Article No. mb981925 *J. Mol. Biol.* (1998) **281**, 253–260

JMB

TGGA Repeats Impair Nucleosome Formation

Hui Cao, Hans R. Widlund, Tomas Simonsson and Mikael Kubista*

Department of Biochemistry and Biophysics, Lundberg Institute, Chalmers University of Technology, Medicinaregatan 9C, SE-413 90, Göteborg Sweden

**Corresponding author*

Nucleosomes, the building blocks of chromatin, are responsible for DNA packaging in eukaryotic cell nuclei. They play a structural role in genome condensation, and influence transcription and replication. Properties of the DNA sequence, such as curvature and flexibility, direct the location of nucleosomes. DNA sequences that position nucleosomes have been identified and rules that govern their properties have been formulated. However, DNA sequences that are refractory to nucleosome formation have been less well characterised and it is possible that they may perturb or alter chromatin structure. Here we identify such sequences by selecting those that refrain from nucleosome formation from a large pool of synthetic DNA fragments with a central region of 146 random base-pairs fitted with adapters for PCR amplification. These were used for *in vitro* salt-induced reconstitution of nucleosomes under thermodynamic equilibrium conditions. Fragments that did not form nucleosomes were purified, amplified by PCR, and the reconstitution was repeated. After 17 rounds of negative selection, the material was highly enriched in sequences reluctant to form nucleosomes. Cloning and sequencing revealed that 35% of the molecules had long repeats of TGGA, and their affinity for histone octamers was about half that of average DNA.

Keywords: nucleosome; TGGA repeat; chromatin; nucleosome-positioning; selection

Introduction

Eukaryotic DNA complexed with basic histone proteins forms chromatin, in which the repeating unit is the nucleosome (Felsenfeld, 1978; Kornberg & Thomas, 1974). In *Saccharomyces cerevisiae* it contains 165 base-pairs (bp) of DNA, in most mammalian cell lines 160 to 200 bp, and in the sea urchin sperm it has as many as 260 bp (van Holde, 1989). In all species a conserved length of 146 bp wrapped around an octamer of histone proteins constitutes the nucleosome core particle. The remaining DNA is associated with histone H1 and connects adjacent nucleosomes in the chromatin filament. This linker DNA is more accessible to regulatory proteins than nucleosomal core DNA where one side of the DNA is hidden, since it faces towards the histone octamer (Simpson, 1990).

Previous studies have shown that the DNA sequence dictates the positioning of DNA around the nucleosome *in vitro* (Thoma & Simpson, 1985). *In vivo* the situation is more complex. Early characterisation of nucleosomal DNA revealed an abundance of AA/TT dinucleotides separated by roughly 10 bp (Drew & Travers, 1985; Muyldermans & Travers, 1994; Satchwell *et al*., 1986). This was interpreted as a nucleosome positioning signal and led to the construction of fragments with alternating A/T and G/C segments that formed very stable nucleosomes (Shrader & Crothers, 1989). Five repeated $(A/T)_3NN(G/C)_3NN$ motifs embedded in a DNA fragment, the so-called TG pentamer, is one of the strongest nucleosome forming elements known. Recently, we identified several other sequences that form exceedingly stable nucleosomes by using a selection strategy (Widlund *et al*., 1997). Fragments having repeated TATAAACGCC motifs, referred to as TATA tetrads, were found to form the most stable nucleosomes. They were even superior to the TG pentamer in nucleosome formation.

Not only sequences with a high affinity for histone octamers have structural and regulatory functions. Sequences that have a low affinity for

Abbreviations used: PCR, polymerase chain reaction; bp, base-pair(s); EMSA, electrophoretic mobility shift assay; RPE, sequencing, repeated primer extension sequencing; all sequences are written 5′-to-3′; ⊖, plim soll.

E-mail address of the corresponding author: Mikael.Kubista@bcbp.chalmers.se

0022–2836/98/320253–08 $30.00/0

histone octamers could also arrange nucleosomes, for example, by forming nucleosome-free regions. Such regions appear in chromatin as DNase I hypersensitive sites that are commonly associated with high transcriptional activity (Wolffe, 1995). Sequences known to disrupt chromatin structure are widely dispersed in eukaryotic genomes. They are found, for example, in telomeric DNA (Blackburn, 1991; Cacchione *et al*., 1997; Rhodes & Giraldo, 1995), in the polymorphic region flanking the human insulin gene (Hammond-Kosack *et al*., 1993), and in the promoter regions of many genes (Albert *et al*., 1997; Batson *et al*., 1993; Lu *et al*., 1992; Shimada *et al*., 1986; Verdone *et al*., 1996).

