Sven and Lilly Lawski Foundation. To whom correspondence may be addressed: Dept. of Pediatric Oncology, Dana-Farber Cancer Inst., Dana 630, Harvard Medical School, 44 Binney St., Boston, MA 02115. Fax: 617-632-2085; E-mail: Hans_Widlund@dfci.harvard.edu.

** To whom correspondence may be addressed: Dept. of Molecular Biotechnology, Lundberg Inst., Chalmers University of Technology, Box 462, Medicinaregatan 9C, SE 405 30 Göteborg, Sweden. E-mail: Mikael.Kubista@molbiotech.chalmers.se.

[1] The abbreviations used are: bp, base pair(s); PCR, polymerase chain reaction; MNase, micrococcal nuclease.

5′ -**TATAAACGCC**-3′
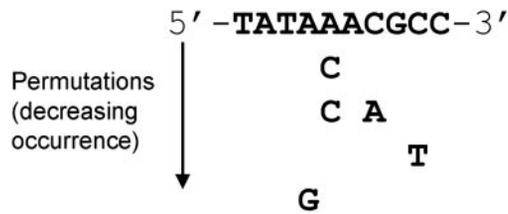C
C A
T
G

Permutations
(decreasing
occurrence)

FIG. 1. **Consensus sequence of the genomic TATAAACGCC repeat sequence used in this work.** Permutations of this repeat are indicated in order of decreased occurrence in the direction of the *arrow*.

secondary structures and DNA-binding proteins) or actively (SWI-SNF, CHRAC, and NURF) redistribute the nucleosomes to various degrees (11, 12).

In a previous study, we selected DNA sequences that form stable nucleosome core particles *in vitro* from a large pool of mouse genomic nucleosomal DNA (13). We found that repeats of a 10-bp consensus sequence (5′-TATAAACGCC-3′) had the highest affinity for binding histone octamers (Fig. 1). This motif also contains alternating A/T- and G/C-rich elements in phase with the helical repeat, much like the artificial TG pentamer. It is therefore of particular interest to characterize the intrinsic curvature and flexibility of this DNA as well as the structure of the nucleosome core particle assembled on this sequence. In addition to forming exceedingly stable nucleosome core particles with high affinity, this sequence also contains multiple putative binding sites for the high mobility group I/Y protein (14). The nucleosome-positioning properties of this sequence could also have important ramifications for the binding and activity of the high mobility group I/Y protein on sequences bearing this motif *in vivo*.

EXPERIMENTAL PROCEDURES

Oligonucleotides—PCR primers used in these experiments were as follows: [+0]-FWD, 5′-CGGAATTCAGATCTTCCTGGGAAAACCCT-GG; [+2]-FWD, 5′-CGGAATTCAGATCTCTTCCTGGGAAAACCCT; [+4]-FWD, 5′-CTCAGATCTTTCTTCCTGGGAAAACCCTGG; [+20]-FWD, 5′-CGACGGTATCGATAAGCTTG; [+0]-REV, 5′-CGGGATCCC-GAGCTGTTTCCTGTGT; [+7]-REV, 5′-GCGGGATCCCGGTTCCCGA-GCTGTTTCCTGTGT; and [+80]-REV, 5′-AATAACCCTCACTA-AAGGGA.

Preparation of Histone H1-depleted Chromatin—Fresh adult chicken blood was suspended in sodium citrate (0.8% final concentration) to prevent clotting. The erythrocytes were spun down (Beckman 5.2 rotor; 3000 rpm for 45 min at 4 °C), and the white coat on top of the red blood cells was removed by suction. The cells were then washed with 0.14 M NaCl, 15 mM sodium citrate, 10 mM Tris-HCl (pH 8), and 0.25 mM 4-(2-aminoethyl)benzenesulfonyl fluoride. After washing, the cells were either stored at −80 °C or used directly. Cells were resuspended and lysed in cold suspension buffer (0.34 M sucrose, 60 mM KCl, 15 mM NaCl, 1 mM CaCl$_2$, 15 mM Tris-HCl (pH 8), 15 mM β-mercaptoethanol, 0.5 mM spermidine, 0.15 mM spermine, and 0.25 mM 4-(2-aminoethyl)benzene-sulfonyl fluoride) containing 0.1% Nonidet P-40 using 10 ml of buffer/ml of red blood cells. The cell suspension was filtered through a funnel packed with glass wool to remove cell debris. The filtrate was centrifuged (Beckman JA10 rotor; 6500 rpm for 45 min at 4 °C), and the nuclei were isolated. The pellet was washed with suspension buffer until it was white. The pellet containing the nuclei was resuspended in 1 ml of suspension buffer/ml of starting volume of red blood cells and either frozen at −80 °C or used immediately. The $A_{260\ nm}$ in 1 M NaOH was ~150/ml of red blood cells. Chromatin was prepared by micrococcal nuclease digestion of nuclei for 15 min at 37 °C using 100 units of enzyme/ml of nuclei. Nuclei were pelleted (4000 × g for 5 min at 4 °C), and the soluble chromatin was extracted with 1 mM EDTA (pH 8) carefully on ice. Insoluble debris was removed by centrifugation (4000 × g for 10 min at 4 °C). The supernatant was made 0.65 M in NaCl carefully on ice. Histone H5/H1 was removed by treatment with 2 g of ion-exchange resin/ml (AG-Dow 50W-2X (Bio-Rad), which was pre-washed with 1 M Tris (pH 8) (twice), followed by 10 mM Tris (pH 8), 1 mM EDTA (pH 8), and 0.7 M NaCl). Depletion of histone H5/H1 was accomplished by carefully stirring the mixture on ice for 10 min and then removal of the ion-exchange resin by centrifugation (1000 × g for 5 min

