J. Mol. Biol. (1997) **267**, 807–817

JMB

Identification and Characterization of Genomic Nucleosome-positioning Sequences

Hans R. Widlund[1], Hui Cao[1], Stina Simonsson[2], Elisabet Magnusson[3] Tomas Simonsson[1], Peter E. Nielsen[4], Jason D. Kahn[5] Donald M. Crothers[6] and Mikael Kubista[1]*

[1]*Department of Biochemistry and Biophysics, The Lundberg Institute, Chalmers University of Technology and Göteborg University, S-413 90, Göteborg Sweden*

[2]*Department of Medical Biochemistry and Microbiology Göteborg University, S-413 90 Göteborg, Sweden*

[3]*Department of Genetics, The Lundberg Institute, Göteborg University, S-413 90, Göteborg Sweden*

[4]*Center for Biomolecular Recognition, Department of Medical Biochemistry and Genetics, Panum Institute DK-2200, Copenhagen N Denmark*

[5]*Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742-2021, USA*

[6]*Department of Chemistry, Yale University, New Haven CT 06511, USA*

**Corresponding author*

Positioned nucleosomes are believed to play important roles in transcriptional regulation and for the organization of chromatin in cell nuclei. Here, we have isolated the DNA segments in the mouse genome that form the most stable nucleosomes yet characterized. In separate molecules we find phased runs of three to four adenine nucleotides, extensive CA repeats, and in a few cases phased TATA tetranucleotides. The latter forms the most stable nucleosome yet characterized. One sequence with CAG repeats was also found. By fluorescence *in situ* hydridization the selected sequences are shown to be localized at the centromeric regions of mouse metaphase chromosomes.

Keywords: nucleosome; centromeres; chromatin; positioning; stability

Introduction

Eukaryotic DNA is complexed with basic histone proteins forming nucleosomes and higher-order chromatin. In the nucleosome core 146 base-pairs (bp) of DNA are wrapped almost two turns around an octamer of histone proteins, hiding one side of the DNA molecule from regulatory proteins. This rotational setting, in combination with the translational positioning of the nucleosome, is thought to be an important factor in transcriptional regulation (Felsenfeld, 1992; Wolffe, 1994; Lewin, 1994). Since nucleosomal organization requires tight bending of the DNA molecule, one might expect certain sequences or sequence features to prevail in nucleosomes, especially in non-coding sequences. A number of naturally occurring sequences that form stable nucleosomes have been identified. Among the best characterized is the somatic 5 S rRNA gene from *Xenopus borealis* (Rhodes, 1985; Hayes *et al.*, 1990), which we use

Abbreviations used: TBP, TATA-box binding protein; HMG-I/Y, α-satellite high mobility group proteins; CENP-B, centromere protein B; FISH, fluorescence *in situ* hybridization.

0022–2836/97/140807–11 $25.00/0/mb970916

here as a reference. Extensive characterization of nucleosomal DNA has revealed an out-of-phase 10 bp periodicity of A·T and G·C base-pairs (Drew & Travers, 1985; Satchwell *et al.*, 1986; Muyldermans & Travers, 1994), which was interpreted in terms of a propensity of A·T regions to compress their minor groove and of G·C regions to compress their major groove. Based on these and related findings for DNA bent by CAP protein (Gartenberg & Crothers, 1988), synthetic DNA molecules were designed with a motif of alternating A·T and G·C base-pairs and found to have a very high affinity for histone octamers (Shrader & Crothers, l989). One of them, the TG-pentamer 5′-(TCGGTGTTAG)$_n$-3′, we use as a second reference.

Most of our current knowledge about nucleosome positioning is derived from studies of non-selected nucleosomal DNA. In this heterogeneous material features important for nucleosomal stability and positioning are buried in a large sequence noise and may be hard to recognize. Here, we have chosen a different approach to identify nucleosome-positioning features. We have created a population of DNA molecules with enhanced nucleosome affinity by selecting those sequences in a genome that form the most stable nucleosomes. These were cloned and found to have accumulated sequence features that could readily be identified as positioning elements. By fluorescence *in situ* hybridization we also show that these sequences are highly abundant in the chromosome centromeric regions.

Results

We have constructed a library of nucleosome core DNA fragments from the mouse genome, that were equipped with adapters for PCR amplification. Those that form the most stable nucleosomes were selected as follows (Tuerk & Gold, 1990; Ellington & Szostak, 1990). An aliquot from the library was mixed with chromatin under conditions where about 10% of the fragments bound histone octamers. These were amplified by PCR and the procedure was repeated. After ten cycles the population was cloned and 87 inserts were sequenced.

The sequenced fragments had distinct features, and were categorized into five groups (Figure 1): two large groups with a high abundance of A-tracts (group 1) and CA-dinucleotides (group 2), two small groups with TATA-tetrads (group 3) and mouse minor satellite sequences (group 4), and a fifth group with no obvious sequence characteristics (group 5). The lengths of the core segments in the selected fragments were 117 to 151 bp (with one aberrant segment of 109 bp), with an average length of 129 bp. A few of similar size were chosen from each group (Table 1) and characterized in more detail (Figure 2).

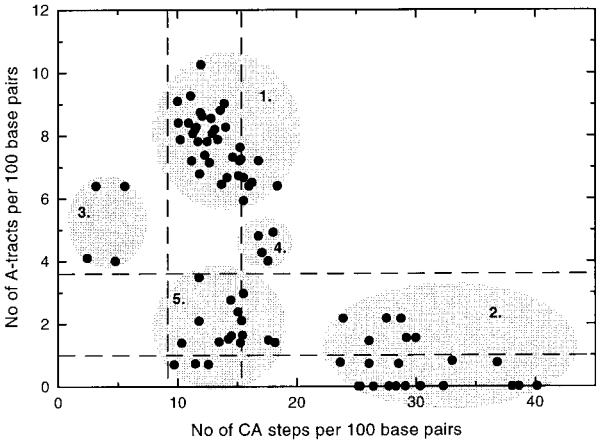

Figure 1. Graph indicating the number of runs of at least three adenine nucleotides and 5′-CA-3′ dinucleotides per 100 base-pairs in the 87 sequenced fragments. This divides them into five groups: A-tracts (1); CA-runs (2); TATA-tetrads (3); mouse minor satellites (4); and those with no evident sequence features (NoSecs) (5). The statistical expectation values ± one standard deviation are indicated. The statistically expected number of A-tracts of at least three consecutive nucleotides in an N base long fragment is:

$$2 \times \left\{ \left(\frac{1}{4}\right)^N + 2\left(\frac{3}{4}\right)\sum_{i=3}^{N-1}\left(\frac{1}{4}\right)^i + \left(\frac{3}{4}\right)\sum_{i=3}^{2N-2}(N-1-i)\left(\frac{1}{4}\right)^i \right\}$$

where the factor of 2 takes into account that both consecutive adenine and thymine nucleotides constitute A-tracts in double-stranded DNA. For $N = 129$, which is the average length of the sequenced fragments, 2.9 A-tracts are expected. The standard deviation, estimated as the square-root of the expectation value (error less than 10% as judged by simulations), is $\sqrt{2.91} = 1.7$. Expressed per 100 bp the expectation value ± one standard deviation is 2.3 ± 1.3. The expected number of CA-dinucleotides is 16 $(=128/8) \pm 4(=\sqrt{16})$, which corresponds to 12.4 ± 3.1 per 100 nucleotides.