Here we identify sequences that are refractory to nucleosome formation. Starting from a population of synthetic random DNA sequences fitted with adapters for PCR we reconstituted nucleosomes and amplified the fragments that were not incorporated. This negative selection yielded a population of fragments that formed nucleosomes to a much lesser extent than average. Cloning and sequencing revealed that they were highly enriched in TGGA repeats.

Results

We have selected the sequences that are most refractory to nucleosome formation from a DNA pool of about 10^{13} different synthetic fragments with the overall sequence: 5′-GTCGTGACTGG-GAAAACCCTGGCG-$(N)_{146}$-TCACACAGGAAA-CAGCTATGAC-3′. The selection was performed by repeated salt-induced nucleosome reconstitutions. Free DNA fragments were mixed with H1-depleted chromatin under conditions where 50 to 75% was incorporated into nucleosomes. Free and incorporated fragments were separated by gel electrophoresis and the free fragments were excised, extracted, amplified by PCR and used as templates in the next round of negative selection. The overall nucleosome stability was measured at several stages during the selection process and was found to decrease successively. A total of 17 cycles was performed, after which the final population was cloned. We sequenced 40 fragments and their inserts were 81 to 126 bp, with a mean of 101 bp, which is considerably less than the inserts in the synthetic starting material, which were 146 bp.

TGGA fragments

Out of the 40 cloned and sequenced fragments, 14 had a high abundance of TGGA repeats (Table 1), and are referred to as TGGA fragments. In most of them the TGGA repeats were arranged in long runs, although a G was occasionally missing, resulting in TGA repeats. This is particularly pronounced in the fragment that we call TGA (Table 1). The average length of the runs of TGGA repeats is 11 bp with the longest being 48 bp.

TGGA repeats are composed of the four dinucleotide steps TG, GG, GA and AT. In the TGGA fragments these account for almost 90% of all dinucleotide steps, which can be compared with the statistical expectation of 25% ($= 4/4^2$) (Table 2).

Table 1. Sequences of the 14 TGGA fragments and the two longest BADSECS

TGGA-fragments

TGGA-1

AGATGGATGGATGATGGATGGATGATGGATAGATGGATGATGGATGGATGGATGATGATGGATGAATAGATGGATGGATGGATGATGGATGGACGATGGATGGATAGATGGATGGATGG

TGGA-2

TGGATCGATGGGTGGATGCGACGGGTGGATCGGATGAGATGATGGATGGATGATCAAGATGATGATGGCATGGGTGGATCGGGTGGATGGATGGCATGAGATGATCGGATGGATGGATGATGAATG

