DNA Sequence Alterations Affect Nucleosome Array Formation of the Chicken Ovalbumin Gene†

Alfred Cioffi,‡ Yamini Dalal,§ and Arnold Stein*,‡

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907

Received February 7, 2004; Revised Manuscript Received April 5, 2004

ABSTRACT: The role of the large amount (more than half of the genome) of noncoding DNA in higher organisms is not well understood. DNA evolved to function in the context of chromatin, and the possibility exists that some of the noncoding DNA serves to influence chromatin structure and function. In this age of genomics and bioinformatics, genomic DNA sequences are being searched for informational content beyond the known genetic code. The discovery that period-10 non-T, A/T, G (VWG) triplets are among the most abundant motifs in human genomic DNA suggests that they may serve some function in higher organisms. In this paper, we provide direct evidence that the regular oscillation of period-10 VWG that occurs in the chicken ovalbumin gene sequence with a dinucleosome-like period facilitates nucleosome array formation. Using a linker histone-dependent in vitro chromatin assembly system that spontaneously aligns nucleosomes into a physiological array, we show that nucleosomes tend to avoid DNA regions with low period-10 VWG counts. This avoidance leads to the formation of an array with a nucleosome repeat equal to half the period value of the oscillation in period-10 VWG, as determined by Fourier analysis. Two different half-period deletions in the wild-type DNA sequence altered the nucleosome array, as predicted computationally. In contrast, a full-period deletion had an insignificant effect on the nucleosome array formed, also consistent with the prediction. An inversion mutation, with no DNA sequences deleted, again altered the nucleosome array formed, as predicted computationally. Hence, a VWG dinucleosome signal is plausible.

The human DNA sequence and the sequences of several other higher organisms are now available or are being completed (*1*). Most of the protein-coding genes have been identified with a high degree of confidence (*2*, *3*), and ultimately the functions of these genes will be elucidated. However, the role of the large amount (more than half of the genome) of noncoding DNA in higher organisms is not well understood (*4*). DNA evolved to function in the context of chromatin, and the possibility exists that some of the noncoding DNA serves to influence chromatin structure and function, including gene regulation. A remaining challenge of DNA research and of genetics is to understand chromatin, the dynamic protein−DNA complex that makes up chromosomes (*5*).

Virtually all of 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 be variable (*6*−*9*). For example, the nucleosome arrays on the repetitive DNA of the transcriptionally silent heterochromatin in *Drosophila* appear to be very regular, and the higher order structure of heterochromatin appears to be highly condensed (*10*). Moreover, variable nucleosome arrangements on DNA from the same nuclei in cells of higher organisms have been inferred by micrococcal nuclease (MNase)[1] digestion of the chromatin and from analysis of the resultant DNA fragment ladders by gel electrophoresis and Southern blotting. Nucleosome spacing periodicities ranging from about 175 to 210 bp have been observed (*11*), and the degree of nucleosome spacing regularity on some sequences differs (*12*). It is plausible that nucleosome arrays displaying different nucleosome arrangements form stretches of chromatin having distinct higher order structures or different modes of flexibility due to variations in the rotational orientation between adjacent nucleosomes (*6*, *13*). Additionally, linker histone association with nucleosomes influences the properties of chromatin (*14*, *15*). It has recently been demonstrated that proper mouse development requires a sufficient amount of linker histone H1 and that mouse tissues containing only half of the physiological amount of histone H1 exhibit shortened nucleosome spacings (*16*). Knowledge of the mechanisms by which nucleosome arrays form should provide insights into how histone modifications and the presence of non-histone chromosomal proteins can lead to functionally important changes in chromatin structure (*17*).

In vitro chromatin assembly systems have provided some useful information on how regular physiological nucleosome spacings might arise. Systems that generate arrays with regular physiologically spaced nucleosomes have been

† This work was supported by U.S. Public Health Service Grant GM62857 (to A.S.).

* To whom correspondence should be addressed. Phone: (765) 494-6546. Fax: (765) 494-0876. E-mail: astein@bilbo.bio.purdue.edu.

‡ Department of Biological Sciences, Purdue University.

§ Basic Sciences Division, Fred Hutchinson Cancer Research Center.

[1] Abbreviations: Wt, wild type; Δ, deletion; bp, base pair; kb, kilobases; VWG, non-T, A/T, G; PCR, polymerase chain reaction; MNase, micrococcal nuclease.

10.1021/bi049717f CCC: $27.50
Published on Web 05/05/2004

developed, based upon the activities found in the extremely rapidly dividing cells of *Drosophila* early embryos (*18*, *19*) or in *Xenopus* eggs (*20*) or oocytes (*21*). These systems require ATP and Mg^{2+}; linker histone is not required. Additionally, a very simple linker histone-dependent in vitro system that spontaneously arranges nucleosomes into arrays with physiologically spaced nucleosomes has been developed; ATP is not required. This system uses polyglutamic acid as a solubilizing agent to allow proper linker histone addition to reconstituted nucleosomes (*22*). An interesting feature of the linker histone-dependent chromatin assembly system is that the DNA base sequence matters (*12*). For example, regular physiologically spaced nucleosomes are not generated on plasmid DNA, cDNAs, or in some regions of vertebrate genomic DNA. However, other regions of vertebrate genomic DNA do assemble into arrays containing highly ordered, physiologically spaced nucleosomes, under the identical conditions used for the sequences that do not order nucleosomes. These results suggest that there could be periodic sequence motifs present in some sequences that facilitate the formation of regularly spaced nucleosome arrays. To find these possible periodic motifs, use was made of the observation, using a machine learning approach, that human genomic DNA contains the triplet non-T, A/T, G (VWG) abundantly repeated at 10 base pair multiples (*23*). It was then suggested that VWG triplets might distort the DNA double helix somewhat, and since the triplets often occur with the period of DNA (about 10 bp), they would be in-phase and might constitute a region of preferred DNA bending, thus facilitating its packaging into nucleosomes (*23*). Experimental support for this idea was obtained by showing that the preferred nucleosome locations in native SV40 chromatin generally correspond to regions of SV40 DNA that are enriched in period-10 VWGs, whereas DNA regions low in period-10 VWGs tend to avoid nucleosomes (*24*).

In the same study (*24*) it was also observed that, for the three sequenced DNAs that were previously assembled into chromatin using the polyglutamic acid system (*25*−*27*), regular long-range oscillations of period-10 VWG counts in a sliding 100 bp window occurred specifically in regions of DNA that ordered nucleosomes. The period of these long-range oscillations, assessed by Fourier analysis, corresponded almost exactly to a value that was equal to twice the measured nucleosome repeat in all cases. Moreover, DNA regions that did not possess a single strong Fourier 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 in our laboratory by studying the mouse adenosine deaminase locus in vitro and in liver nuclei (Dalal et al., unpublished observations). Although 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 because linker histones can readily induce the formation of a regular nucleosome array on DNA that is devoid of sequence information, for example, on poly[d(A-T)]·poly[d(A-T)] (*22*). Thus, when chromatin is assembled at a physiological core histone to DNA ratio, 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.

The chicken ovalbumin gene assembles into chromatin with a regular nucleosome array having a spacing periodicity of approximately 190 bp in a linker histone-dependent in vitro chromatin assembly system (*25*). This result correlates with the presence of a strong predominant Fourier amplitude having a dinucleosome-like period value of 2×190 bp for the oscillations in period-10 VWG counts in the DNA sequence (*24*). In this study, we have analyzed this correlation more thoroughly by mapping nucleosome positions with respect to the period-10 VWG oscillations in a region of the gene. Additionally, to obtain more direct evidence for the postulated dinucleosome period VWG signal, we have examined the effects of two small deletion mutations and an inversion mutation predicted to alter the nucleosome array and of one larger deletion that should not. The results of these experiments support the hypothesis that the dinucleosome-like period-10 VWG signal contributes to the formation of ordered nucleosome arrays and suggest that the chicken ovalbumin gene DNA sequence may play a role in the organization of its chromatin, in addition to its known coding function.

MATERIALS AND METHODS

Computational Analysis. GenBank genomic DNA sequences for the chicken ovalbumin gene (accession number J00895) were analyzed for variations in period-10 VWG (non-T, A/T, G) contents as described previously (*24*). Briefly, the occurrences of VWG 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 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 previously described. The regularity and the period of the long-range period-10 VWG oscillations were assessed by Fourier analysis using the window size stated. An ideal sinusoidal oscillation (for comparison with the computed oscillating curve) was created in Microsoft PowerPoint by constructing a grid of base pair number (*x* axis) and average periodic VWG/CWB count (*y* axis), with the *x* axis displaced to the average *y* value of the oscillating curve described above. A sinusoidal curve of one full period was then traced using PowerPoint, spanning the desired number of base pairs (380) and amplified vertically to encompass the highest peak and the lowest valley of the VWG/CWB oscillation. This unit sine wave was then duplicated, joined end to end, and superimposed on the computed curve for the best fit possible.

DNA Constructs. The wild-type 4.5 subclone was a 4.5 kb *Dra*I fragment of the chicken ovalbumin gene (GenBank numbers 2844−7314) inserted into the *Sna*BI site of pUC19. To make the 189 bp and the 381 bp deletion mutants, two separate pairs of *Bsp*DI sites were created in the wild-type 4.5 construct at nucleotide numbers 4846 and 5035, or 4654 and 5035, respectively, using the Stratagene QuikChange

mutagenesis kit. To create the Δ189 deletion mutant, the appropriate construct was cut with *Bsp*DI at nucleotide numbers 4846 and 5053, blunted, religated, and transformed in strain DH5α bacteria (Stratagene) using standard conditions. The Δ381 subclone was created the same way, but using the *Bsp*DI sites at 4654 and 5035 instead. The 1050 bp inversion mutant was constructed by creating a *Bsp*DI site at nucleotide number 5835 in the wild-type 4.5 kb subclone as above, then digesting the construct with *Hin*dIII (nucleotide number 4785) and *Bsp*DI, blunting, religating, and transforming as above. The wild-type 1.9 subclone was a 1.9 kb *Pvu*II fragment of the chicken ovalbumen gene (GenBank numbers 4639−6540) inserted into the *Eco*RI site of pBR322 (*25*). A 201 bp region was deleted by cutting the wild-type 1.9 subclone with *Sna*BI (nucleotide number 5621) and *Xba*I (nucleotide number 5822), then religating, and transforming as above. Highly pure supercoiled DNA was prepared either by the standard maxiprep technique using CsCl−ethidium bromide banding (*28*) or by using the Wizard maxiprep kit (Promega). Residual RNA was removed by ultracentrifugation through 1 M NaCl (*28*). DNA samples were generally stored in 10 mM Tris-HCl (pH 8.0) and 1 mM Na_2EDTA at 4 °C.