Adenine runs

The most common feature among the sequenced fragments are runs of three or more consecutive adenine (or thymine) nucleotides (group 1). Out of the 87 sequenced fragments, 38 have at least eight such runs, which is more than twice the statistically expected abundance (Figure 1). The fragments are highly homologous, having 70% sequence identity, and all are identified as different mouse major satellite sequences by Blast search at NCBI. They form nucleosomes of a stability comparable to the TG-pentamer (Figure 2), which, in combination with the large abundance of major satellite sequences in the mouse genome (Prashad & Cutler, 1976), explains why they appear so frequently in the selected material.

Figure 3A shows a bar graph of the positions of the midpoints of the adenine runs in these sequences. A periodicity, determined to 9.7 bp by Fourier transform analysis, is seen, suggesting an

Table 1. Fragments selected for characterization. Their sequence features are indicated, as well as the putative binding sites for TBP and HMG-I/Y (5′-TATAAA-3′) and for CENP-B (consensus sequence 5′-ATTCGTTGGAAACCGGC-3′; in the selected fragment the underlined thymine is replaced by a cytosine). The complete set of sequenced fragments is available *via* the Internet at: http://www.bcbp.gu.se/mbg or at http://130.241.160.80

A-tracts A-1, 125bp
AGGGACATGGAATATGGAGAGAAAACTGAAAATCACGGAAAATGAAAAATACACACTTTAGGACGTGAAATATAGCGAGGAAAACTGAAAAAGGTGGAATATTTAGAAATGTCCACTGTAGGACG
A-2, 124bp
ATGGCTAGAAAACTGAAAATCATGGAAAAAGAGAAACATCCACTTGACGACTTGAAAAATGTCGAAATCACTAAAAAACGTGAAAAATGAGAAATGCACACTGAGGGACCTGGAATATGGCGAG
CA-runs - CA-1 (125bp)
TACACACACACCACATCATGCATACACACACATCAATGCAATGCATACACACATACATACACATACTAACACATACACTCACACACACGCAGAAATTATGCATGCATCATCGACATTGGCACGCA
CA-2 (124bp)
GAGAGTAACACAGGCACAGGTGTGGAGAGTAACACAGGCACAGGTGTGGGAGAGTGACACACAGGCACAGGTGAGGAGAGTACACACAGGCACAGGTGTGGAGAGCACACACAGGTGCGGAGAG
CA-3 (139bp)
CACTGTGACAACAATGTGGAACATTGTGACATCACAGTGGCGAACAGTGACGGAACAGTAAAGGAGTCTGACAGTACAGTGGAAAACAGTGACGGAACTGTGGAGCACTGTGATTACACCATGGGACATGTTGCACCAC
CAG-runs - CAG (132bp)
AGCAGCAGCAGCAACAGTAGTAGAAGCAGCAGCACTAACGACAGCACAGCAGTAGCAGTAATAGAAGCAGCAGCAGCAGTAGCAGTAGCAGCAGCAGCAGCAGCAATTTCAACAACAGCAGCAGCAGCT
TATA-tetrads - TATA (126bp)
AGGTCTATAAGCGTCTATAAGCGTCTATAAACGTCTATAAACGTCTATAAACGCCTATAAACGCCTATAAACGCCTATACAAGCCTATAAACGCCTATACACGTCTATGCACGACTATACACGTCT
Mouse minor satellites - mins (123bp)
ATTTGTAGAACAGTGTATATCAATGAGCTACAATGAAAATCATGGAAAATGATAAAAACCACACTGTAGAACATATTAGATGAGTGAGTTACACTGAAAAACACATCCGTTGGAAACCGGCAT
Fragments with no sequence features - NoSecs-1 (125bp)
TTGGTCACTGTGTTAAGTAGGGTGTGGTGGCATTCACCTGTAGTCCTAACATTCGTGAGGTAGAGGTAGGAGGGTCAGAAATTCAAAGTTATCTTCAGCTGCTCTATAGTTCAAGGCCGGGGCTG
NoSecs-2 (127bp)
AGGGAGGGTTCTGACCTCAGAACTCAGGAGGTGGATCAGAGCCCCAGACTGCTAGACACATGCCCTGAAAGAGGAAAGCTTGCCTGCAGAGAGTGCTCTGACCACTGGAACTCAGGAAAGAGCTAGT
NoSecs-3 (124bp)
GGGCTGTAGAATCTGATGGAGGTGTAGGATGGATGGACAGTATGACAAAAGGGTACTAGCCTGGGACAGCAGGATTGGTGGAAAGGTTACAGGCAGGCCCAGCAGGCTCGGACGCTGTATAGAG
NoSecs-4 (144bp)
TGAGCTGTGCGGAAGCAAAAGGAAAGCGATTAACGGTGTGCTGCATGATAGTAAGTTATGCCGCGGACAAGCAGCGTCGAAGCACCGGCGATGGCACAGCTGGCACTAAGGCCGGTGGTATCAGAAGCTCCTGACAGCTTAGGG

advantage of having the adenine runs roughly in phase with the DNA helical repeat. Comparing the numbers of runs of different lengths we find a large overrepresentation of three and four consecutive adenine nucleotides: two adenines are followed by a third six times more often than expected statistically and three adenine nucleotides are followed 2.5 times more often than expected by a fourth adenine (Figure 3B). There is no overrepresentation of single adenine nucleotides, A-A dinucleotide steps, or runs of five or more adenine nucleotides.

Phased A-tracts bend DNA (Wu & Crothers, l984), but the fragments we have selected are not appreciably bent as judged by gel mobility studies (Figure 2). Their gel migration anomaly (R_L) is about 1.2, which is far less than observed for extensively curved A-tracts ($R_L > 2$; Haran *et al.*, 1994). Out of an average of ten adenine runs per fragment, nine have the same polarity, which suggests there may be an advantage in having them in a polar arrangement.

CA-dinucleotides

Of the sequenced fragments, 23 are enriched in 5′-CA-3′ (equivalent to 5′-TG-3′) dinucleotide steps (group 2). In some, these steps are completely dominant, comprising up to 80% of all nucleotides. The CA steps are not phased, but they have a strong tendency to form long runs (Figure 3C). For example, the expected occurrence of CACACA stretches in 23 fragments is about three, but here we find 320. The four nucleotides in this set are present in about equal amounts, so the expected probability that a given CA-stretch is extended by another nucleotide is 0.25. The observed frequency is far higher, particularly for longer stretches (Figure 3D).

There is a pronounced oscillation in the length dependence, which is out-of phase for stretches beginning with adenine and with cytosine, respectively. Maxima are observed when there is a cytosine in the 3′ end, which implies that it is the CA step, not the AC step, that is important for nucleosome positioning. This is also suggested by the 25% larger number of CA steps than AC steps in these sequences.

