

Long-range oscillation in a periodic DNA sequence motif may influence nucleosome array formation

Yamini Dalal, Tomara J. Fleury, Alfred Cioffi and Arnold Stein*

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

Received December 7, 2004; Revised and Accepted January 14, 2005

ABSTRACT

We have experimentally examined the characteristics of nucleosome array formation in different regions of mouse liver chromatin, and have computationally analyzed the corresponding genomic DNA sequences. We have shown that the mouse adenosine deaminase (MADA) gene locus is packaged into an exceptionally regular nucleosome array with a shortened repeat, consistent with our computational prediction based on the DNA sequence. A survey of the mouse genome indicates that <10% of 70 kb windows possess a nucleosome-ordering signal, consisting of regular long-range oscillations in the period-10 triplet motif non-T, A/T, G (VWG), which is as strong as the signal in the MADA locus. A strong signal in the center of this locus, confirmed by *in vitro* chromatin assembly experiments, appears to cooperate with weaker, in-phase signals throughout the locus. In contrast, the mouse odorant receptor (MOR) locus, which lacks locus-wide signals, was representative of ~40% of the mouse genomic DNA surveyed. Within this locus, nucleosome arrays were similar to those of bulk chromatin. Genomic DNA sequences which were computationally similar to MADA or MOR resulted in MADA- or MOR-like nucleosome ladders experimentally. Overall, we provide evidence that computationally predictable information in the DNA sequence may affect nucleosome array formation in animal tissue.

INTRODUCTION

DNA of higher organisms is non-coding to a large extent, and is packaged into chromatin. The possibility exists that the DNA sequence can influence chromatin structure and function

in ways not yet appreciated (1). Virtually all the DNA in the nucleus of a eukaryotic cell is packaged into nucleosome arrays. These arrays are condensed into higher-order chromatin structures, which appear to vary (2–5). The cause of apparent variability in chromatin structure and its relationship with the function is not clear. It is plausible that nucleosome arrays displaying different nucleosome arrangements form stretches of chromatin having distinct higher-order structures (6,7) or different modes of flexibility (8) due to variations in the rotational orientation between adjacent nucleosomes (2). Knowledge of the mechanisms by which distinctive higher-order chromatin structures intrinsically form should provide new insights into how histone modification and the presence of non-histone chromosomal proteins can remodel chromatin structure in a functional way (9). For example, instead of first unfolding a generic solenoid-like structure and subsequently rearranging it into some other type of chromatin architecture that is active, remodeling may simply involve making subtle alterations to an already distinctive structure.

Results obtained using a linker histone-dependent *in vitro* chromatin assembly system suggested that sequence motifs could be present in some sequences that facilitate the formation of regularly spaced nucleosomes (6). It was also observed that regular oscillations of period-10 non-T, A/T, G (VWG), a motif that is very abundant in vertebrate genomes (10), occurred specifically in regions of DNA that ordered nucleosomes *in vitro* (11). The period of these oscillations, assessed by Fourier analysis, corresponded almost exactly to a value that was equal to twice the measured nucleosome repeat in all cases for chromatin assembled *in vitro*. Moreover, DNA regions that did not possess a single strong Fourier (period) peak did not order nucleosomes into regular arrays. These observations suggested the hypothesis that nucleosome ordering by linker histones might be facilitated by a dinucleosome period signal consisting of regular period-10 VWG oscillations. Further support for this hypothesis was obtained by making small alterations in the chicken ovalbumin gene sequence, which affected nucleosome array formation *in vitro* in a computationally predictable way (12). Although

*To whom correspondence should be addressed. Tel: +1 765 494 6546; Fax: +1 765 494 0876; Email: astein@bilbo.bio.purdue.edu

Present address:

Yamini Dalal, Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

© The Author 2005. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oupjournals.org

the finding of an apparent dinucleosome period signal, rather than a mononucleosome period signal, was unexpected, it is quite consistent with linker histone-dependent nucleosome ordering (13). Linker histones should be able to easily order the intervening nucleosomes that exist between other nucleosomes whose positions are restricted to some extent by the presence of the dinucleosome period signal. Additionally, it was shown that nucleosomes tend to avoid the DNA regions that have low counts of period-10 VWG, presumably because they are less bendable (12).

In this report, we have examined nucleosome arrays in different regions of mouse liver chromatin. We show that the *mouse adenosine deaminase* (MADA) gene on chromosome 2 is contained in a genomic region with a very regular nucleosome array, possessing a nucleosome repeat that is ~11 bp shorter than that of bulk mouse liver chromatin. The high degree of regularity and the value of the nucleosome repeat are consistent with the results of a computational analysis of 70–100 kb of DNA sequence in this region. A survey of the mouse genome indicates that <10% of the mouse genomic DNA is predicted to contain nucleosome arrays that have a similar regularity. An anonymous genomic DNA region that was computationally similar to MADA genomic sequences was also found to be MADA-like experimentally. *In vitro* chromatin assembly experiments confirmed our computational prediction of a strong period-10 VWG signal in the center of the locus. Moreover, the *in vitro* experiments revealed variations in nucleosome array formation on isolated 3–10 kb-insert subclones of the MADA gene that correlated very well with our computational predictions using small 3 kb windows. A more typical locus, resembling much of the genome computationally, is exemplified by the mouse odorant receptor (MOR) locus on chromosome 7. This large region lacks long-range in-phase period-10 VWG oscillations and resembles the bulk chromatin experimentally in the local regions that lack signals. When we selected an anonymous genomic DNA region that, like the MOR locus, lacked period-10 VWG signals, it also possessed nucleosome ladders similar to those of the bulk chromatin.

This study attempts to address some of the larger questions in DNA research such as the purpose of the vast amounts of non-coding DNA in the genomes of higher organisms, the organization of chromosomal domains and the rules that govern their formation. Our results suggest that it may be possible to predict computationally the regions of the genome having distinctive chromatin structures.

MATERIALS AND METHODS

Computational analysis

Genomic DNA sequences for 29 807 bp of the MADA gene (accession number U73107), 192 082 bp of the MADA gene with flanking DNA (accession number AL591490) and 63 797 bp of the MOR locus (accession number AF071080) were retrieved from GenBank. The chromosome coordinates for the 100 kb (signal-containing) sequence #1 was: build 33, chromosome 2, 4,239,454–4,339,454; and the accession number of the 100 kb (signal-lacking) sequence #2 was: build 33, chromosome 6, NT_039340, AC023286, nucleotides 8057–108 057. Sequences were analyzed for their variations in

period-10 VWG content as described previously (11). Briefly, the occurrences of VWG/CWB with a periodicity from 10.00 to 10.33 were counted in a sliding 102 bp window, ± 51 bp from each VWG position. These histogram data were then averaged in a sliding 60 bp window (5 bp increments) to generate a continuous oscillating curve of the average period-10 VWG count versus GenBank nucleotide number. The total number of VWG/CWB occurrences in a sliding 600 bp window was also computed, and used to apply a small correction for the presence of VWG-poor or VWG-rich regions, as described previously. The regularity and the periods of the long-range period-10 VWG oscillations were assessed by Fourier analysis using the window size and the nucleotide numbers stated.

DNA constructs

Cosmids C2.2 and C3.5 (containing the full-length mouse ADA gene) and subclones EE3.6 and pKSH8.8 were kindly provided by Dr Kellems (14). The EE3.6 construct provided was a 3.6 kb EcoRI fragment inserted into the EcoRI site of the cloning vector Bluescript. The pKSH8.8 construct provided was an 8.8 kb fragment inserted into the Hind III site of Bluescript. The M6.5 construct was a 6.5 kb EcoRI–SphI fragment of C2.2 inserted into the EcoRI/SphI sites of pUC19. The M10.4 construct was a 10.4 kb EcoRI fragment of C2.2 inserted into the EcoRI site of pUC19. Plasmids containing MOR subclones were provided by Catherine Farrell and Gary Felsenfeld (NIDDKD, NIH, Bethesda, MD). These constructs were used to transform the DH5 α bacterial strain (Invitrogen) using standard conditions. Highly pure supercoiled DNA was prepared either by standard maxiprep technique with CsCl–ethidium bromide banding (15), or by using the Wizard Maxiprep Kit (Promega). Residual RNA was removed by ultracentrifugation through 1 M NaCl (15). DNA samples were generally stored in 10 mM Tris–HCl, pH 8.0 and 1 mM Na₂EDTA at 4°C.