In Vitro Chromatin Assembly. Pure supercoiled DNA was first reconstituted with core histones using salt gradient dialysis as previously described (*29*). Briefly, 20 μg of DNA at a concentration of 1000 μg/mL was mixed with purified core histones, obtained by salt extraction from chicken erythrocyte nuclei, at ratios ranging from 0.8 to 1.00 μg of core histones/μg of DNA in 1 M NaCl, 10 mM Tris-HCl (pH 8.0), and 1 mM Na_2EDTA for 30 min, followed by dialysis against 0.8 M NaCl, 10 mM Tris-HCl (pH 8.0), and 1 mM Na_2EDTA and against 0.6 M NaCl, 10 mM Tris-HCl (pH 8.0), and 1 mM Na_2EDTA for 2 h each, followed by 2 h against 20 mM Tris-HCl (pH 7.2) and 0.2 mM Na_2EDTA. The chromatin then was incubated overnight at 37 °C with linker histone H5 (0.5 μg of H5/μg of DNA), at a final concentration of 100 μg/mL DNA, 2 mg/mL polyglutamic acid (pH 7.2), 0.15 M NaCl, 20 mM Tris-HCl (pH 7.2), and 0.2 mM Na_2EDTA. Prior to digestion with micrococcal nuclease (MNase), the salt was usually dialyzed out.

Micrococcal Nuclease Digestion. The quality and extent of the nucleosome arrays formed were assessed by MNase digestion of the chromatin in 1 mM $CaCl_2$ and 5.0 units of MNase/μg of DNA for 30 s−2 min at 37 °C. Digestion was terminated by plunging the sample into 0.2 mg of proteinase K/mL, 10 mM EDTA, and 0.1% SDS and incubating for 1 h at 37 °C. Samples were then extracted once with phenol−chloroform−isoamyl alcohol (25:24:1) and once with chloroform and then ethanol precipitated 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. In the experiments involving the Δ189 bp and the Δ381 bp deletion mutants, after MNase digestion of the chromatin, the purified DNA fragments were digested with *Ava*II and *Stu*I (at sites 2.4 kb apart) in order to eliminate the detection of DNA fragments that were located far from the regions deleted.

Gel Electrophoresis, Southern Blots, and Hybridizations. Nucleosome oligomer size DNA fragments (about 0.2 μg/lane for blotting or 1.5 μg/lane for ethidium bromide staining) were resolved on 1.5% agarose gels in 1 × TBE using a 13.5 cm × 13.5 cm bridge gel apparatus, run for an average of 4 h at 100 V. The gels were then processed for Southern blots. λ DNA cut with *Afl*III and labeled with [α-^{32}P]dATP was used as size markers for blots. A 123 bp ladder (Invitrogen) was used as size markers for stained gels. DNA fragments were transferred after electrophoresis to charged nylon membranes (GeneScreen Plus, PerkinElmer) by standard capillary blot conditions; transfer was generally overnight. DNA was fixed to the membrane wet by UV cross-linking. All prehybridizations, hybridizations, and washes were performed using the membrane manufacturer's recommendations in a Hybaid rotary oven at 68 °C. The dried blots were then exposed for desired intensity, generally from 1 h to several hours using BioMax MR film (Kodak). The wild-type probe (557 bp) was obtained by digesting the wild-type 4.5 kb subclone with *Pst*I; the Δ189 probe (368 bp) was obtained by digesting the Δ189 mutant DNA with *Pst*I, and the Δ381 probe (176 bp) was obtained by digesting the Δ381 mutant DNA also with *Pst*I. The 200 bp wild-type probe for the inversion experiment was obtained by PCR of the wild-type sequence from nucleotide numbers 5200 to 5400 and the inversion probe for the same experiment from PCR of the corresponding 200 bp sequence of the inversion mutant DNA. All probe purifications were performed by running the digested DNA on agarose gels and staining the gel to visualize the band, and the DNA fragment of interest was then excised from the gel and purified using the Qiaex gel purification kit (Qiagen). Pure denatured DNA probe fragments (10−20 ng) were labeled with [α-^{32}P]dATP by random priming (*30*). The labeled DNA was then denatured, chilled on ice, and supplemented with denatured salmon sperm DNA (for a final mixture concentration of 50 μg/mL). Final probe concentrations in the hybridization buffer were 50−100 ng/mL. Hybridization was generally carried out for 8−16 h at 68 °C in the Hybaid rotary oven, and washes were as recommended by the hybridization oven manufacturer. The stained gels were soaked in a dilute solution of ethidium bromide (0.45 μg/mL) for 30 min, destained overnight in deionized water, and then photographed with Polaroid type 55 film under UV illumination.

Indirect End Labeling. The purified DNA fragments obtained after MNase digestion (chromatin, 4.3 units/μg of DNA, 30 s; naked DNA, 0.4 unit/μg of DNA, 30 s) were digested to completion either with *Xmn*I for the 5′ end-labeling experiment or with *Xba*I for the 3′ end-labeling experiment. Purified DNA was then electrophoresed and blotted as before. The 221 bp probe for the 5′ indirect end-labeling experiment was obtained by digesting the wild-type 4.5 subclone with *Xmn*I (nucleotide number 3077) and *Msc*I (nucleotide number 3298). The 202 bp probe for the 3′ indirect end-labeling experiment was obtained by digesting the wild-type 4.5 subclone with *Xba*I (nucleotide number 5822) and *Sna*BI (nucleotide number 5620). The DNA bands detected by these probes extended inward toward the region of interest from either the *Xmn*I site (5′ end) or the *Xba*I site (3′ end). Hybridizations and washes were carried out as per the membrane manufacturer's recommendation. For each autoradiogram to be analyzed, the distance from the base of the loading well to the center of each lane band was measured to approximately 0.1 mm. [α-^{32}P]dATP-labeled λ *Afl*III digest bands of known base pair size were used as adjacent markers. Differences between the cutting sites on the naked DNA and

the chromatin were used to infer nucleosome positions, where possible, as described in the text.

Nucleosome Repeat Analysis. Individual nucleosome oligomer bands were sized using a calibration curve based on the molecular weight markers. For the calibration curve, the logarithms of the marker fragment sizes (bp) were plotted against gel migration distances, and the curve was fit to a third degree polynomial. Nucleosome oligomer migration distances (band midpoints) were converted to base pairs using the polynomial coefficients. The slope of the best-fit straight line of the plot of nucleosome oligomer size versus nucleosome oligomer (or band) number gives the nucleosome repeat length (*31*). Curve fitting was performed using Microsoft Excel.

RESULTS

Nucleosomes on the Chicken Ovalbumin Gene Spontaneously Rearrange into a Spaced Orderly Array in the Presence of Linker Histone. Figure 1A shows a map of the gene, the locations of the central 4.5 kb DNA fragment subclone and the 1.9 kb fragment subclone studied, restriction sites that are referred to, hybridization probes used, the regions that were deleted in the three deletion mutants, and the region inverted in the inversion mutant. Figure 1B shows the "MNase ladders" obtained from linker histone H5-containing chromatin assembled in vitro. The DNA fragments from two digestion times (30 s, lane 1; 1 min, lane 2) were run on an agarose gel and subjected to Southern hybridization using the Wt probe (Figure 1A). The ladders closely resemble those obtained from native chromatin. It can be seen in lane 1 that, for example, the 7-mer band is slightly lower than the 1399 bp DNA marker (M), consistent with a nucleosome repeat of somewhat less than 200 bp. In the absence of linker histone (data not shown), five to six bands that were multiples of about 150 bp could be seen, superimposed upon a high background. Figure 1C shows an analysis of the MNase ladders; the slope of the curve gives the nucleosome repeat length. The nucleosome oligomer bands had lengths that were multiples of a unit repeat of 193 ± 5 bp when the 12 bands visible in Figure 1B, lane 1, were analyzed. Analysis of the more extensive digest (Figure 1B, lane 2) gave a good fit to the 10 bands visible with a unit repeat also of 193 ± 5 bp. Table 1 gives the fit parameters of the ladders for including different combinations of consecutive points. It can be seen that the ladders can most generally be described as reflecting a nucleosome array with a unit repeat ranging from 187 to 200 bp. The negative *y*-intercepts (Table 1) for the more extensively digested sample are expected and arise from the exonuclease activity associated with MNase on prolonged digestion (*31*).

The Formation of a Regular Nucleosome Array Correlates with the Presence of Regular Oscillations of Period-10 VWG in the DNA Sequence. Figure 2A shows the oscillations of the period-10 VWG counts in a 102 bp sliding window across the entire 4.5 kb subclone, computed as described in the Materials and Methods section. To roughly assess the regularity of the oscillations, we have superimposed a perfect sinusoidal oscillation with a period of 380 bp. Other periodicity values differed more from the curve computed from the DNA sequence under study. It can be seen that although one peak, located approximately between (Gen-