TGGA-3

TGGATGGATGGATGGATGGATGGGTGGATGGATGAGATGATGGATGGATGATAAGATGATGGATGGATGGGTGGATAGGCGGATGGATGGTTGAGATGATGGAGGGATGGATGATAGATG

TGA

ATCATGGTGATGTTGATGATGATGATGATGGTGATGATGATGATGGTGATGGCGATGATGATGATGATGGATGGCGATGGTGATAATGATATGGGGAGTAGG

TGGATGGATGGATGGATGGGTGGATGGATGAGATGATGGATGGATGATAAGGTGATGGATGGATGGGTGGATGGGTGGATGGATGGATGGAGATGATGGACGGGTGCATGATGGATG

TGGATAGGTGGATGGGTAGATGGATGAATGAATGAGTAGATGGATGGATGGATGATGGATGGATGGATGGATGGATGGATGG

TGGATCGATGGATGGATGGGTGGATGGATCAGATGATGGATGTGATGATAAGATCGATCGGATGGATGGGTGGATGGCCGGATGGATGGATGAGATGACGGACGGATGGGTGATGGATG

TGGATGGATGGATGAATGGGTGGATGGATGAGATGATGGATGGATGATAAGATGATGGATGGATGGTGGATGGGCGGATGGATGGATGAGATGATGGATGGATGATGGATG

ATAGATGGATGAGTGGATGGATGGATGGGTGGATGGATGGACGGGTGGATGGGTGGATGGGTAGATGGATGGATGGATGAGTGGATGGATGGATGGGTGGATGGGTGGATGG

ATAGATGGATGAGTGGATGGATGGGTGGATGGATAGATGGGTGGATGGGTGGATGGGTGGATGGATGATGGATGGATGAGTGGATGGATGGATGGATGGGTGGATGGGTGGACGG

ATAGATGGATAAGTGGATGGATGGATGGGTGGATGGATGGATGGGTTGATGGGTGGATGGGTGGATGCATGGATAAGTGGATGGATGGACGGATGGGTGGATGGGTGGATGG

AAAGGGTGGATGGAATGGGATGGAATGGGGTGGAATGGATGGACGGATGGATGGGGTGGATAGAATGGATGGATGGATGGACGAATGGATGTGTGGATGGATGGATGGATGGATGGT

AGGTGGATGGATGGTGGATAGATGGATGGATGGACGGATGGATGGGTGGATAAATGGATGGATGGATGGACGAATGGATGTGTGGATGGATGGATGGATGGATGGATGG

ATAGATGGATGAGTGGGTGGATGGATGGGTGGATGGATGGATGGGTGGATGGGTGGATGGGCGGATGGATGATGGATGGATGAGTGGATGGATGGATGGATGGGTGGATGGGTGGATG

BADSECS

BADSECS-1

CACCACATAGTTTGGTGAAGCCGGCTTCCAGCAAGCACAGTCGACTAACAAAAACGAAATATCTGCCCCGGGAGACGAGCTCCTTGGATTACCGCTCTTGCCAACATG

BADSECS-2

TCTAGAGTGTACAACTATCTACCCTGTAGGCATCAAGTCTATTTCGGTAATCACTGCAGTTGCATCATTTCGATACGTTGCTCTTGCTTCGCTAGCAACGGACGATCGTACAAGCAC

TGA and TGGA steps are highlighted and arrows indicate boundaries between successive runs. The complete set of sequences is available at http://www.bcbp.chalmers.se/mbg.

Table 2. The prevalence as percentage of the four most abundant di-, tri- and tetranucleotide steps in the TGGAfragments

TG/CA	25.6	TGG/CCA	17.0	TGGA/TCCA	13.3
GG/CC	21.8	GGA/TCC	14.1	GGAT/ATCC	13.3
GA/TC	21.7	GAT/ATC	19.4	GATG/CATC	17.2
AT/AT	20.0	ATG/CAT	18.4	ATGG/CCAT	13.5
Total	89.1		68.9		57.3

The trinucleotides TGG, GGA, GAT and ATG that make up TGGA repeats, account for 69% of all trinucleotide steps (statistical expectation 6.25%), and the tetranucleotides TGGA, GGAT, GATG and ATGG comprise 57% of all tetranucleotide steps (statistical expectation 1.6%).

The inserts in the TGGA fragments were 86 to 126 bp, with an average of 114 bp. This is considerably longer than in the other selected fragments, but still shorter than in the starting material (146 bp). Three of the longest fragments, TGGA-1 (123 bp), TGGA-2 (126 bp), TGGA-3 (120 bp), and the one containing TGA repeats (107 bp), were chosen for further characterisation.

BADSECS fragments

The other selected fragments (26 out of 40) had no obvious sequence features and are referred to as BADSECS. Their inserts were 81 to 117 bp with a mean of 93 bp. The two longest, BADSECS-1 (108 bp) and BADSECS-2 (117 bp), were chosen for further characterisation.

Nucleosome affinities

The fragments chosen for further characterisation (Table 3) were reconstituted into nucleosomes and their affinity for histone octamers was studied by electrophoretic mobility shift assay (EMSA). They were compared with the synthetic starting material, and also with our previously constructed material from the mouse genome (Widlund *et al.*, 1997).

Figure 1 shows a typical gel where the selected sequences are compared with the starting material. The ratio between the affinity constants for the histone octamer of any two fragments in the gel can be calculated directly from the band intensities:

$$K_A/K_B = (I_A^{Nucl}/I_A^{free})/(I_B^{Nucl}/I_B^{free})$$

where the I values are the amounts of material, measured through radioactivity, in the form of either nucleosomes or free fragments. For example, from this experiment the $K_{TGGA\text{-}3}/K_{SM}$ ratio can be estimated to be $(21{:}79)/(50{:}50) = 0.27$. The ratio between affinity constants is related to the difference in the fragments' affinity for histone octamers as:

$$\Delta(\Delta G^{\ominus})_{A\rightarrow B} = -RT\ \ln(K_A/K_B)$$

which corresponds to the change in free energy for the hypothetical reaction:

$$A_{free} + B_{nucl} \rightarrow A_{nucl} + B_{free}$$

Assuming that thermodynamic equilibrium is established at room temperature (Shrader & Crothers, 1989; Lowary & Widom, 1997), this corresponds to a $\Delta G^{\ominus}_{TGGA} - \Delta G^{\ominus}_{SM}$ of about 3.2 kJ mol^{-1}.

The measured affinities varied among the ten performed experiments as shown in Figure 2. The average values are shown in Table 3. The four TGGA fragments were roughly equivalent, and had distinctly lower affinity for histone octamers than any other of the analysed fragments, and they formed about 3 kJ mol^{-1} less-stable nucleosomes than the synthetic starting material. The two BADSECS fragments did not form particularly weak nucleosomes. In fact, BADSECS-2 formed nucleosomes almost as stable as the starting material. We also found that the synthetic starting material had the same affinity for histone octamers as mouse nucleosomal DNA (Widlund *et al.*, 1997).