In vitro chromatin assembly

Pure supercoiled DNA was first reconstituted into chromatin at physiological core histone to DNA ratios, using salt dialysis as described previously (16). Briefly, 20 μ g of DNA at a concentration of 200 μ g/ml was mixed with purified core histones, obtained by salt extraction from chicken erythrocyte nuclei, at ratios ranging from 0.70–0.90 μ g core histones/ μ g DNA in 1 M NaCl, 10 mM Tris–HCl, pH 8.0 and 1 mM Na₂EDTA for 30 min, followed by dialysis at room temperature against 0.8 M NaCl, 10 mM Tris–HCl, pH 8.0 and 1 mM Na₂EDTA and 0.6 M NaCl, 10 mM Tris–HCl, pH 8.0 and 1 mM Na₂EDTA for 2 h each, followed by 2 h against 20 mM Tris–HCl, pH 7.2 and 0.2 mM Na₂EDTA. The chromatin was then incubated overnight at 37°C, either with or without the linker histone H5 (0.5 μ g H5/ μ g DNA), at a final concentration of 100 μ g/ml DNA, 2 mg/ml polyglutamic acid, pH 7.2, 0.15 M NaCl, 20 mM Tris–HCl and 0.2 mM Na₂EDTA. Prior to digestion with micrococcal nuclease (MNase), the salt was usually dialyzed out.

MNase digestion

The quality and extent of the nucleosome arrays formed *in vitro* were assessed by MNase digestion of the chromatin in 1 mM CaCl₂, 4.5 U MNase/ μ g DNA for 15 s–2 min at 37°C. Digestion was terminated by plunging the sample into 0.2 mg

proteinase K/ml, 10 mM EDTA and 0.1% SDS and it was then incubated for 15–30 min at 50°C. Samples were adjusted to 0.1 M Tris–HCl, pH 8.0 and 1% SDS, and extracted once with phenol/chloroform/isoamyl alcohol (25:24:1) and again with chloroform, subsequently precipitated with ethanol after adding NaCl to 0.15 M from a concentrated stock solution. The dried DNA pellet was dissolved in 1× gel loading buffer for gel electrophoresis. The degree of order in nucleosome arrays in native chromatin was determined by partial MNase digestion of the DNA in the chromatin of liver cell nuclei. Nuclei were prepared from mouse liver (17). Nuclei containing ~1 mg of DNA were gently pelleted and resuspended in 1 ml of 0.1 M NaCl, 10 mM Tris–HCl (pH 8.0) and 1 mM EDTA. After equilibration for 5 min at 37°C, 0.1 M CaCl₂ was added to a final concentration of 2 mM, then 30 U of MNase was added, and the sample was digested for 0.5, 1.5, 3 or 5 min (or as stated otherwise). After deproteinization, the nucleic acid was treated with RNase A and prepared for electrophoresis. The DNA fragments were subjected to agarose gel electrophoresis, blotted, and regions of the DNA of interest were detected by Southern hybridization.

Electrophoresis

Nucleosome oligomer size DNA fragments (~2.5 µg/lane for *in vitro* experiments, or ~10 µg/lane for DNA from nuclei) were resolved on 1.5% agarose gels in 1× TBE using a 13.5 × 13.5 cm bridge gel apparatus run for 3 h 20 min at 100 V. The gels were then either stained in a dilute ethidium bromide solution for direct visualization by UV, or further processed for Southern blots. Lambda DNA cut with AflIII, labeled with ³²P, was used as size markers.

Southern blots and hybridizations

DNA fragments were transferred after electrophoresis to charged nylon membranes (GeneScreen Plus; PerkinElmer) under standard capillary blot conditions and the transfer was generally overnight. DNA was fixed to the membrane wet by UV-cross linking. All probes used were restriction fragments from the specified subclones, generally chosen for their location near the center of the 3 kb window used for the Fourier in the computational analysis. MADA probe *a* was a 631 bp EaeI fragment from clone EE3.6, probe *b* was a 516 bp AvalI fragment from clone M6.5, probe *b'* was an 820 bp HindIII/PstI fragment from clone M6.5, probe *c* was a 743 bp EcoNI fragment of pKSH8.8 and probe *d* was an 814 bp BanI fragment from clone M10.4. MOR probe *b* was a 658 bp SfeI fragment (AF071080 nucleotide numbers 29 219–33 342), probe *c* was a 740 bp HindIII fragment (nucleotide numbers 49 148–49 887) and MOR probe *d* was a 425 bp AflIII fragment (nucleotide numbers 53 775–54 199), each obtained from subclones. The probe from roughly the center of sequence #1 was a 450 bp PCR-amplified fragment from mouse liver genomic DNA using the following primers: FP, AGCAGC-CAGGTAAAGCATCCTACA; RP, CAGCAGCTTTGGC-TGGTATGCAAT. The probe from roughly the center of sequence #2 was a 493 bp PCR-amplified fragment from mouse liver genomic DNA using the following primers: FP, ACGTGAGCTGATGTGACTTGGGAT; RP, TCCACCTCT-ATCAAAGCCCAGTGT.

To purify the restriction fragments or the PCR-amplified DNA fragments to be used as probes, the DNA was run on an agarose gel, the gel was stained to visualize the band of interest, and the DNA fragment was excised from the gel. The DNA was purified from the gel slice using the Qiaex gel purification kit (Qiagen). The pure denatured DNA probe fragment (25–50 ng) was labeled with [α -³²P]dATP by random priming (18), the labeled DNA was denatured, chilled on ice and supplemented with denatured salmon sperm DNA (at a final concentration of 50 µg/ml). Final probe concentrations in the hybridization buffer were 50–100 ng/ml for the *in vitro* experiments and 2.5 ng/ml (MADA) or 5 ng/ml (MOR, sequence #1, and sequence #2) for nuclei. All pre-hybridizations, hybridizations and washes were performed in a Hyb-Aid oven at 65°C using the manufacturer's recommendations, except for the high stringency washes. The high stringency washes for the MADA probe were done in 0.1× SSC and 0.1% SDS at 65°C followed by a wash in 0.1× SSC at room temperature. For the MOR probe, the high stringency washes were done as above, but at 63°C. The blots were exposed to Biomax MR film (Kodak) from 30 min to 2 h for *in vitro* experiments or from 1 to 3 days using a Kodak Biomax MS intensifying screen for nuclei. To assess the specificity of hybridization, in MNase digests of nuclei, purified mouse genomic DNA digested with appropriate restriction enzymes (usually PstI and HindIII) were included in lanes labeled D.

Nucleosome repeat analysis

The extent and quality of the nucleosome array was determined by the pattern obtained from MNase digestion. The individual oligomer bands were sized using a calibration curve based on the standard molecular weight markers. The slope of the best-fit straight line gives the nucleosome repeat length (19).

RESULTS

The MADA gene is packaged into a distinctive and highly regular nucleosome array in mouse liver nuclei

Figure 1 shows a map of the MADA gene, with ~20 kb of upstream sequence and ~20 kb of downstream sequence, extending 70 kb overall. Probes and clones used in this study are also indicated. MNase cuts the DNA in chromatin between the nucleosomes. Hence, the more distinct the bands of the (nucleosome) ladder of DNA fragments arising from excised nucleosome oligomers are, the more regular is the nucleosome array. Figure 2A shows typical nucleosome ladders obtained from bulk chromatin. About 10 bands can be detected. However, there is a considerable background above the 3mer. Figure 2B shows the same digest probed with hybridization probe *b'*, located roughly in the center of the gene. Probe *b* detected repetitive DNA and could not be used for digests of nuclei. It can be seen that the background is lower and the bands are more distinct than those of the bulk chromatin. Probes *c* and *d* gave very similar results to those of probe *b'* (data not shown, see Table 1 for a summary of the results), suggesting that the whole gene has a similar nucleosome arrangement. As a control, purified high molecular weight DNA was digested with MNase under the same conditions as used for nuclei (except that ~400 times less

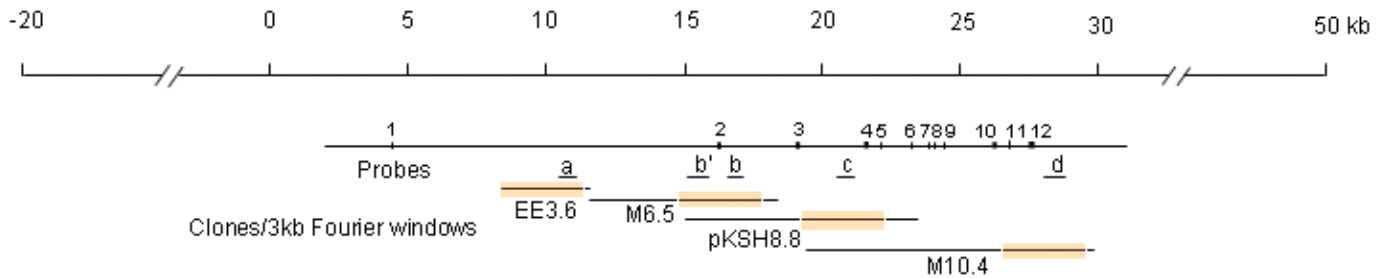


Figure 1. Map of the 70 kb MADA gene locus, including hybridization probes, subclones and computational windows used in the analysis. Numbering corresponds to the sequence in GenBank file accession number U73107. Exons are depicted as numbered black boxes. Positions of hybridization probes: *a*, *b'*, *b*, *c* and *d* are indicated. Positions of subclones: EE3.6, M6.5, pKSH8.8 and M10.4 are indicated, with the 3 kb computational windows in orange.