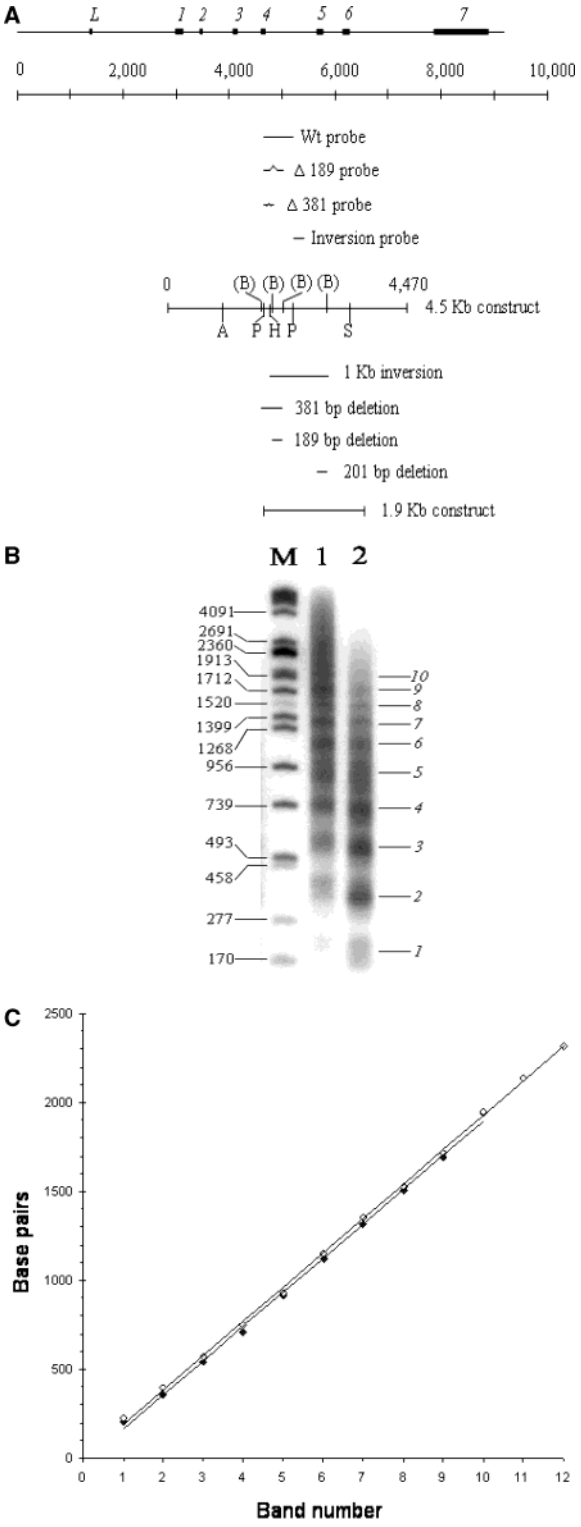

FIGURE 1: Nucleosome array formation on the wild-type chicken ovalbumin gene. (A) Map of the gene, locations/sizes of probes, locations of the 4.5 and 1.9 kb subclones, and DNA regions deleted or inverted. Exons are represented as black boxes labeled L and 1−7. The scale below the gene map is in GenBank nucleotide numbers. The 4.5 kb subclone is renumbered (0−4, 470 bp); restriction sites on this subclone are A, *Ava*II; P, *Pst*I; H, *Hin*dIII; S, *Stu*I; (B), *Bsp*DI (created). (B) Southern blot showing MNase ladders detected with the Wt probe. Lanes: 1, 30 s digest; 2, 1 min digest (nucleosome oligomer bands are indicated); M, DNA size markers with fragment sizes indicated. (C) Plots of nucleosome oligomer sizes versus nucleosome oligomer numbers for the 30 s (□) and the 1 min (■) digests. The straight line fit parameters are reported in Table 1.

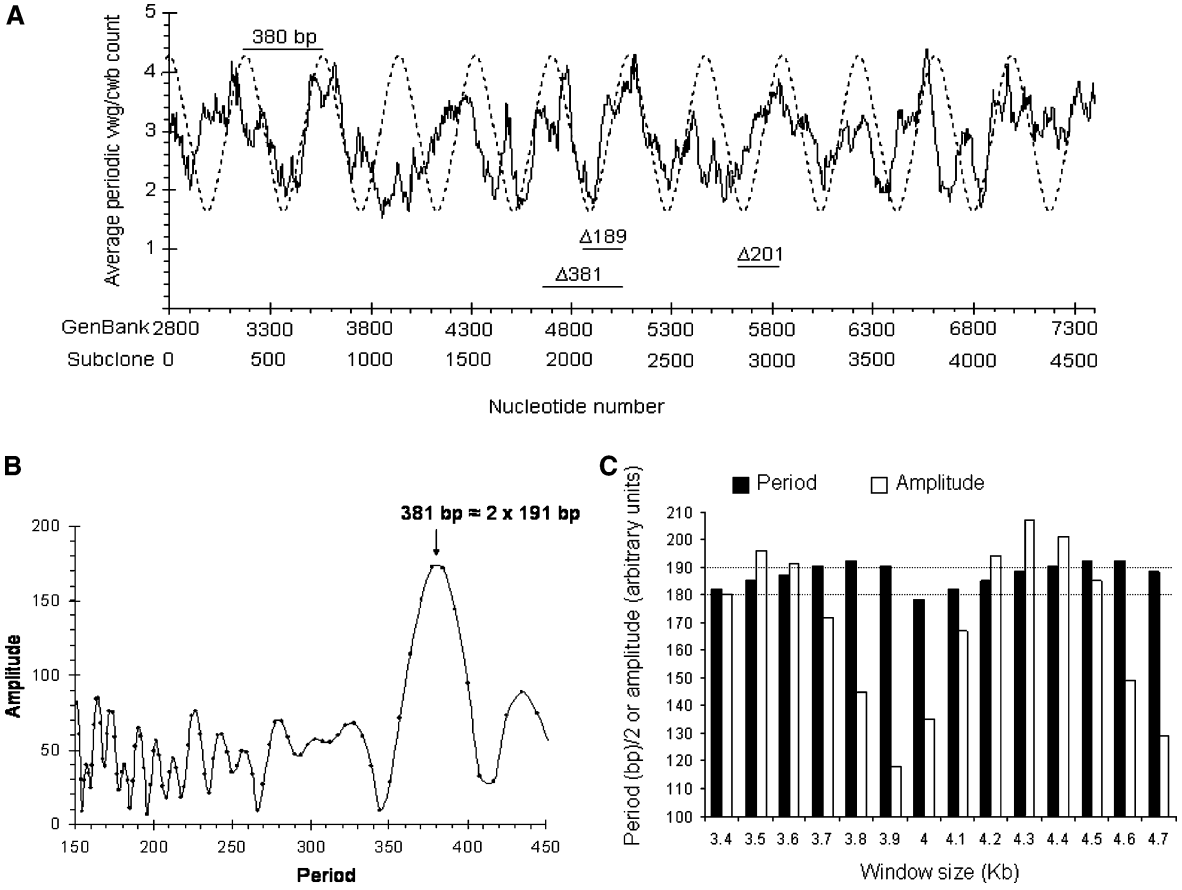

FIGURE 2: Computational analysis of the wild-type sequence. (A) Plot of the period-10 vwg/cwb count (solid curve) in a sliding 102 bp window versus nucleotide number (GenBank or subclone); the counts were averaged as described in the text. The dotted curve is a perfect sinusoidal oscillation with a 380 bp period. The locations and sizes of the three deletions (analyzed subsequently) are indicated. (B) Fourier transform of the solid curve shown in (A) in a 3.7 kb window centered on the Wt probe (Figure 1A). The arrow denotes the period value (bp) of the amplitude maximum of the predominant peak at 381 bp. (C) Histogram representation of the amplitude and period values of the predominant Fourier peaks obtained using windows ranging from 3.4 to 4.7 kb; all windows were centered on the Wt probe. The half-period range from 180 to 190 bp is indicated by dotted lines.

Table 1: Wild-Type Nucleosome Band Plot Analysis

30 s				1 min			
band no.	slope	*y*-intercept	R^2	band no.	slope	*y*-intercept	R^2
1−12	193	−3.09	0.9992	1−10	193	−29.14	0.9984
1−11	193	−1.82	0.9990	1−9	189	−16.38	0.9990
1−10	191	3.47	0.9989	1−8	189	−15.38	0.9985
1−9	189	11.55	0.9989	1−7	187	−9.03	0.9982
2−12	195	−18.10	0.9994	2−10	196	−53.90	0.9990
3−12	196	−33.15	0.9995	3−10	198	−69.90	0.9989
4−12	198	−45.25	0.9994	4−10	200	−88.12	0.9988

Bank) nucleotide numbers 3800 and 4100 on the idealized sinusoidal oscillation, is not present in the curve computed from the DNA sequence and the peaks located approximately between nucleotide numbers 4200 and 4400, or between nucleotide numbers 6400 and 6600, do not completely superimpose with the idealized sinusoidal oscillation, most of the other peaks correspond rather well with those of the perfect sinusoidal oscillation (although several small and narrow apparently spurious peaks are present). To better assess the regularity of the period-10 VWG oscillations, we computed the Fourier transform of the curve generated from the DNA sequence in Figure 2A. The Fourier (Figure 2B), computed in a 3.7 kb window centered on the probe, and ranging from (GenBank) nucleotide numbers 3075 to 6775, shows a strong predominant amplitude at a period of 381 bp. This analysis indicates the presence of a fairly regular oscillation in the curve of period-10 VWG versus nucleotide number with a period of 381 bp located roughly in the middle of the gene. Dividing this dinucleosome-like period by 2 (381 bp/2 $\approx$ 191 bp) gives a nucleosome repeat consistent with the value (193 $\pm$ 5 bp) obtained experimentally (Figure 1C). Figure 2C shows the effect of varying the window size for the Fourier while keeping the window centered on the hybridization probe. The values obtained for the period of the predominant oscillation divided by 2 agree well with the experimentally determined range of nucleosome repeats (Table 1).