Table 3. Fragments and populations characterised

Fragment	Length (bp)	Sequence feature	Affinity (kJ/mol)
Mouse nucleosomal DNA	204	Mixed	≈0
Starting material	204	Random	0
Final population	160	Enriched	2.5
TGGA-1	123 + 58	TGGA	2.7
TGGA-2	126 + 58	TGGA	3.2
TGGA-3	120 + 58	TGGA	3.1
TGA	107 + 58	TGA	3.2
BADSECS-1	108 + 58	None	1.3
BADSECS-2	117 + 58	None	0.5

Affinities are expressed relative to the starting material. Length includes adapters.

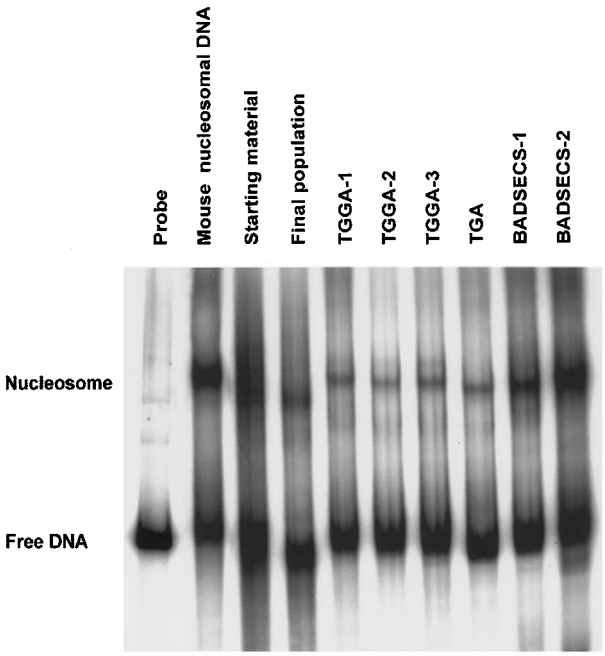

Figure 1. A typical EMSA showing histone octamer affinity of the fragments in Table 3. Lanes from left to right; 1, probe; 2, mouse nucleosomal DNA; 3, starting material; 4, final population; 5 to 7, TGGA repeats; 8, TGA repeat; 9 and 10, BADSECS fragments.

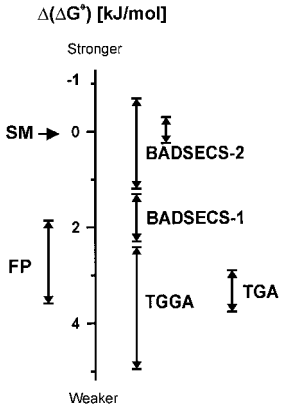

Figure 2. Changes in free energy upon nucleosome formation expressed relative to the synthetic starting material: $\Delta(\Delta G^{\circ}) = -RT \quad \ln[(I_i^{Nucl}/I_i^{free})/(I_{SM}^{Nucl}/I_{SM}^{free})]$, where I^{Nucl} and I^{free} are the amounts of nucleosomal and free DNA, calculated from radioactivity in each band, for the fragment of interest and the starting material (Shrader & Crothers, 1989). Mouse nucleosomal DNA was prepared from the mouse genome (Widlund *et al.*, 1997). SM is the synthetic starting material and FP the final population of the negative selection. BADSECS-1 and BADSECS-2 are the fragments with no obvious sequence features. TGGA represents the group of the three characterised TGGA fragments, and TGA is the fragment with TGA repeat. The data are based on ten independent reconstitutions with H1-depleted chromatin at a concentration of 5 A_{260}.

Gel mobility

The gel mobilities of the selected fragments were also analysed by gel electrophoresis. That they all migrate faster than the starting material (Figure 3) can be attributed to their shorter lengths. There are, however, some variations that cannot be explained solely by fragment length. For example, TGGA-1 migrates slower than both TGGA-2 and TGGA-3, although it has an intermediate length; the TGA repeat exhibits a substantially higher mobility than BADSECS-2 even though the length difference is only 1 bp; TGGA-2 and BADSECS-2 have the same mobility despite a length difference of 10 bp.

Discussion

Inefficient negative selection

The starting material for the negative selection was synthetic DNA with a central region of 146 random base-pairs fitted with adapters for PCR amplification. The adapters had the same sequences as those used in the selection of nucleosome forming sequences (Widlund *et al.*, 1997). The reason why we performed the negative selection starting from synthetic fragments is that the genomic starting material, due to the preparation procedure, would be depleted in sequences refractory to nucleosome formation.