MNase/ μ g DNA was used). Figure 2C shows the result of probing this digest with MADA probe *b'*. It can be seen that the digestion pattern does not resemble a nucleosome ladder; discrete bands are not present. Moreover, the digestion pattern is very similar to that of bulk purified DNA, as detected by ethidium bromide (EtBr) staining (data not shown). Thus, the regular extended ladder seen in Figure 2B arises from the regular nucleosome arrangement on this region of DNA in mouse liver nuclei, and not from a regular arrangement of preferred cutting sites on the DNA.

Densitometer scans (Figure 2D) of lane 4 from the bulk chromatin (Figure 2A) and the MADA probe (Figure 2B) confirm that the bands above the 2mer are more distinct and the background signal is lower for the MADA probe than for the bulk chromatin. For example, for nucleosome oligomer bands 5–8, the superimposed densitometer tracings show that at equivalent background levels the peaks are greater for the MADA probe than for bulk chromatin. Moreover, for nucleosome oligomer bands 3 and 4, at equivalent peak heights the background signal is lower for MADA than for bulk chromatin. Additionally, more bands can be resolved for the MADA probe than for the bulk chromatin, despite the greater resolution of the photographic film used for the bulk chromatin DNA detection than the X-ray film used for the Southern blot.

While comparing the MADA ladders with bulk chromatin ladders, it needs to be taken into account that gel staining (bulk chromatin) detects DNA fragments in proportion to their weight, whereas Southern hybridization (MADA probe) detects fragments in proportion to their number. Thus, the larger the fragment detected on a stained gel, the more its intensity is over-represented on a number basis. To compare the ladders using the same basis, the raw data densitometer scans from Figure 2D were digitized, and the bulk chromatin ladder scan was corrected to reflect detection based on the fragment number rather than on fragment weight. The monomer was excluded from the analysis because it was under-represented in both the stained gel and the blot due to technical problems. It is now clear (Figure 2E) that the chromatin of the MADA gene is also more resistant to nuclease digestion than bulk chromatin. Assuming that the higher-order chromatin structures are equally accessible to the nuclease, irregular nucleosome arrays having occasional long linkers would be expected to be more susceptible to digestion than regular nucleosome arrays lacking long linkers.

On detailed analysis of the ladders (data not shown) it can furthermore be deduced that the nucleosome repeat on the

MADA gene is ~ 11 bp shorter than that of bulk chromatin. The bulk chromatin repeat is 194 ± 5 bp, whereas the MADA gene repeat is 183 ± 5 bp (Table 1). It can be seen that the 7mer of the MADA ladder (Figure 2B) is very close to 1268 bp marker position, consistent with 7×181 bp = 1267 bp, whereas, the 7mer of the bulk chromatin (Figure 2A) is larger (between the 1268 and the 1399 bp markers), consistent with 7×194 bp = 1358 bp. All of these observed differences (shorter nucleosome repeat length, more distinct bands and higher resistance to digestion) are consistent with the MADA gene locus containing a distinctive nucleosome arrangement.

Computational analysis of the DNA sequence predicts that the MADA locus is significantly different from most loci in mouse genomic DNA

We investigated whether the above results in mouse liver could be understood in terms of regular dinucleosome period oscillations in period-10 VWG, similar to what we found previously for chromatin assembled *in vitro* on relatively small (~ 5 kb length) DNA fragments (11,12). To determine whether there were regular oscillations in period-10 VWG in the vicinity of the gene-centered probe, and to determine the number of base pairs the regular oscillations extended, we performed the Fourier analysis in windows of increasing sizes, all centered on probe *b'*. Interestingly, we found that a predominant amplitude with a period in the range of 360–370 bp occurred for all windows tested from 3 to 100 kb in size. At windows greater than 100 kb, there was no longer a single predominant periodicity. The results for the 50, 70 and 100 kb windows are shown in Figure 3A–C, respectively. Whereas in the mid-size 50 kb window the non-physiological peaks with periods of ~ 310 , 330 and 450 bp made appreciable contributions to the Fourier transform, in the larger 70 kb window the 370 bp period peak was $>70\%$ larger than any of the other peaks, indicating that there is a strong ~ 370 bp period oscillation throughout the MADA locus. This predominant 370 bp period peak persisted up to the 100 kb window. This computational result is consistent with the observed 183 ± 5 bp nucleosome repeats throughout the gene locus, since 370 bp/ $2 = 185$ bp. Hence, the computational result suggests that there should be an ordered nucleosome array throughout this locus with a 185 bp repeat, very similar to what was observed experimentally. These computational results also suggest that a 70–100 kb window is the most appropriate window size for assessing the nucleosome array-forming potential of genomic DNA sequences.

