Article No. jmbi.1999.2686 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 288, 213–229

JMB

Sequence Motifs and Free Energies of Selected Natural and Non-natural Nucleosome Positioning DNA Sequences

A. Thåström[1,2], P. T. Lowary[1], H. R. Widlund[2], H. Cao[2], M. Kubista[2] and J. Widom[1,3]*

[1]*Department of Biochemistry Molecular Biology, and Cell Biology, Northwestern University, Evanston IL 60208-3500, USA*

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

[3]*Department of Chemistry Northwester University Evanston, IL 60208, USA*

**Corresponding author*

Our laboratories recently completed SELEX experiments to isolate DNA sequences that most-strongly favor or disfavor nucleosome formation and positioning, from the entire mouse genome or from even more diverse pools of chemically synthetic random sequence DNA. Here we directly compare these selected natural and non-natural sequences. We find that the strongest natural positioning sequences have affinities for histone binding and nucleosome formation that are sixfold or more lower than those possessed by many of the selected non-natural sequences. We conclude that even the highest-affinity sequence regions of eukaryotic genomes are not evolved for the highest affinity or nucleosome positioning power. Fourier transform calculations on the selected natural sequences reveal a special significance for nucleosome positioning of a motif consisting of ~10 bp periodic placement of TA dinucleotide steps. Contributions to histone binding and nucleosome formation from periodic TA steps are more significant than those from other periodic steps such as AA (=TT), CC (=GG) and more important than those from the other YR steps (CA (=TG) and CG), which are reported to have greater conformational flexibility in protein-DNA complexes even than TA. We report the development of improved procedures for measuring the free energies of even stronger positioning sequences that may be isolated in the future, and show that when the favorable free energy of histone-DNA interactions becomes sufficiently large, measurements based on the widely used exchange method become unreliable.

Keywords: affinity; chromatin; histone; nucleosome; positioning

Introduction

DNA sequences that bias the positioning of nucleosomes (nucleosome-positioning sequences) are important for many reasons, both *in vivo* and *in vitro*. *In vivo* the biased positioning of nucleosomes has important roles in basal gene regulation, DNA compaction, and chromosome segregation. Nucleosome positioning sequences are routinely used in studies of the structure and function of chromatin *in vitro* because they allow the preparation of homogeneous nucleosomes and nucleosome oligomers. Future studies will use nucleosome positioning sequences to experimentally explore the coupling between nucleosome positioning and chromosome function *in vivo*. Substantial effort has been directed toward analyzing nucleosome positioning sequences, with the goal of deducing the DNA sequence rules and motifs that confer nucleosome positioning ability.

Nucleosome positioning refers to the position the DNA helix adopts with respect to the histone core. Translational positioning refers to the extent to which the histone octamer selects a particular stretch of 147 bp of DNA in preference to other stretches of the same length. Rotational positioning is a degenerate form of translational positioning in which a set of translational positions, differing by integral multiples of the DNA helical twist are all occupied in preference to the set of other possible locations. Thus, nucleosome positioning sequences

Abbreviations used: BZA, benzamidine hydrochloride; PMSF, phenylmethylsulfonyl fluoride.

E-mail address of the corresponding author: j-widom@nwu.edu

0022-2836/99/170213–17 $30.00/0

are DNA sequences that bias their own packing into nucleosomes by preferential positioning of the histone octamer.

DNA sequence-directed nucleosome positioning achieves equilibrium (Widom, 1999); and see below), so the sites of preferential positioning will be those having minimum free energy. Thus nucleosome-positioning is statistical, not precise: all possible positions along a DNA sequence will be occupied to an extent greater than 0, with the occupancy at each position linked to the free energy in accord with Boltzmann's law. There exists a particular mathematical relationship between the free energy of histone-DNA interactions measured in competitive nucleosome reconstitution experiments, and the time- or ensemble-averaged probability of occupancy of the preferred site (Lowary & Widom, 1997).

In vivo, many factors may contribute to the free energy of nucleosome positioning (Yao *et al*., 1993). The fact that DNA sequence is not the only determinant of nucleosome positioning is evident from the fact that differing cell types within a single organism, which all have the same genomic DNA sequences, nevertheless may differ even in bulk chromatin properties such as the nucleosome repeat length. However, many cases are known in which the favored positioning of nucleosomes *in vitro* is the same as the dominant positioning found *in vivo* (e.g. Fragoso *et al*., 1995), implying that DNA sequence preferences are significant. In addition to DNA sequence preferences, other interactions that may contribute significantly to nucleosome positioning include: nucleosome-nucleosome interactions arising from higher-order chromatin folding, competition from other DNA-binding proteins, and statistical effects from filling-in by nucleosomes of a region bounded by points that either attract or repel nucleosomes.

In the nucleosome the DNA is bent very tightly around the histone octamer (Luger *et al*., 1997a). Thus, other things being equal, a particular DNA sequence that happens (owing to its sequence) to be bent to a degree appropriate for nucleosome packaging will have a higher affinity for the histone octamer (and thus an increased nucleosome-positioning power), since less mechanical work (and less free energy) will be expended to wrap the DNA on the histone surface. Alternatively, a DNA sequence that has no particular static bendedness but which instead is more easily bendable (lower bending force constant) will similarly exhibit increased affinity and nucleosome positioning power. Similarly, the histone octamer controls the local DNA helical twist (Luger *et al*., 1997a); thus DNA sequences that have the appropriate position-dependent helical twist, or other sequences that are simply more readily twisted (lower twisting force constant) will exhibit increased affinity and nucleosome positioning power. Finally, certain DNA sequences may have increased affinity and positioning power simply because by happenstance they make more or better bonds (with histones or solvent, net, in the process of nucleosome formation) compared to other DNA sequences. In general, positioning sequences bias nucleosome stability and dynamics (Widom, 1998).

Several studies have been performed to look for or create nucleosome positioning sequences (see Simpson, 1991; Turnell & Travers, 1992; Widlund *et al*., 1997; Widom, 1996; Lowary & Widom, 1998 and references therein). Early work (Satchwell *et al*., 1986) suggested that fragments of DNA preferentially orient themselves so that G + C-rich regions are found where the major groove is compressed (i.e. minor groove faces out, away from the histone surface) and A + T-rich regions are found where the minor groove is compressed (faces in). Based on these results, Shrader & Crothers (1989, 1990) constructed fragments with alternating A + T and G + C-rich segments in phase with the DNA helical repeat. Molecules based on this repeat formed stable nucleosomes. The A·T base-pairs are believed to be preferred where the minor groove is in contact with the core histones and the G·C base-pairs where the minor groove faces outward. Maximal nucleosome binding occurs in fragments in which these anisotropically flexible regions are repeated with a significantly shorter period than the helical repeat of free DNA. Shrader and Crothers also showed that it is possible to reproduce the binding energies of natural nucleosome positioning sequences by using short oligomers of these DNA sequences. Longer oligomers yielded fragments that are stronger histone binders than natural sequences that were known at the time. The fragment that formed one of the most stable nucleosomes in their study is the so-called TG pentamer ($5'$-TCGGTGTTAGAGCCTGTAAC)$_5$-$3'$), which is now often used in studies of chromatin structure and function *in vitro* and *in vivo*.

New rules for sequence-directed positioning continue to be discovered, suggesting that not all are known (Widom, 1996). Indeed, our recent studies (Lowary & Widom, 1998) revealed a set of new sequence rules and motifs having much greater statistical significance, and presumably much greater nucleosome positioning power, than for those already known.

Our laboratories have been engaged in systematic analyses of the rules for sequence-dependent nucleosome positioning using SELEX procedures. Widlund *et al*. (1997) selected the highest-affinity nucleosome positioning sequences from the mouse genome. The selected sequences exhibit some striking characteristics such as polar runs of phased A-tracts, phased TATA-tetrads, and also the CAG repeat correlated with neurodegenerative diseases. The TATA sequences bind the histone octamer with even higher affinity than the synthetic TG-pentamer. Lowary & Widom (1998) carried out a selection on a very large pool of (non-natural) chemically synthetic random DNA. The sequences isolated in that study form the most stable nucleosomes so far characterized, and reveal a wealth of new sequence rules and motifs that presumably

are responsible for the sequences' higher affinity. Finally Cao *et al*. (1998) carried out a negative selection experiment on non-natural DNA, to isolate sequences that have low affinity for the histone core. This study revealed a set of sequence motifs that appear to disfavor nucleosome formation. One such sequence class, the TGGA repeat, is noteworthy in that long TGGA repeats (20-100 repeats) are present in the genomes of higher eukaryotes including humans. In other recent studies, we have quantified the free energy of bulk genomic DNA compared to chemically synthetic random sequence DNA (Lowary & Widom, 1997), and we have used Fourier transform methods to analyze non-random periodic DNA sequence motifs that may be correlated with nucleosome positioning in genomic DNA (Widom, 1996).

While the studies of selected natural and non-natural nucleosome positioning sequences yield many new insights, it is especially interesting to compare the two sets of molecules. Thus the present project compares the strongest nucleosome-positioning DNA sequences selected from the entire mouse genome (Widlund *et al*., 1997) with those selected from pools of non-natural synthetic random DNA having even greater diversity (Lowary & Widom, 1998; Cao *et al*., 1998). Key goals include (i) assessing the extent to which the strongest natural positioning sequences are evolved for highest histone-binding affinity through direct side-by-side comparison of the free energies of selected natural and non-natural positioning sequences; and (ii) assessing whether the new DNA sequence rules and motifs for sequence-directed nucleosome positioning deduced from analysis of selected non-natural sequences are also features of the best natural positioning sequences. Additional technical goals include: (iii) developing improved procedures for measuring the free energy of even stronger positioning sequences that may be developed in the future; and (iv) comparing two different procedures for measuring free energies.

Results and Discussion

Direct comparison of free energies for selected natural and non-natural sequences

A primary aim of this study is to compare the relative free energies for histone binding in nucleosome reconstitution of the selected natural and non-natural sequences, in a direct side-by-side competition experiment in the identical competitive environment.