at 4 °C). The histone H5/H1-depleted chromatin was then dialyzed against 0.25 mM EDTA (pH 8) with 1 mM β-mercaptoethanol. The histone content and the integrity of the chromatin were monitored by 15% SDS-polyacrylamide gel electrophoresis. In the nucleosome reconstitution experiments, linker histone-depleted chromatin at $A_{260\ nm}$ = 20 was used as the histone donor.

Salt-induced in Vitro Reconstitution of Nucleosomes—Nucleosomes were reconstituted by stepwise dilution using long histone H1-depleted chromatin as the histone donor by a modified procedure (6). The reaction mixture contained 6.5 μg of histone donor chromatin, 1 μg of probe, and, when selective pressure was applied, 10 μg of calf thymus DNA in a volume of 10 μl. The sample was incubated at high salt (1 M NaCl, 20 mM Tris (pH 7.2), and 0.1% Nonidet P-40) for 30 min at 37 °C prior to dilution by low salt buffer (20 mM Tris (pH 7.2) and 0.1% Nonidet P-40). The salt concentration was lowered by three additions of low salt buffer at 20-min intervals. The sample was then kept at 37 °C for an additional 60 min. Reconstituted nucleosomes were analyzed by electrophoresis on a native 5% (19:1) polyacrylamide gel run in 1× TBE (50 mM Tris, 50 mM boric acid, and 10 mM EDTA) at 4 °C. Prior to loading the samples, glycerol was added to 5% (v/v) (without dyes). The gel was run at 4 watts for 1 h, dried, and analyzed by a PhosphorImager and/or autoradiographed.

Generation of Radiolabeled DNA—Plasmid pHCn41 (a derivative of pCRScript) carrying the appropriate insert of the TATAAACGCC repeat sequence was introduced into *Escherichia coli* DH5α cells grown in LB/ampicillin medium. The plasmid was purified using a QIAGEN flow column. Plasmid DNA was extracted once with phenol and chloroform. The plasmid was digested with either *Eco*RI or *Bam*HI. The DNA overhangs were filled using the Klenow fragment of DNA polymerase and [α-^{32}P]dATP (as well as [α-^{32}P]dGTP for *Bam*HI-digested DNA). Radiolabeled plasmid DNA was digested with a second restriction endonuclease, electrophoresed on a 6% (19:1) polyacrylamide gel, and recovered by electroelution. Radiolabeled PCR probes were generated either by the incorporation of [α-^{32}P]dNTPs (for cyclization experiments) or by phosphorylation using T4 polynucleotide kinase and [γ-^{32}P]ATP (for footprinting and affinity electrophoretic mobility shift assay experiments).

Footprinting Experiments—Reconstituted nucleosomes and free DNA were treated with the hydroxyl radical according to previously described methods (15), extracted with phenol and chloroform, and ethanol-precipitated. Micrococcal nuclease (MNase) footprinting was performed by adding 0.1 unit of MNase/μl of the reconstitution mixture in 5 mM CaCl$_2$. The reaction was allowed to proceed for 1 min at room temperature and stopped by the addition of EDTA, and the sample was put on ice. The samples was extracted with phenol and chloroform and then ethanol-precipitated. MNase footprinting was performed on both restriction fragments and PCR-amplified DNA. Samples were dissolved in formamide loading buffer and denatured at 95 °C for 5 min prior to loading on a 8% denaturing polyacrylamide gel. (A + G)-specific Maxam-Gilbert sequencing markers were used. Following electrophoresis, the gel was dried and analyzed on a PhosphorImager.