The CA-rich sequences form the least stable nucleosomes of the five groups, but they still show considerably stronger binding to histone octamers than the 5 S rRNA gene (Figure 2). They exhibit normal gel mobilities, and should not be bent in solution.

CAG-triplets

One of the CA-rich fragments is distinguished by having the CA dinucleotides separated by a guanine residue forming CAG-repeats. Extended 5′-CAG-3′ (or 5′-CTG-3′) repeats are associated with several human diseases such as myotonic dystrophy and Huntington's disease (Mahadevan *et al.*, 1992; Huntington's Disease Collaborative Research Group, 1993), and have previously been shown to form unusually stable nucleosomes and to alter the local chromatin structure (Wang *et al.*, 1994; Wang & Griffith, 1995). In our material we find one such sequence (Table 1). It has 26 CAG triplets, all of which have the same polarity. They do not constitute a single long run, but are dispersed in several short repeats of one to five triplets. Its nucleosome affinity is somewhat higher than that of the other fragments enriched in the CA-dinucleotide (Figure 2).

TATA-tetrads

Four fragments with a high content (seven to 12) of TATA tetrads form a third group. They have a very high internal homology with a 10 bp consensus repeat, 5′-TATAA(A/C)CG(T/C)C-3′, and

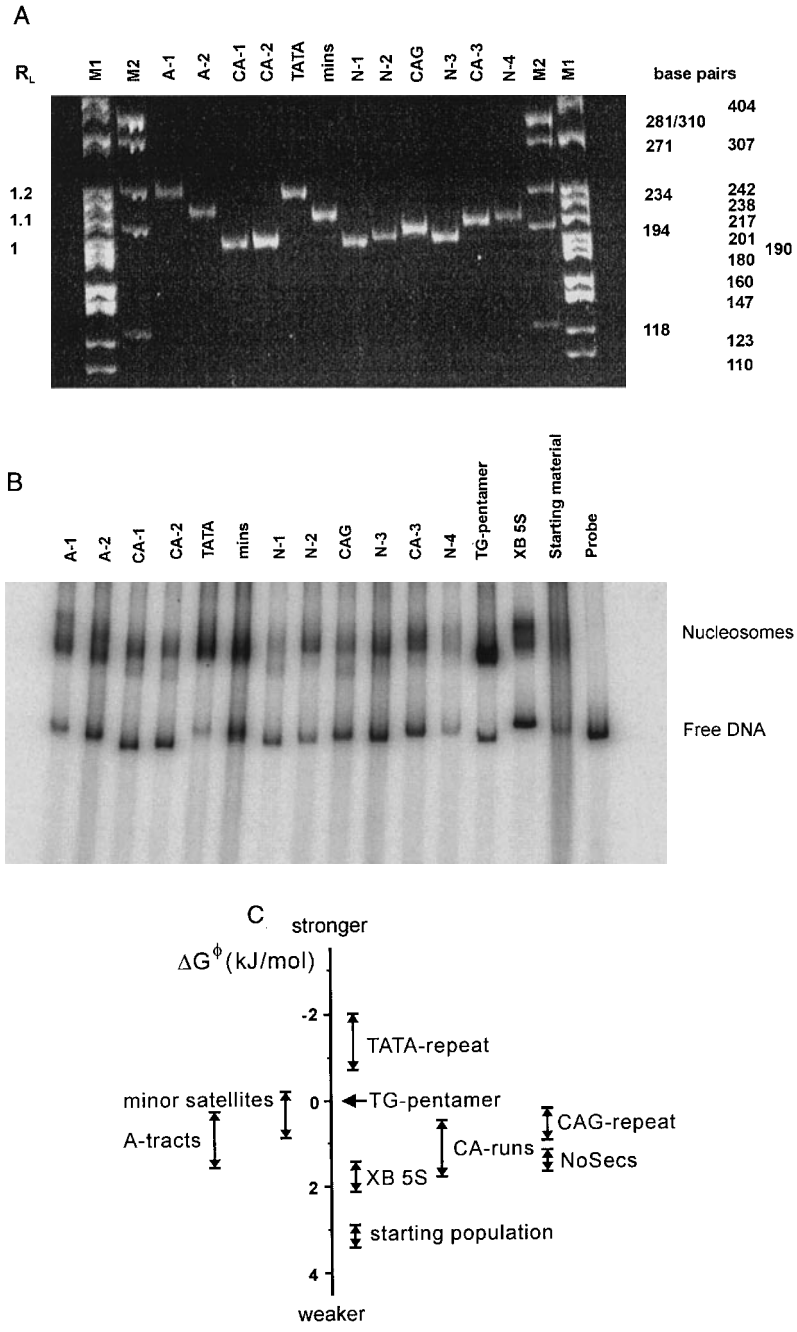

Figure 2. Electrophoretic characterization of the fragments in Table 1. A, Polyacrylamide gel electrophoresis reflecting the electrophoretic mobilities of the free fragments. Lanes M1 and M2 are pBR322 *Msp*I and ϕX174 *Hae*III digest, respectively. The same amount of DNA was loaded in each lane. The length in bp is indicated at the right. B, Gel shift reflecting the histone octamer affinities of the fragments. For some of them additional nucleosome bands can be discerned, suggesting multiple translational positions. C, Changes in free energy upon forming nucleosomes expressed relative to the TG-pentamer: $\Delta G^\circ = -RT\ln(f_i/f_{TG})$, where f_{TG} and f_i are the ratios between the intensities of the nucleosome and the DNA bands for the TG-pentamer and the fragment of interest, respectively (Shrader & Crothers, 1989). The data are averages from at least ten gels with separately constructed material.

form the most stable nucleosomes of the tested fragments (Figure 2). This makes them the strongest nucleosome-positioning sequences so far characterized.

About half of the TATA tetrads are followed by two adenine nucleotides, constituting multiple binding sites (TATA-boxes) for the TATA-box binding protein (TBP), which is a constituent of the general transcription factor TFIID. This stretch is also a putative binding site for the α-satellite high mobility group proteins HMG-I/Y (Solomon *et al.*, 1986).

Mouse minor satellite sequences

Four of the sequenced fragments (group 4) show strong homology with mouse minor satellite sequences (Kipling *et al.*, 1994). They have a somewhat elevated content of CA-dinucleotides, and contain also four runs of adenine nucleotides. In contrast to the fragments with extensive CA runs, the CA-dinucleotides here are non-contiguous and present to about the same extent in both orientations. Their affinity for histone octamers is about the same as that of the TG-pentamer, and they are not bent in solution (Figure 2). They all have at the