The dialysis method (see Materials and Methods) is especially suited to this goal. In this method radiolabeled tracer DNA competes with a large excess of unlabeled competitor DNA for binding to limiting amounts of histone octamer. The mixture is equilibrated initially in conditions in which histones and DNA have negligible affinity for each other and hence are in rapid exchange; subsequent slow, reversible dialysis to physiological and then even lower salt concentrations allows nucleosomes to form, equilibrate between DNA molecules and at multiple positions along each DNA molecule, ultimately ''freezing-in'' the population for analysis. The resulting nucleosomal and free DNA species are resolved by native gel electrophoresis and quantified by phosphorimager. The dialysis method facilitates a direct comparison of free energies for many sequences since both the competitive solution environment and the process by which the competition is ultimately ''frozen'' for analysis can be made literally identical for many samples simultaneously. This is accomplished as follows: (i) We produce a single reaction mixture containing limiting histone octamer, excess DNA competitor, and initial 2 M NaCl solution, in sufficient amount for use in all of the desired reactions at once. This mixture is aliquoted into different tubes, and tracer quantities of the differing radiolabeled DNA fragments to be analyzed are added. (ii) As an additional precaution, we thoroughly equilibrate the samples against the initial 2 M NaCl solution prior to dialysis down to lower [NaCl]. (iii) We place all of the dialysis buttons (inside dialysis bags) into a single vessel of dialysis buffer, so that the samples will have an identical history during reconstitution.

The results of one such experiment for a variety of the selected natural and non-natural sequence clones is illustrated in Figure 1(a) the large number of clones analyzed in any one experiment required multiple gels). Clones denoted with letters or numbers refer to the selected natural or non-natural sequences, respectively (for further information, see (Cao *et al*., 1998; Widlund *et al*., 1997; Lowary & Widom, 1998).

Most clones yield a set of shifted nucleosomal complexes, which in previous work from other laboratories have been shown to correspond to nucleosomes located at differing positions on the DNA. This is well known behavior, and occurs whenever the length of available DNA exceeds the 147 bp of the nucleosome core particle. (Evidence that these bands do not arise from dinucleosomes or other non-nucleosomal aggregates has been presented elsewhere (Lowary & Widom, 1997, 1998), and includes the comigration of the labeled reconstituted products with natural nucleosomes on sucrose gradients that are capable of resolving nucleosomes from dinucleosomes or other non-nucleosomal species.) Background-subtracted counts are obtained for a region encompassing the whole set of shifted complexes and the unshifted (free) DNA, using appropriately sized and positioned integration regions. Affinities measured in this way (or by the widely-used exchange method, see below) are macroscopic, i.e. net, over the entire molecular length, reflecting occupancy of each of the possible positions in accord with Boltzmann's law.

The results from many such experiments are summarized quantitatively in Table 1. The free

energies are expressed as differences relative to the free energy for the 5 S gene positioning sequence, which was included in each experiment as an additional sample (although it was run on a different gel). We emphasize, however, that this choice of reference does not affect the rank order or the quantitative free energy differences between any members of the sequences studied.

The procedure itself is robust despite issues such as multiple nucleosomal species, and these differences in free energy and affinity are significant. The differences in free energy between samples are large in comparison to the standard deviation for repeated measurements, ~0.3 kcal mol^{-1}. The standard deviations between measurements are small in comparison to the mean energy of thermal fluctuations, RT, ~0.6 kcal mol^{-1}, whereas the differences between samples can be large in comparison to thermal energies. The corresponding affinities differ by amounts that are often of biological significance. Finally, the robustness of the procedure is evident from the fact that different investigators,

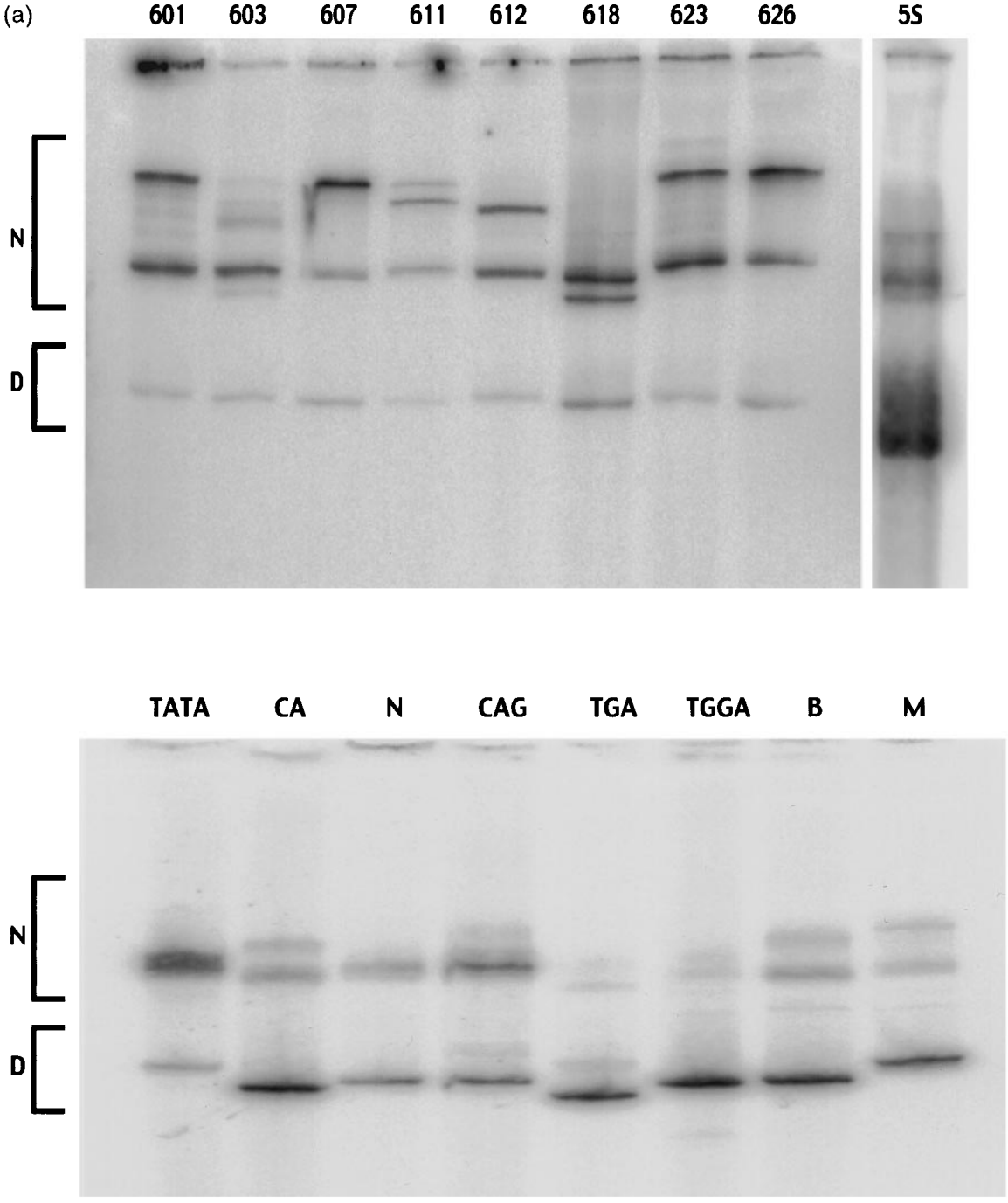

Figure 1 (*legend opposite*)