Cyclization Kinetics—Labeled DNA molecules were generated using PCR amplification with [α-^{32}P]dATP and [α-^{32}P]dGTP. A set of five different primers ([+0]-FWD, [+2]-FWD, [+4]-FWD, [+0]-REV, and [+7]-REV) was used in pairs to create a set of molecules differing in length over one helical turn, in steps of 2–3 bp. These DNA molecules were purified and cleaved on both sides using *Bgl*II and *Bam*HI. These two enzymes leave cohesive ends upon cleavage, making it possible to ligate the ends by either bimolecular association or unimolecular circularization. Ligation was carried out in a volume of 70 μl of 50 mM Tris (pH 7.6), 10 mM MgCl$_2$, 1 mM ATP, 1 mM dithiothreitol, and 5% polyethylene glycol 8000 with 1 unit of ligase at room temperature. The DNA concentration in each sample was 60 nM as measured by absorption spectrophotometry. Samples of 12 μl were removed at 30 s and 1, 3, and 10 min. The ligation reaction was stopped by the addition of 8 μl of stop buffer (0.4 mg/ml proteinase K, 30% glycerol, 75 mM EDTA (pH 8), and 0.25% bromphenol blue) and then by heating at 70 °C for 5 min. Bal-31 exonuclease was added to aliquots of the trial samples after ligation to verify the presence of cyclized products. Samples were electrophoresed on a native 6% (30:1) polyacrylamide gel run in 0.5× TBE at 6 watts for 2 h at room temperature. The gel was dried and analyzed using a PhosphorImager.

RESULTS

We previously selected the TATAAACGCC repeat sequence from a mouse genomic library of DNA associated with nucleosome core particles (13). This 180-bp sequence has a very high
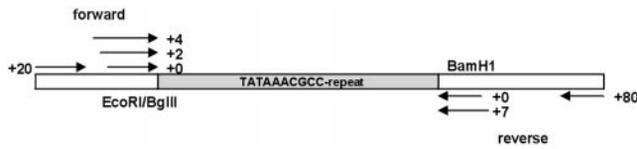
FIG. 2. **Schematic drawing of the different TATAAACGCC repeat constructs used.** Indicated are restriction sites and the primers used for PCR.

affinity for the histone octamer, ~350-fold higher than random sequence DNA (16).

Intrinsic Curvature and Flexibility of the TATAAACGCC Repeat Sequence—To characterize the nucleosome core particle formed on the TATAAACGCC repeat sequence, we first evaluated the DNA sequence for intrinsic curvature and flexibility. Previously, we reported that this sequence displayed a gel migration anomaly of $R_L = 1.2$, indicating moderate intrinsic curvature of the DNA. We conducted further gel migration assays using TBMg buffer (17) to more closely mimic physiological conditions. This yielded a gel migration anomaly of $R_L = 1.3$ (data not shown), indicating that the presence of divalent Mg^{2+} does not significantly alter the intrinsic curvature of the sequence. Sequence-dependent DNA flexibility has been suggested to play an important role in the positioning of nucleosome core particles. Cyclization kinetics is the method of choice for determining the intrinsic flexibility of a given DNA sequence (Ref. 18 and thoroughly described in Ref. 19). This method analyzes the equilibrium between monomer DNA molecules and either linear dimer DNA molecules or circularized monomer DNA molecules. A comparison of the rate of forming linear dimer DNA *versus* the rate of forming circularized monomer DNA yields the probability of ring closure (*J*-factor). The probability depends on two mechanical properties of the DNA, the torsional flexibility and the writhe.

For our cyclization experiments, we have used a set of PCR-generated DNA molecules (with a total length of ~180 bp) that have cohesive ends and that differ in length in steps of 2–3 bp over a complete turn of duplex DNA (Fig. 2). Fig. 3*A* displays a native gel on which the products of a typical cyclization reaction were separated. We found that the TATAAACGCC repeat sequence has an overall helical repeat of 10.5 bp/turn as determined by the cyclization maximum (Fig. 3, *A* and *C*). This is in accordance with the expected helical repeat for mixed sequence B-DNA.

The *J*-factors for these fragments yield further information regarding the flexibility of the sequence. The high cyclization probability (*J*-factor = 400 nM, as shown in Fig. 3*C*) is indicative of high intrinsic flexibility for this sequence. This is ~25-fold higher than the *J*-factor for mixed sequence DNA of the same length (20). Moreover, the small amplitude of 300 nM in the *J*-factor over one helical turn (Fig. 3*C*) also suggests that the TATAAACGCC repeat sequence is prone to twist flexibility. This could play a very important role in positioning the sequence with a helical twist of 10.5 bp/turn in a nucleosome core particle, where the average helical twist is 10.2 bp/turn (21, 22).

Footprinting Studies—Nucleosome core particles reconstituted on the TATAAACGCC repeat sequence were footprinted by hydroxyl radicals as pure species (Fig. 4). This method is well suited for characterizing the translational and rotational positioning of the DNA in the core particle. The location of the pseudodyad axis of the nucleosome, as well as the positioning of the minor groove with respect to the histone octamer at any given point in the sequence, can readily be determined by hydroxyl radical footprinting (21).