The purpose of the negative selection was to enrich the population in sequences that are not incorporated into nucleosomes. To achieve high selection efficiency only a small fraction of fragments should remain unbound in each cycle; most should be incorporated into nucleosomes. It proved difficult to find such conditions. Using H1-depleted chromatin as histone source in the reconstitution did not result in incorporation of more than 50 to 75% of the fragments into nucleosomes; as much as 25 to 50% was retained for amplifica-

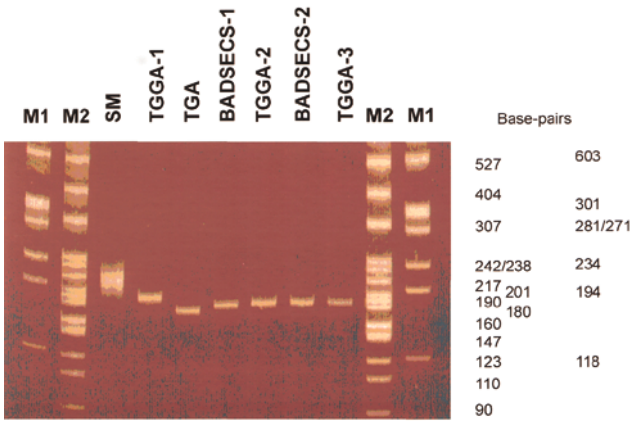

Figure 3. Polyacrylamide gel electrophoresis showing the electrophoretic mobilities of the free fragments. Lanes M1 and M2 are the markers pBR322/*Msp*I and ϕX174/*Hae*III. Lengths of markers in base-pairs are indicated at the right.

tion in each cycle. This made the negative selection much less efficient when compared with the selection we performed earlier (Widlund *et al*., 1997). We had to make 17 cycles before reaching what we considered a satisfactory degree of enrichment.

Shorter sequences form less-stable nucleosomes

The selected fragments are substantially shorter than the starting material. The lengths of the inserts lengths in the TGGA fragments were 86 to 126 bp, with a mean of 114 bp, and in the BADSECS they were 81 to 107 bp, with a mean of 93 bp. This is considerably shorter than the starting material insert (146 bp) and suggests that negative selection favours shorter fragments. Shorter fragments are indeed expected to form less-stable nucleosomes owing to entropic (= statistical) effects. If all base-pairs in a fragment are equivalent and n of them constitute a binding site, an m base-pair long fragment will have $(m - n + 1)$ binding sites (McGhee & von Hippel, 1974). If the microscopic affinity constant for one of these sites is β; the macroscopic affinity constant for the fragment is $K = \beta(m - n + 1)$ (Samuelsson *et al*., 1994). For nucleosomes n is normally considered to be 146†. Including adapters the lengths of the TGGA fragments characterised are about 180 bp (178, 181 and 184), and the length of the starting material is 204 bp. Hence, owing to shorter length, they should bind nucleosomes $(204 - 146 + 1)/(180-146+1) = 1.7$ times more weakly than the starting material. They bind 2.5–5 $(\exp^{\Delta G^{\circ}/RT})$ times more weakly (Figure 2). Hence, a substantial part of the difference in affinity should be due to their sequences.

The BADSECS fragments characterised were 166 and 175 bp, and they should bind histone octamers 2 to 2.8 times more weakly than the starting material, owing to their shorter lengths. For these sequences their shorter lengths account for the whole difference in affinity relative to the starting material, and their sequences may not be important. The rest of the BADSECS are even shorter, which suggests they too have been selected because of shorter length and not because of sequence.

In our selection of nucleosome positioning sequences we also found the selected fragments to be shorter than the starting material (Widlund *et al*., 1997). This may seem contradictory. However, in these fragments the base-pairs were not equivalent and each fragment had a small number of preferred binding sites. Increased length does not then provide additional preferred binding sites, and only marginally affects overall affinity through entropy changes. On the other hand, it leads to destabilisation of the nucleosome, owing to electrostatic repulsion between the fragment ends that are brought together by wrapping around the histone octamer.

But where do the shorter fragments come from in the negative selection? The synthetic starting material was purified by gel electrophoresis and was homogeneous in length. In our selection we find 14 TGGA fragments that all are shorter than the starting material. Since there is no reason why the starting material should contain only shorter TGGA fragments, we conclude that they had the correct length from the beginning and shortened during the selection process. Although the BADSECS may originally have been contaminating shorter sequences, it is likely that they too have become shorter during the PCR. We do not imply that PCR in general shortens DNA, but we think that the combination of selection pressure for shorter fragments in the reconstitution procedure with occasional errors made by the *Taq* polymerase may have produced such a bias.