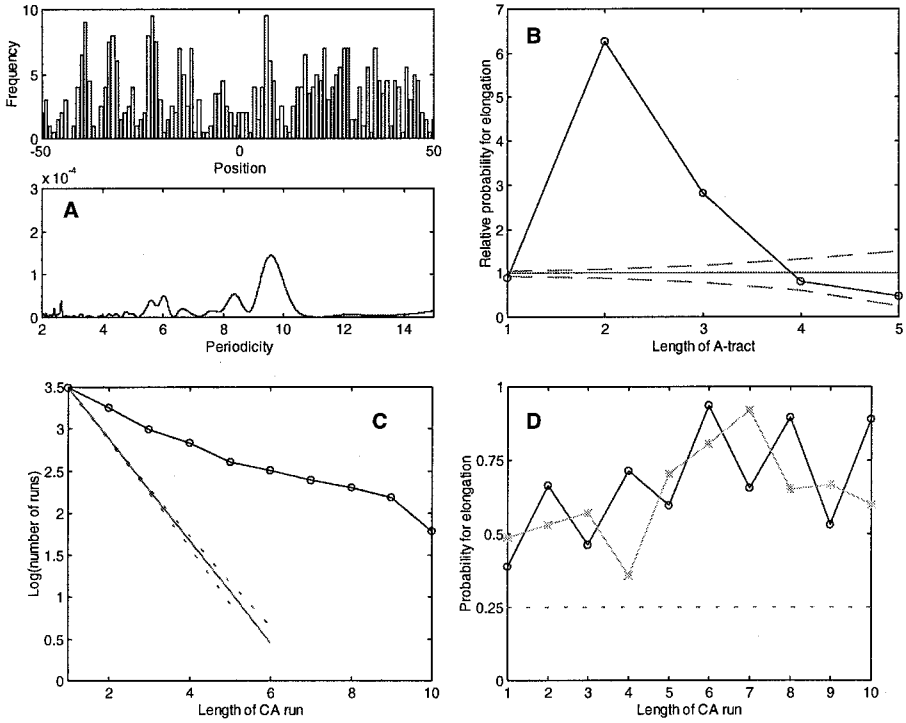

Figure 3. A and B, Properties of the A-tracts in group 1. A, Top: Number of A-tracts, defined by their midpoints, at different positions in the fragments after alignment. Bottom: Fourier analysis revealing a periodicity of 9.7 bp, obtained as the maximum in the power spectral density of the discrete 2^{15} point Fourier transform of the distribution above. B, The ratio between observed and expected frequencies that a stretch of n adenine residues is followed by an additional adenine residue (calculated as the ratio between the number of A-tracts longer than n divided by the number of A-tracts of exactly length n, and the total number of adenine residues divided by the total number of non-adenine residues). Broken lines are expectation values ± one standard deviation. C and D, Properties of the CA-dinucleotides in group 2. C, Comparison of observed and expected number of CA-runs of different lengths. Broken lines are expectation values ± one standard deviation. The expected number of CA repeats of length n is: $2 \times 0.25^n \times$ (total number of bases in the fragments $- (n - 1) \times$ number of fragments), where the last term takes into account the finite length of the fragments. Note the logarithmic scale. Note also that a run such as CACACA contains three CA steps, two CACA steps and one CACACA step, adding up to a total of ten CA steps. D, The probability that a tract of a certain length is continued. Tracts beginning with A (–○–) and tracts beginning with C (–∗–) are shown separately. The two curves oscillate out-of-phase having maxima at lengths where there are cytosine nucleotides at the 3′-end. The probability that an AC tract is extended to ACA is the ratio between the observed number of tracts (NOT_C)ACA divided by the number of tracts (NOT_C)AC(ANY_NUCLEOTIDE). This can be expressed as $({}^{n}\text{ACA-}{}^{n}\text{CACA})/({}^{n}\text{AC}\cdot{}^{n}\text{CAC})$, where ${}^{n}\text{X}$ is the observed number of tracts of type X. In the analyzed set all nucleotides are present in about the same amounts, hence the statistical probability that any tract is extended by one step is 0.25.

3′-end a putative binding site for the centromere protein B (CENP-B) which localizes to the central domain of the centromeres of mouse and human chromosomes (Masumoto *et al.*, 1989; Kitagawa *et al.*, 1995).

Fragments with no evident sequence characteristics

For 18 of the 87 fragments (group 5) we did not find features we could identify as potential nucleosome-positioning elements. These sequences have a normal abundance of both A-tracts and CA-repeats (Figure 1), and statistical analysis revealed no over-representation of any di- tri- or tetranucleotides. Nor could we detect any phasing between potential structural elements. This, however, does not mean that they are devoid of nucleosome-positioning features. The four fragments that were characterized in more detail have moderate gel anomalies ($R_L \leqslant 1.2$) and form nucleosomes of considerable stability (Figure 2). None of the fragments in this group was found among deposited sequences in available databases.

Fluorescence *in situ* hybridization

A fluorescent replica of the selected population was created by PCR and used for fluorescence *in situ* hybridization (FISH) with mouse metaphase chromosomes. As seen in Figure 4, the selected sequences are highly abundant in the centromeric regions.

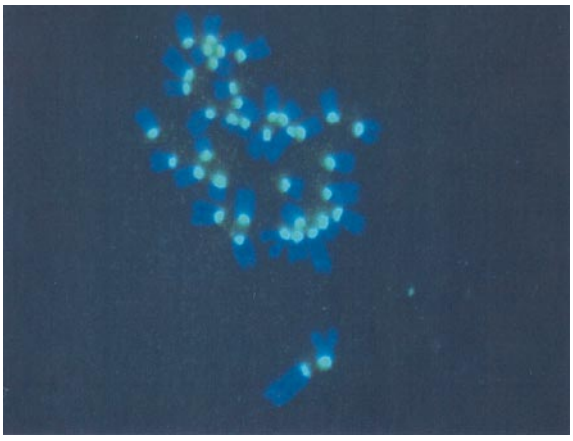

Figure 4. Fluorescence *in situ* hybridization of the selected population with mouse metaphase chromosomes. A large abundance of the selected sequences is seen at the centromeric regions (note that mouse chromosomes lack the short upper chromatid arms).

Discussion

Sequences that form exceedingly stable nucleosomes in the mouse genome have been selected

The 146 base-pairs in the nucleosome core can be arranged in $4^{146} \approx 10^{88}$ different ways. This is an incredibly large number and it is not possible to identify with certainty the sequence that forms the most stable nucleosome. Here we have sought to select the sequences in the mouse genome that form the most stable nucleosomes. The genome has about 3×10^9 bp, which is vanishingly small compared to 10^{88}, and for statistical reasons is not even expected to contain any of the more efficient nucleosome-forming sequences. However, the selected sequences will be the strongest among those with potential biological significance.

Sequences were selected by ten cycles of salt-induced reconstitution of nucleosomes under selective pressure, which yielded a population with a strongly enhanced affinity for histone octamers. Sequencing revealed over 70% sequence identity within groups 1 to 4. In group 5 two fragments with identical sequences were encountered. This suggests that the selection had produced a population with a relatively small number of highly similar sequences.

To test the reproducibility of the selection procedure it was done twice using separately constructed starting materials. Both selections produced sequences of the main groups 1, 2 and 5, though in somewhat different amounts. Sequences in groups 3 and 4 were obtained in only one selection. This indicates some instability in the procedure, which could be due to different amounts of the various sequences being present in the two starting materials owing to subtle differences in preparation. Thus, although the selections have generated a population that forms nucleosomes of considerably enhanced stability, and has accumulated sequence features, there is a risk that some potentially important features have been missed. However, the sequences which were not isolated in the selection are unlikely both to form strong nucleosomes and to be abundant in the genome, and should not have general regulatory or structural functions, though we cannot exclude the possibility that some sequences are preferentially lost in the PCR.