The hydroxyl radical footprint of the nucleosome core particle positioned on the TATAAACGCC repeat sequence, shown
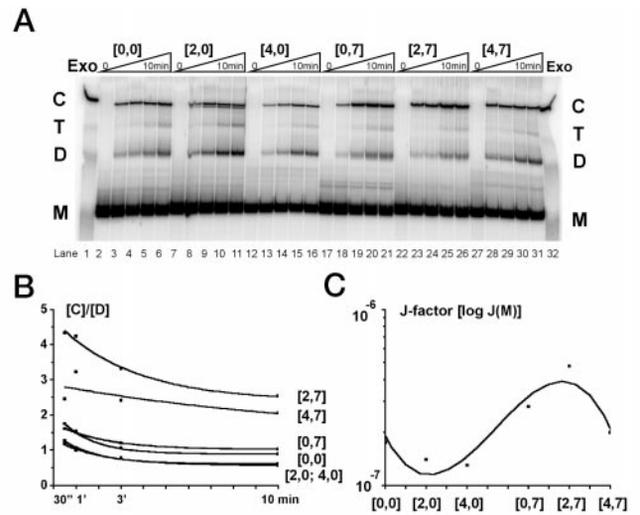

FIG. 3. *A*, cyclization kinetics. *Lanes 1* and *32*, Bal-31 exonuclease digest (*Exo*); *lanes 2–6*, construct [+0,+0] at 0 and 30 s and 1, 3, and 10 min; *lanes 7–11*, construct [+2,0] at 0 and 30 s and 1, 3, and 10 min; *lanes 12–16*, construct [+4,0] at 0, 30 s, 1 min, 3 min and 10 min; *lanes 17–21*, construct [0,+7] at 0, 30 s, 1 min, 3 min and 10 min; *lanes 22–26*, construct [+2,+7] at 0, 30 s, 1 min, 3 min and 10 min; *lanes 27–31*, construct [+4,+7] at 0, 30 s, 1 min, 3 min and 10 min. *Band M*, monomer; *band D*, linear dimer; *band T*, trimer; *band C*, cyclized monomer. To verify the identity of the cyclized monomer product, we recovered the DNA from one gel and tried to cleave it with *Bgl*II and *Bam*HI. We did not observe any cleavage, suggesting that *band C* is the cyclized monomer. *B*, plot of [*C*]/[*D*] for the individual fragments at the sampled time points. *C*, *J*-factors, calculated as follows,

$$J = 2M_0 \lim_{t \to 0} \frac{[C]}{[D]}, \qquad \text{(Eq. 1)}$$

for a slow proceeding parallel reaction (34), where $M_0 = 60$ nM. The maximum (400 nM) is found at +9 (construct [+2,+7]), and the minimum between +2 and +4. Note the small amplitude of only 300 nM. The average number of base pairs/turn is calculated as the total fragment length at cyclization maximum (179 bp) divided by an integer number of turns, here being 17 as the only appropriate choice. This yields an average of 10.52 bp/turn.
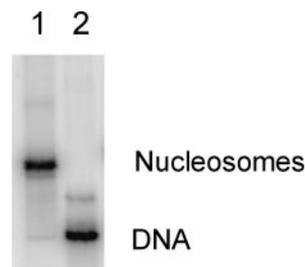

FIG. 4. **Nucleosome reconstitution on the TATAAACGCC repeat sequence.** *Lane 1*, salt-induced nucleosome reconstitution mixture; *lane 2*, free probe DNA. Note in *lane 1* that there is a faint band below the main nucleosome band, possibly reflecting an additional weak translational position.