The Arrangement of Nucleosomes with Respect to the Oscillations in Period-10 VWG across the DNA Sequence. If the formation of the ordered nucleosome array with a 193 bp periodicity is being caused by the oscillation in period-10 VWG with a 2 $\times$ 193 bp period, one would expect that at least some degree of nucleosome positioning with respect to the DNA sequence should occur. Possible nucleosome positioning with respect to the DNA base sequence was determined by the standard indirect end-label technique, using a short probe (*32*). In this experiment, nucleosome positions can be inferred from differences between cuts in the naked DNA and cuts in the chromatin, both mapped from a convenient restriction site where the short probe terminates. Figure 3A shows the cutting sites for naked DNA (D) and chromatin (C) after light MNase digestion and subsequent

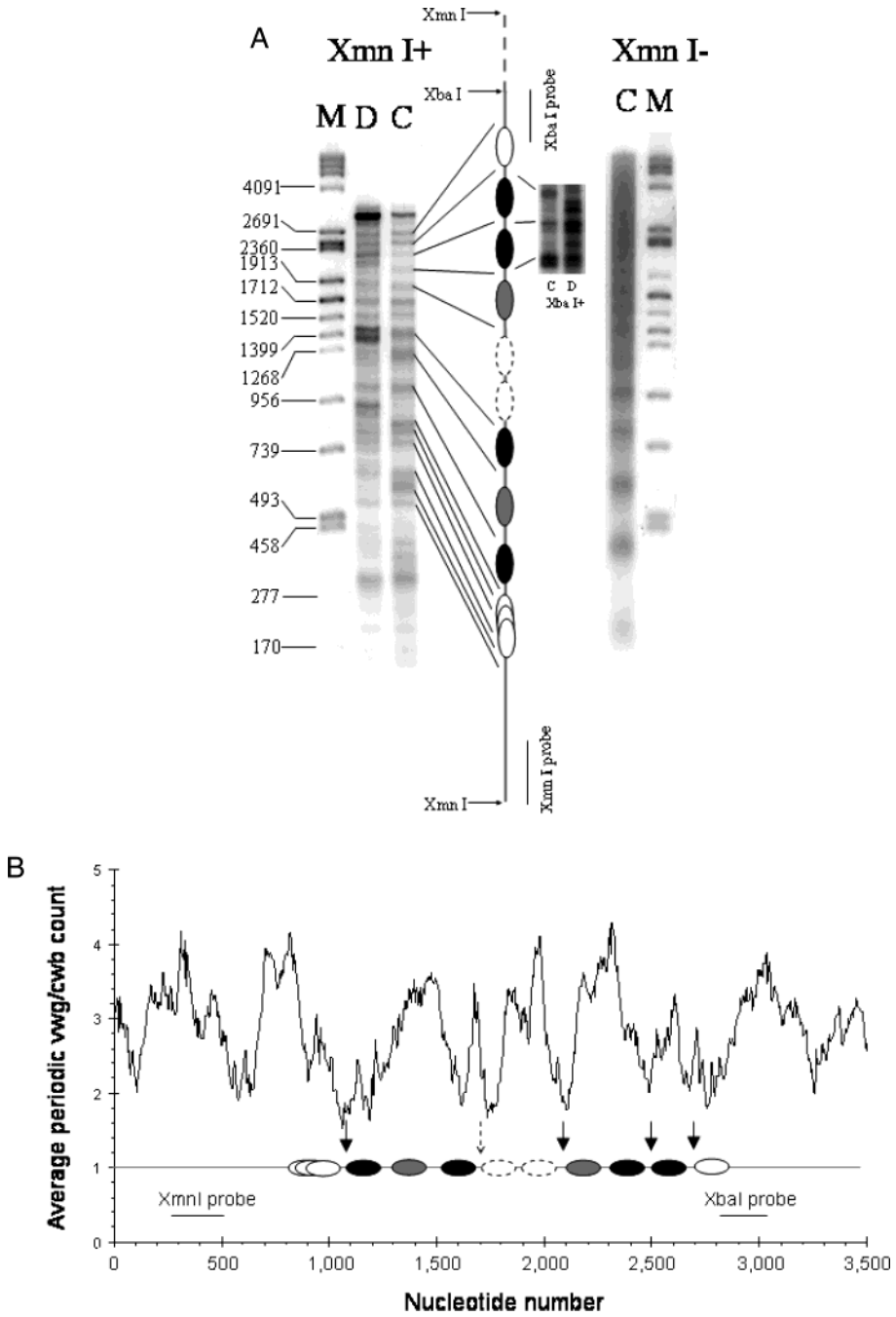

FIGURE 3: Assessment of nucleosome positioning with respect to the DNA sequence. (A) Indirect end-label experiment. MNase-produced DNA fragments from chromatin (C) or naked DNA (D) before (*Xmn*I−) or after (*Xmn*I+) digestion of the purified DNA fragments to completion with *Xmn*I or *Xba*I (*Xba*I+). Lanes M show DNA size markers with fragment sizes indicated. The line and oval diagram depicting assigned relative nucleosome positions is based upon the chromatin and DNA digestion patterns, as described in the text. Black, gray, or white ovals denote high, intermediate, or low confidence positioning, respectively. The dashed ovals indicate a region where there is room for up to two nonpositioned nucleosomes. The *Xmn*I and the *Xba*I reference sites and the probes used are shown. (B) Inferred nucleosome positions with respect to the period-10 VWG oscillations. The solid arrows denote DNA regions where the period-10 VWG counts are low (valleys); the dashed arrow denotes a possible valley region where a sharp spurious peak occurs.

cutting of the DNA fragments with *Xmn*I. The chromatin digest DNA fragment without *Xmn*I digestion (*Xmn*I−) is also included, showing that a regular MNase ladder (with a 193 bp repeat length) was produced. All samples were run on the same gel, and the blot was probed with the short probe indicated. Inferred nucleosome positions are shown using three levels of confidence. For the highest confidence level, we have required that a preferred MNase cutting site on the naked DNA be protected and flanked by cutting sites in the chromatin that are from 146 to 220 bp apart, the size range of a nucleosome. For the intermediate confidence level, we have required that there be a pair of cutting sites in the chromatin that are from 146 to 220 bp apart, with at least one of the sites occurring at a position that does not correspond to a preferred naked DNA cutting site or, alternatively, being present at the same position as in the naked DNA but having a greater intensity in the chromatin sample. For the low confidence level, we have simply required the presence of a pair of cutting sites in the chromatin that are from 146 to 220 bp apart. For regions

where there are no differences between the cutting sites in the naked DNA and chromatin, and where the cutting sites are not between 146 and 220 bp, we assume that nucleosomes may be present but that they are not positioned with respect to the DNA sequence. On the basis of the above criteria, four of the nucleosomes could be assigned positions on the DNA with high confidence (black ovals), two with an intermediate level of confidence (gray ovals), and two with low confidence (white ovals), one of which appears to have three alternative positions. Additionally, it is likely that one or two nonpositioned nucleosomes occupy the central approximately 400 bp region of the mapped DNA shown, where no differences are apparent between the naked DNA and the chromatin bands (dashed ovals). The bands on the chromatin digest that are taken to correspond to nucleosome edges are indicated by lines to the corresponding nucleosome ovals. To better compare the naked DNA and chromatin cutting sites at distances greater than 1900 bp from the *Xmn*I site, another probe and another restriction site (*Xba*I), located 2746 bp away from the *Xmn*I site, were used to assess nucleosome positioning from the other direction. This datum, shown to the right of the nucleosome ovals, confirms, at a high confidence level, that two nucleosomes are positioned in this region of the DNA.

Figure 3B shows the relationship between the positioned nucleosomes and the oscillations in period-10 VWG. It appears that nucleosomes generally seem to avoid the period-10 VWG valleys (arrows). For example, the unusually broad valley at approximately (subclone) nucleotide number 1050 falls between nucleosomes; the same is true for the valleys at about 2100, 2500, and 2700 (solid arrows). The sharp narrow peak at around 1700 (dashed arrow) appears to be a spurious one in a region that ideally should be a valley (see Figure 2A). The chromatin in Figure 3A shows a broad band in this region at a position that was not a preferred DNA cutting site, suggestive of a nucleosome edge.

Disruptive Effect of a Deletion Mutation of a Size Corresponding to a Half-Period Oscillation of the Presumed DNA Signal. If the formation of an ordered nucleosome array on the chicken ovalbumin gene is facilitated by the presence of fairly regular dinucleosome period oscillations in period-10 VWG occurring in the DNA sequence (Figure 2A), then making an approximately half-period (189 bp) deletion in the DNA (see Figure 1A for the deletion position) should disturb this presumed signal. This effect is illustrated in Figure 4A. It can be seen that the period-10 VWG oscillations after the deletion become more irregular and more out of phase with the idealized 2×190 bp period sinusoidal curve. Irregularity in the period-10 VWG oscillation was introduced at the region of the deletion, and new relationships were formed between the somewhat irregularly spaced peaks and valleys that flank the deletion. Thus, MNase-produced DNA fragments from chromatin assembled on the deletion mutant that span the deletion site may no longer be multiples of a 190 bp unit repeat. Figure 4B shows the Fourier transform of the period-10 VWG oscillations of the 189 bp deletion mutant (Figure 4A) compared to that of the wild-type sequence (Figure 2A). The predominant 381 bp peak of the wild-type sequence is replaced by two different peaks, one at 396 bp and the other at 354 bp. The presence of two dinucleosome period peaks, instead of a single peak, would be predicted to cause an aberration in the nucleosome array formed on this sequence, according to the dinucleosome signal hypothesis that we propose.

To see if this deletion mutant assembles into chromatin possessing an aberrant MNase ladder, we performed the necessary experiment. A hybridization probe from near the center of the Fourier window was prepared from the deletion mutant DNA (see Figure 1A). To reduce the proportion of MNase-produced DNA fragments that did not flank the deletion site (and hence should be ordinary), we cut the DNA purified from the MNase digest of the chromatin with *Ava*II and *Stu*I, generating a smaller fragment (2.4 kb for the wild type) that contained the deleted region. Figure 4C, upper part, shows the MNase ladders from the 189 bp deletion mutant compared to the wild-type chromatin, both treated in the same way and probed with their corresponding probes. The differences between the wild-type and mutant chromatin samples are subtle, but significant. Densitometer scans of these two lanes are shown in Figure 4C, lower part, so that the band patterns can be better compared. The typical nucleosome oligomer-like bands of the wild-type chromatin are numbered in the usual way, whereas bands are numbered sequentially from the smallest to the largest for the deletion mutant. It can be seen that, although most of the wild-type nucleosome oligomer bands are also present in the deletion mutant, the positions of bands 5 and 6 differ slightly from the apparently corresponding bands of the deletion mutant (7 and 8) (Figure 4C, upper part). Additionally, the fairly intense band 6 of the deletion mutant is not present in the wild type, and the low-intensity band 4 of the deletion mutant (see scan) is also not present in the wild type. The band between the seventh and eighth of the deletion mutant was not counted because it appears to arise from a fragment having one end truncated by a restriction enzyme. This artificial truncation of a few of the MNase fragments by restriction enzyme cutting is expected, and it appears to occur for the wild-type sequence as well, giving a fragment that is 189 bp larger than the one produced from the deletion mutant. This fragment, although less prominent, can be seen on the gel of the wild type just below the seventh band and on the densitometer scan as a shoulder on the right side of the seventh band. Plots of band size versus band number (Figure 4D) show that the wild-type plot is linear, the bands close to being multiples of a 195 ± 5 bp unit repeat, whereas the deletion mutant plot clearly is not linear. If the deletion mutant band between 7 and 8 is counted and the less prominent shoulder on the wild-type band 7 is ignored (due to its lower relative intensity), the deletion mutant will be even more aberrant. Interestingly, points $1-3$ for the deletion mutant fall on a straight line having a slope of 177 bp, whereas points $6-11$ fall on a straight line having a slope of 198 bp. These two slopes have values that are equal to the predicted value of the period of each predominant Fourier peak of the deletion mutant divided by 2 (Figure 4B).