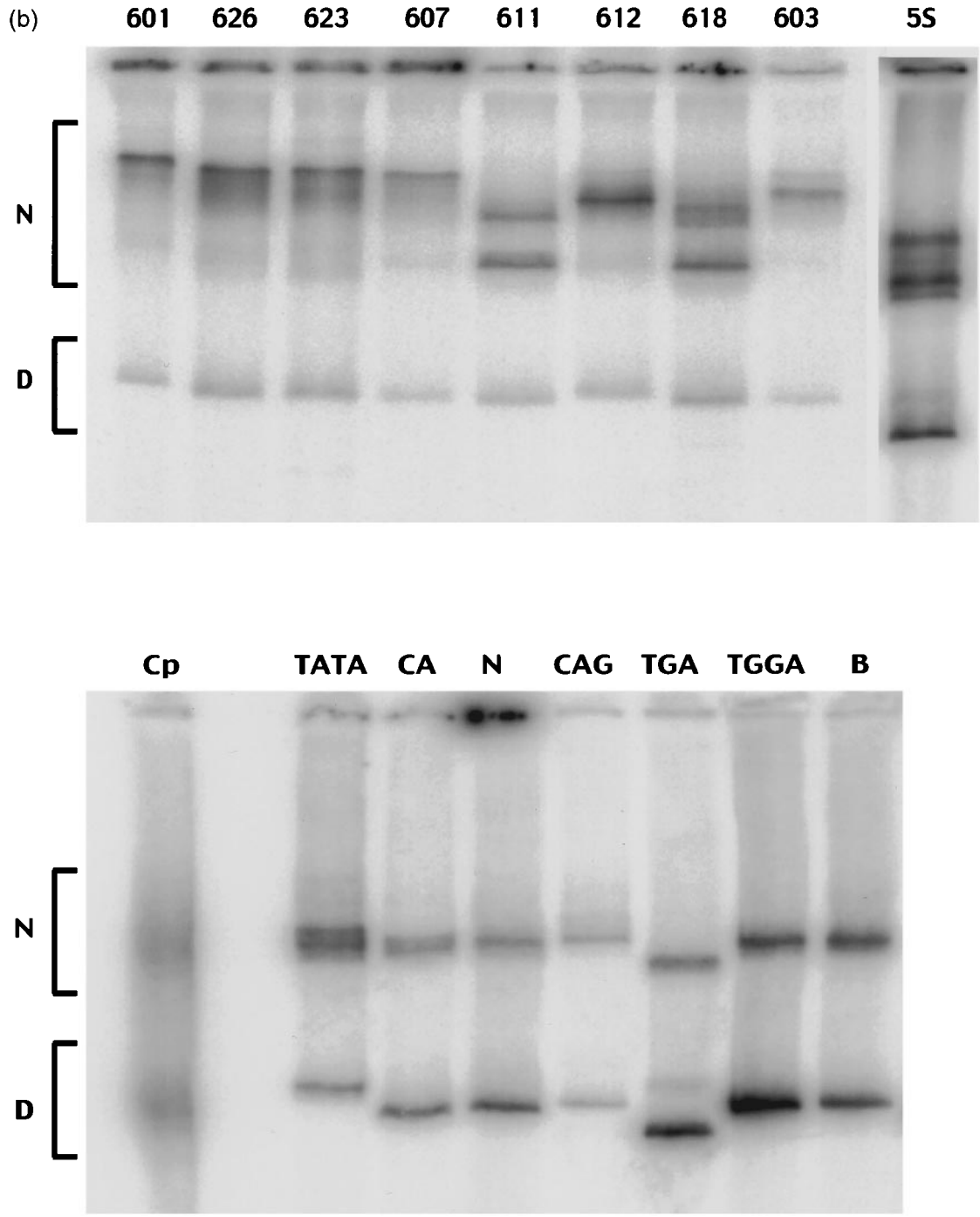

Figure 1. Competitive nucleosome reconstitution assays. Shown are phosphorimages of the products of competitive nucleosome reconstitution resolved by native gel electrophoresis. Brackets labeled D denote the mobilities of free (unincorporated) radiolabeled tracer DNA, and brackets labeled N denote the approximate range of mobilities of tracer DNA incorporated into nucleosomes. Experiments of this kind routinely yield a set of shifted nucleosomal products, corresponding to differing positions of the histone octamer along the DNA (see the text). (a) Results obtained using the dialysis method; all samples are prepared with the identical competitive solution environment. Every independent experiment additionally included control and reference samples in which the radiolabeled tracer DNA was core particle DNA (a key control, i.e. tracer = competitor; see the text) and in which the radiolabeled tracer DNA was derived from the 5 S RNA gene natural nucleosome positioning sequence (our internal reference sequence, used for comparison with other studies). For each independent experiment, analysis of all of the samples required multiple native gels; only two gels are shown. Top, various high affinity non-natural sequences (Lowary & Widom, 1998). Bottom, various high affinity natural sequences (TATA, CA, N (NoSecs), CAG, M (mouse minor satellite); Widlund *et al.*, 1997); and ''antiselected'' non-natural sequences (TGA, TGGA, B (BadSecs); Cao *et al.*, 1998). Also shown are the results obtained for the 5 S sequence; this was run at the same time as the other test sequences, with the identical competitive environment and dialysis history, but the products were analyzed on a different gel. The free energies are expressed as differences relative to the free energy for the 5 S gene positioning sequence. We emphasize, however, that this choice of reference does not affect the rank order or the quantitative free energy differences between any members of the sequences studied. (b) Results obtained using the exchange method. Lanes labeled as for (a); additionally shown are results with core particle DNA as tracer (Cp).

Table 1. Free energies measured by dialysis and exchange

Clone	$\Delta\Delta G^{o}_{D}$ (kcal mol^{-1})		$\Delta\Delta G^{o}_{E}$ (kcal mol^{-1})	
607	−2.89±0.44	($n = 5$)	−1.26±0.37	($n = 4$)
611	−2.83±0.22	($n = 6$)	−1.37±0.23	($n = 4$)
623	−2.75±0.28	($n = 4$)	−1.30±0.28	($n = 4$)
601	−2.74±0.33	($n = 6$)	−1.29±0.29	($n = 4$)
612	−2.70±0.46	($n = 6$)	−1.41±0.12	($n = 4$)
626	−2.60±0.25	($n = 6$)	−1.19±0.19	($n = 4$)
603	−2.46±0.37	($n = 4$)	−1.22±0.39	($n = 4$)
618	−2.15±0.26	($n = 6$)	−1.22±0.28	($n = 4$)
TATA	−1.82±0.29	($n = 6$)	−0.78±0.46	($n = 4$)
CAG	−0.78±0.15	($n = 6$)	−0.36±0.46	($n = 4$)
NoSecs	−0.37±0.11	($n = 6$)	0.27±0.47	($n = 4$)
L. variegatus 5 S RNA gene	(0.00)	($n = 6$)	(0.00)	($n = 4$)
BadSecs	0.29±0.11	($n = 6$)	0.07±0.41	($n = 4$)
CA	0.31±0.19	($n = 6$)	0.32±0.26	($n = 4$)
Mouse Minor satellite	0.35±0.03	($n = 3$)	(n.d.)	
TGA	1.15±0.28	($n = 6$)	0.62±0.51	($n = 4$)
TGGA	1.22±0.24	($n = 6$)	0.97±0.36	($n = 4$)

Clones are listed in order of decreasing affinity (nucleosome positioning power) as measured by the dialysis method. Free energies are reported relative to those for the 5 S sequence, which is included in every experiment and has a difference free energy defined as 0 in each experiment. Note, however, that this choice of reference does not affect the rank order or the quantitative free energy differences between any members of the sequences studies. $\Delta\Delta G^{o}_{D}$ are free energies measured using the dialysis method; $\Delta\Delta G^{o}_{E}$ are free energies measured using the exchange method. Values given are the mean ± one standard deviation (n trials). n.d., not determined. Clones are denoted as in Figure 1. The TATA sequence is the highest-affinity natural sequence, yet its affinity is much lower than that of many of the non-natural sequences.

studying different preparations of the same DNA molecules at different times, obtain results that are in good quantitative agreement (e.g. compare the present results in Table 1 with those of Lowary & Widom, 1998).

Free energy (positioning power) of highest-affinity natural sequences

Evidently, strongest nucleosome positioning is not an evolved attribute of even the strongest-positioning regions of a natural eukaryotic genome. Many of the selected non-natural sequences have a negative free energy that is ∼1 kcal mol^{-1} greater than that for the single most favorable selected natural sequence (TATA), and ∼2 kcal mol^{-1} greater than that for the second most favorable natural sequence (CAG); these free energy differences correspond to affinities that are ∼sixfold and ∼38-fold greater, respectively. Note also that non-natural sequences having even higher affinity than those discovered to date surely exist: our selection on non-natural sequences examined only 10^{-120} of the set of all possible sequences of that length.

While strongest nucleosome positioning is not an evolved attribute of even the strongest-positioning regions of a natural eukaryotic genome, nevertheless specific genomic sequences can have much greater nucleosome-positioning power than the bulk of the genome. The two strongest of the selected natural sequences, TATA and CAG, have affinities that are ∼27-fold and ∼fourfold greater than that of the well-studied 5 S gene nucleosome positioning sequence, which in turn has several-fold higher affinity than bulk genomic sequences (Widlund *et al.*, 1997). In contrast, in an earlier study we showed that >95% of bulk genomic DNA has a free energy for histone binding and nucleosome positioning that differs insignificantly (0 ± 0.3 kcal mol^{-1}) from that of arbitrary chemically synthetic random DNA sequences (Lowary & Widom, 1997).

Taken together, these studies reveal a picture of the eukaryotic genome in which the bulk of the genome has insignificant nucleosome-positioning power (at the level of individual nucleosomes), while rare specific regions have a substantially increased positioning power, although one that is modest in comparison to what is possible. It will be of great interest to correlate the genomic regions having greatest positioning power with the underlying genetic organization of the chromosome. A comparable selection experiment on the yeast genome, which may facilitate such a genetic analysis, is presently underway in our laboratory.

Fourier transform analysis of selected natural sequences

Our earlier study of selected non-natural sequences revealed a set of DNA sequence rules and motifs that were enriched for by the selection process and thus are presumed responsible for the increased affinity of these selected sequences. Some of these enriched sequence motifs were in the form of sequence periodicities. Surprisingly, the most significant periodic signal was a ∼10 bp periodic spacing of TA dinucleotides; the strength of the fundamental and harmonics in the periodic signal for TA made this the most strongly non-random periodic signal detected in that analysis. Previous studies had noted ∼10 bp sequence periodicities for various dinucleotides in natural nucleosomal DNA, but the ∼10 bp-periodic signal from TA

appeared to be insignificant in comparison to others such as those from AA (=TT), CC (=GG), or GC (Satchwell *et al.*, 1996; Ioshikhes *et al.*, 1992; Bina, 1994; Staffelbach *et al.*, 1994; Bolshoy, 1995). Moreover, in entire genomic DNA sequences, the signal from TA does not occur with statistical significance at all, whereas statistically significant signals from AA (=TT) and others are readily detected. Therefore it is particularly interesting to test whether the periodic TA signal and other new sequence motifs are specifically enriched in the highest-affinity regions of natural genomes. As described elsewhere (Widom, 1996), Fourier transformation is a particularly valuable tool for analyzing sequence signals in natural DNA.

The first step in our analysis is to create a database of sequences that contains only one species from each sequence family. By keeping only one member of each distinct sequence family, sequence motifs that are present simply by happenstance will be diminished by averaging over the entire database, whereas sequence motifs that contributed to the physical selection will be enhanced. Previous related analyses have typically considered all distinct sequences; by eliminating sequences that, while distinct, are nevertheless quite similar, our analysis is a more conservative one. The initial set of 87 distinct sequences (Widlund *et al.*, 1997) was pared down to yield a non-redundant database of 39 (see Lowary & Widom, 1998, and Materials and Methods).

Fourier analysis was carried out for each of the ten distinct dinucleotides. The experimental spectra are overlaid with the mean and standard deviation of random expectation, derived from equivalent calculations on randomized sequences. The resulting power spectra are shown in Figure 2 panels a to j. (Since both the forward and reverse-complement of each sequence are considered, the results for any particular dinucleotide are identical also for its reverse-complement.)