Nucleosome stability is determined mainly by the central 120 base-pairs

The starting material was core DNA fragments obtained from nucleosomes by micrococcal nuclease treatment, flanked by adapters of 24 and 22 bp. Nucleosome core DNA is in general very homogeneous in length, and sequencing of our starting material revealed it was 146 (±3) bp. Still, the core segments in the selected population were 117 to 151 bp, with an average of 129 bp. Some of the length dispersion may have arisen from accumulated errors during the many PCR amplifications, but may also reflect a heterogeneity in the starting material that has been brought forward by the selection. Although the origin of the large spread is unclear, it indicates that the length of the interacting segment is not crucial for stability, at least when it is longer than about 120 bp. A similar conclusion has been reached previously (Hayes *et al.*, 1990).

The selected fragments that were sequenced had an average length of the central region that is substantially shorter than the expected length of nucleosome core DNA, suggesting the selection is biased towards shorter fragments. Non-specific binding affinity increases in general with increasing fragment length, because longer fragments have more potential binding sites (McGhee & von Hippel, 1974; Jansen *et al.*, 1993). This is also likely to be the case for non-specific nucleosome formation. The bias towards shorter fragments in the selected material suggests that binding to these is specific, and that the longer ones are somehow destabilized. With the adapters the fragments are 163 to 191 bp, and it is possible that wrapping around the histone octamer brings the fragment ends into proximity, which leads to destabilization owing to electrostatic repulsion.

Polar runs of three to four adenine nucleotides in phase with the helical repeat give rise to high nucleosome stability, possibly by exploiting a novel feature of adenine tracts

Sequences in group 1 have polar runs of three to four adenine nucleotides roughly in phase with the helical repeat, suggesting they give rise to the very high nucleosome stability of these

sequences. Various phased repeats of A + T-rich segments in nucleosomal DNA have been noted previously and proposed to be important nucleosome-positioning signals (Trifonov & Sussman, 1980; Satchwell *et al*., 1986; Lowman & Bina, 1990). The suggested reason was their propensity to compress the minor groove, which simplifies bending around the histone octamer when they are in phase with the helical repeat. Indeed, this was the basis for the construction of the TG-pentamer (Shrader & Crothers, 1989). However, this mechanism does not account for some pertinent features of the adenine runs in the fragments we have selected. These are exclusively homopolymeric, three or four nucleotides long, and tend to have like polarities.

Owing to the high curvature of nucleosome core DNA, one might expect bent sequences to form more stable nucleosomes, and phased A-tracts are well known to induce considerable curvature in DNA (Wu & Crothers, 1984). However, our sequences are not extensively bent in solution as evidenced by their only moderate migration anomaly ($R_L \leqslant 1.2$, Figure 2B). The reason is presumably that the A-tracts are too short. In our sequences they are only three or four nucleotides, while five are required to induce substantial bending (Koo *et al*., 1986). Such long adenine runs are not overrepresented in our sequences (Figure 3B). We also note that the adenine runs in our sequences are phased by 9.7 bp, which is significantly less than the 10.35 bp repeat in extensively bent A-tract DNA (Drak & Crothers, 1991).

Having ruled out minor groove compression in A + T-rich regions and static bending in A-tracts as the most plausible causes for the high nucleosome stability of these sequences, then what property of the adenine runs might be responsible? As already mentioned, the adenine runs in the fragments we have selected show a striking polarity. Out of ten runs, nine have the same polarity. Statistically, this is highly unlikely and one might suspect an advantage in having them in a polar arrangement. However, there is also a possibility that such sequences are overrepresented in the genome. Serendipitously the adapters we used have fortuitous runs of three or four consecutive adenine nucleotides (see Materials and Methods), and all 38 fragments in this group turned out to be joined with the adenine-containing strand of the core region to the adenine-containing strands of the adapters. Since there is no obvious reason for such a preference in the starting material, this bias was probably generated by selection. The influence of the adapters on the selection of the fragments in this group is also evident from their somewhat shorter average length of the core region (123 bp) compared to that of the other selected fragments (129 bp). Hence, whether sequences with polar runs of adenine nucleotides are overrepresented in the genome or not there seems to be an advantage in having them in a polar arrangement when forming nucleosomes.

Deformation of DNA that affects the two strands symmetrically with respect to the base-pair dyad, such as base-pair roll or base-pair opening, is independent of sequence polarity, while a deformation that affects them asymmetrically, such as base-pair tilt, is reversed. As a consequence, elements that produce symmetric deformations should be positioned with a groove facing towards the histone octamer, and elements producing asymmetric deformations should be positioned with a phosphodiester backbone facing the histone octamer. The adenine runs in the selected sequences are polar and hence expected to produce an asymmetric deformation. This is different from general A + T-rich regions which are thought to position nucleosomes by compressing their minor grooves (Shrader & Crothers, 1989), and the expected wrapping of prebent A-tract DNA, which is also bent towards the minor groove of the adenine tract (Zinkel & Crothers, l987). In fact, the reason adenine runs longer than four nucleotides are not overrepresented in our fragments could be that their natural bend towards the minor groove is incompatible with the nucleosome-induced asymmetric deformation that seems to be important here.

The first symmetric deformation that comes to mind is base-pair tilt. Although a small tilt component has been deduced for AA-dinucleotide steps in free DNA (Bolshoy *et al*., 1991), there is today no hard evidence for such deformation in A-tracts. Another possibility is that short A-tracts favor a large roll angle immediately adjacent to the end of the tract, as is seen in crystal structures of oligonucleotides containing A-tracts (Nelson *et al*., 1987; Coll *et al*., 1987; DiGabriele *et al*., 1989; Edwards *et al*., 1992; DiGabriele & Steitz, 1993). Presently we cannot distinguish between these and other possible asymmetric deformations, and conclude that a hitherto unrecognized property of adenine tracts seems to be responsible for the high nucleosomal stability of these sequences.

Previous analyses of genomic DNA revealed out-of-phase 10 bp periodicities of AA- and TT-dinucleotides (Trifonov & Sussman, 1980; Lowman & Bina, 1990). Although isolated AA-dinucleotide steps are not overrepresented in our fragments, the $A_{3\text{-}4}$ runs we observe do of course embrace such steps. The out-of-phase arrangement of AA and TT steps in genomic DNA is consistent with the polar arrangement of the $A_{3\text{-}4}$ runs in the selected sequences, since the reversed polarity every fifth base-pair leads to deformations towards the same macroscopic side of the DNA helix. In fact, the few T runs that are interspersed among the $A_{3\text{-}4}$ runs in our fragments tend to be out of phase.