To exclude the possibility that aberrant bands 4 and 6 of the deletion mutant arose simply from being end fragments, originating selectively from this construct by the restriction enzyme cutting, we examined whether the Wt and Δ189 probes used could detect the approximately 700 bp (band 4) or 850 bp (band 6) if they originated from the *Ava*II (A) or *Stu*I (S) ends of the 4470 bp subclone (Figure 1A). It is clear from the diagram that only the 850 bp fragment (band 6) could be detected and that it would have to arise from the

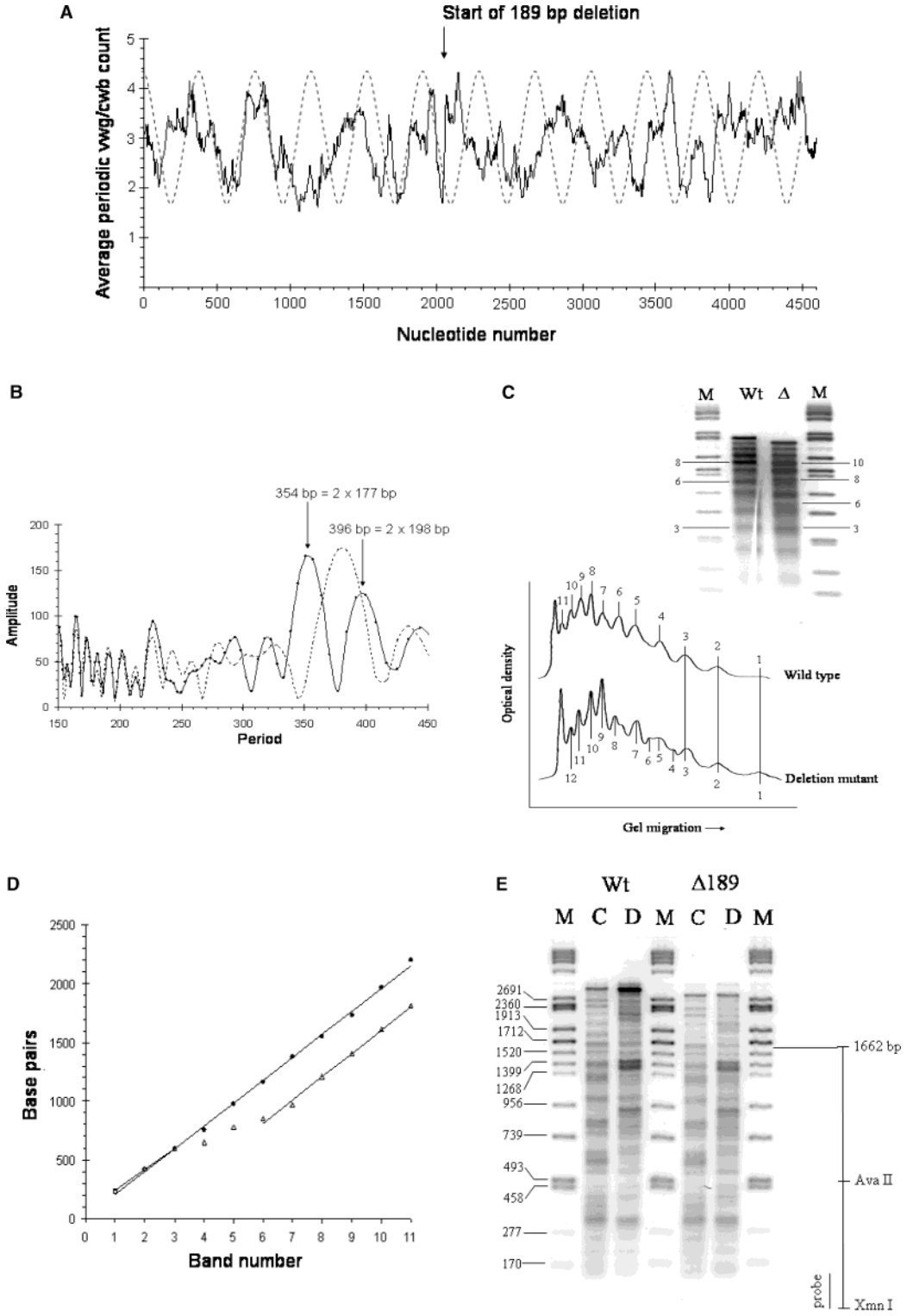

FIGURE 4: Nucleosome array formation on the 4.5 kb subclone containing a disruptive half-period (189 bp) deletion. (A) Plot of the period-10 vwg/cwb count (solid curve) in a sliding 102 bp window versus the deletion mutant subclone nucleotide number; the counts were averaged as described in the text. The dotted curve is a perfect sinusoidal oscillation with a 380 bp period. The start point of the region where the 189 bp deletion was made is indicated. (B) Fourier transforms of the wild-type sequence period-10 VWG oscillation (dashed curve) and deletion mutant oscillation shown in (A) (solid curve) in a window (2.4 kb for the Wt sequence) corresponding to the region between the *Ava*II and *Stu*I sites (Figure 1A). The arrows denote the period values (bp) of the amplitude maxima of the two predominant peaks for the deletion mutant sequence. (C) (Upper part) Southern blot showing MNase ladders for the wild-type (Wt) and deletion mutant (Δ) chromatin. Selected nucleosome oligomer bands detected are numbered sequentially from the bottom of the gel. Lanes M contained the same size markers as in Figure 1B. The blots were hybridized with the Wt or Δ 189 probes (Figure 1A), and the two blots from the same gel were aligned using the markers. (Lower part) Densitometer lane scans of the Wt and Δ lanes above; bands are numbered as above. (D) Plot of DNA fragment size (bp) versus band number to assess the linearity of the relationship for the wild-type (●) and deletion mutant (Δ) sequences. (E) Indirect end-label experiment comparing the MNase cut sites in naked DNA (D) and chromatin (C), measured from the *Xmn*I reference site, for the wild-type (Wt) and deletion mutant (Δ189) sequences. Lanes M contain size markers; the fragment sizes in bp are indicated. The positions of the probe and the *Xmn*I and *Ava*II restriction sites are indicated.

*Ava*II end; the probes could not hybridize to a fragment of length less than about 1000 bp from the *Stu*I end and could not hybridize to a fragment of length less than about 800 bp from the *Ava*II end. To determine whether there could be an 850 bp fragment produced specifically from the deletion mutant with one end cut by MNase and the other by *Ava*II, we performed an indirect end-label experiment comparing the wild-type and deletion mutant MNase cutting patterns (Figure 4E). In this experiment, we utilized the *Xmn*I reference site and the *Xmn*I probe used previously. The *Xmn*I site is 812 bp to the left of the *Ava*II site. Thus, a band located 850 bp from the *Ava*II site would be 850 bp + 812 bp = 1662 bp from the *Xmn*I site. Figure 4E shows that the MNase cuts are essentially identical from before the *Ava*II site until 1700 bp, where the deletion begins. Thus, the 850 bp fragment (deletion mutant band 6 in Figure 4C), arising selectively in the deletion mutant, could not have been generated by cleavage at the *Ava*II end because it should have also been present in the wild-type chromatin, as all of the other MNase cuts detected in the indirect end-label experiment were. These considerations indicate that both of the aberrant bands 4 and 6 of the deletion mutant chromatin were produced by MNase cuts at both ends and that they do not fall on a simple ladder as multiples of a unit repeat because the pattern is aperiodic, as predicted.

Restoring Effect of a Deletion Mutation of a Size Corresponding to a Full-Period Oscillation of the Presumed DNA Signal. Whereas a half-period (189 bp) deletion should disrupt the Fourier significantly, a full-period deletion should have a minimal effect on the Fourier and on the characteristics of the nucleosome array formed, according to the dinucleosome oscillation hypothesis that we propose. To further test this hypothesis, we made a full-period (381 bp) deletion that encompassed the disruptive 189 bp deletion to see if a regular nucleosome array could be restored. The position of the 381 bp deletion is indicated in Figure 1A. Figure 5A shows that the period-10 VWG oscillations now more nearly resemble the perfect 380 bp period of the idealized oscillation, in contrast to what was obtained for the 189 bp deletion mutant (Figure 4A). Also, in contrast with the 189 bp deletion, the 381 bp deletion mutant Fourier now closely resembles the Fourier of the wild type (Figure 5B). Figure 5C, upper part, shows the result of assembling this deletion mutant into chromatin, digesting with MNase, cutting the purified DNA with *Ava*II and *Stu*I, and hybridizing the blot of the resulting electrophoresed DNA fragments with a probe, prepared from this deletion mutant, that was located near the center of the Fourier window. Nine MNase-produced prominent bands were detected; the approximately 2000 bp *Ava*II−*Stu*I restriction fragment was also detected. The densitometer scan (Figure 5C, lower part) shows that the nine numbered bands are the predominant ones and that they appear to generate a ladder. For a restriction fragment this small (2 kb), one should expect a greater contribution from fragments containing the *Ava*II or *Stu*I ends than obtained previously and that these fragments should not necessarily fall on the ladder. The minor bands visible on the autoradiogram and the corresponding shoulders on some of the densitometer scan peaks reflect this expectation. Figure 5D shows that a plot of the predominant DNA fragment sizes versus their band numbers is linear with a slope of 197 ± 5 bp and a +8 bp *y*-intercept, indicating that the bands form a regular MNase ladder with a 197 ± 5 bp unit repeat, in agreement with the 2 × 193 bp period predominant Fourier amplitude predicted (Figure 5B).