in Fig. 5, reveals several interesting features. This sequence is capable of positioning a nucleosome core particle with a predominantly unique translational position. The DNA pseudodyad axis of symmetry is easily identified by the altered helical repeat of 10.7 bp/turn for the three helical turns around the histone dyad axis. The rest of the nucleosome core particle has a helical repeat of 10.0 bp/turn. This is in perfect agreement with other well characterized nucleosome core particles with unique transitional positioning (21). In addition, the DNA is in a single rotational position about the histone octamer. For the A-rich strand, the AA dinucleotides are positioned where
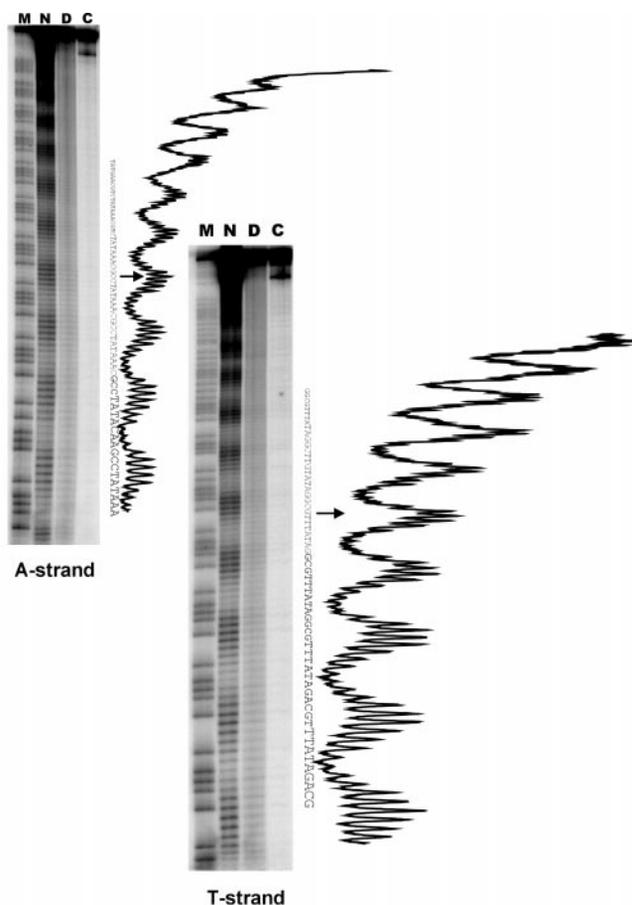
FIG. 5. **Hydroxyl radical footprinting and densitometric analysis of the dyad region.** *Lane M,* marker Maxam-Gilbert (A + G)-specific sequencing reaction; *lane N,* hydroxyl radical footprint of the nucleosome; *lane D,* hydroxyl radical cleavage pattern of naked DNA; *lane C,* control untreated DNA. The nucleosome dyad (*arrow*) is pinpointed by the change of periodicity in the cleavage intensity from 10.0 to 10.7 nucleotides/turn (see "Results").

FIG. 6. **MNase and hydroxyl radical footprinting over the ends of the nucleosome core DNA.** *Lane M,* marker Maxam-Gilbert (A + G)-specific sequencing reaction; *lane N,* hydroxyl radical footprint of the nucleosome; *lane D,* hydroxyl radical cleavage pattern of naked DNA; *lane C,* control untreated DNA; *lane Mn,* products of the MNase footprinting reaction. The nucleosome boundaries are marked, and there is a clear alteration from the periodical 10-bp cleavage at the ends.

the minor groove faces inward (toward the histone octamer), and the CC dinucleotides are positioned such that the minor groove faces outwards (away from the histone octamer). For the opposite, T-rich strand, the TA dinucleotide steps are located where the minor groove faces inward, and the CG step has the minor groove facing outwards. An offset of 2 bp in the 3′-direction is seen between the two strands, which is typical for minor groove binding. These results are consistent with the expected phasing of A/T- and G/C-rich sequences in the nucleosome core particle (4, 5, 6). Furthermore, this rotational setting, with the A/T-rich regions having the minor groove buried toward the histone octamer, restricts the binding of A/T minor groove-recognizing proteins, *e.g.* the high mobility group I/Y protein (14). Finally, it is also interesting to note that the naked DNA itself shows a cleavage pattern somewhat similar to that of the nucleosome, but less intense, as would be expected for a repeated sequence containing alternating A/T- and G/C-rich regions. Thus, the TATAAACGCC repeat DNA appears to be predisposed to formation of the structure it adopts when complexed with the histone octamer. We have also used MNase footprinting to define the boundaries of the nucleosome and as an independent measure of translational positioning of the core particle. Fig. 6 shows the periodic (every 10 bp) MNase cleavage of the DNA within the central 120 bp of the nucleosome core particle. This periodic cleavage is not observed at the very ends of the nucleosome core particle. This is consistent with the DNA in this region being loosely associated with the histone octamer (21, 22). These studies confirm and comple-
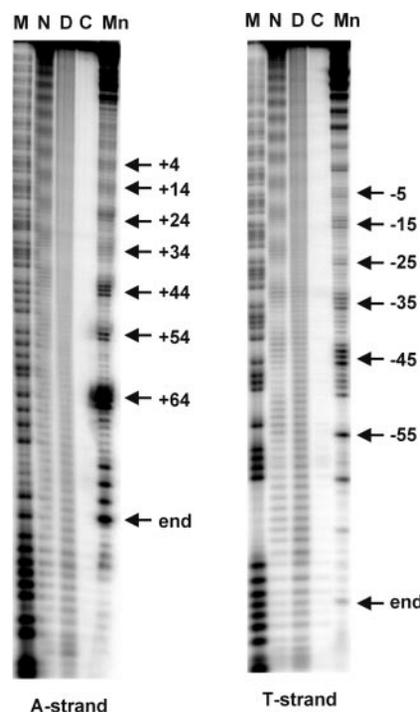
ment the results of hydroxyl radical footprinting (Fig. 7).