The spectrum for AA (=TT) shows a single large statistically significant peak at a periodicity of $\sim 0.1\ \mathrm{bp}^{-1}$ (i.e. ~ 10 bp in real-space). Such a periodicity for this dinucleotide has been discovered before in unselected natural nucleosomal

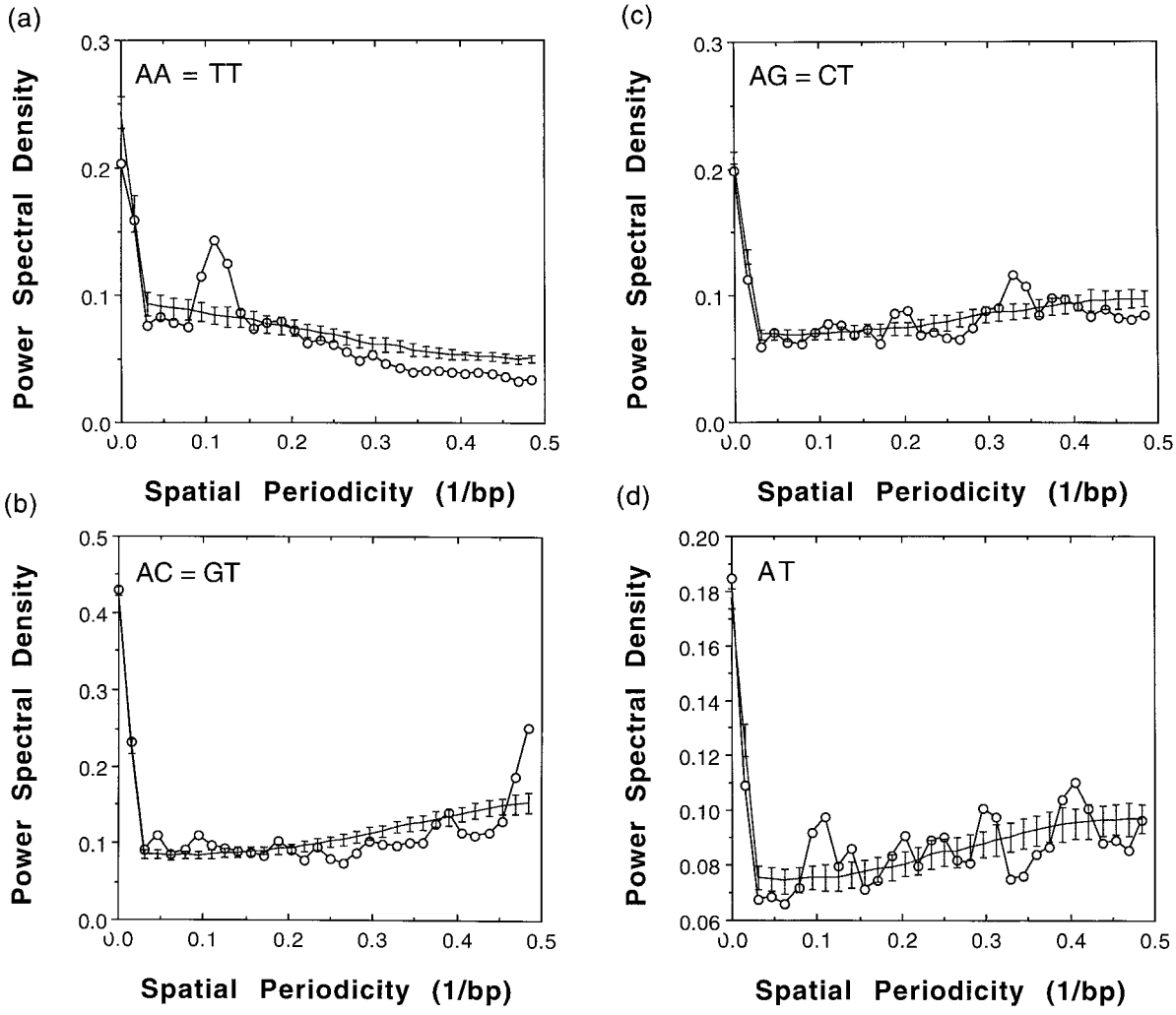

Figure 2 (*legend on page* 221)

DNA, in unbiased samples of entire genome sequences, and in the selected non-natural DNA (see Satchwell *et al*., 1986; Ioshikhes *et al*., 1992; Bina, 1994; Staffelbach *et al*., 1994; Bolshoy, 1995; Widom, 1996; Lowary & Widom, 1998 and references therein). Its presence in the selected non-natural DNA confirmed the interpretation that this periodicity for this dinucleotide contributes favorably to nucleosome formation. Interestingly, however, the present analysis of selected natural sequences, like the earlier analyses of whole genome sequences and unselected nucleosomal DNA, reveals only the single peak at $\sim$0.1 bp^{-1}, whereas the selected non-natural sequences reveal two peaks; a strong fundamental at the same $\sim$0.1 bp^{-1} periodicity, together with a strong second harmonic at $\sim$0.2 bp^{-1}. Evidently, the organization of AA (=TT) dinucleotides in the selected natural DNA is less strongly periodic than in the selected non-natural molecules. This correlates with the relatively lower affinity of the natural molecules as compared to the non-natural ones.

The spectrum for GA (=TC) reveals a weakly significant peak at $\sim$0.1 bp^{-1}, comparable to that detected in the selected non-natural sequences and in bulk genomic sequences. The presence of this same motif in the selected natural positioning sequences reinforces the conclusion that it contributes significantly to the affinity of DNA for histones in nucleosomes.

The spectra for several other dinucleotides (AG = CT, CC = GG, GC) reveal either no peaks having much statistical significance (measured as multiples of the standard deviation), or only a single peak at a periodicity of 0.33 bp^{-1}, which is attributable to codons.

Special significance of the TA step

The most important new result is for the dinucleotide TA. Our earlier Fourier transform analysis for TA steps in the selected non-natural sequences revealed a strong fundamental peak at a periodicity of $\sim$0.1 bp^{-1} together with unprecedented strong higher order harmonics at $\sim$0.2, 0.3, and

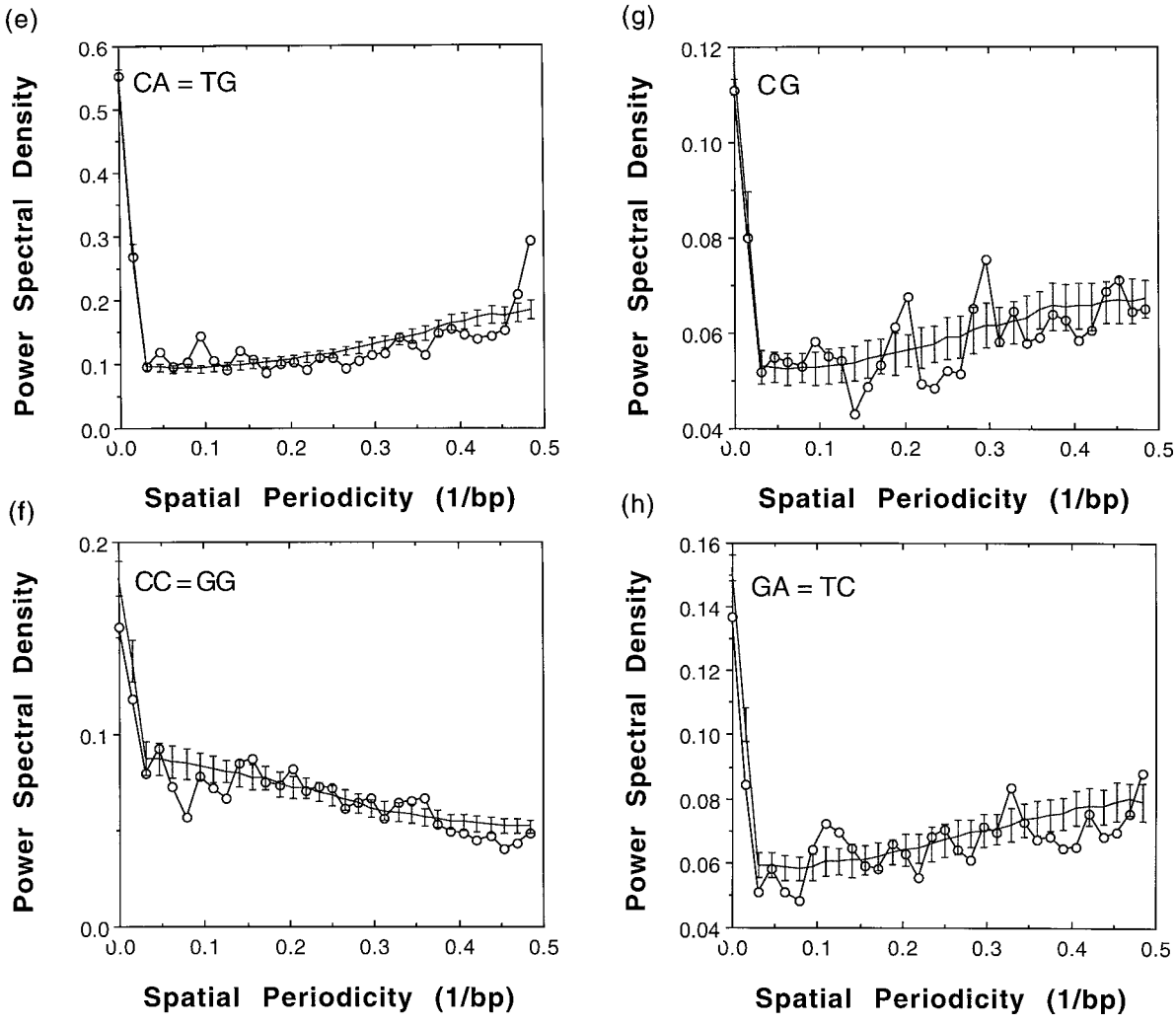

Figure 2 (*legend opposite*)

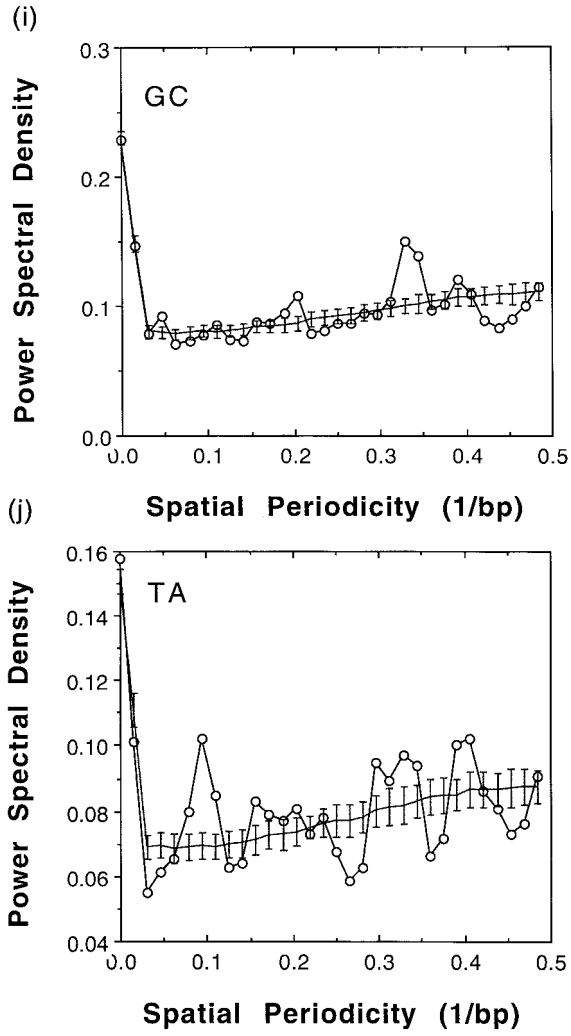

Figure 2. Fourier transform analysis (Lowary & Widom, 1998; Widom, 1996) of the selected natural sequences (Widlund *et al.*, 1997). Continuous lines and (○) datapoints: experimental transforms; fainter lines and datapoints with error bars: mean values and standard deviations for random expectation, obtained from transforms on randomized versions of the same sequences. See Materials and Methods for details. Panels a-j, results obtained for the ten distinct dinucleotide encodings. Since the reverse complements of every sequence are included in the calculations, the transform for any particular encoding is identical to that for the reverse-complement encoding.