The TGGA fragments were not selected because of guanine tetraplex formation

The TGGA repeats in the selected fragments exhibit a typical motif for the formation of guanine tetraplexes (Sen & Gilbert, 1988; Sundquist & Klug, 1989). These are exceedingly stable structures, which would hardly incorporate into nucleosomes. They are built up of square coplanar arrays of four guanine bases associated *via* a cyclic Hoogsteen hydrogen bonding arrangement that involves N1, N7, O6 and N2 of each guanine base (Guschlbauer *et al*., 1990). They are further stabilised by the precise co-ordination of cations, especially potassium, in the cavities formed between the guanine tetrads (Williamson, 1994).

Guanine tetraplex formation is a frequent nuisance in PCR where the structure causes premature chain termination. However, this can be avoided by the use of potassium-free buffers (Woodford *et al*., 1995) or 7-deazaguanosine instead of guanosine (Woodford *et al*., 1994). We tested both but saw no effects, suggesting that the TGGA repeats do not form tetraplexes under our experimental conditions (data not shown).

Runs containing exclusively TG, GG, GA and AT dinucleotide steps refrain from nucleosome formation

When broken up into dinucleotide steps TGGA repeats are composed of the four steps TG, GG, GA and AT. The only DNA sequences that can be formed using these steps exclusively are $T(G)_nA$ repeats ($n \geqslant 1$) and poly(dG). This suggests these sequences may adopt structures that impair nucleosome formation, and that these are inter-

† Growing evidence indicates that the nucleosome affinity is dominated by the central 120 bp. Therefore, $n = 120$ may be more appropriate for nucleosome core particles (Hayes *et al*., 1990; Widlund *et al*., 1997; Luger *et al*., 1997). This would make the length dependence less important and is more consistent with the data presented here.

rupted by any other dinucleotide step than TG, GG, GA and AT.

In addition to TGGA steps, the selected fragments also contain TGA steps. The TGA fragment, for example, contains TGA steps almost exclusively. There are, however, no runs with more than two consecutive guanine residues. This could mean that such repeats have other properties and have been lost during the selection, or, alternatively, that they were under-represented in the starting material owing to low synthetic yield.

Recently, several telomeric sequences that have a low propensity to form nucleosomes were found (Cacchione *et al*., 1997). These sequences could all be written $(G_nT_mA_{[0\ or\ 1]})_x$, which suggests that the dinucleotide steps GG, GT, TT, TA and AG may also constitute a group that, when uninterrupted, forms a DNA structure that impairs nucleosome formation.

Genomic nucleosomal DNA does not exhibit enhanced stability

Comparison of the random synthetic starting material used here with the mouse nucleosomal DNA (Widlund *et al*., 1997) shows no significant difference in nucleosome stability (Figure 2, Table 3). In accordance with earlier observations (Lowary & Widom, 1997), we conclude that eukaryotic nucleosomal DNA does not exhibit enhanced nucleosome stability.

TGGA repeats are present in higher eukaryotic genomes

Since the starting material was synthetic, one may ask whether the selected sequences have any biological significance. In fact, they may not even exist in genomic DNA. Data bank searches identified TGGA repeats in genomes of higher eukaryotes. There are as many as 249 TGGA repeats in the upstream region of the human gene encoding myelin basic protein (Boylan *et al*., 1990). A region based on 165 TGGA repeats is located 1100 to 1750 bp upstream from the human myoglobin gene (Weller *et al*., 1984), and the intron between exon 12 and exon 13 within the human red blood cell anion exchanger 1 gene contains 31 TGGA repeats (Schofield *et al*., 1994). Yet another TGGA repeat region is located at the 3′ end of the second intron in the class II Eb gene of mouse (Saha *et al*., 1993). Notably, we did not find any TGGA repeat regions in prokaryotic genomes. In the lower eukaryote *S. cerevisiae*, whose genome has been fully sequenced, no repeat longer than three TGGA exists.

Materials and Methods

Preparation of H1-depleted chromatin

H1-depleted chromatin was prepared from mouse Ehrlich ascites cells as described (Widlund *et al*., 1997).

Construction of starting material

DNA fragments with 146 bases of random sequence fitted with adapters, 24 and 22 bases in length, were synthesised and purified by denaturing gel electrophoresis. Using the same primers as described (Widlund *et al*., 1997), 31 and 27 bp in length, about 10^{13} molecules of synthetic DNA were amplified by PCR under standard conditions (2.5 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.8), 50 mM KCl, *Taq* polymerase 3 units/100 μl). The 204 bp product, which had restriction sites for *Eco*RI and *Bam*HI at each end, was purified by agarose gel electrophoresis, electroeluted, and precipitated with ethanol. It was then dissolved in 20 mM Tris-HCl buffer (pH 7.6) and used as starting material for the negative selection.