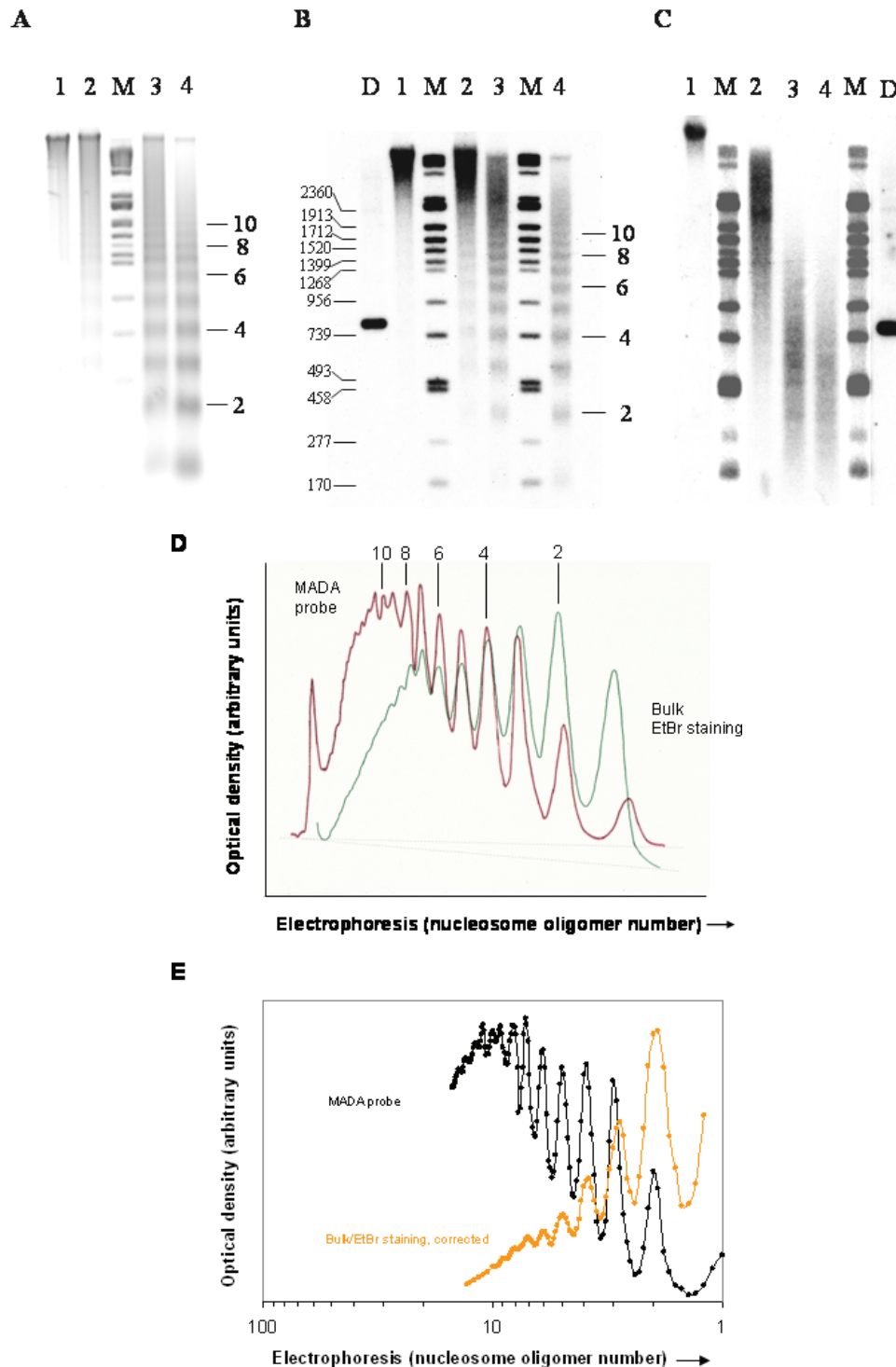


Figure 2. Nucleosome ordering on the MADA locus compared with that of the bulk chromatin in liver nuclei. (A) Total DNA from nuclei digested with MNase for 0.5, 1.5, 3 or 5 min (lanes 1–4, respectively) was stained with EtBr. Nucleosome oligomer band positions for the 2mer, 4mer, 6mer, 8mer and 10mer are indicated. The lane labeled M contained size markers [the sizes are given in (B)]. (B) Portions of the same digests as shown in (A); lanes 1–4 were run on another agarose gel under the same conditions, the gel was blotted and the membrane was hybridized to MADA gene probe *b'* (see Figure 1). Nucleosome oligomer bands are indicated as in (A). Lane D, Hind III/Pst I digest of purified mouse genomic DNA to assess hybridization specificity, as described in Materials and Methods. The sizes of the size marker bands (lanes M) are indicated. (C) Purified genomic DNA isolated from mouse liver was digested with MNase (0.01 U/ μ g DNA) for 0, 1, 2 or 3 min (lanes 1–4, respectively). The size markers (lanes M) were the same as those used in (A) and (B). (D) Raw data densitometer lane scan tracings of lane 4 from (A), bulk chromatin (green tracing), and (B), MADA probe (red tracing). Even integer nucleosome oligomer peaks are numbered. The dashed gray lines indicate baselines for each scan. (E) Digitized bulk chromatin densitometer lane scan from (B) (orange curve) corrected to detection by number of DNA fragments (as in blotting), instead of detection by fragment weight (as in EtBr staining). Optical density values at every position of the tracing were multiplied by 2/(the fragment size in nucleosomal units), thus correcting each value by a factor proportional to the size difference from the nucleosome dimer (midpoint of the dimer peak). The monomer peak was excluded from the analysis. The uncorrected digitized MADA probe densitometer lane scan from (B) is also shown for comparison (black curve).

To determine how common MADA-like long-range oscillations in period-10 VWG are in the genome, we surveyed mouse genomic DNA by arbitrarily selecting 1 Mb or greater regions from each chromosome (except Y), and computing Fourier transforms (Fts) in 100 kb windows, sliding 50 kb at

a time. Thus, ~20 Mb were sampled using ~400 (independent) windows. Five categories of Fts were apparent. Approximately 10% of the Fts exhibited MADA-like peaks in the physiological dinucleosome range (from 340 to 420 bp) that were $\geq 40\%$ higher than the peaks at other periods, as shown in Figure 3C.

We assessed the statistical significance of our sampling using the cumulative binomial distribution function. In our sampling there were 40 successes (Fts with amplitudes in the physiological dinucleosome period range that were at least 40% greater than the amplitudes of the other peaks) out of 400 trials (giving 10%). The cumulative binomial distribution function computes the probability of obtaining the 40 or fewer successes by chance for any given true probability of success (i.e. the true probability value of success per trial genome-wide). This function, at a 95% confidence level, establishes that the true genomic probability of success is between 7.8 and 12.8% and, at 99% confidence level that the true probability is between 7.0 and 14.0%. These confidence intervals do not change appreciably if the sample size were to be doubled. Therefore, our sampling estimate of 10% MADA-like Fts in the genome is statistically significant, with the confidence levels stated.

Table 1. Summary of nucleosome ladder analyses

Chromatin	Probe	Repeat length (bp)		Window (kb)
		Experimental (± 5 bp)	Predicted	
Mouse liver	EtBr	194	–	–
	MADA b'	183	185	70
	MADA c	182	185	70
	MADA d	183	185	70
<i>In vitro</i>	MADA a	160	162	3
	MADA b	180	179	3
	MADA c	–	Aperiodic	3
	MADA d	200	200	3
Mouse liver	MOR b	194	No signal	3 and 70
	MOR c	196	No signal	3 and 70
Mouse liver	Seq. 1	181	179	100
	Seq. 2	192	No signal	3 and 100

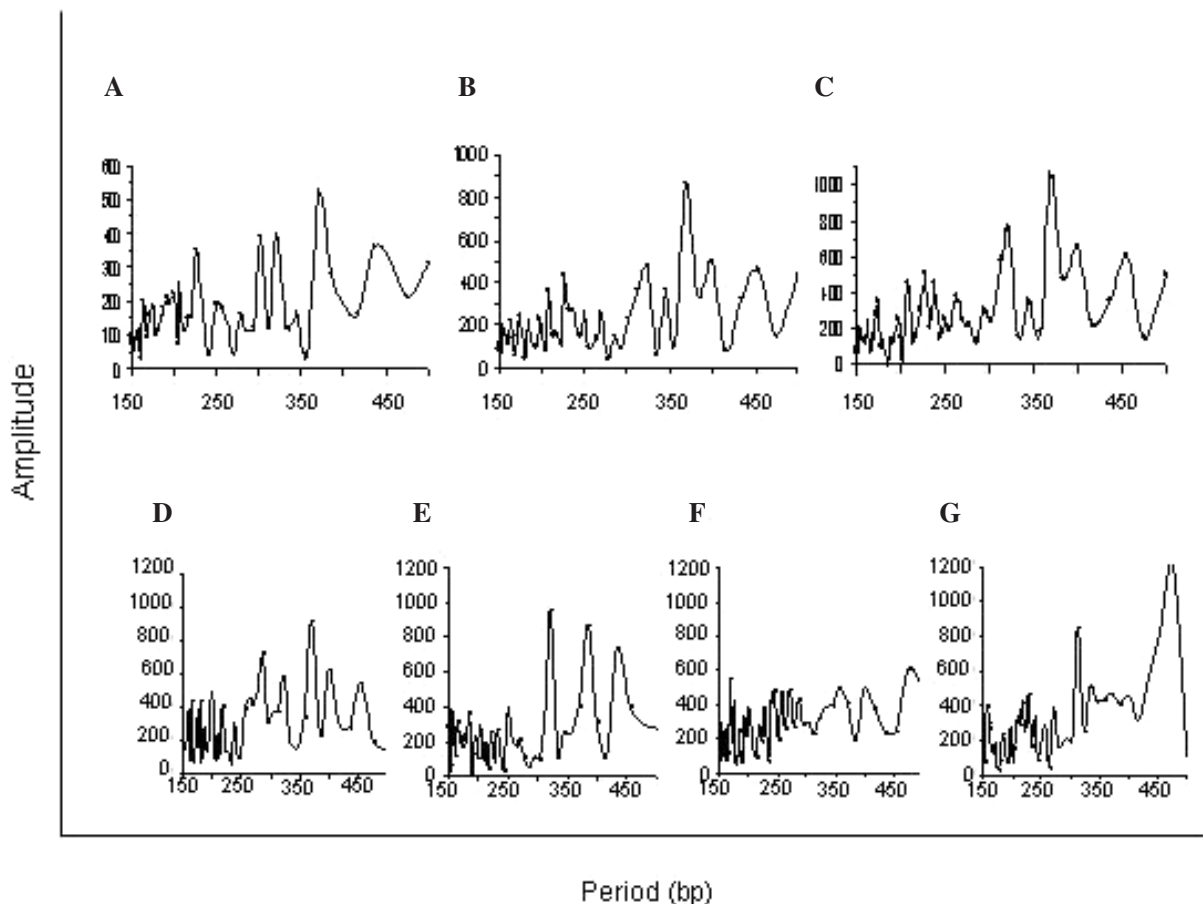


Figure 3. Fourier transforms of the long-range period-10 VWG oscillations in large windows for the MADA locus (A–C) and a survey of mouse genomic DNA (D–G). Amplitude tests are period-10 VWG counts (abbreviated ‘count’ in the text). (A) MADA locus, 50 kb window, centered on probe *b'*, with nucleotide numbers 90 000–140 000 (GenBank file U73107). (B) MADA locus, 70 kb window, centered on probe *b'*, with nucleotide numbers 80 000–150 000. (C) MADA locus, 100 kb window, centered on probe *b'*, with nucleotide numbers 65 500–165 500. (D) Representing ~18% of mouse genomic DNA; a predominant peak in the typical dinucleosome range (340–420 bp) is present, but it is <40% higher than other peaks. (E) Representing ~30% of mouse genomic DNA; a peak is present in the typical dinucleosome range, but it is lower than other peaks. (F) Representing ~25% of mouse genomic DNA; noise-level peaks throughout. (G) Representing ~15% of mouse genomic DNA; noise-level peaks in the typical dinucleosome range, with higher amplitude peaks present that are not in the typical dinucleosome range.