0.4 bp^{-1}, indicative of a more highly periodic positioning of TA steps than for any other of the ten distinct dinucleotide steps. Figure 2 shows this signal to be the dominant feature also in the selected natural sequences, as measured by the fraction of the total spectral power for each dinucleotide that is attributable to the ~10 bp-periodicity fundamental plus its harmonics. The spectrum for TA reveals a strong fundamental at ~0.1 bp^{-1}, together with harmonics at 0.2 (weak), 0.3, and 0.4 bp^{-1}, very similar although with a lower signal-to-noise ratio to the results obtained on the non-natural sequences. (An additional peak at 0.33 bp^{-1} is attributable to codons, while a weak peak at ~0.15 bp^{-1} is of unknown origin and significance.) The relatively lower signal-to-noise ratio of the fundamental signal at 0.1 bp^{-1} and its higher orders from these selected natural sequences compared to the selected non-natural ones again correlates with the relatively lower affinity of the natural molecules as compared to the non-natural ones.

A potential concern is that, among the selected natural sequences, the ''phased TATA'' sequence family might dominate the Fourier transform for the TA dinucleotide. In this case the transform for TA could be revealing a feature that is unique to that sequence family, not necessarily reflecting a property that is general to selected (i.e. high affinity) nucleosome positioning sequences. Since only one such family member is represented in the non-redundant database, this possibility is unlikely. Nevertheless, we tested whether the one phased TATA sequence in the non-redundant dataset was responsible for the appearance of this signal by deleting this sequence from the database and recomputing the power spectrum for TA. The resulting spectrum was very similar (not shown), revealing the fundamental and its higher orders, very slightly weakened. Evidently, the periodic placement of TA steps is a general property common to many of the selected sequences, not just the phased TATA sequence, correlating with and responsible for their higher affinity.

This finding of a strong periodic signal for TA in the selected natural sequences leads to two important conclusions. First, it confirms the special significance of the TA step for DNA sequence-dependent nucleosome positioning that was detected in the analysis of the selected non-natural sequences. It is noteworthy that in this behavior the TA step is distinguished from other YR steps, all of which are believed to be particularly flexible (Olson *et al.*, 1998). Indeed, of the possible YR steps, the others (CA = TG and CG) are actually found to be more conformationally flexible in protein-DNA complexes than is TA. Second, it shows that regions that are enriched in ~10 bp periodic placement of TA steps, and that because of this have a relatively higher affinity for histones and stronger nucleosome positioning power, do exist in natural genomes, despite the fact that no such signal is present at a statistically significant level for this step in unbiased surveys of two entire eukaryotic genomes. Our new results imply that the absence of such a signal for TA in bulk genomic DNA correlates with (and, indeed, may be a major cause of) the fact that bulk genomic DNA has a free energy for histone binding and nucleosome positioning that differs negligibly from that of arbitrarily chosen chemically synthetic random DNA sequences.

New results for other dinucleotides

The results for other dinucleotides reveal a set of additional features that may warrant further analysis. The spectrum for AT bears a striking resemblance to that for TA, although with reduced statistical significance. A set of four peaks are observed, which can be interpreted as a fundamental at $\sim 0.1\ bp^{-1}$ together with the second, third, and fourth harmonics. In important contrast to the case for TA, no statistically significant periodicities were detected for AT in the analysis of selected non-natural DNA. However, a small but significant peak at $0.1\ bp^{-1}$ for this dinucleotide can be detected in bulk genomic sequences (Widom, 1996).

The spectrum for CA (=TG) reveals a relatively strong and statistically significant peak at $\sim 0.1\ bp^{-1}$. As for AT, a small but significant peak for this dinucleotide can be detected also in bulk genomic sequences, yet it is not significantly enriched in the selected non-natural sequences. It might be noteworthy that sequences having extensive CA repeats were one of the classes enriched in the selected natural DNA, but this seems unlikely to be related to the appearance of this peak in the power spectrum, for several reasons. Only one such sequence is retained in the non-redundant database; in any case extensive CA repeats would not lead to this periodicity; and finally extensive direct CA repeats can be equivalently viewed as repeats of the dinucleotide AC, yet no such peak is detected in the spectrum for AC (=GT-see above). Evidently this peak arises from the organization of CA dinucleotides in strings other than simple CA repeats.

The spectrum for CG reveals a pair of weakly significant peaks at ~ 0.2 and $0.3\ bp^{-1}$. These are not statistically significant features in analyses of bulk genomic sequences, nor are they found in the selected non-natural sequences. One possibility is that they could simply have arisen by happenstance, exaggerated by the modest size of the database.

Finally, the spectra for AC (=GT) and for CA (=TG) additionally reveal a peak at the limiting sampling frequency, $0.5\ bp^{-1}$. Such a peak for these dinucleotides is not a feature of the selected non-natural sequences, nor is it detected in unbiased surveys of entire eukaryotic genomes. It is likely that it arises from the extensive CA repeat sequence, which would be expected to contribute a strong peak at that periodicity for both CA and AC.

How should the existence and significance of these peaks for (AC + TG), AT and CA (=TG) (and perhaps also for CG) be understood? One plausible interpretation is as follows. The selection study on non-natural DNA represents an unbiased survey of all possible short range motifs such as these, yet these do not appear as statistically significant features. this suggests to us that they contributed relatively weakly or not at all to high affinity binding to histones and nucleosome positioning. In that case, their presence as significant features in the selected natural sequences would be attributed to their physical linkage to other sequence motifs that do contribute significantly to histone binding affinity and nucleosome positioning. In other words, natural genomes do not consist of random strings of nucleotides; rather, the requirements of biological function lead to the non-random juxtapositions or linkages of many particular sequence motifs. Physical selection of any one of these motifs will simultaneously lead to the recovery of other unselected but physically linked motifs. In accord with this interpretation, qualitative analysis of the sequences in the non-redundant data set reveals examples in which sequences that are enriched in CA dinucleotides spaced at integral multiples of 10 bp (but not in direct CA repeats) are also enriched in 10 bp periodic placement of the dinucleotide TA, which we now know to strongly contribute to histone binding affinity and nucleosome positioning (see above). Thus the sequences may have survived selection because of the presence of the correlated TA dinucleotides which confers higher affinity; the CA signal would be enriched simultaneously because these two ''traits'' are physically linked. Similarly, the signal for AT could arise as a consequence of periodic AT steps present in the same molecules as those having periodic TA steps. In any case this is only an interpretation; further studies will be required to definitively assess the significance of these motifs.

More stringent competition assay for higher-affinity sequences

A problem for the analysis of future further-improved (non-natural or natural) positioning sequences is that our existing selected sequences already have such high affinity that they are at or near the limits of our competitive assay to measure. Because of this high affinity, despite the very high concentration of competitor DNA used, essentially all of the tracer is incorporated into nucleosomes, so that the ratio [DNA in nucleosomes]/[free DNA] has enormous error in the denominator. In particular, the small amount of free DNA that exists may arise predominantly from electrophoresis-induced dissociation rather than equilibrium free DNA. A simple solution to this problem; namely, further increasing the [competitor DNA], is impractical for two reasons. First, the required concentration would soon become prohibitively high. Second, we find experimentally that when the concentration of competitor is too high (for a given histone concentration), nucleosomes fail to assemble at all (P.T.L. and J.W., unpublished). Presumably, the problem is a kinetic one, arising from the quaternary nature of the reconstitution process (DNA + 2 H2A/H2B + $H3_2H4_2$). While nucleosome formation may be energetically favorable, it may become kinetically infeasible when there are too many different mol-

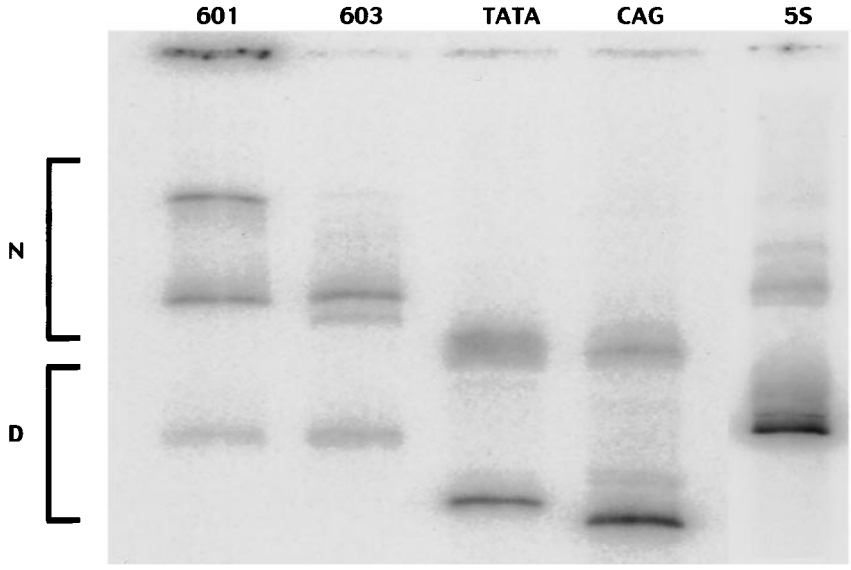

Figure 3. More-competitive dialysis assay. The 5 S gene nucleosome positioning sequence is used as the competitor, at an 18:1 molar ratio with histone octamer. See Materials and Methods and Figure 1 for other details. The more-competitive environment strongly shifts the balance of products from nucleosomes toward free DNA (compare the same clones in Figure 1(a)), allowing more-accurate measurements of the equilibrium constants and free energies for high-affinity clones, yet the relative free energies are unchanged (see the text). In this experiment the track labeled 5 S serves as both the key control (tracer-competitor) and as a reference standard for relative free energies. Relatively faint ghost bands having mobilities slightly larger than the main labeled DNA are seen in the tracks for CAG and the 5 S sequence (also in Figure 1). These are present in the tracer DNA itself, and presumably reflect a small amount of aberrant PCR product, as these samples were prepared by PCR followed by labeling with kinase, without prior purification of the desired PCR products as was done in our earlier work (Lowary & Widom, 1998). These species are omitted from the quantitative analysis and negligibly alter the results, as can be seen by the good agreement of the present results with those of Lowary and Widom.

ecules competing (kinetically) for the limiting numbers of histone subunits.