Multiple Translational Positioning—The TATAAACGCC repeat sequence has a 350-fold higher affinity for the histone octamer compared with mixed sequence DNA, as reported in our earlier work (13, 16). Since this is a consensus repeat sequence, we were interested in determining whether the core particle formed with a longer DNA fragment would be positioned over the center of the repeat if flanked by other sequence. We compared the original 180-bp fragment with a longer, 280-bp fragment in translational positioning and overall affinity of binding the histone octamer. Fig. 8 shows the high resolution electrophoretic mobility shift assay for nucleosome core particles reconstituted on the two fragments. Both fragments bind the histone octamer with similar overall affinity (difference $\Delta(\Delta G)$ = 300 cal/mol as compared with $\Delta(\Delta G)$ = 4500 cal/mol relative to random sequence DNA). It is evident from Fig. 8 that the longer, 280-bp fragment shows at least six different translational positions in contrast to the predominant single (perhaps two?) translational positions seen for the 180-bp fragment. We have used hydroxyl radical and MNase footprinting to characterize the ensemble of translational positions for the 280-bp fragment. We still found that the dominant translational position is the same as that previously seen for the 180-bp fragment (Fig. 9). The ensemble of translational positions also shows the same rotational setting of the DNA about the histone octamer.

DISCUSSION

Our results clearly show that both intrinsic flexibility and curvature contribute to making the TATAAACGCC repeat DNA a high affinity nucleosome-positioning sequence. The moderate gel migration anomaly (R_L = 1.2–1.3) is probably a reflection of reduced macroscopic curvature caused by the increased flexibility of the DNA fragment. The shorter, 180-bp fragment containing this motif is able to uniquely position a
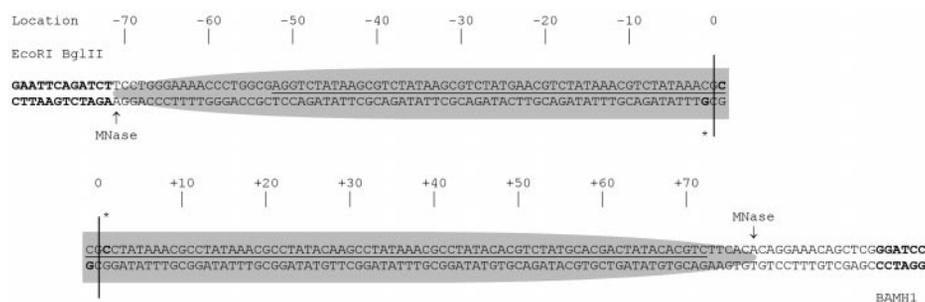
FIG. 7. **Location of the TATAAACGCC repeat in the nucleosome.** *Asterisks* indicate the locations of the nucleosome pseudodyad on each strand, based on densitometric analysis of hydroxyl radical cleavage patterns. The *lines* indicate the base pair falling on the dyad, and the numbers of bases from this point are marked. Marked by *arrows* are MNase cleavage sites at the ends of the nucleosome. The location of the nucleosome on the sequence is marked by the *shaded area*.
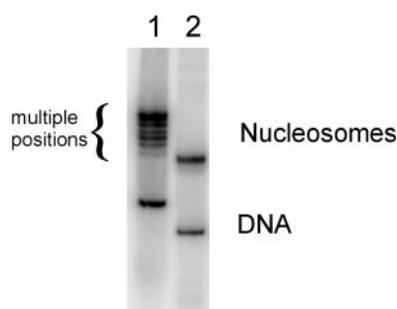

FIG. 8. **Electrophoretic mobility shift assay of nucleosomes formed on two different length DNA fragments.** Shown are the results from the nucleosome reconstitution with the TATAAACGCC repeat sequence. *Lane 1*, the 280-bp fragment; *lane 2*, the 183-bp fragment. The affinity (total over all translational positions) for the long fragment is about equal ($\Delta(\Delta G) = 300$ cal/mol) to that of the short fragment, as measured by densitometry. In this experiment an additional 1 μg of competitor calf thymus DNA/μl was present.

nucleosome core particle as seen by footprinting with hydroxyl radicals and MNase. In the longer, 280-bp fragment, where the central TATAAACGCC repeat is offset by 80 bp with respect to the center of the fragment, we observed multiple translational settings. However, there is a strong preference for the central TATAAACGCC repeat, which is not located at the center of the 280-bp fragment. This goes to further demonstrate the strong nucleosome-positioning properties of this sequence.