Extensive runs of CA-dinucleotides give rise to high nucleosome stability, possibly by undergoing a conformational change

In Figure 3C the probability that a given CA-tract is continued is seen to increase with increas-

ing length, eventually exceeding 90% for tracts ending with a cytosine. This implies that some property of CA-runs is becoming more pronounced with increasing tract length. This is a characteristic feature for sequences undergoing allosteric conformational changes (Chaires, 1986; Samuelsson *et al*., 1994), and suggests that CA-runs may have different conformations. When incorporated into nucleosomes, they can adapt a conformation that is more prone to curve. Bimodal stability of CA steps is suggested by crystallographic observations. In crystals, CA steps have either a positive roll, small positive slide and small twist and t, g^-, g^- conformation about the bonds C3′-O3′, O3′-P and P-O5′, respectively, or a negative roll, very large positive slide and large twist along with a g^-, t, g^- conformation (Bansal, 1996). We do not know what conformation is important for nucleosome stability, but the prominent dinucleotide repeat apparent in Figure 3C suggests it should have a dinucleotide structural unit. A further indication that a non-standard conformation is important is the pronounced polarity of the CA-dinucleotide steps in the CA-runs. The frequency of CA steps followed by another CA step is 0.37, while it is 0.12 that a CA step is followed by a TG. The frequency of TG steps followed by CA is only 0.06.

Each fragment has several runs of CA-dinucleotides. These are not phased and they have no particular polarity, implying that they do not need to be coordinated in space. Hence, the important conformation of CA-runs in the nucleosome is not expected to have asymmetric features, such as directed bends. Compared to canonical DNA it should rather have an enhanced (symmetric) flexibility. The CA step has a low stacking energy (Breslauer *et al*., 1986), and should be susceptible to conformational changes, particularly when arranged in long runs. Indeed, cyclization studies have shown that the flexibility of CA repeats increases in the series CA < CAC < CACA (Harrington, 1993).

CAG repeats form stable nucleosomes, which might lead to their extension

One of the CA-rich sequences is different in having a CAG repeat instead of a CA-repeat. It forms somewhat more stable nucleosomes than the tested CA-repeat-containing sequences, thus forming nucleosomes of considerable stability, as also noted previously (Wang *et al*., 1994; Wang & Griffith, 1995; Godde & Wolffe, 1996). With only one fragment we cannot perform statistics, but we note that the CAG triplets in each repeat have the same polarity.

Extensive CAG repeats have been associated with several human neurodegenerative disorders (Barinaga, 1996), and one wonders if this could be a consequence of their high affinity for histone octamers. There is no direct relation, since the repeats are found in coding regions where they give rise to polyglutamine stretches (CTG codons in the complementary strand). If these become too long disease may arise, possibly because these stretches bind and inactivate glyceraldehyde-3-phosphate dehydrogenase. The reason CAG repeats tend to grow in length is not known, but might be related to their affinity for histone octamers. Polymerases do slip, and one may speculate that the probability of slip increases when the DNA is bound tightly to histone octamers. This could lead to a vicious cycle: fragments containing CAG repeats form stable nucleosomes where DNA polymerase slips, extending the CAG stretch which increases the nucleosome stability.

The prime nucleosome-forming sequence

The ''TATA-tetrad'' fragments in group 3 form the most stable nucleosomes so far characterized (Figure 2). They have a distinct 10 bp consensus repeat, 5′-TATAA(A/C)CG(T/C)C-3′, suggesting that all base-pairs are important for its nucleosome-forming properties. With adenine in position 5 and cytosine in position 9, the repeat begins with an A·T region followed by an A_3-tract and it ends with a G·C region. The reason behind the extreme nucleosomal stability of these fragments could be that they combine the out-of-phase repeat of A·T and G·C segments found in the TG-pentamer (Shrader & Crothers, 1989), with phased $A_{3\text{-}4}$ runs as found in the A-tract fragments (group 1). They could bind the histone octamers with the A·T segments facing in with the minor grooves, the A_3-tracts facing in with the backbone of the adenine-containing strand and the G·C segments facing in with the major grooves.

The distinct consensus suggests that the precise order of nucleotides in the T·A and G·C regions may also be essential. TAT forms with the A-tract two successive TA-steps which are known to be highly unstable and prone to deformation (Calladine, 1982; Dlakic & Harrington, 1995). The CGCC arrangement we cannot presently rationalize.

It is interesting that sequences that give rise to high nucleosome stability contain multiple TATA-boxes (5′-TATAAA-3′). The four selected fragments have three to six. One function of TATA-boxes is to interact with TBP, which significantly distorts the DNA structure (Kim *et al*., 1993a,b). TBP binding is severely inhibited by nucleosome formation, and depends on the orientation of the TATA sequence relative to the surface of the histone core (Imbalzano *et al*., 1994). Since the TATA-boxes in these fragments are in phase with the helical repeat, either none or all of them, depending on the rotational setting, should be accessible for TBP binding. Although we do not know if these fragments interact with TBP under any conditions, they certainly have intriguing sequence properties for being involving in transcriptional regulation.

Nucleosome-positioning signals

The DNA in the nucleosome must adapt to a defined highly curved path determined by the basic histone proteins, where it is subjected to several sharp kinks (Richmond *et al*., 1984) and has a helical repeat that is altered from that in solution (Hayes *et al*., 1990). It is reasonable that some sequences or sequence features more readily adapt to these requirements than others, thus being more readily incorporated into nucleosomes. From our results it is clear that there is not one single nucleosome-positioning signal, but many. Some are distinct sequence features, such as polar runs of three to four adenine nucleotides in phase with the helical repeat, extensive CA and CAG runs, and out-of-phase repeats of G·C and A·T regions, while others seem to be more complex. The minor satellite sequences (group 4) contain four adenine runs and have a somewhat elevated number of CA dinucleotides (Figure 1). But the adenine runs are not phased and the CA steps are isolated, and can hardly account for the high nucleosomal stability of these fragments. They will most likely have other positioning elements that have escaped our notice. The sequences in group 5 all seem to be different. Since the only criterion for grouping them was that they do not fit in any of the other groups, they may indeed be different and have unique positioning signals. So far we have been unable to identify these signals. We note that previously identified nucleosome-positioning sequences, such as the 5 S rRNA genes, do not fall into our first four groups, and we can presently not identify their positioning signals either.

Chromosomal location of nucleosome-positioning sequences

Our fragments form more stable nucleosomes than the 5 S rRNA gene from *X. borealis* and some even more stable than the TG-pentamer (Figure 2). This high affinity for histone octamers suggests that they may have structural roles in the arrangement of chromatin. As visualized by FISH the chromosome centromeres are very rich in these sequences (Figure 4). Both mouse minor satellite sequences, and mouse major satellite sequences, to which the A-tracts belong, have previously been localized to the centromeric regions (Prashad & Cutler, 1976; Wong & Rattner, 1988). The minor satellite sequences we have isolated have a binding site for the CENP-B protein and could serve as anchors for the chromatin structure. The TATA-containing sequences could bind HMG-I/Y proteins, which are associated with α-satellite sequences in the centromeres.

Although the high abundance of sequences forming stable nucleosomes in the centromeres could be interpreted as an accumulation through evolution, suggesting they have important functions here, there is an alternative; they may have been depleted from the remaining parts of the chromosomes. Presently we cannot exclude this possibility.