Disruptive Effect of a 201 bp Deletion Mutation in Another Region of the Gene. We next examined the effects of another small, approximately half-period (201 bp) deletion, located 584 bp downstream of the previous one, using a smaller sized (1.9 kb for the wild type) subclone than the one used previously (Figure 1A). Due to the presence of irregularities in the period-10 VWG oscillation across the gene, one might expect to see somewhat different results depending upon where a deletion is made. Furthermore, the smaller sized subclone should be more sensitive to sequence changes than the one used above because a larger proportion of the DNA sequence is now altered. Also, the chance that possible effects on nucleosome array formation may arise from more distal ovalbumin gene sequences is eliminated. Figure 6A shows the Fourier transforms for the wild type and the 201 bp deletion mutant. The wild-type sequence in this region has a strong (for a 1.9 kb window) predominant 357 bp period oscillation in period-10 VWG. For the deletion mutant, the predominant Fourier amplitude is shifted to the left to a period of 314 bp. Additionally, a smaller and broader peak in the dinucleosome range appears at about 390 bp. Figure 6B shows the results obtained after assembling the wild-type and mutant constructs into chromatin, digesting with MNase, gel electrophoresis of the resulting DNA fragments, and staining the gel with ethidium bromide. A regular approximately 180 ± 5 bp ladder is apparent for the wild-type chromatin, with the 6-mer running slightly faster than the 1107 bp size marker (lanes 2 and 3) and very close to 6 × 180 = 1080 bp. The 180 ± 5 bp nucleosome repeat is consistent with the computational prediction of 357 bp/2 ≈ 179 bp. In contrast, the deletion mutant chromatin gave a less distinct ladder than the wild type for the 1 min digestion points (compare lanes 1 and 4), consistent with the presence of two Fourier peaks in the dinucleosome range (Figure 6A). However, for the more extensively digested (3 min) chromatin samples (lane 6), a more distinct ladder with an approximately 158 ± 5 bp repeat appears for the deletion mutant, with the 7-mer now running at the position of the 1107 bp size marker, since 7 × 158 bp = 1106 bp. The 158 ± 5 bp nucleosome repeat is consistent with the predominant Fourier peak at 314 bp, since 314 bp/2 = 157 bp. It is reasonable to assume that the majority of chromatin molecules that possessed the fairly regular 158 bp nucleosome array were slightly more resistant to digestion than the molecules that had longer and more heterogeneous repeats. Hence, the more resistant molecules would become enriched in the solution after the molecules with longer nucleosome linkers had been digested away. In the absence of linker histone (not shown) the MNase digests of the above samples revealed closely similar DNA fragment patterns, reflecting close packed nucleosomes.

Effect of an Approximately 1000 bp Inversion Mutation That Alters the Presumed Signal in DNA. Another way of altering the wild-type DNA sequence, but without deleting portions, is by making an inversion. The 1050 bp inversion indicated in Figure 1A was made in our 4.5 kb subclone. This inversion caused some alteration of the wild-type Fourier, as shown in Figure 7A. The predominant 384 bp = 2 × 192 bp amplitude of the wild-type sequence (dashed

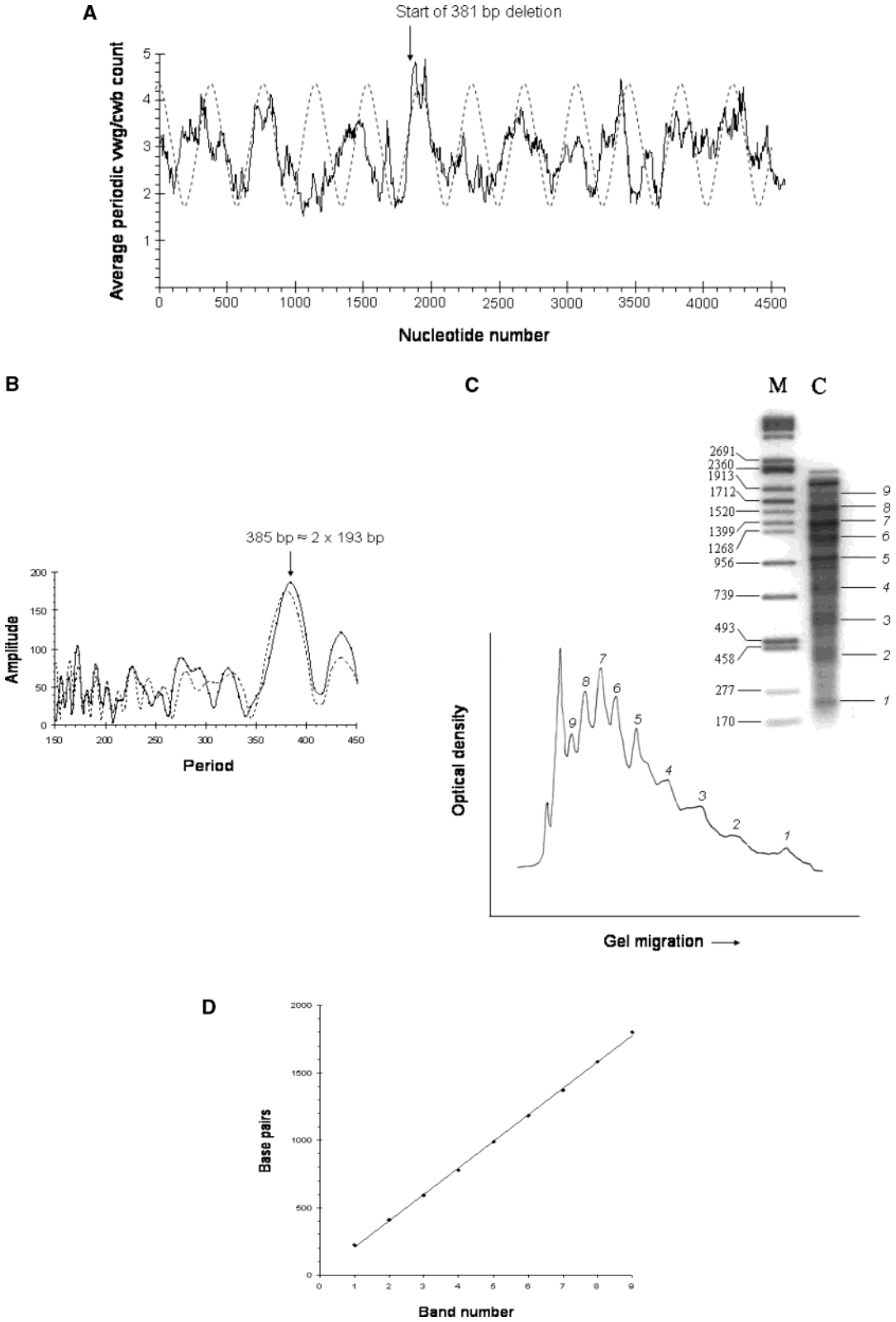

FIGURE 5: Nucleosome array formation on the 4.5 kb subclone containing a restorative full-period (381 bp) deletion. (A) Plot of the period-10 vwg/cwb count (solid curve) in a sliding 102 bp window versus the deletion mutant subclone nucleotide number; the counts were averaged as described in the text. The dotted curve is a perfect sinusoidal oscillation with a 380 bp period. The start point of the region where the 381 bp deletion was made is indicated. (B) Fourier transforms of the wild-type sequence period-10 VWG oscillation (dashed curve) and deletion mutant oscillation shown in (A) (solid curve) in a window (2.4 kb for the Wt sequence) corresponding to the region between the *Ava*II and *Stu*I sites (Figure 1A). The arrow denotes the period value (bp) of the amplitude maxima of the predominant peak for the deletion mutant sequence. (C) (Upper part) Southern blot showing the MNase ladder for the deletion mutant chromatin (lane C). The predominant bands detected are numbered sequentially from the bottom of the gel (see text). Lane M contains size markers; fragment sizes are indicated. The blots were hybridized with the Δ381 probe (Figure 1A). (Lower part) Densitometer lane scan of lane C above; bands are numbered as above. (D) Plot of DNA fragment size (bp) versus band number to assess the linearity of the relationship. The best fit straight line shown had a slope of 197 bp and a *y*-intercept of +8 bp; the R^2 value of the fit was 0.9993.

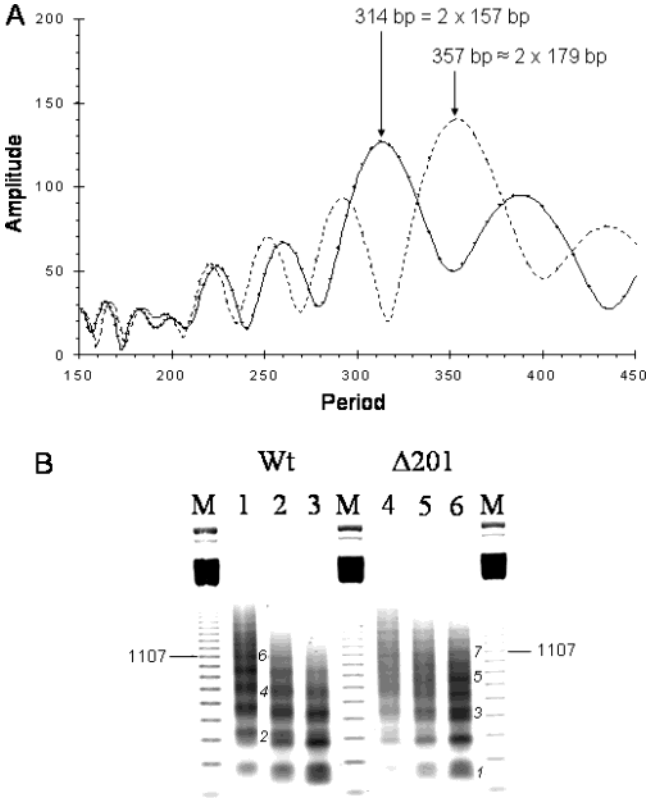

FIGURE 6: Nucleosome array formation on the 1.9 kb subclone containing a disruptive half-period (201 bp) deletion. (A) Fourier transforms of the wild-type sequence period-10 VWG oscillation (dashed curve) and deletion mutant oscillation (solid curve) in a window (1.9 kb for the Wt sequence) corresponding to the subclone. The arrows denote the period values (bp) of the amplitude maxima of the predominant peak for the wild-type and the deletion mutant sequences. (B) MNase ladders, detected by ethidium bromide staining, for the wild-type (Wt) and deletion mutant (Δ201) chromatin. Lanes 1 and 4 are 1 min digests, lanes 2 and 5 are 2 min digests, and lanes 3 and 6 are 3 min digests. Lanes M contain 123 bp ladder size markers; the 1107 bp marker band is indicated. Selected nucleosome oligomer bands are numbered.

curve) was reduced in the inversion mutant (solid curve), and a new 330 bp = 2 × 165 bp peak becomes nearly as prominent as the 384 bp peak. Figure 7B shows how the two most prominent Fourier amplitudes depend on the window size, which was centered on the inversion probe shown in Figure 1A. The value of the larger period fluctuated, ranging between 180 and 210 bp, whereas the value of the smaller period remained fairly constant at about 165 bp. Thus, the computational prediction, according to the dinucleosome period hypothesis, is that there should be two predominant periodicities in the nucleosome array assembled on the inversion mutant sequence: one having a periodicity of approximately 192 bp and the other with a periodicity of about 165 bp.