Reconstitution of nucleosomes

DNA fragments were digested with *Eco*RI and 3′-labeled using [α-^{32}P]dATP and Klenow fragment. A 1 pmol sample of radiolabelled fragments was mixed with H1-depleted chromatin (4 μg in cycles 1 to 8 and 12 μg in cycles 9 to 17) in 1 M NaCl, 20 mM Tris-HCl (pH 7.6), 0.1% (v/v) Nonidet P-40 and 100 μg/ml bovine serum albumin. The mixture was incubated for 30 minutes at +37°C and then diluted to 100 mM NaCl by three additions, made at 20-minute intervals, of low salt buffer at room temperature. The samples were mixed with 1/6 volume of 30% (v/v) glycerol and loaded on a 1.5 mm native 6% (acrylamide to bisacrylamide, 75:1 (w/v)) polyacrylamide gel in 0.5 × TBE buffer. The gel was run at 4 W and room temperature for two hours. The radioactivity in the nucleosome and free DNA bands was quantified using a storage phosphorus screen and analysed by a phosphor-imager (Molecular Dynamics).

Negative selection

A 1 pmol sample of labelled DNA fragments was reconstituted into nucleosomes by mixing with H1-depleted chromatin. The ratio of nucleosomes to free DNA was adjusted by varying the concentration of chromatin to obtain 50 to 75% incorporation. Free DNA fragments and nucleosomes were separated by polyacrylamide gel electrophoresis in 0.5 × TBE buffer at room temperature for two hours. The free fragments were recovered by electro-elution, precipitated with ethanol and used as template in the next round of PCR. After 17 cycles the population was cloned and sequenced.

Cloning and sequencing

The final population was digested with *Eco*RI/*Bam*HI and ligated into pCR-Script SK+ vector (Stratagene) between the *Eco*RI/*Bam*HI sites, transformed into *Escherichia coli* DH5α cells and spread on agar plates. The colonies containing inserts were picked by blue/white screening and grown in LB medium. The presence of plasmid inserts was verified by PCR and prepared by flow column purification (Qiagen). Approximately 0.2 μg plasmid DNA was used for repeated primer extension sequencing (RPE sequencing) (40 cycles using Amersham ThermoSequenase with 2 μl of the N-mixes and 10 pmol of fluorescent primer in a total volume of 8 μl). The

samples were analysed on an A.L.F. DNA sequencer (Pharmacia).

Gel migration analysis

Samples were produced by PCR amplification of cloned fragments and analysed on a native 8% (acrylamide to bisacrylamide, 30:1 (w/v)) polyacrylamide gel run in 0.5 × TBE buffer at 4°C. The gel was stained with ethidium bromide and photographed.

Acknowledgements

The authors thank Professor Donald M. Crothers for valuable discussions and Professor Peter E. Nielsen's lab for providing mouse Ehrlich ascites cell nuclei. The Swedish Cancer Foundation and the Swedish Natural Sciences Foundation supported this work. H.R.W. has a fellowship from the Sven and Lilly Lawski foundation.

References

Albert, T., Mautner, J., Funk, J. O., Hortnagel, K., Pullner, A. & Eick, D. (1997). Nucleosomal structures of c-myc promoters with transcriptionally engaged RNA polymerase II. *Mol. Cell. Biol.* **17(8)**, 4363–4371.

Batson, S. C., Rimsky, S., Sundseth, R. & Hansen, U. (1993). Association of nucleosome-free regions and basal transcription factors with in vivo-assembled chromatin templates active in vitro. *Nucl. Acids Res.* **21(15)**, 3459–3468.

Blackburn, E. H. (1991). Structure and function of telomeres. *Nature,* **350(6319)**, 569–573.

Boylan, K. B., Ayres, T. M., Popko, B., Takahashi, N., Hood, L. E. & Prusiner, S. B. (1990). Repetitive DNA $(TGGA)_n$ 5′ to the human myelin basic protein gene: a new form of oligonucleotide repetitive sequence showing length polymorphism. *Genomics,* **6(1)**, 16–22.

Cacchione, S., Cerone, M. A. & Savino, M. (1997). In vitro low propensity to form nucleosomes of four telomeric sequences. *FEBS Letters,* **400(1)**, 37–41.

Drew, H. R. & Travers, A. A. (1985). DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* **186(4)**, 773–790.

Felsenfeld, G. (1978). Chromatin. *Nature,* **271(5641)**, 115–122.

Guschlbauer, W., Chantot, J. F. & Thiele, D. (1990). Four-stranded nucleic acid structures 25 years later: from guanosine gels to telomer DNA. *J. Biomol. Struct. Dynam.* **8(3)**, 491–511.