Materials and Methods

Preparation of genomic nucleosome core DNA

Chromatin was prepared from mouse *Ehrlich asceites* cells (Kubista *et al*., 1985). Histone H1 was removed by ion-exchange chromatography (in 0.6 M NaCl using BioRad AG 50W-X2, 100 to 200 mesh) and the material was dialyzed against 1 mM EDTA, 10 μM 4-(2-aminoethyl)benzenesulfonyl flouride (AEBSF). Nucleosome core particles were produced by micrococcal nuclease (3 units/μl, in 4.5 mM $CaCl_2$ for two minutes at 37°C), and purified by 3% (w/v) agarose gel electrophoresis.

The 3′-phosphate groups were removed by alkaline phosphatase, blunt ends were generated with T4 DNA polymerase and 5′-phosphate groups were added using T4 polynucleotide kinase. The fragments were then ligated into long polymers with an excess of a 135 bp adapter sequence from pUC18 having a single *Eco*RI site at position 35 (1 unit/μl ligase in 10% (w/v) PEG8000 in a total volume of 40 μl for 16 hours at 4°C). After *Eco*RI digestion, 281 bp fragments, containing the core DNA flanked by two different parts of the adapter fragment, were purified by 5% (w/v) polyacrylamide (acrylamide to bis-acrylamide, 20:1, w/w) gel electrophoresis. These were finally amplified by PCR using primers to crop the flanking adapters to 24 and 22 bp, respectively, to give a starting population of fragments with the basic structure: 5′-GTCGTGACTGG-GAAAACCCTGGCG-$(N)_{146}$-TCACACAGGAAACAGC-TATGAC-3′, where $(N)_{146}$ are the core DNA fragments originating from the mouse genome. Analysis of this population by denaturing PAGE showed it was 192 (±3) bp.

Reconstitution of nucleosome

A 1 pmol amount of radiolabeled fragments was mixed with 3.5 μg of H1-depleted chromatin in 1 M NaCl, 20 mM Tris, 0.1% (v/v) Nonidet P-40 and 100 μg/ml bovine serum albumin (BSA). The mixture was incubated for 30 minutes at 37°C and diluted to 100 mM NaCl by three successive additions (separated by 20 minutes) of a low-salt buffer at room temperature. Samples were analyzed in a native 5% polyacrylamide (acrylamide to bis-acrylamide, 75:1) gel in the presence of 0.01% Nonidet P-40, 0.5 × TBE buffer at 4W at room temperature for two hours. As references, the TG-pentamer (165 bp) and the 5 S rRNA gene from *X. borealis* (218 bp), restricted from plasmids, were used. The amount of radioactivity was quantified using a phosphorimager (Molecular Dynamics).

Selection procedure

About 10^9 molecules, corresponding to the number of base-pairs in the mouse genome, were PCR amplified about 1000-fold, purified by agarose gel electrophoresis and used as the starting material for the selection. It was mixed with H1-depleted chromatin and allowed to compete for histone octamers under conditions where about 10% formed nucleosomes (adjusted by the chromatin concentration). These were purified by polyacrylamide gel electrophoresis and the

DNA was extracted and amplified by PCR, whereafter the procedure was repeated. After ten cycles the enriched population was cloned and sequenced. The whole selection procedure, including the construction of starting material, was repeated once.

Cloning and sequencing

The selected material was ligated into pCR-Script SK+ vector (Stratagene) between the *Eco*RI and *Bam*HI sites, cloned into *Escherichia coli* and grown in LB medium. Plasmids were prepared by flow column purification (Qiagen) and 0.2 μg was used for repeated primer extension sequencing (40 cycles using ThermoSequenace (Amersham) and 2 μl of the N-mixes with 10 pmol of fluorescent primer (Pharmacia) in a total volume of 8 μl). The samples were analyzed on an A.L.F. DNA sequencer (Pharmacia).

Anomalous migration

Samples were produced by PCR amplification of cloned fragments (2.5 mM $MgCl_2$, 10 pmol of each primer and 3 units of *Taq* polymerase in 100 μl), and analyzed in a native 8% polyacrylamide (acrylamide to bis-acrylamide, 30:1) gel run in 0.5 × TBE buffer at 4°C. The gel was stained with ethidium bromide and photographed.

Fluorescence *in situ* hybridization

Probes were produced by incorporation of fluorescein-12-dCTP by PCR amplification of an aliquot of the selected material. Typically 40 μM Fl-12-dCTP (DuPont), 2 μM dCTP, 80 μM d(A,G,T)TP, 2.5 mM $MgCl_2$ and 6 units/100 μl *Taq* polymerase were used in the reaction. The product was extracted with phenol and chloroform, precipitated with ethanol and resuspended in hybridization buffer (1 g Dextran/100 ml, 50% (w/v) formamide and 2 × SSC). Metaphase chromosomes from normal mouse fibroblast cells were prepared and spread on slides (Martinsson *et al*., 1982). The probe was hybridized overnight and then washed with SSC buffer and counterstained with 4′,6-diamidino-2-phenylindole (DAPI). Photographs were taken with a charge-coupled device (CCD) camera (Hamamatsu) through a fluorescence microscope (Leica).

Acknowledgments

The authors thank Daniela Rhodes for critically reading the manuscript. We are grateful to Alan Wolffe for plasmid XP-10, Prasad Kuduvalli and Thomas Tullius for plasmid XP-11, both with the 5 S somatic rRNA gene from *X. borealis*. We thank Leif Lundh for advice on sequencing and Petra Tomic for her work on the project. H.R.W. receives a grant from the Sven and Lilly Lawski Foundation. This project is funded by the Swedish Cancer Foundation and the Swedish Natural Sciences Research Council.

References

Bansal, M. (1996). Structural variations observed in DNA crystal structures and their implications for protein-DNA interactions. In *Biological Structure and Dynamics, Proceedings of the Ninth Conversation* (Sarma, R. H. & Sarma, M. H., eds), vol. 1, pp. 121–134, Adenine Press, Schenectady, NY.

Barinaga, M. (1996). An intriguing new lead on Huntington's disease. *Science,* **271**, 1233–1234.

Bolshoy, A., McNamara, P., Harrington, R. E. & Trifonov, E. N. (1991). Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl Acad. Sci. USA,* **88**, 2312–2316.

Breslauer, K. J., Frank, R., Blocker, H. & Markey, L. A. (1986). Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA,* **83**, 3746–3750.

Calladine, C. R. (1982). Mechanics of sequence-dependent stacking of bases in B-DNA. *J. Mol. Biol.* **161**, 343–352.

Chaires, J. B. (1986). Allosteric conversion of Z DNA to an intercalated right-handed conformation by daunomycin. *J. Biol. Chem.* **261**, 8899–8907.

Coll, M., Fredrick, C. A., Wang, A. H. & Rich, A. (1987). A bifurcated hydrogen bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl Acad. Sci. USA,* **84**, 8385–8389.

Diekmann, S. (1992). Analyzing DNA curvature in polyacrylamide gels. In *Methods in Enzymology: DNA Structures. Part B: Chemical and Electrophoretic Analysis of DNA* (Lilley, D. M. J. & Dahlberg, J. E., eds), vol. 212, pp. 30–46, Academic Press, San Diego.