To test this prediction, we assembled the inversion mutant into chromatin, digested the chromatin with MNase, electrophoresed the resulting DNA fragments, and hybridized the blot to a probe prepared from the deletion mutant (Figure 1A) that was located in the center of the Fourier window. Figure 7C again shows the typical 193 ± 5 bp ladder for the wild-type sequence, as was obtained previously (Figure 1B). However, for the inversion mutant chromatin, it is clear that the DNA fragment sizes for the nucleosome oligomers of size greater than the 4-mer deviate significantly from those of the wild-type chromatin. Figure 7D shows that the plot of nucleosome oligomer size versus nucleosome oligomer number is not linear. Thus, the inversion mutant nucleosome oligomers are not multiples of a unit repeat. The first six oligomers are well represented by a straight line with a slope of 166 bp, whereas oligomers five to nine are represented fairly well by a straight line with a slope of 193 bp. These two slopes agree well with the computational predictions of 165 and 192 bp, respectively (Figure 7A).

DISCUSSION

We have shown that nucleosomes deposited by salt gradient dialysis on a 4.5 kb subclone, taken from roughly the center of the chicken ovalbumin gene, spontaneously align into a physiologically spaced, well-ordered nucleosome array when incubated with purified linker histone H5 in the presence of the solubilizing agent polyglutamic acid. The fact that plasmid DNA sequences, cDNAs, and some other cloned eukaryotic DNA sequences do not generate ordered, spaced nucleosome arrays under identical conditions (*12*, *25*) suggests that there are motifs present within the chicken ovalbumin gene sequence that facilitate this linker histone-dependent nucleosome alignment. Potentially, there could be a variety of sequence motifs hidden in this genomic DNA sequence that can contribute to the formation of a 193 ± 5 bp nucleosome array; understandably, such motifs would not be easy to find. Using a hidden Markov model and a supercomputer to examine a very large number of unspecified patterns in human DNA, Baldi et al. (*23*) serendipitously found that the single motif non-T, A/T, G (VWG), repeated at 10 bp intervals, was highly overrepresented. Because of the 10 bp periodicity, it was suggested that this periodic motif might be a preferential DNA-bending signal involved in nucleosome positioning. This idea was supported experimentally in a study from our laboratory (*24*), and it was additionally found that the period-10 VWG motif oscillated with a dinucleosome-like periodicity on some genomic DNA sequences. Figure 2B shows that the chicken ovalbumin gene DNA sequence possesses a strong predominant dinucleosome-like 381 bp periodicity for a 3.7 kb window centered on a probe located in the center of the gene. This computational result correlates well with the experimentally determined 193 ± 5 bp nucleosome repeat, since 381 bp/2 ≈ 191 bp. Moreover, the computational result does not change appreciably for window sizes ranging from 3.4 to 4.7 kb (Figure 2C). Large predominant Fourier amplitudes with periods of approximately 2 × 190 bp are well represented. Some variation in the Fourier transform is expected as additional sequences are included, particularly when only portions of peaks or valleys are included in a given window (see Figure 2A). Furthermore, shifting the probe center position 400 bp to the right did not change the results significantly (not shown). Therefore, the probes used for this study (Figure 1A) are representative of a substantial portion of the gene.

If the DNA sequence is to exert an influence on nucleosome array formation, it is reasonable to expect that the nucleosomes exhibit at least some degree of positioning with respect to the DNA sequence. However, each nucleosome of the array does not have to be uniquely positioned. Moreover, nucleosome positioning with respect to the DNA sequence does not have to be very precise to simply generate

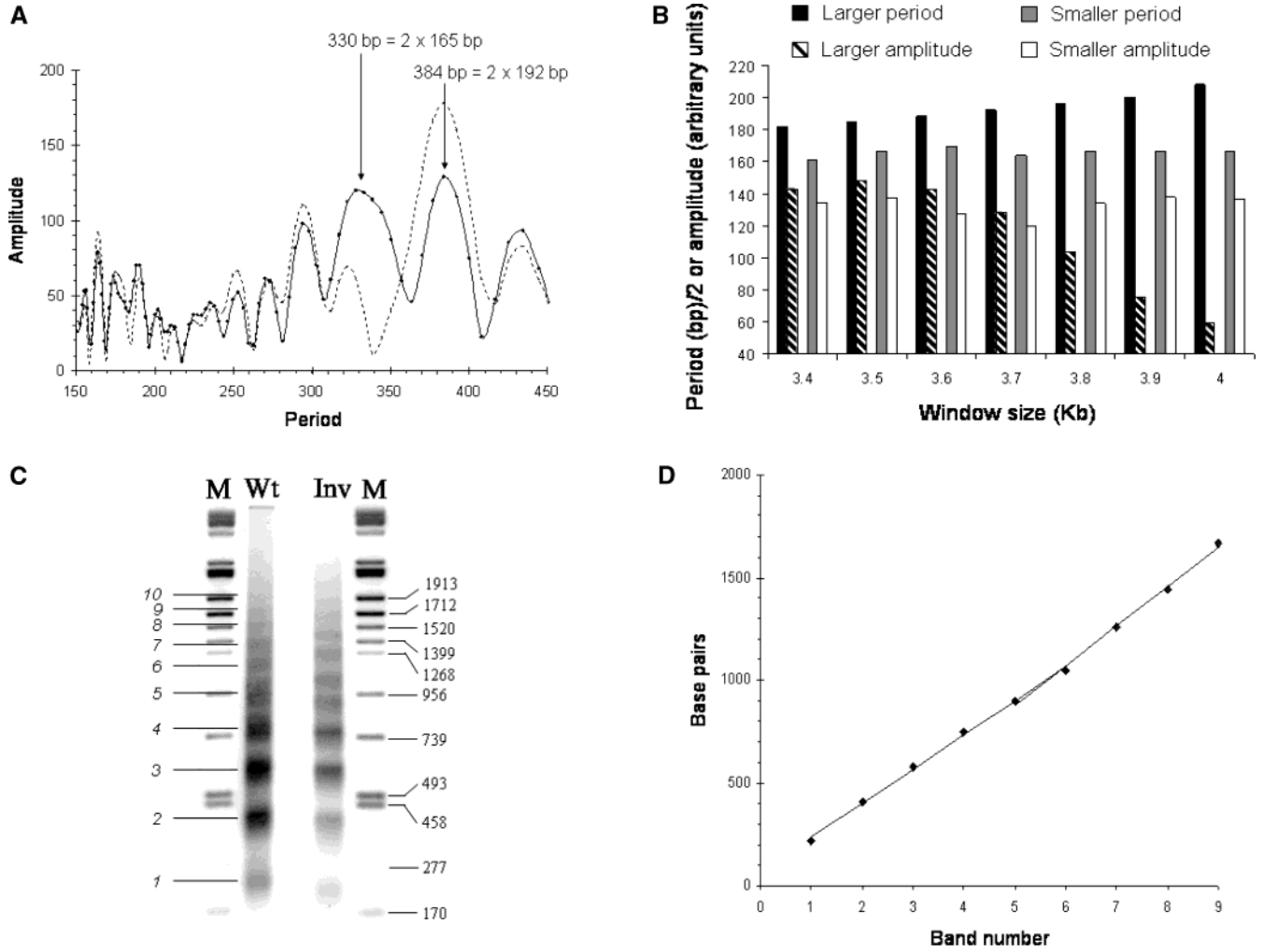

FIGURE 7: Nucleosome array formation on the 4.5 kb subclone containing a 1050 bp inversion. (A) Fourier transforms of the Wt sequence period-10 VWG oscillation (dashed curve) and inversion mutant oscillation (solid curve) in a 3.7 kb window. The arrows denote the period values (bp) of the amplitude maxima of the two predominant peaks for the inversion mutant sequences. (B) Histogram representation of the amplitude and period values of the predominant Fourier peaks obtained using windows ranging from 3.4 to 4.7 kb; all windows were centered on the inversion probe (Figure 1A). (C) Southern blot showing MNase ladders for the wild type (Wt) and inversion mutant (Inv) chromatin. The nucleosome oligomer bands of the Wt ladder are numbered. Lanes M contain size markers; fragment sizes are given. The blots were hybridized with the Wt or inversion probes (Figure 1A), and the two blots from the same gel were aligned using the markers. (D) Plot of DNA fragment size (bp) versus band number to assess the linearity of the relationship. Best fit straight lines for band numbers 1−6 and 5−9 are shown (see text).

a typical nucleosome ladder. The indirect end-label experiment shown in Figure 3A illustrates that at least four nucleosomes in the vicinity of the *Xmn*I site are definitely positioned with respect to the DNA sequence (black ovals). Indirect end-label data do not permit high-precision nucleosome placement on DNA for the inferred positions. For example, the black nucleosome oval representing the 147 bp core of the nucleosome, positioned with high confidence, approximately 940 bp from the *Xmn*I reference site (Figure 3A) could lie anywhere between the two chromatin MNase cutting sites that flank the well-protected preferred naked DNA cutting site, and these sites are 180 bp apart. Similarly, the adjacent nucleosome (gray oval, denoting intermediate confidence level) could be anywhere within a 210 bp region.

Some insight into how the period-10 VWG oscillations serve to promote the formation of a regular array is obtained from Figure 3B. The nucleosome positioning data suggest that nucleosomes may not necessarily have a preference for forming on the presumably more bendable DNA located at regions of high period-10 VWG counts (peaks). Rather, nucleosomes may tend to avoid forming on regions of DNA with low period-10 VWG counts (valleys), constituting the less bendable DNA. Thus, the period-10 VWG valleys should correspond to nucleosome linkers, which can be cut by MNase, as observed. For regular period-10 VWG oscillations with a period of 380 bp, avoidance of the valley regions should promote the formation of a regular array with two nucleosomes for each 380 bp, or a nucleosome repeat of 380 bp/2 = 190 bp, also as observed (see Table 1).