Hammond-Kosack, M. C., Kilpatrick, M. W. & Docherty, K. (1993). The human insulin gene-linked polymorphic region adopts a G-quartet structure in chromatin assembled in vitro. *J. Mol. Endocrinol.* **10(2)**, 121–126.

Hayes, J. J., Tullius, T. D. & Wolffe, A. P. (1990). The structure of DNA in a nucleosome. *Proc. Natl Acad. Sci. USA,* **87(19)**, 7405–7409.

Kornberg, R. D. & Thomas, J. O. (1974). Chromatin structure; oligomers of the histones. *Science,* **184(139)**, 865–868.

Lowary, P. T. & Widom, J. (1997). Nucleosome packaging and nucleosome positioning of genomic DNA. *Proc. Natl Acad. Sci. USA,* **94(4)**, 1183–1188.

Lu, Q., Wallrath, L. L., Allan, B. D., Glaser, R. L., Lis, J. T. & Elgin, S. C. (1992). Promoter sequence containing $(CT)_n \cdot (GA)_n$ repeats is critical for the formation of the DNase I hypersensitive sites in the *Drosophila hsp26* gene. *J. Mol. Biol.* **225(4)**, 985–998.

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

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(2)**, 469–489.

Muyldermans, S. & Travers, A. A. (1994). DNA sequence organization in chromatosomes. *J. Mol. Biol.* **235(3)**, 855–870.

Rhodes, D. & Giraldo, R. (1995). Telomere structure and function. *Curr. Opin. Struct. Biol.* **5(3)**, 311–322.

Saha, B. K., Shields, J. J., Miller, R. D., Hansen, T. H. & Shreffler, D. C. (1993). A highly polymorphic microsatellite in the class II Eb gene allows tracing of major histocompatibility complex evolution in mouse. *Proc. Natl Acad Sci. USA,* **90(11)**, 5312–5316. (Published erratum appears in *Proc. Natl Acad Sci. USA*, 1993 Sept. 15; **90(18)**, 8757).

Samuelson, 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(4)**, 659–675.

Schofield, A. E., Martin, P. G., Spillett, D. & Tanner, M. J. (1994). The structure of the human red blood cell anion exchanger (EPB3, AE1, band 3) gene. *Blood,* **84(6)**, 2000–2012.

Sen, D. & Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature,* **334(6180)**, 364–366.

Shimada, T., Inokuchi, K. & Nienhuis, A. W. (1986). Chromatin structure of the human dihydrofolate reductase gene promoter. Multiple protein-binding sites. *J. Biol. Chem.* **261(3)**, 1445–1452.

Shrader, T. E. & Crothers, D. M. (1989). Artificial nucleosome positioning sequences. *Proc. Natl Acad. Sci. USA,* **86(19)**, 7418–7422.

Simpson, R. T. (1990). Nucleosome positioning can affect the function of a cis-acting DNA element in vivo. *Nature,* **343(6256)**, 387–389.

Sundquist, W. I. & Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature,* **342(6251)**, 825–829.

Thoma, F. & Simpson, R. T. (1985). Local protein-DNA interactions may determine nucleosome positions on yeast plasmids. *Nature,* **315(6016)**, 250–252.

van Holde, K. E. (1989). *Chromatin*, Springer Series in Molecular Biology, Springer-Verlag, New York.

Weller, P., Jeffreys, A. J., Wilson, V. & Blanchetot, A. (1984). Organization of the human myoglobin gene. *EMBO J.* **3(2)**, 439–446.

Verdone, L., Camilloni, G., Di Mauro, E. & Caserta, M. (1996). Chromatin remodeling during *Saccharomyces cerevisiae* ADH2 gene activation. *Mol. Cell. Biol.* **16(5)**, 1978–1988.

Widlund, H. R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P. E., Kahn, J. D., Crothers, D. M. & Kubista, M. (1997). Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.* **267(4)**, 807–817.

Williamson, J. R. (1994). G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 703–730.

Wolffe, A. P. (1995). *Trans*-acting factors, DNase I sensitivity, DNase I hypersensitivity sites and chromosomal architecture. In *Chromatin: Structure and Function* (Wolffe, A., ed.), 2nd edit., pp. 199–208, Academic Press, London.

Woodford, K. J., Howell, R. M. & Usdin, K. (1994). A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *J. Biol. Chem.* **269(43)**, 27029–27035.

Woodford, K., Weitzmann, M. N. & Usdin, K. (1995). The use of K(+)-free buffers eliminates a common cause of premature chain termination in PCR and PCR sequencing. *Nucl. Acids Res.* **23(3)**, 539.

Edited by T. Richmond

(Received 6 November 1997; received in revised form 27 April 1998; accepted 30 April 1998)