We developed an alternative approach to this problem, in which we increase the affinity (i.e. free energy) of the competitor DNA. We used purified restriction fragments containing the 5 S positioning sequence, which, until recently, was the highest affinity natural nucleosome positioning sequence known. An example of such an experiment is shown in Figure 3. In this much more competitive environment, the measured equilibrium for our present best non-natural selected sequences (defined as [DNA in nucleosomes]/[free DNA]) is ~1 (Figure 3, data not shown); thus this improved assay would sensitively detect future sequences that bind with higher affinity than our best existing sequences. Moreover, as studies in the future continue to generate further-improved molecules, one can continue to adapt this assay in parallel, simply by progressing to ever-higher affinity competitor.

The results from Figure 3 are represented quantitatively in Table 2. Two lines of evidence show that the system is behaving properly. Importantly, we find that despite the much-more competitive environment in this experiment (much lower incorporation of radiolabeled tracer into nucleosomes), the resulting difference free energies between various clones are equivalent within experimental error to those obtained using the less-competitive chicken erythrocyte nucleosome core particle DNA as competitor (Figure 1(a), Table 1). Also shown in Figure 3 is the result of a control experiment in which the radiolabeled tracer is identical to the competitor, although now both are the 5 S sequence rather than chicken erythrocyte core particle DNA as in our earlier studies. As must be the case if the system is equilibrating properly, the fraction of counts in nucleosomes *versus* free DNA agrees with that expected given the stoichiometries of histone octamer and DNA.

Comparison of dialysis and exchange methods for free energy measurement

It is of interest to compare the results obtained using the dialysis method for measurement of DNA-histone interaction free energies with the results of an alternative approach based on histone octamer exchanger. In this method, H1- (and H5-) depleted chromatin serves as the source of both histone octamer and competitor DNA. A specific DNA fragment to be analyzed is added as radiolabeled tracer; the mixture is incubated in 1 M NaCl at 37 °C, and then diluted in steps down to 0.1 M NaCl at room temperature. Subsequent analysis is identical to that used in the dialysis method. The exchange method used here is identical to that

Table 2. Free energies measured in more-competitive assay

Clone	$\Delta\Delta G^{o}_{SS}$ (kcal mol^{-1})
601	−2.89
603	−2.33
TATA	−1.44
CAG	−0.53

$\Delta\Delta G^{o}$ values measured using the 256 bp *Eco*RI fragment of the *L. variegatus* 5 S RNA sequence (Simpson & Stafford, 1983) as competitor. Results are reported relative to those obtained with a 217 bp derivative of the 5 S sequence (Protacio & Widom, 1996) used as the radiolabeled tracer for an internal standard.

used in our earlier study (Widlund *et al*., 1997), and derives from a method developed previously by Shrader & Crothers (1989, 1990). (One difference is that Widlund *et al*. used long, H1-depleted chromatin (7.5 μg) as both the competitor and histone donor, while Shrader & Crothers used defined amounts of H1-depleted chromatin (2.5 μg) as the histone donor with an excess of chicken erythrocyte DNA (20 μg) as competitor.)

The exchange method at first would appear to offer clear advantages and potential disadvantages compared to the dialysis method. Key advantages of the exchange method are: (i) It is much faster. (ii) It avoids the difficult sample handling steps associated with loading and unloading radioactive samples in microdialysis buttons. Potential disadvantage include: (i) 1 M NaCl is less than the [NaCl] required to drive dissociation of DNA from the histone octamer. While histone octamer is believed to equilibrate freely between differing DNA molecules in 1 M NaCl, the extent to which this equilibration is complete for all possible tracer DNA sequences, especially the highest-affinity sequences, is not certain. (ii) Each step of dilution is inherently non-equilibrium (irreversible); the detailed path taken by the system may vary between samples and even between molecules within the same tube.

Examples of results obtained using the exchange method are illustrated in Figure 1(b) and are summarized quantitatively along with those obtained by the dialysis method in Table 1. Several points are apparent. For the lower-affinity sequences (Table 1, bottom), the two methods are in good quantitative agreement. Strikingly, however, the results from the two methods diverge for the higher affinity sequences. The exchange method exhibits a plateau in negative free energy at around 1.3 kcal mol^{-1} (relative to the 5 S sequence), meaning that it is insensitive to differences between sequences with negative free energies (affinities) higher than that. This result implies that, for the higher affinity sequences, at least one of these two methods does not yield true equilibrium results. It raises the question of which of the two methods may be more reliable, and indeed whether either of the two methods yields results reflective of a true equilibrium.

Equilibrium measurements of relative free energies

The relative incorporation of radiolabeled tracer DNA into nucleosomes could, in principle, be governed either by equilibrium free energies or it could be kinetically determined. For the lower-affinity sequences analyzed in previous studies, several classic tests for equilibrium suggest that these experimental methods do yield equilibrium results. Apparent equilibrium constants are insensitive to times of incubation or other details of the exchange or dialysis processes, and most importantly, they do not depend on the direction from which the equilibrium is approached, a standard test for equilibrium (Shrader & Crothers, 1989, 1990; Widom, 1999). While these results establish that the system does reach a true equilibrium, there remains some question as to exactly at what salt concentration this equilibrium applies (Drew, 1991). The partitioning of histones between differing DNA molecules becomes frozen in as [NaCl] is decreased much below ~1 M but, because nucleosomes are mobile in physiological ionic conditions (Meersseman *et al*., 1992; Pennings *et al*., 1991; Varga-Weisz *et al*. 1995; Ura *et al*., 1995), the histones remain free to equilibrate along those DNA molecules even in physiological ionic conditions. Thus nucleosomes resulting from competitive reconstitutions are equilibrated at physiological ionic strength on DNA molecules that were chosen at equilibrium at somewhat higher [NaCl] (Widom, 1999). (It should also be noted that, contrary to common belief, exchange of histone octamers between differing DNA molecules occurs at least to some degree even in physiological ionic conditions; P.T.L. & J.W., unpublished. Thus it could prove to be the case that the distribution of histone octamers even between differing DNA molecules may be equilibrated at ionic conditions close to physiological.) Our new finding (for the lower affinity sequences) that the two different methods give quantitatively equivalent results adds further evidence in support of the conclusion that for such sequences the final distribution of products is an equilibrium one.

For the higher affinity sequences, at least one of the two methods used for Figure 1 and Table 1 cannot be yielding equilibrium results, since the apparent equilibrium values yielded by the two methods differ. For several reasons, we consider that the dialysis method is the one more likely to yield equilibrium values for these (apparently) high-affinity sequences.

One possibility is that, for whatever reason, perhaps when the tracer sequences have exceptionally high affinity, the histone octamers may fail to equilibrate adequately in the 1 M NaCl used for the initial equilibration in the exchange method (in contrast, our dialysis procedure starts at 2 M NaCl). This possibility was tested by carrying out the exchange procedure starting at 2 M NaCl instead of 1 M; however the results using this altered exchange procedure did not differ from those using the original 1 M procedure (H.R.W., unpublished).

An alternative possibility is that histone transfer to the fragments with very high affinity occurs less efficiently than expected thermodynamically. It is plausible that the affinity for histone octamers of the histone source increases when the source is depleted of histone octamers, i.e. the source has a minor population of less tightly bound histone octamers that are readily transferred, while a major population is more tightly bound. Considering that the histone octamer transfer takes place at a rather high ionic strength, H1-depleted chromatin is

expected to be highly condensed (Widom, 1998), possibly aggregated. Perhaps the fraction that is readily transferred is present in less condensed or aggregated short fibers, while most of the material is firmly bound; as the more favorable histone octamer source runs out of available octamers, the ability of the system to supply further octamer as needed by the remaining tracer DNA would be diminished. This would affect mainly the high-affinity DNA sequences because these will bind more of the histone octamer. In any case, the two methods differ chiefly in two respects: the histone source and the conditions for transfer. One or both of these must be significant, and our data appear to favor the former.

Other reasons suggesting that the dialysis method is likely to be yielding equilibrium values even for the high-affinity sequences include: (i) it involves a gradual change of state from the initial strongly dissociative conditions to the final strongly stabilizing conditions; this gradual change of state is expected to be literally always at quasi-equilibrium. (In contrast, the exchange method involves a set of non-equilibrium rapid salt concentration jumps.) (ii) It correctly yields the expected free energy equal to zero when the tracer is identical to the competitor. (iii) It yields relative free energies that are invariant with the concentration or affinity of the competitor; i.e. that do not vary regardless of whether the majority of the tracer is incorporated into nucleosomes (low-affinity competitor) or whether very little tracer is incorporated (high-affinity competitor) (Tables 1 and 2).

In summary, for the lower-affinity sequences, the results obtained with the exchange method of free energy measurement are in agreement with those obtained by dialysis. Considerable data, including the agreement between the two methods, support the conclusion that both experiments yield true equilibrium free energies. For the highest-affinity sequences, the results obtained from the two methods differ; but the available data suggest that the dialysis method likely yields true equilibrium free energies, and thus is to be preferred.