There is clearly a correlation between our computational and our experimental results. According to the dinucleosome period hypothesis for period-10 VWG oscillations ("dinucleosome VWG hypothesis"), the existence of a strong predominant Fourier amplitude predicts that an ordered nucleosome array should form spontaneously in the presence of linker histone in our in vitro system. The predicted value of the nucleosome repeat should be one-half the value of the period of the predominant Fourier peak. This is indeed what we have observed. In addition to the data presented here, our laboratory has consistently obtained strong cor-

relations, without exceptions, between computational predictions and experimental results using a variety of vertebrate DNA sequences (ref *24* and unpublished observations). In this work we have provided more direct evidence for the dinucleosome VWG hypothesis by making relatively small sequence variations that alter the Fourier transform of the curve of the averaged period-10 VWG counts versus nucleotide number. The Fourier is altered because the period-10 VWG oscillation is not perfect, causing some new relationships between peaks and valleys to be created by the alteration. In each case, the experimental results were very consistent with our computational predictions.

For the 189 bp deletion mutant, the single predominant Fourier peak of the wild-type sequence is lost and splits into two: one at a slightly higher period and one at a slightly lower period (Figure 4B). Experimentally, the MNase ladder becomes more complex, deviating from a simple periodic ladder pattern (Figure 4C), and the two new periodicities predicted can be detected within the overall aperiodic MNase ladder (Figure 4D). However, by making a full-period 381 bp deletion, the wild-type Fourier (Figure 5B) and MNase ladder (Figure 5C) were essentially restored, even though this deletion fully encompassed the previous disruptive one (Figure 1A). Because a nucleosome unit repeat-sized (189 bp) deletion interfered with the formation of a regular array, these data are not consistent with any signal having a mononucleosome (189 bp) periodicity, nor are they consistent with a signal having the sequence motifs within the 189 bp that were deleted being essential for nucleosome array formation. In contrast, these data are consistent with the dinucleosome VWG hypothesis.

Similar to the 189 bp deletion, the 201 bp deletion altered the Fourier transform of the period-10 VWG oscillations versus nucleotide number curve. In this case, the predominant Fourier amplitude shifted to the left considerably (Figure 6A). Hence, a very short nucleosome repeat is predicted for this mutant (at longer digestion times). Experimentally, this is again what we observed (Figure 6B). Finally, for a 1050 bp inversion mutation, the single predominant Fourier peak at 384 bp (2×192 bp) of the wild-type sequence was diminished, and a second peak, nearly as intense, at 330 bp (2×165 bp) appeared (Figure 7A). Therefore, our computational prediction is that the inversion mutant chromatin should not generate a simple MNase ladder but should generate DNA fragments having both periodicities. This is once more what we observed (Figure 7C,D). In this experiment no regions of DNA were deleted; only the relationships between some sequences changed. In summary, all of these mutagenesis experiments suggest that nucleosome array formation is facilitated by dinucleosome period oscillations in period-10 VWG and that it is unlikely that some other type of signal in DNA is instead responsible.

In our in vitro system, nucleosome array formation is solely a consequence of the chemistry of histone−DNA interactions, and clearly the DNA sequence influences nucleosome array formation in the presence of linker histone. In contrast, the chromatin assembly extracts derived from the extremely rapidly dividing *Drosophila* early embryos are ATP-dependent, linker histone is not required, and nucleosome array formation does not depend on the DNA sequence (*33*). Thus, there is more than one way to generate a physiological nucleosome array, and many factors could affect nucleosome array formation. The mechanism used to form nucleosome arrays might depend on the developmental stage, the chromosomal locus, or the cell type. In vivo, the mechanism of histone deposition and redistribution on replicating DNA differs considerably from salt gradient dialysis (*11*). Nevertheless, it is likely that the same chemistry that causes linker histone to align nucleosomes into ordered arrays in our in vitro system contributes to nucleosome array formation in vivo for chromatin that contains linker histones. Recently, it has been shown that linker histones affect nucleosome spacing in the mouse and are essential for development (*16*). Additionally, clear differences between nucleosome arrays have been found in vivo. For example, it has been demonstrated that the repetitive DNA of the highly condensed *Drosophila* heterochromatin possesses nucleosome arrays that are more ordered and that have a different repeat length than the nucleosomes of bulk chromatin (*10*). Also, it has been demonstrated that, in the mouse, the constitutively expressed adenosine deaminase gene is packaged into chromatin possessing a more ordered nucleosome array, with a different repeat length than the bulk chromatin, which correlates with an unusually strong dinucleosome VWG signal throughout its 70 kb locus (Dalal et al., unpublished observations). An intriguing idea is that distinctive higher order chromatin structures form from distinctive arrays of nucleosomes (*10*, *12*). Therefore, the potential to form particular chromatin structures may actually be encoded in the DNA sequence.

Fifty years ago the structure of DNA was elucidated (*34*). It took another eight years before the genetic code was deciphered (*35*). Could there possibly be another level of DNA sequence information beyond the triplet-based genetic code in higher organisms? If so, it would have had to withstand long-term selective pressure. Now, in the age of genomics and bioinformatics, it is possible to search genomic DNA for further information. The discovery that period-10 VWG triplets are perhaps the most abundant motif in human genomic DNA (*23*) suggests that this motif may serve a function. If it can be established that this VWG repeat has a role to play in the structural organization of chromatin, it may well be that these six triplets, existing predominantly in noncoding DNA, constitute another level of genetic information beyond the known triplet code.

ACKNOWLEDGMENT

We thank Dr. B. W. O'Malley for providing the chicken ovalbumin DNA genomic clone, Chad Pitschka for assistance with Excel Macros, and Tomara Fleury for technical assistance with some of the experiments.

REFERENCES

1. Collins, F. S., Morgan, M., and Patrinos, A. (2003) The human genome project: lessons from large-scale biology, *Science 300*, 286−290.
2. Venter, J. C., et al. (2001) The sequence of the human genome, *Science 291*, 1304−1351.
3. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome, *Nature 409*, 860−921.
4. 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.

5. Rennie, J. (2003) A conversation with James D. Watson, *Sci. Am. 288*, 66−69.
6. 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. U.S.A. 90*, 9021−9025.
7. 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. U.S.A. 91*, 5277−5280.
8. Woodcock, C. L., and Horowitz, R. A. (1995) Chromatin organization re-evaluated, *Trends Cell Biol. 5*, 272−277.
9. van Holde, K. E., and Zlatanova, J. (1995) Chromatin higher order structure: chasing a mirage?, *J. Biol. Chem. 270*, 8373−8376.
10. 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.
11. van Holde, K. E. (1989) *Chromatin*, Springer-Verlag, New York.
12. Liu, K., and Stein, A. (1997) DNA sequence encodes information for nucleosome array formation, *J. Mol. Biol. 207*, 559−573.
13. Stein, A., Dalal, Y., and Fleury, T. J. (2002) Circle ligation of *in vitro* assembled chromatin indicates a highly flexible structure, *Nucleic Acids Res. 20*, 5103−5109.
14. Howe, L., Iskandar, M., and Ausio, J. (1998) Folding of chromatin in the presence of heterogeneous histone H1 binding to nucleosomes, *J. Biol. Chem. 273*, 11625−11629.
15. Georgel, P. T., and Hansen, J. C. (2001) Linker histone function in chromatin: dual mechanism of action, *Biochem. Cell. Biol. 79*, 313−316.
16. Fan, Y., Nikitina, T., Morin-Kensicki, E. M., Zhao, J., Magnuson, T. R., Woodcock, C. L., and Skoultchi, A. I. (2003) H1 linker histones are essential for mouse development and affect nucleosome spacing *in vivo*, *Mol. Cell. Biol. 23*, 4559−4572.
17. Tsukiyama, T. (2002) The *in vivo* functions of ATP-dependent chromatin remodeling factors, *Nat. Rev. 3*, 422−429.
18. Becker, P. B., and Wu, C. (1992) Cell-free system for assembly of transcriptionally repressed chromatin from *Drosophila* embryos, *Mol. Cell. Biol. 12*, 2241−2249.
19. Bulger, M., Ito, T., Kamakaka, R. T., and Kadonaga, J. T. (1995) Assembly of regularly spaced nucleosome arrays by *Drosophila* chromatin assembly factor 1 and a 56-kDa histone-binding protein, *Proc. Natl. Acad. Sci. U.S.A. 92*, 11726−11730.
20. Laskey, R. A., Mills, A. D., and Morris, N. N. (1977) Assembly of SV40 chromatin in a cell-free system from *Xenopus* eggs, *Cell 10*, 237−243.
21. Almouzni, G., and Méchali, M. (1988) Assembly of spaced chromatin: involvement of ATP and DNA topoisomerase activity, *EMBO J. 7*, 4355−4365.
22. Stein, A., and Bina, M. (1984) A model chromatin assembly system: factors affecting nucleosome spacing, *J. Mol. Biol. 178*, 341−363.
23. 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.
24. Stein, A., and Bina, M. (1999) A signal encoded in vertebrate DNA that influences nucleosome positioning and alignment, *Nucleic Acids Res. 27*, 848−853.
25. Lauderdale, J. D., and Stein, A. (1992) Introns of the chicken ovalbumin gene promote nucleosome alignment *in vitro*, *Nucleic Acids Res. 20*, 6589−6596.
26. Liu, K., Lauderdale, J. D., and Stein, A. (1993) Signals in chicken β-globin DNA influence chromatin assembly *in vitro*, *Mol. Cell. Biol. 13*, 7596−7603.
27. Jeong, S., and Stein, A. (1994) DNA sequence affects nucleosome ordering on replicating plasmids in transfected Cos-1 cells and *in vitro*, *J. Biol. Chem. 269*, 2197−2205.
28. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
29. 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.
30. Feinberg, A. P., and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction fragments to high specific activity, *Anal. Biochem. 132*, 6−13.
31. 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.
32. Wu, C. (1980) The 5′-ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I, *Nature 286*, 854−860.
33. Blank, T. A., and Becker, P. B. (1996) The effect of nucleosome phasing sequences and DNA topology on nucleosome spacing, *J. Mol. Biol. 260*, 1−8.
34. Watson, J. D., and Crick, F. H. C. (1953) Molecular structure of nucleic acids, *Nature 171*, 737−738.
35. Crick, F. H. C., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961) General nature of the genetic code for proteins, *Nature 192*, 1227−1232.

BI049717F