Using a stricter criterion, the percentage of MADA-like loci in the mouse genome is even smaller. For example, the percentage of MADA-like Fts in the mouse genome that resembles the one shown in Figure 3B, a 70 kb window with a dinucleosome period range peak >50% higher than other period peaks, is <5%. Thus, the mouse genome computational analysis indicates that the MADA locus is significantly different than most other loci, consistent with the experimental results of Figure 2.

In addition to the ~10% of 100 kb length mouse genomic DNA sequences that have MADA-like Fts, four other categories of Fts were found: Approximately 18% of the Fts sampled had predominant peaks in the physiological dinucleosome range that were <40% higher than the peaks at other periods, as in Figure 3D. Approximately 30% of the Fts sampled exhibited physiological dinucleosome range peaks that were lower than other peaks (Figure 3E), ~25% of the Fts sampled exhibited noise-level peaks (as in Figure 3F), and ~15% of the Fts sampled exhibited noise-level physiological dinucleosome range peaks, with higher amplitude non-physiological dinucleosome range peaks present (Figure 3G). The statistical significance of these observations was higher than that found for the MADA-like Fts (because of the larger percentage of 'success' values). Combining the last two categories, both containing noise-level signals in the physiological dinucleosome range (represented by Figure 3F and G), into a single category, representing ~40% of mouse genomic DNA, would make this category the largest of the four considered, contrasting significantly with the uncommon MADA locus.

To test the predictive power of our computational methods, we selected one of the MADA-like 100 kb sequences found in our sampling of the mouse genome (anonymous sequence 1). The Ft closely resembled that of the MADA locus, predicting a 179 bp nucleosome repeat, a value distinguishable from that of the bulk chromatin (~195 bp). We then selected PCR primers from roughly the center of this anonymous locus and amplified a 450 bp segment for use as a hybridization probe. Portions of the same mouse liver digest as shown in Figure 2 were electrophoresed, blotted and probed, as shown in Figure 2. A detailed analysis of the ladders (data not shown) indicated that the nucleosome repeat was 181 ± 5 bp (see Table 1), very similar to the MADA result.

Randomization of the wild-type MADA sequence in the 70 kb locus eliminates the strong signal

It was of interest to know how frequently strong in-phase long-range period-10 VWG oscillations, similar to those in the MADA locus, might occur in random 70 kb DNA sequences. To answer this question, we chose to randomize the wild-type MADA sequence computationally. Thus, the base content would remain the same as the wild-type sequence. One hundred different randomized sequences were generated by randomly selecting 100 bp (using a random number generator) and moving them to the end of the sequence 17 000 times, as in shuffling a deck of cards. We found that sequences thus generated possessed no significant similarity to the wild-type sequence by NCBI's Blast2. The results of this computational study showed that none of the sequences generated possessed a signal that resembled the wild-type signal (conservatively taken as having a predominant Fourier peak in the

physiological dinucleosome range with an amplitude of >40% larger than the other peaks). Therefore, such sequences do not appear to occur often by chance. From the binomial distribution, there is a 95% confidence level that no more than 3% of the fantastically large number (70 000!) of sequence permutations have signals as strong as the wild-type MADA signal.

Localization of a region of the MADA locus where the signal is strong and correlation of computational predictions with chromatin assembly on MADA subclones *in vitro*

To determine whether the intense 370 bp period Fourier amplitude in a 70 kb window (Figure 3B) consists of either (i) many small in-phase oscillations that are spread evenly throughout the 70 kb locus, or (ii) several large in-phase oscillations that are present only in certain discrete regions, Fouriers were computed across the MADA locus in a sliding minimal (3 kb) window. We found that the signals were not uniformly distributed. The strongest signal was located close to probe *b*, which had an amplitude of 200 count at a period of 360 bp (Figure 4A). Deleting this 3 kb sequence, or replacing it with 3000 Ns (A or C or G or T nucleotides), in the 70 kb window reduced the percentage by which the amplitude of the 370 bp peak exceeded the amplitude of the next highest peak from 72 to 25%. However, the 370 bp peak was still predominant. Thus, the 3 kb region around probe *b* (Figure 1) contributes significantly to the large (70 kb) window result, but other sequences distributed throughout the locus must also contribute. A somewhat weaker 360 bp period signal (data not shown) was found at approximately nucleotide number 5000 on the map shown in Figure 1. Additionally, Figure 4A shows that a weak non-physiological 323 bp signal was found near probe *a*, a moderate 167 count amplitude broad peak at 400 bp was found near probe *d*, and an aperiodic signal was found near probe *c*.

The predictions from the computational analysis shown in Figure 4A are that for the isolated 3 kb DNA regions examined computationally above, probe *a* should detect a weak 160 bp (323 bp/2) ladder, probe *b* should detect a strong 180 bp (357 bp/2) ladder, probe *c* should detect an aperiodic series of bands and probe *d* should detect a moderately strong 200 bp (400 bp/2) ladder. There are obviously significant differences between chromatin assembly *in vivo* and chromatin assembly in our *in vitro* system. *In vivo*, small (3 kb) regions of DNA containing variations (from what is computed using a 70–100 kb window) in signals for nucleosome array formation would be expected to have only subtle effects on nucleosome ordering due to the presence of a large sequence independent chromatin assembly machinery (20) and the presence of signals in the flanking DNA. This expectation is consistent with our experimental results of probing throughout this locus (Table 1). However, *in vitro*, the influence of variations in DNA sequence on the nucleosome organization of small isolated (3–10 kb insert) subclones should be much greater because both the cellular chromatin assembly machinery and the flanking DNA regions are absent.

To test how well our computational predictions for local variations in DNA signals are borne out *in vitro*, the four 3–10 kb insert subclones, EE3.6, M6.5, pKSH8.8 and M10.4 (see Figure 1) were assembled into chromatin *in vitro*, and digested