Conclusions

We have quantified the energetic significance of those sequences in the mouse genome having the highest affinity for histone binding and (equivalently) the strongest capability for DNA sequence-directed nucleosome positioning. The strongest natural positioning sequences have affinities that are ~sixfold or more lower than those possessed by many of the selected non-natural sequences. Evidently, even the strongest positioning regions of eukaryotic genomes are not evolved for the highest-possible affinity and nucleosome positioning power.

In earlier studies we showed that the bulk of eukaryotic genomic DNA (>95%) contributes negligibly (0 ± 0.3 kcal mol^{-1}; Lowary & Widom, 1997) to histone binding and nucleosome positioning (at the level of individual nucleosomes), yet strong signals attributable to roles in nucleosome positioning (Lowary & Widom, 1998) are readily detectable encoded in genomic DNA (Widom, 1996). Those findings implied that a small subset of eukaryotic genomic DNA is evolved or constrained for roles in nucleosome positioning. Our SELEX experiment on the mouse genome (Widlund *et al*., 1997) identifies these sequences and allows their properties to be investigated. Taken together, these findings imply that rare specific genomic regions have a positioning power that is substantially increased compared to bulk genomic sequences, yet modest in comparison to what is possible. Such free energies cannot yield ''precise'' nucleosome positioning (Lowary & Widom, 1997), yet may contribute in essential ways to gene regulation when this is considered as a problem of coupled chemical equilibria (Polach & Widom, 1995; Polach & Widom, 1996; Widom, 1998).

Fourier transform calculations on the selected natural sequences reveal a special significance for nucleosome positioning of a motif consisting of ~10 bp periodic placement of TA dinucleotide steps. In earlier studies of unselected natural nucleosomal DNA, dinucleotide steps such as AA (=TT) and CC (=GG) were found to be the most strongly periodic at this wavelength (Satchwell *et al*., 1986; Ioshikhes *et al*., 1992; Bina, 1994; Staffelbach *et al*., 1994; Bolshoy, 1995). In contrast, we find that, in the selected highest-affinity natural sequences, the TA step is the most strongly periodic (measured as the fraction of the total power present in the 10 bp period fundamental or its higher-order harmonics). A special significance for the TA step was first revealed in our selection experiment on non-natural DNA (Lowary & Widom, 1998). Surprisingly, such a signal was only weakly detected in studies of DNA sequences present in natural nucleosomes, and it not statistically significant in bulk genomic DNA. Evidently, the strength (Fourier amplitude) of this signal correlates directly with the affinity of the sequences.

An important related conclusion is that all ''YR'' steps are not equivalent for histone binding and nucleosome formation, despite all having greater than average flexibility as measured by the dispersion of DNA helical parameters such as twist and roll angles for these steps in large databases of solved structures (Olson *et al*., 1998). In both the selected natural and non-natural sequences, the periodic signal from TA greatly exceeds that from CA (=TG) or CG. While the mechanism through which periodic placement of TA steps may contribute to histone-binding affinity is not known, the available data suggest that unusual DNA structures or dynamics may play a prominent role (Lowary & Widom, 1998).

We report the development of a new, more-competitive procedure for measuring free energies in nucleosome reconstitution. This procedure should

be continuously adaptable to ever-higher-affinity sequences, as these are developed in future studies.

Finally, we show that when the favorable free energy of histone-DNA interactions becomes sufficiently large, measurements based on the widely-used exchange method become unreliable.

Materials and Methods

Preparation of DNA, chromatin, and histones

Plasmids were prepared using a Quantum Prep Plasmid Miniprep Kit (BioRad). 1.3 μg of plasmid was digested with *Bam*HI and *Eco*RI, chloroform-phenol extracted, ethanol precipitated, and dissolved in 0.1 × TE (TE is 10 mM Tris-HCl (pH 7.6), 1 mM EDTA, pH 8.0). The samples were subjected to 15 cycles of PCR in 100 μl reactions containing 2 mM dNTPs, 5 μM each primer, 50 nM Tris-HCl (pH 9.0 at 25°C), 20 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, and 20 units Tfl DNA polymerase. PCR was repeated as needed. Amplified fragments were purified in 1.5% (w/v) agarose gels, extracted using a DNA Extraction Kit (Amicon), ethanol precipitated dissolved in 0.1 × TE. DNA was labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP.

Histone octamer was purified as described (Feng *et al.*, 1993). H1- (and H5-) depleted chromatin for use as competitor and histone donor was purified as described (Widlund *et al.*, 1997).

Dialysis method for measurement of histone-DNA interaction free energy in nucleosome reconstitution

In this method (Lowary & Widom, 1997, 1998) radiolabeled tracer DNA competes with a large excess of unlabeled competitor DNA for limiting amounts of histone octamer. Competitive reconstitution is achieved by gradual dialysis, using a double-dialysis procedure, with a dialysis ''button'' inside a dialysis bag to ensure that the dialysis is gradual (i.e. a reversible, quasi-equilibrium process). The dialysis starts from a solution containing 2 M NaCl, in which histones and DNA have negligible affinity for each other. As dialysis proceeds, the [NaCl] sweeps slowly through conditions in which histone-DNA interactions become increasingly favorable. Nucleosomes form and equilibrate among the possible DNA molecules and the possible positions along those molecules. As the [NaCl] drops to a final value that is well below physiological, nucleosomes become strongly stabilized and increasingly immobile, ultimately freezing-in the distribution of products that had equilibrated at somewhat higher [NaCl]. Evidence that this procedure results in authentic nucleosomes, and that the system does in fact achieve and then freeze-in an equilibrium distribution, is summarized in our earlier papers (Lowary & Widom, 1997, 1998). (Note also that a closely analogous procedure was used to create the nucleosomes which were used for the determination of the high resolution crystal structure of the nucleosome (Luger *et al.*, 1997a,b).

The products of competitive reconstitutions are separated by native polyacrylamide gel electrophoresis and quantified by phosphorimager. The ratio of radiolabeled tracer DNA incorporated into nucleosomes to free tracer DNA defines an equilibrium constant and a corresponding free energy for histone binding of the tracer DNA, which are valid for that specific competitive environment. Difference free energies can be obtained by subtraction of the free energies for differing radiolabeled tracer molecules. For comparison with other studies, it can be useful to report free energies as differences relative to the free energy of a particular molecule that is used in common by many laboratories; however any arbitrary sequence may be chosen as a reference without affecting the magnitude (or the rank order) of the relative affinities. In general we include a sequence derived (Protacio & Widom, 1996) from the well-studied sea urchin 5 S RNA gene nucleosome positioning sequence (FitzGerald & Simpson, 1985; Simpson & Stafford, 1983; Simpson *et al.*, 1985) for use as a reference, since this sequence has been widely used by many laboratories.

An important control that we routinely carry out in these experiments is a reaction in which the radiolabeled tracer DNA is identical to the competitor. In this case the tracer has an exactly-known free energy relative to the competitor, which is necessarily equal to zero. Thus if the system is equilibrating properly, the ratio of tracer in nucleosomes to free DNA will equal the molar ratio of histone octamer to competitor DNA present in the experiment. Using the dialysis method we find that this expected result is in fact obtained (Lowary & Widom, 1997, 1998; and see below).

For the studies reported here, tracer amounts of radiolabeled fragments were mixed with 2 μg histone octamer, 37.5 μg chicken erythrocyte core particle DNA, 2 M NaCl, 0.5 × TE, 0.5 mM PMSF and 1.0 mM BZA in a 50 μl reaction. NaCl, buffer, protease inhibitors, histone octamer, and competitor DNA were first combined in a single ''mother'' reaction mixture, so that the identical competitive environment would exist for all of the samples at once; appropriate volumes were aliquoted out, and radiolabeled tracer added. Samples were loaded into microdialysis buttons (the dialysis membrane is carefully equilibrated in the reconstitution buffer prior to assembly of the buttons), placed inside another dialysis bag, equilibrated in the starting buffer (with its 2 M NaCl) and then slowly dialyzed down to 0.5 × TE (plus PMSF and BZA) as described (Lowary & Widom, 1998). Reconstitution reactions were repeated multiple times on different days.

Exchange method for measurement of histone-DNA interaction free energy in nucleosome reconstitution

For reconstitution by exchange (Widlund *et al.*, 1997), both the histone octamer and competitor DNA are supplied together in the form of H1- (and H5-) depleted long chromatin. Radiolabeled tracer DNA is added to the H1-depleted chromatin in a moderate [NaCl] that allows exchange of histone octamer from the donor chromatin onto the tracer DNA. After incubation for a period of time at elevated temperature (which, in addition to the elevated [NaCl] acts to facilitate the equilibration of histone octamer between donor chromatin and tracer DNA), the system is diluted in steps to lower [NaCl] at room temperature, again freezing-in the equilibrium. The products of the competitive reconstitution reactions are analyzed as described below.

Tracer amounts of labeled DNA fragment were dissolved in 2.5 μl of 4 × reconstitution buffer (4 × buffer is: 4 M NaCl, 40 mM Tris (pH 7.6), 0.4% (v/v) Nonidet P-40 and 100 μg/ml bovine serum albumin (BSA)). H1-depleted chromatic (7.5 μg in 7.5 μl) was added to yield a final concentration of 1 M NaCl in a final volume of 10 μl. The samples were incubated at 37°C for 30 minutes and then diluted to 0.1 M NaCl by three stepwise

additions of 30 μl volumes of low-salt buffer (20 mM Tris, 0.1% Triton X-100 and 100 μg/ml BSA) at room temperature with 20 minutes between additions (Widlund *et al*., 1997). This procedure results in a final [NaCl] of 0.1 M. Reconstitution reactions were repeated multiple times on different days.

Test of the reconstitution procedure

A key test for proper behavior of the reconstruction procedures is to verify that, when the tracer DNA is identical to the competitor, the fraction of tracer incorporated into nucleosomes is equal to the molar ratio of histone octamer to total DNA (competitor + tracer, ≈competitor) present (Lowary & Widom, 1997, 1998). The equilibrium dialysis procedure allows this exact test to be carried out, simply by using a small amount of the very same competitor DNA sample as the labeled tracer. for the exchange method, the tracer DNA will differ slightly from the competitor. Both are prepared from the same cell type and in principle contain the same representation of genomic sequences, but the tracer needs to be relatively short (e.g. bulk nucleosome core particle DNA) so as to give a defined band in nucleosome reconstitution, whereas as performed here, the competitor DNA and histone donor are in the form of longer chromatin fragments.