DiGabriele, A. D. & Steitz, T. A. (1993). A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *J. Mol. Biol.* **231**, 1024–1039.

DiGabriele, A. D., Sanderson, M. R. & Steitz, T. A. (1989). Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract. *Proc. Natl Acad. Sci. USA,* **86**, 1816–1820.

Dlakic, M. & Harrington, R. E. (1995). Bending and torsional flexibility of G/C-rich sequences as determined by cyclization assays. *J. Biol. Chem.* **270**, 29945–29952.

Drak, J. & Crothers, D. M. (1991). Helical repeat and chirality effects on DNA gel electrophoretic mobility. *Proc. Natl Acad. Sci. USA,* **88**, 3074–3078.

Drew, H. R. & Travers, A. A. (1985). DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* **186**, 773–790.

Edwards, K. J., Brown, D. G., Spink, N., Skelly, J. V. & Neidle, S. (1992). Molecular structure of the B-DNA dodecamer $d(CGCAAATTTGCG)_2$. An examination of propeller twist and minor groove water structure at 2.2 Å resolution. *J. Mol. Biol.* **226**, 1161–1173.

Ellington, A. D. & Szostak, J. W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature,* **346**, 818–822.

Felsenfeld, G. (1992). Chromatin as an essential part of the transcriptional mechanism. *Nature,* **355**, 219–224.

Gartenberg, M. R. & Crothers, D. M. (1988). DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature,* **333**, 824–829.

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

Haran, T. E., Kahn, J. D. & Crothers, D. M. (1994). Sequence elements responsible for DNA curvature. *J. Mol. Biol.* **244**, 135–143.

Harrington, R. E. (1993). Studies of DNA bending and flexibility using gel electrophoresis. *Electrophoresis,* **14**, 732–746.

Hayes, J. J., Tullius, T. D. & Wolffe, A. P. (1990). The structure of DNA in a nucleosome. *Proc. Natl Acad. Sci. USA,* **87**, 7405–7409.

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.

Imbalzano, A. N., Kwon, H., Green, M. R. & Kingston, R. E. (1994). Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature,* **370**, 481–485.

Jansen, K., Nordén, B. & Kubista, M. (1993). Sequence dependence of 4′,6-diamidino-2-phenylindole (DAPI)-DNA interactions. *J. Am. Chem. Soc.* **115**, 10527–10530.

Kim, Y., Geiger, J. H., Hahn, S. & Siegler, P. B. (1993a). Crystal structure of a yeast TBP/TATA-box complex. *Nature,* **365**, 512–520.

Kim, Y., Nikolov, D. B. & Siegler, P. B. (1993b). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature,* **365**, 520–527.

Kipling, D., Wilson, H. E., Mitchell, A. R., Taylor, B. A. & Cooke, H. J. (1994). Mouse centromere mapping using oligonucleotide probes that detect variants of the minor satellite. *Chromosoma,* **103**, 46–55.

Kitagawa, K., Masumoto, H., Ikeda, M. & Okazaki, T. (1995). Analysis of protein-DNA and protein-protein interactions of the centromere protein B (CENP-B) and properties of the DNA-CENP-B complex in the cell cycle. *Mol. Cell. Biol.* **15**, 1602–1612.

Koo, H.-S., Wu, H.-M. & Crothers, D. M. (1986). DNA bending at adenine-thymine tracts. *Nature,* **320**, 501–506.

Kubista, M., Härd, T., Nielsen, P. E. & Nordén, B. (1985). Structural transitions of chromatin at low salt concentrations: a flow linear dichroism study. *Biochemistry,* **24**, 6336–6342.

Lewin, B. (1994). Chromatin and gene expression: constant questions, but changing answers. *Cell,* **79**, 397–406.

Lowman, H. & Bina, M. (1990). Correlation between dinucleotide periodicities and nucleosome positioning on mouse satellite DNA. *Biopolymers,* **30**, 861–876.

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jasen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., Leblond, S., Earle-MacDonald, J., De Jong, P. J., Wieringa, B. & Korneluk, R. G. (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. *Science,* **255**, 1253–1255.

Martinsson, T., Tenning, P., Lundh, L. & Levan, G. (1982). Methotrexate resistance and double minutes in a cell line from the SEWA mouse ascites tumor. *Hereditas,* **97**, 123–137.

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. & Okazaki, T. (1989). A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell. Biol.* **109**, 1963–1973.

McGhee, J. D. & von Hippel, P. H. (1974). Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* **86**, 469–489.

Muyldermans, S. & Travers, A. A. (1994). DNA sequence organization in chromatosomes. *J. Mol. Biol.* **235**, 855–870.

Nelson, H. C., Finch, J. T., Luisi, B. F. & Klug, A. (1987). The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature,* **330**, 221–226.

Prashad, N. & Cutler, R. G. (1976). Percent satellite DNA as a function of tissue and age of mice. *Biochim. Biophys. Acta,* **418**, 1–23.

Rhodes, D. (1979). Nucleosome cores reconstituted from poly (dA-dT) and the octamer of histones. *Nucl. Acids Res.* **6**, 1805–1816.

Rhodes, D. (1985). Structural analysis of a triple complex between the histone octamer, a Xenopus gene for 5S RNA and transcription factor IIIA. *EMBO J.* **4**, 3473–3482.

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

Samuelsson, P., Jansen, K. & Kubista, M. (1994). Long-range interactions between DNA-bound ligands. *J. Mol. Recog.* **7**, 241–253.

Satchwell, S. C., Drew, H. R. & Travers, A. A. (1986). Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191**, 659–675.

Shrader, T. E. & Crothers, D. M. (1989). Artificial nucleosome positioning sequences. *Proc. Natl Acad. Sci. USA,* **86**, 7418–7422.

Solomon, M. J., Strauss, F. & Varshavsky, A. (1986). A mammalian high mobility group protein recognizes any stretch of six A·T base pairs in duplex DNA. *Proc. Natl Acad. Sci. USA,* **83**, 1276–1280.

Trifonov, E. N. & Sussman, J. L. (1980). The pitch of chromatin is reflected in its nucleotide sequence. *Proc. Natl Acad. Sci. USA,* **77**, 3816–3820.

Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science,* **249**, 505–510.

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

Wang, Y.-H., Amirhaeri, S., Kang, S., Wells, R. D. & Griffith, J. D. (1994). Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. *Science,* **265**, 1709–1712.

Wolffe, A. P. (1994). Transcription: in tune with the histones. *Cell,* **77**, 13–17.

Wong, A. K. C. & Rattner, J. B. (1988). Sequence organization and cytological localization of the minor satellite of mouse. *Nucl. Acids Res.* **16**, 11645–11661.

Wu, H.-M. & Crothers, D. M. (1984). The locus of sequence-directed and protein-induced DNA bending. *Nature,* **308**, 509–513.

Zinkel, S. S. & Crothers, D. M. (1987). DNA bend direction by phase sensitive detection. *Nature,* **328**, 178–181.

Edited by T. Richmond

(Received 17 October 1996; received in revised form 14 January 1997; accepted 17 January 1997)