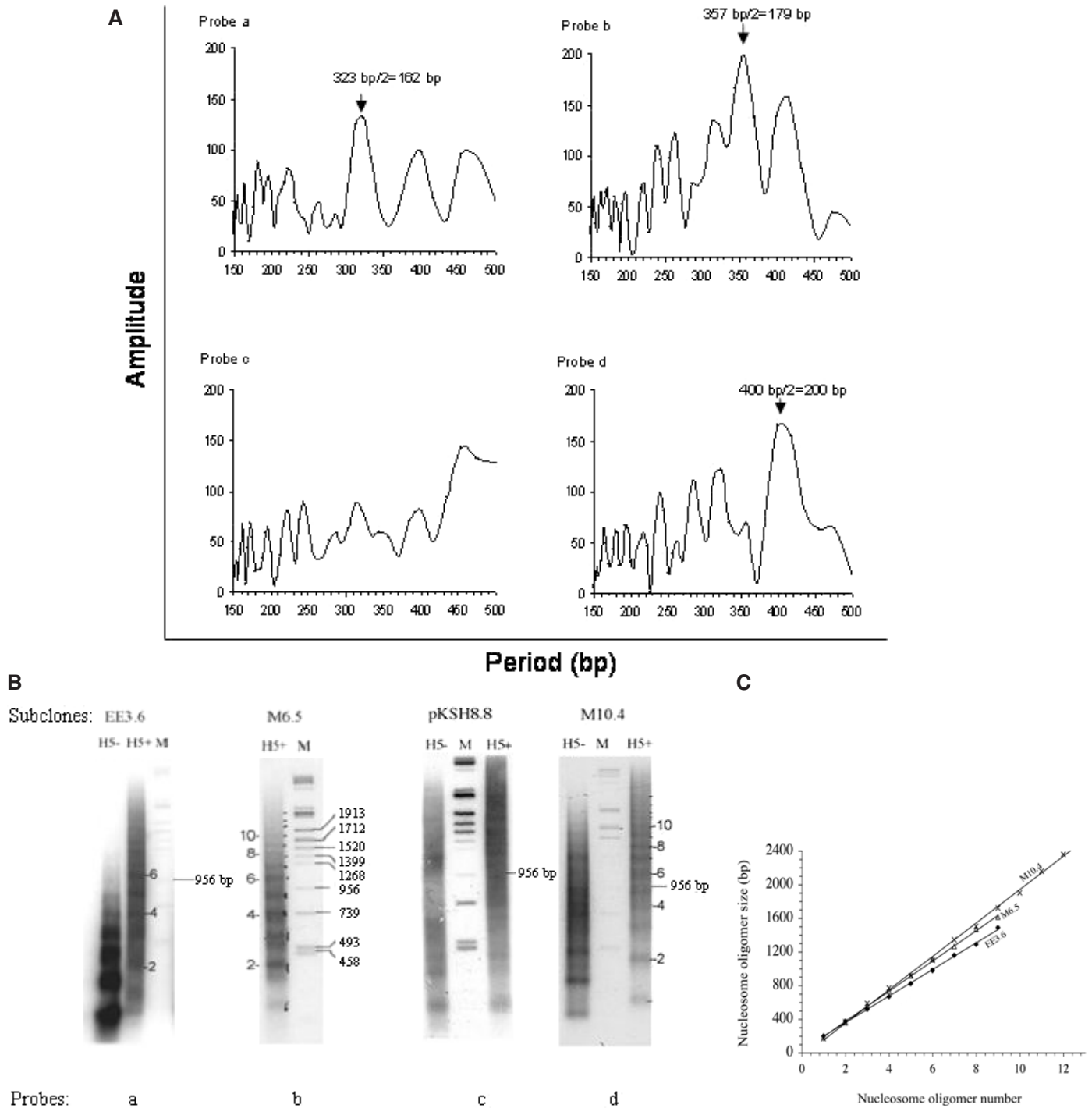


Figure 4. Computational analysis and *in vitro* experimental determination of nucleosome array characteristics in local regions of the MADA locus. (A) Fts in 3 kb windows, approximately centered on probes *a–d*, respectively (see Figure 1). Nucleotide numbers (GenBank file U73107) for the Fourier windows were: 8200–11 200, 14 800–17 800, 19 390–22 390 and 26 485–29 485 for probe *a–d* regions, respectively. Arrows denote predominant Fourier Amplitudes; the periods of their occurrence/2 give predicted nucleosome repeat. (B) Nucleosome ladders from chromatin assembled *in vitro* on 3–10 kb subclones. The purified DNA from MNase digests of chromatin assembled *in vitro* in the presence or absence of linker histone H5 on the subclone indicated was run on agarose gels, blotted and hybridized to probe *a, b, c* or *d*, as indicated; autoradiograms are shown. The subclone used is indicated at the top of each of the four autoradiograms; the hybridization probe used is indicated at the bottom (see Figure 1 for probe location). Lanes labeled H5+ denote assembly reactions performed in the presence of linker histone H5; lanes labeled H5– denote portions of the same assembly reactions used in H5+ that were incubated without linker histone. Nucleosome oligomer bands are identified for the H5+ samples for probes *a, b* and *d*. Lanes labeled M contained DNA size markers; marker fragment sizes are indicated for the M6.5 subclone. The 956 bp size marker is indicated in all autoradiograms, and is referred to in the text. (C) Plots of nucleosome oligomer DNA size versus nucleosome oligomer number are shown. For a regularly spaced nucleosome arrangement, a straight line should be obtained, reflecting multiples of a unit repeat. The value of the unit repeat (the nucleosome spacing periodicity or repeat length) is given by the slope of the line, with units of base pairs/nucleosome. Plots and straight-line fits were done with Microsoft Excel. Subclones: M10.4, probe *d* (cross symbols); M6.5, probe *b* (open triangles); EE3.6, probe *a* (closed diamonds); the H5+ lanes of (B) were analyzed. Repeat lengths are reported in Table 1.

with MNase. The nucleosome ladders were then assessed by Southern hybridization using probes *a–d* that were approximately centered in each of the 3 kb windows (see Figure 1) that generated the Fts shown in Figure 4A. The nucleosome ladders obtained are shown in Figure 4B. The linker histone H5-containing samples for probes *a*, *b* and *d* gave ladders reflecting regular nucleosome arrangements. However, the nucleosome spacing periodicities differed among the probes. Probe *a* detected bands with high background that were multiples of 160 bp for subclone EE3.6. It can be seen in Figure 4B that the 6mer of the ladder for this subclone (probe *a*) runs approximately at the 956 bp marker band position, consistent with $6 \times 160 \text{ bp} = 960 \text{ bp}$. In contrast, Figure 4B shows a strong ladder for subclone M6.5 (probe *b*) where the 6mer is larger than 956 bp. In this ladder the 3mer is split, possibly due to the presence of 3mers with slightly different lengths being detected by this probe. A moderately strong ladder was detected on subclone M10.4 by probe *d*. Here, the 6mer is even larger than that found for probe *b*, and the 5mer runs at $\sim 956 \text{ bp}$. A detailed analysis of these ladders is shown in Figure 4C. The straight-line fits of the plots of nucleosome oligomer size versus nucleosome oligomer number demonstrate that the ladder bands are multiples of unit repeats, and that the repeat lengths (measured from the slopes of the lines) are $160 \pm 5 \text{ bp}$ for subclone EE3.6 (probe *a*), $180 \pm 5 \text{ bp}$ for subclone M6.5 (probe *b*) and $200 \pm 5 \text{ bp}$ for subclone M10.4 (probe *d*). All four of the linker histone H5-containing samples gave the nucleosome spacing periodicities that were predicted computationally (Figure 4A) as one half of the predominant Fourier (dinucleosome) peak period value (see Table 1). Moreover, the quality of each ladder correlated with the amplitude value of the corresponding predominant Fourier peak. In the absence of linker histone (see the H5–lane in Figure 4A), closely-packed nucleosome bands were evident. In contrast with the other probes, probe *c* (in subclone pKSH8.8) detected an aperiodic MNase digestion pattern (Figure 4B, pKSH8.8, probe *c*), again consistent with our computational prediction (Figure 4A).

The MOR locus is more representative of mouse genomic DNA than the MADA locus regarding nucleosome array formation

The MOR locus on chromosome 7 is computationally more typical of mouse genomic DNA. The 63 kb region containing this locus has a Fourier (Figure 5A) possessing only small amplitudes in the physiological dinucleosome region similar to those shown in Figure 3F and G which, when combined, make up $\sim 40\%$ of the mouse genome. A more detailed computational study revealed that most 3 kb regions of the MOR locus, as well as for other loci, exhibited noise-level Ft amplitudes, as shown for the MOR probe *c* region (Figure 5B). Thus, the MOR probe *c* region should be representative of much of the mouse genome. Experimentally, portions of the same MNase digests of mouse liver chromatin used for examining the MADA locus were used to assess the nucleosome ladders detected by MOR probe *c* (Figure 5C). It can be seen that probe *c* detected nucleosome ladders that were similar to those found for bulk chromatin, although the background signal was somewhat lower (compare with Figure 2A). Seven bands can be detected in the 5 min digest (lane 3),

and analysis of the ladder gives a $196 \pm 5 \text{ bp}$ repeat (Table 1). The densitometer scan of this lane (Figure 5D) is similar to that of the bulk chromatin scan (Figure 2D) in that the band intensities drop off significantly after the 3mer. Thus, MOR probe *c* detects nucleosome ladders in liver nuclei that are similar to those of bulk mouse liver chromatin, consistent with the Ft in this region of the MOR locus being highly representative of mouse genomic DNA (data not shown).

In addition, we probed two other regions of the MOR locus from MNase digests of mouse liver chromatin. MOR probe *b*, located $\sim 20 \text{ kb}$ upstream of probe *c*, corresponded to another 3 kb region lacking a period-10 VWG signal. The MNase ladders detected from this region were again very similar to those detected by probe *c* (data not shown, see Table 1 for a summary of the results). The second 3 kb region, detected by MOR probe *d*, was located $\sim 5 \text{ kb}$ downstream of probe *c*. This region corresponded to a moderate period-10 VWG signal (160 count Fourier amplitude, 360 bp period) predicting a 180 bp repeat. Experimentally, the MNase ladders detected by this probe exhibited some subtle differences from the ladders detected by probes *b* and *c*, such as higher background and a slightly shorter nucleosome repeat (data not shown).