Analysis

Independent of which procedures were used for the reconstitutions, all samples were analyzed in 5% native gels (acrylamide to biscarylamide, 30:1) containing 1/3 × TBE (30 mM Tris-borate, 0.66 mM EDTA). The gels were prerun for one hour at room temperature at 150 V. Samples were loaded in 3% Ficoll and TE, while running. Gels were run at room temperature at 150 V, dried, and quantified by phosphorimager. Background-subtracted counts are obtained for a region encompassing the whole set of shifted complexes and the unshifted (free) DNA, using appropriately sized and positioned integration regions. This procedure is robust, as can be seen from the fact that different investigators, studying different preparations of the same DNA molecules at different times, obtain results that are in good quantitative agreement (e.g. compare the present results in Table 1 with those of Lowary & Widom, 1998). Equilibrium constants (K_{eq}) were calculated from the ratio of (background corrected) counts in the set of nucleosomal products to (background corrected) counts in free DNA. ΔG° values were calculated from $\Delta G^{\circ} = -RT \ln(K_{eq})$ (R is the gas constant, T the absolute temperature).

With both methods, for comparison with other studies it is useful to compute $\Delta\Delta G^{\circ}$ values as differences between ΔG° for any particular clone relative to an internal reference 5 S sequence ($\Delta\Delta G^{\circ} = \Delta G^{\circ}_{clone} - \Delta G^{\circ}_{5S}$). Note, however, that the choice of (or even the use of) an internal reference does not change the relative free energies between any pairs of clones, nor the ranked order of affinities, measured in a given experiment. We used a sequence derived from the sea urchin (*Lytechinus variegatus*) 5 S RNA gene (Protacio & Widom, 1996) as a reference in parallel reconstitution reactions prepared with the identical buffer plus competitor mix. All free energies are reported relative to this reference molecule.

More-stringent method for analysis of higher-affinity sequences

We developed a new, more stringent procedure based on the dialysis method which should prove useful for analysis of future even-higher affinity sequences. The procedure duplicates our dialysis method (see above) except that we use only 0.25 μg histone octamer and include 4.86 μg of the *L. variegatus* 5 S RNA gene (256 bp *Eco*RI fragment; Simpson & Stafford, 1983) as a stronger competitor in each 50 μl reconstitution reaction. Analogous to our other dialysis procedure, proper behavior of the reconstitution procedure for these more-stringently competitive reactions is verified by running parallel reactions that incorporate a small amount of radiolabeled 5 S DNA as the tracer. We find that the fraction of labeled tracer in this case (i.e. when tracer equals competitor) is that expected given the molar ratio of histone octamer to unlabeled 5 S DNA supplied.

Fourier transform analysis

An initial database of 87 selected genomic sequences (Widlund *et al*., 1997) was screened as described (Lowary & Widom, 1998) using the GCG program GAP to identify and eliminate closely related sequences. The natural sequences reveal a more-continuous distribution of relatedness than do the non-natural sequences, for which the statistics of the exceptionally diverse synthesis allow a sharp distinction between molecules that derive from a common parent and those that do not. For the selected natural sequences, we chose a cutoff for relatedness that pared the database down to 39 files. In this non-redundant database the two most-closely related files retain greater than random homology, 62% identity in a central 131 bp region, *versus* 25% identity expected by happenstance, yet they are also plainly from two very distinct genomic regions. Their similarity may chiefly reflect similar features that allowed both to be enriched for during the physical selection for highest histone-binding affinity. For each of the ten distinct dinucleotide steps, all sequences in the non-redundant dataset and their reverse-complements were encoded 1 at positions corresponding to the first nucleotide of that step, and 0 elsewhere (i.e. the string …GAAAAT… would be represented as …011100… for the dionucleotide AA).

Calculations were carried out using the same programs used in our study of selected non-natural nucleosome positioning sequences (Lowary & Widom, 1998). Power spectra were calculated using the program SPCTRM (Press *et al*., 1986) in consecutive half-overlapping 64 bp-long segments, and were summed for all segments from the encoded sequence and its reverse-complement, for all sequences in the non-redundant dataset. Hence, results for any particular dinucleotide are identical also for the reverse-complement of that dinucleotide. Additional calculations and checks establishing the validity of these computer programs are discussed in an earlier paper (Widom, 1996).

To evaluate the statistical significance of the transforms, power spectra were calculated as above in 100 separate runs over randomized versions of the non-redundant dataset, and the mean and standard deviation in the power spectral density from the 100 random runs were evaluated at each reciprocal lattice point. To create the randomized sequences, the number of occurrences of a particular dinucleotide within each real sequence was

counted, and then an equal number were placed at random locations chosen by the random number generator RAN1 (Press *et al*., 1986), with proper handling of permitted and forbidden adjacencies. Both the ''forward'' and reverse-complement of each randomized sequence were included in the calculations. Evidence that these computer programs work as intended, and that the statistical analysis is appropriate, is discussed in our earlier papers (Widom, 1996; Lowary & Widom, 1998).

Acknowledgments

We thank Jeff Anderson, Gu Li, Kevin Polach, and Reine Protacio for numerous helpful discussions. The computer programs PILEUP and GAP are provided in the Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI. We acknowledge with gratitude the use of instruments in the Keck Biophysics Facility, which was established with a generous grant from the W.M. Keck Foundation. J.W. acknowledges research support from the NIH.

References

Bina, M. (1994). Periodicity of dinucleotides in nucleosomes derived from simian virus 40 chromatin. *J. Mol. Biol.* **235**, 198-208.

Bolshoy, A. (1995). CC Dinucleotides contribute to the bending of DNA in chromatin. *Nature Struct. Biol.* **2**, 446-448.

Cao, H., Widlund, H. R., Simonsson, T. & Kubista, M. (1998). TGGA-repeats impair nucleosome formation. *J. Mol. Biol.* **281**, 253-260.

Drew, H. R. (1991). Can one measure the free energy of binding of the histone octamer to different DNA sequences by salt-dependent reconstitution? *J. Mol. Biol.* **219**, 391-392.

Feng, H.-P., Scherl, D. S. & Widom, J. (1993). Lifetime of the histone octamer studied by continuous-flow quasielastic light scattering: test of a model for nucleosome transcription. *Biochemistry*, **32**, 7824-7831.

FitzGerald, P. C. & Simpson, R. T. (1985). Effects of sequence alterations in a DNA segment containing the 5 S RNA gene from *Lytechinus variegatus* on positioning of a nucleosome core particle *in vitro*. *J. Biol. Chem.* **260**, 15318-15324.

Fragoso, G., John, S., Roberts, M. S. & Hager, G. L. (1995). Nucleosome positioning on the MMTV LTR results from the frequency-biased occupancy of multiple frames. *Genes Dev.* **9**, 1933-1947.

Ioshikhes, I., Bolshoy, A. & Trifonov, E. N. (1992). Preferred positions of AA and TT dinucleotides in aligned nucleosomal DNA sequences. *J. Biomol. Struct. Dynam.* **9**, 1111-1117.

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

Lowary, P. T. & Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19-42.

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

Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M. Y. & Richmond, T. J. (1997b). Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **272**, 301-311.

Meersseman, G., Pennings, S. & Bradbury, E. M. (1992). Mobile nucleosomes - a general behavior. *EMBO J.* **11**, 2951-2959.

Olson, W. K., Gorin, A. A., Lux, X.-J., Hock, L. N. & Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA*, **95**, 11163-11168.

Pennings, S., Meersseman, G. & Bradbury, M. E. (1991). Mobility of positioned nucleosomes on 5 S rDNA. *J. Mol. Biol.* **220**, 101-110.

Polach, K. J. & Widom, J. (1995). Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J. Mol. Biol.* **254**, 130-149.

Polach, K. J. & Widom, J. (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J. Mol. Biol.* **258**, 800-812.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986). *Numerical Recipes*, Cambridge University Press, Cambridge.

Protacio, R. U. & Widom, J. (1996). Nucleosome transcription studied in a real-time synchronous system: test of the lexosome model and direct determination of effects due to histone octamer. *J. Mol. Biol.* **256**, 458-472.

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

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

Shrader, T. E. & Crothers, D. M. (1990). Effects of DNA sequence and histone-histone interactions on nucleosome placement. *J. Mol. Biol.* **216**, 69-84.

Simpson, R. T. (1991). Nucleosome positioning: occurrence, mechanisms, and functional consequences. *Prog. Nucl. Acid Res. Mol. Biol.* **40**, 143-183.

Simpson, R. T. & Stafford, D. W. (1983). Structural features of a phased nucleosome core particle. *Proc. Natl Acad. Sci. USA*, **80**, 51-55.

Simpson, R. T., Thoma, F. & Brubaker, J. M. (1985). Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: a model system for study of higher order structure. *Cell*, **42**, 799-808.

Staffelbach, H., Koller, T. & Burks, C. (1994). DNA structure patterns and nucleosome positioning. *J. Biomol. Struct. Dynam.* **12**, 301-325.

Turnell, W. G. & Travers, A. A. (1992). Algorithms for prediction of histone octamer binding sites. *Methods Enzymol.* **212**, 387-400.

Ura, K., Hayes, J. J. & Wolffe, A. P. (1995). A positive role for nucleosome mobility in the transcriptional activity of chromatin templates: restriction by linker histones. *EMBO J.* **14**, 3752-3765.

Varga-Weisz, P. D., Blank, T. A. & Becker, P. B. (1995). Energy-dependent chromatin assembly and nucleosome mobility in a cell-free system. *EMBO J.* **14**, 2209-2216.

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

Widom, J. (1996). Short-range order in two eukaryotic genomes: relation to chromosome structure. *J. Mol. Biol.* **259**, 579-588.
Widom, J. (1998). Structure, dynamics, and function of chromatin *in vitro*. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 285-327.
Widom, J. (1999). Equilibrium and dynamic nucleosome stability. *Methods Mol. Biol.* in the press.
Yao, J., Lowary, P. T. & Widom, J. (1993). Twist constraints on linker DNA in the 30 nm chromatin fiber: implications for nucleosome phasing. *Proc. Natl Acad. Sci. USA*, **90**, 9364-9368.

Edited by T. Richmond

(Received 10 August 1998; received in revised form 4 March 1999; accepted 8 March 1999)