The TATAAACGCC Repeat Is Highly Flexible—The *J*-factor of 400 n$_M$ observed for this sequence compare well with those reported for other highly flexible molecules. Highly flexible sequences, such as CTG and CGG repeats involved in triplet expansion diseases, have *J*-factors in the range of 200–300 n$_M$ (20). These sequences have also been reported to have high affinity for histone octamers *in vitro* (13, 23, 24). Interestingly, the highly flexible CTG repeat was reported to be favored at the dyad, suggesting that high flexibility, especially in twist, is essential for nucleosome positioning over that region (25). This can explain the multiple translational positions that we observed for the longer, 280-bp fragment. On this longer fragment, the DNA is able to occupy multiple positions because it is able to adjust the pseudodyad twist at multiple positions.

The artificial TG-5 nucleosome-positioning sequence failed to position a nucleosome *in vivo* (7). This was later reinvestigated using a modified version of TG-5 containing an extra base pair in the center to accommodate the constraints at the dyad (8). Not surprisingly, in retrospect, this sequence did not position a nucleosome *in vivo* because the additional base pair could only partially satisfy the need for overwinding at the dyad axis. Hydroxyl radical footprinting of several nucleosome core particles has shown a distinct overwinding of the DNA to 10.7
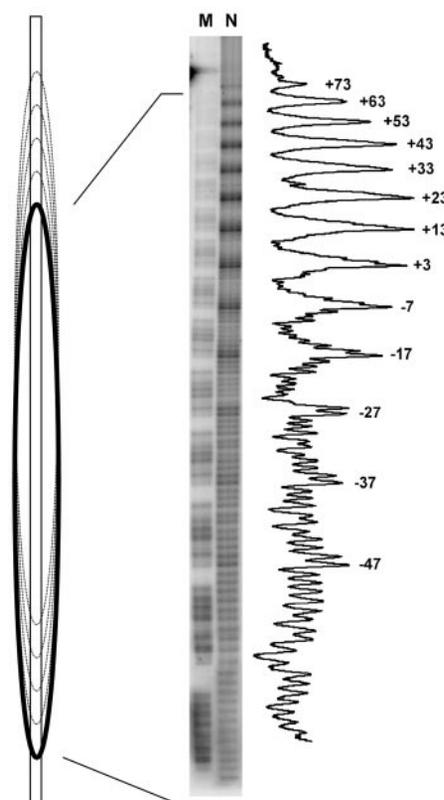

FIG. 9. **Hydroxyl radical footprinting and densitometry of the nucleosome formed on the 280-bp fragment.** *Lane M*, (A + G)-specific sequencing reaction; *lane N*, hydroxyl radical-treated nucleosomes. Indicated are the positions of the peaks in the cleavage pattern relative to the original TATAAACGCC repeat. To the left is shown a schematic representation of the nucleosomal setting on the 280-bp fragment with the dominant position marked with *solid lines* in the off-center position (same as the 180-bp fragment).

bp/turn over the three helical turns at the dyad axis (21). The high resolution x-ray crystal structure of the nucleosome core particle also indicated large differences in helical twist over the entire length of nucleosomal DNA (22). In addition, the DNA is severely distorted at ± 1.5 helical turns from the dyad axis (22), and particularly this site is predominantly cleaved by the enediyne calicheaminicin γ_1^I (26). This further demonstrates that the DNA around the dyad is overwound and structurally deformed; and therefore, DNA molecules capable of adjusting to these nucleosomal restraints are preferred. Consequently, the CTG and TATAAACGCC repeats would both rank as excellent potential candidates for nucleosome positioning *in vivo* due to the high degree of twist flexibility allowed for in these sequences.

The TATAAACGCC Motif Contains Elements Necessary for a Nucleosome-positioning Sequence—If one were to construct a nucleosome-positioning sequence based on all the information available from studies of naturally occurring and artificial nucleosomal sequences, the TATAAACGCC repeat would certainly qualify as a very good candidate for this purpose. Statistical sequencing of nucleosomal DNA from chicken erythrocytes by Travers and co-workers (4) and various data base analyses have shown the importance of phased AA and GG dinucleotides to allow bending of the DNA in the nucleosome. Evidence from footprinting studies and x-ray crystallography have indicated the requirement for overwinding, and hence twist flexibility, in the dyad region (21, 22). It is also known that TA and CG dinucleotides can accommodate high torsional stress owing to their bi-stable nature (27–29). To be accommodated within the nucleosome TATAAACGCC repeat motif satisfies all of these requirements because of the presence of alternating AA and CC/CG dinucleotide steps phased with the helical repeat, which would allow for bending of the DNA in the nucleosome (4, 5, 30). High twist flexibility, potentially due to the presence of TA (or CTA (10)), would allow for the torsional stress due to underwinding and especially overwinding in the central dyad region of the nucleosome. We note, in addition, that the TATA-binding protein, which recognizes a similar motif, the TATA box 5′-TATA(A/T)AA-3′, severely distorts the DNA upon binding (31, 32). The role of both intrinsic curvature and inherent flexibility has been ascribed to the functionality of the TATA box (33) and correlates well with our observations.