To further test the predictive power of our computational methods, we selected yet another anonymous 100 kb region of mouse genomic DNA which computationally lacked period-10 VWG signals, thus resembling the MOR locus. We then selected PCR primers from roughly the center of this locus (which also lacked VWG signals), and amplified a 493 bp segment (anonymous probe 2) for use as a hybridization probe. As described above, portions of the same mouse liver digest as shown in Figure 2 were electrophoresed, blotted and probed, as shown in Figure 2. A detailed analysis of the ladders (data not shown) indicated that the nucleosome repeat was $192 \pm 5 \text{ bp}$ (see Table 1), similar to the MOR probe *c* and the bulk chromatin results.

DISCUSSION

It seems likely that the large amount of non-coding DNA present in the genomes of higher organisms serves some purpose, and that chromatin structure mediates the effects of the DNA. We have recently shown that nucleosome arrays formed in our linker histone-dependent *in vitro* chromatin assembly system are strongly influenced by oscillations in the abundant period-10 VWG motif present in genomic DNA in a predictable way (12). Nucleosome alignment into a physiological array is dependent upon the presence of linker histones. Linker histone H1 has been recently shown to be essential for proper mouse development (21). Additionally, higher-order chromatin structures might be influenced significantly by the characteristics of the nucleosome arrays contained within them (2). Here, we have provided evidence that computationally predictable long-range, dinucleosome period, oscillations in period-10 VWG may influence nucleosome array formation in mouse liver chromatin.

We observed that hybridization probes from the MADA locus (Figure 1) detected a distinctive and highly regular nucleosome array throughout the locus (Figure 2). Computational analysis indicated that this locus contains a strong centrally located signal embedded in a large genomic DNA sequence context possessing weak, but in-phase, signals. This model is

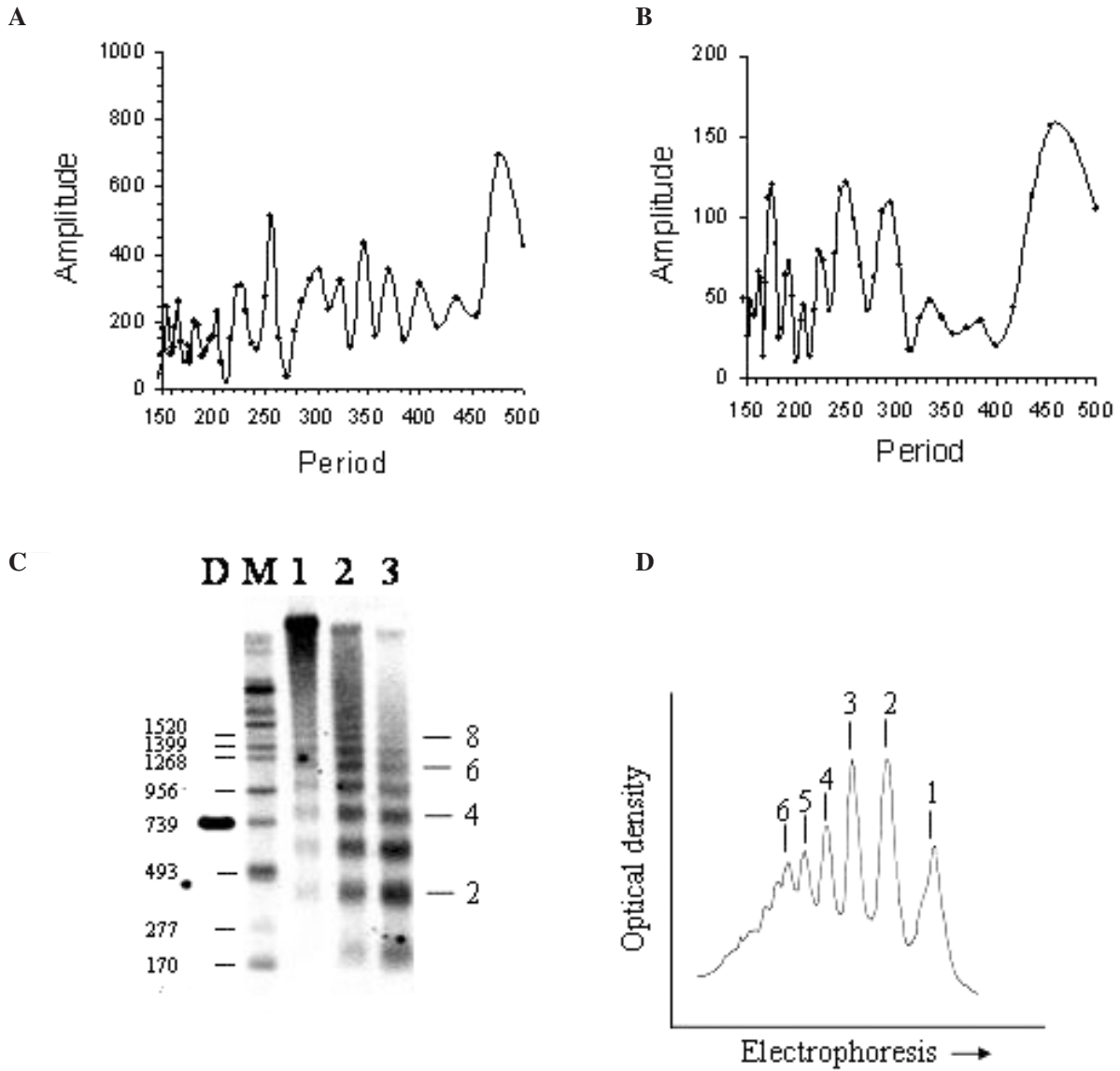


Figure 5. Computational analysis and experimentally determined nucleosome ladders from mouse liver nuclei in the MOR locus. (A) Ft for the whole (63 797 bp) MOR locus (GenBank file AF071080). (B) Ft in a 3 kb window containing probe *c*. Nucleotide numbers for this window were 48 000–51 000. (C) Nucleosome ladders from the probe *c* region. Portions of the same digests used in Figure 2A were run on a gel, blotted and hybridized to MOR probe *c*. Lanes are labeled as in Figure 2A. Bands corresponding to nucleosome oligomers 2, 4, 6 and 8 are identified at the right. Size marker (lane M) bands are identified. (D) Lane 3 densitometer scan; optical density peaks arising from the indicated nucleosome oligomer bands are shown.

supported by our *in vitro* chromatin assembly experiments (Figure 4B), which necessarily must use small (less than ~10 kb) genomic DNA inserts, showing that the central region possessing the strong signal (probe *b*) indeed gave the strongest nucleosome ladder. Several other small regions of the gene possessing weaker signals or lacking signals gave weak, variant or aberrant ladders, all correlating with the Fourier predictions for these small isolated regions. We suggest that in mouse liver weak in-phase DNA signals, existing throughout this locus, cooperate with the strong probe *b* region signal to give the highly ordered nucleosome array and the shorter repeat that we observed when probing throughout the locus.

The functional consequences of the distinctive chromatin structure found for the MADA locus and ~10% of the genome are not currently clear. However, it is reasonable to speculate

that the chromatin structures encompassing these sequences may have a direct bearing on their functions. The MADA gene is a housekeeping gene that is expressed at a moderate level in liver (22). Additionally, the (G+C) content of the MADA locus is 49%, well above the mouse genome average value of 42% (23). We examined the 40 computationally MADA-like windows that we have found thus far to see whether either of these characteristics were representative of those windows. We found that the 40 MADA-like 100 kb windows contained a variety of different kinds of sequences, whose average properties were close to genome average values. For example, 15 windows contained no annotated genes, 17 contained at least a portion of one gene, 5 contained at least a portion of two genes and 3 contained at least a portion of three genes, giving an average value of 0.9 genes per 100 kb window, a value close

to the genome average (23). Moreover, the genes were of varied types, and some were tissue-specific while others were constitutively expressed. Additionally, ~25% of the MADA-like 100 kb windows contained >80% annotated repetitive DNA. Finally, the (G+C) content ranged from 35.7 to 47.3%, with an average value of 40.4%, close to the genome average (42%). Thus, the computationally MADA-like 100 kb windows found thus far were certainly not all housekeeping genes or sequences of high G+C content, but seemed to be typical of mouse genomic DNA, suggesting that distinctive chromatin structures can occur for a variety of different types of sequences.