Conclusions—The TATAAACGCC repeat is a highly flexible sequence with intrinsic curvature. *In vitro*, this sequence is able to position a nucleosome core particle with very high affinity, ~350-fold higher than random sequence DNA. The major part of this increased affinity is most likely due to its 25-fold higher *J*-factor. On 180-bp fragments, it reconstitutes a uniquely positioned nucleosome core particle. On longer fragments, it yields multiple translational positions, but maintains the predominant position of the pseudodyad axis of symmetry. These results, coupled with the high twist flexibility, suggest that repeats of this sequence motif could position a nucleosome *in vivo*.

REFERENCES

1. Wolffe, A. P. (1994) *Cell* **8,** 13–16
2. Simpson, R. T. (1991) *Prog. Nucleic Acid Res. Mol. Biol.* **40,** 143–184
3. Satchwell, S. C., Drew, H. R., and Travers, A. A. (1986) *J. Mol. Biol.* **191,** 659–675
4. Drew, H. R., and Travers, A. A. (1985) *J. Mol. Biol.* **186,** 773–790
5. Rhodes, D. (1979) *Nucleic Acids Res.* **6,** 1805–1816
6. Shrader, T. E., and Crothers, D. M. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 7418–7422
7. Tanaka, S., Zatchej, M., and Thoma, F. (1992) *EMBO J.* **11,** 1187–1193
8. Patterton, H.-G., and Simpson, R. T. (1995) *Nucleic Acids Res.* **23,** 4170–4179
9. Polach, K. J., and Widom, J. (1995) *J. Mol. Biol.* **254,** 130–149
10. Lowary, P. T., and Widom, J. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 1183–1188
11. Weintraub, H. (1983) *Cell* **32,** 1191–1203
12. Imbalzano, A. N. (1998) *Crit. Rev. Eukaryotic Gene Expression* **8,** 225–255
13. Widlund, H. R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P. E., Kahn, J. D., Crothers, D. M., and Kubista, M. (1997) *J. Mol. Biol.* **267,** 807–817
14. Solomon, M. J., Strauss, F., and Varhavsky, A. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 1276–1280
15. Dixon, W. J., Hayes, J. J., Levin, J. R., Weidner, M. F., Dombroski, B. A., and Tullius, T. D. (1991) *Methods Enzymol.* **208,** 380–413
16. Thåström, A., Lowary, P. T., Widlund, H. R., Cao, H., Kubista, M., and Widom, J. (1999) *J. Mol. Biol.* **288,** 213–229
17. Dlakic, M., and Harrington, R. E. (1995) *J. Biol. Chem.* **270,** 29945–29952
18. Shore, D., Langowski, J., and Baldwin, R. L. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 4833–4837
19. Crothers, D. M., Drak, J., Kahn, J. D., and Levene, S. D. (1992) *Methods Enzymol.* **212,** 3–29
20. Bacolla, A., Gellibolian, R., Shimizu, M., Amirhaeri, S., Kang, S., Ohshima, K., Larson, J. E., Harvey, S. C., Stollar, B. D., and Wells, R. D. (1997) *J. Biol. Chem.* **272,** 16783–16792
21. Hayes, J. J., Tullius, T. D., and Wolffe, A. P. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 7405–7409
22. Luger, K., Mäder, A., Richmond, R., Sargent, D. F., and Richmond, T. J. (1997) *Nature* **389,** 251–260
23. Wang, Y. H., Amirhaeri, S., Kang, S., Wells, R. D., and Griffith, J. D. (1994) *Science* **265,** 669–671
24. Wang, Y. H., and Griffith, J. (1995) *Genomics* **25,** 570–573
25. Godde, J. S., and Wolffe, A. P. (1996) *J. Biol. Chem.* **271,** 15222–15229
26. Kuduvalli, P. N., Townsend, C. A., and Tullius, T. D. (1995) *Biochemistry* **34,** 3899–3906
27. el Hassan, M. A., and Calladine, C. R. (1996) *J. Mol. Biol.* **259,** 95–103
28. el Hassan, M. A., and Calladine, C. R. (1998) *J. Mol. Biol.* **282,** 331–343
29. Hunter, C. A., and Lu, X.-J. (1997) *J. Mol. Biol.* **265,** 603–619, 1997
30. Bolshoy, A. (1995) *Nat. Struct. Biol.* **2,** 446–448
31. Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) *Nature* **365,** 512–520
32. Kim, J. L., Nikolov, D. B., and Burley S. K. (1993) *Nature* **365,** 520–527
33. de Souza, O. N., and Ornstein, R. L. (1998) *Biopolymers* **46,** 403–415
34. Taylor, W. H., and Hagerman, P. J. (1990) *J. Mol. Biol.* **212,** 363–376