A possible example of chromatin architecture having a role in chromosome function is imprinting. Genomic imprinting is traditionally defined as an epigenetic process leading to parental origin-dependent gene expression. About 60 imprinted genes have thus far been discovered. It has been recently suggested that imprinting may be a feature of large chromatin domains with their own domain-wide characteristics (24). Thus, it was of interest to computationally examine these loci. Strikingly, computational analysis of 26 mouse imprinted gene loci (data not shown) revealed strong, MADA-like signals in 21 out of the 26 loci (81%), despite the relatively rare occurrence (10%) of such signals in the mouse genome. This finding suggests that imprinted gene loci may in fact have distinctive chromatin structures.

A more typical locus, in terms of chromatin structure, in mouse genomic DNA is represented by the MOR locus on chromosome 7. Computationally, we estimate that ~40% of mouse genomic DNA is similar to this locus in its nucleosome array forming potential. A 70 kb window encompassing this locus contains no significant (period-10 VWG) signals for nucleosome array formation, although a few 3 kb windows within the locus exhibit moderate strength signals. We found that the nucleosome arrays formed in two local MOR regions which lacked signals (probes *b* and *c*) were similar to that of the bulk chromatin. In contrast, the probe *d* region, located ~5 kb away from probe *c*, possessed some subtle differences with probes *b* and *c* in nucleosome array formation that were consistent with the presence of a localized moderate strength signal in this region.

We interpret the above results in the following way. The similarity to bulk chromatin in localized MOR regions lacking signals suggests that the typical ~195 bp repeat observed in mouse liver (Table 1), and most other tissues, arises largely from the cellular chromatin assembly machinery, which is essentially sequence independent. The subtly different nucleosome array formed on the MOR probe *d* region suggests that an isolated localized VWG signal has to work against the 'default' cellular chromatin assembly machinery, but still has some influence on nucleosome array formation. In contrast, for the MADA locus, in-phase period-10 VWG signals exist throughout the whole locus, and these signals cooperate and thereby dominate nucleosome array formation, yielding a highly ordered and distinctive arrangement.

To test the predictive power of our computational analysis, we selected two anonymous genomic regions based upon computation. One was predicted to be MADA-like giving a short 179 bp repeat and the other lacked signals and was thus predicted to give the bulk chromatin repeat. The experimental results were in agreement with our predictions (see Table 1).

We are currently searching the mouse genome computationally for atypical nucleosome repeats and preparing hybridization probes from these regions to test our computational predictions experimentally. These experiments involve carefully establishing signal threshold values, and determining the possible influences of the presence of non-physiological Fourier (periodicity) peaks. If the presence of distinctive chromatin structures can in fact be predicted computationally, as all evidence indicates so far, it will then be of interest to perform a large-scale computational analysis to see if the regions of the genome predicted to have distinctive chromatin structures display any characteristics in common besides imprinting. Additionally, it might be possible to use a more detailed assay, such as indirect end-labeling, to experimentally focus on the small genomic regions with distinctive chromatin architectures that were found computationally.

Overall, we have provided evidence that long-range oscillation in the period-10 triplet motif VWG, existing mostly in non-coding DNA, affects nucleosome array formation in animal tissue. Signals encoded in genomic DNA that direct the formation of distinctive nucleosome arrangements, which may influence chromatin structure and function, may be computationally predictable.

ACKNOWLEDGEMENTS

We thank Drs Rodney Kellems and Ping Xu (Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX) for providing us with the MADA genomic DNA constructs; we thank Drs Catherine Farrell and Gary Felsenfeld, (NIDDKD, NIH, Bethesda, MD) for providing us with the mouse olfactory receptor clones. We would also like to thank Chad Pitschka for the assistance with Excel Macros and Visual Basic, and Liqin Zhu for help with the genomic PCR experiments. This work was supported by NIH grant GM62857, NIGMS to A.S. The Open Access publication charges for this article were waived by Oxford University Press.

REFERENCES

- Collins,F.S., Green,E.D., Guttmacher,A.E. and Guyer,M.S. (2003) A vision for the future of genomics research. *Nature*, **422**, 835–847.
- Woodcock,C.L., Grigoryev,S.A., Horowitz,R.A. and Whitaker,N. (1993) A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. *Proc. Natl Acad. Sci. USA*, **90**, 9021–9025.
- Zlatanova,J., Leuba,S.H., Yang,G., Bustamante,C. and van Holde,K. (1994) Linker DNA accessibility in chromatin fibers of different conformations: a reevaluation. *Proc. Natl Acad. Sci. USA*, **91**, 5277–5280.
- Woodcock,C.L. and Horowitz,R.A. (1995) Chromatin organization re-viewed. *Trends Cell Biol.*, **5**, 272–277.
- van Holde,K. and Zlatanova,J. (1995) Chromatin higher order structure: chasing a mirage? *J. Biol. Chem.*, **270**, 8373–8376.
- Liu,K. and Stein,A. (1997) DNA sequence encodes information for nucleosome array formation. *J. Mol. Biol.*, **270**, 559–573.
- Sun,F.-L., Cuaycong,M.H. and Elgin,S.C.R. (2001) Long-range nucleosome ordering is associated with gene silencing in *Drosophila melanogaster* pericentric heterochromatin. *Mol. Cell. Biol.*, **21**, 2867–2879.
- Stein,A., Dalal,Y. and Fleury,T.J. (2002) Circle ligation of *in vitro* assembled chromatin indicates a highly flexible structure. *Nucleic Acids Res.*, **30**, 5103–5109.
- Tsukiyama,T. (2002) The *in vivo* functions of ATP-dependent chromatin remodeling factors. *Nature Rev. Mol. Cell. Biol.*, **3**, 422–429.

10. Baldi,P., Brunak,S., Chauvin,Y. and Krogh,A. (1996) Naturally occurring nucleosome positioning signals in human exons and introns. *J. Mol. Biol.*, **263**, 503–510.
11. Stein,A. and Bina,M. (1999) A signal encoded in vertebrate DNA that influences nucleosome positioning and alignment. *Nucleic Acids Res.*, **27**, 848–853.
12. Cioffi,A., Dalal,Y. and Stein,A. (2004) DNA sequence alterations affect nucleosome array formation of the chicken ovalbumin gene. *Biochemistry*, **43**, 6709–6722.
13. Stein,A. and Bina,M. (1984) A model chromatin assembly system. Factors affecting nucleosome spacing. *J. Mol. Biol.*, **178**, 341–363.
14. al-Ubaidi,M.R., Ramamurthy,V., Maa,M.C., Ingolia,D.E., Chinsky,J.M., Martin,B.D. and Kellems,R.E. (1990) Structural and functional analysis of the murine adenosine deaminase gene. *Genomics*, **7**, 476–485.
15. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual, 3rd edn*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
16. Jeong,S., Lauderdale,J.D. and Stein,A. (1991) Chromatin assembly on plasmid DNA *in vitro*: apparent spreading of nucleosome alignment from one region of pBR327 by histone H5. *J. Mol. Biol.*, **222**, 1131–1147.
17. Hewish,D.R. and Burgoyne,L.A. (1973) Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun.*, **52**, 504–510.
18. Feinberg,A.P. and Vogelstein,B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.*, **132**, 6–13.
19. Thomas,J.O. and Thompson,R.J. (1977) Variation in chromatin structure in two cell types from the same tissue: a short DNA repeat length in cerebral cortex neurons. *Cell*, **10**, 633–640.
20. Henikoff,S., Furuyama,T. and Ahmad,K. (2004) Histone variants, nucleosome assembly, and epigenetic inheritance. *Trends Genet.*, **20**, 320–326.
21. Fan,Y., Nikitina,T., Morin-Kensicki,E.M., Zhao,J., Magnuson,T.R., Woodcock,C.L. and Skoultschi,A.I. (2003) H1 linker histones are essential for mouse development and affect nucleosome spacing *in vivo*. *Mol. Cell. Biol.*, **23**, 4559–4572.
22. Winston,J.H., Hong,L., Datta,S.K. and Kellems,R.E. (1996) An intron 1 regulatory region from the murine adenosine deaminase gene can activate heterologous promoters for ubiquitous expression in transgenic mice. *Somat. Cell Mol. Genet.*, **22**, 261–278.
23. Mouse Genome Sequencing Consortium (2002), Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
24. Paldi,A. (2003) Genomic imprinting: could the chromatin structure be the driving force? *Curr. Top. Dev. Biol.*, **53**, 115–138.