

# Nucleosome packaging and nucleosome positioning of genomic DNA

(chromatin/gene regulation/transcriptional activation)

P. T. LOWARY\* AND J. WIDOM\*†‡

Departments of \*Biochemistry, Molecular Biology, and Cell Biology, and †Chemistry, Northwestern University, Evanston, IL 60208-3500

Communicated by Roger D. Kornberg, Stanford University School of Medicine, Stanford, CA, December 18, 1996 (received for review July 18, 1996)

**ABSTRACT** The goals of this study were to assess the extent to which bulk genomic DNA sequences contribute to their own packaging in nucleosomes and to reveal the relationship between nucleosome packaging and positioning. Using a competitive nucleosome reconstitution assay, we found that at least 95% of bulk DNA sequences have an affinity for histone octamer in nucleosomes that is similar to that of randomly synthesized DNA; they contribute little to their own packaging at the level of individual nucleosomes. An equation was developed that relates the measured free energy to the fractional occupancy of specific nucleosome positions. Evidently, the bulk of eukaryotic genomic DNA is also not evolved or constrained for significant sequence-directed nucleosome positioning at the level of individual nucleosomes. Implications for gene regulation *in vivo* are discussed.

DNA in nucleosomes is tightly bent in comparison to its persistence length, a length scale of DNA stiffness (1). Substantial mechanical work must be done against the bending stiffness to package DNA in nucleosomes; this mechanical work is done at the expense of the favorable free energy of histone–DNA interactions and subtracts from the net thermodynamic stability of the particles. The free energies involved are surprisingly large. Taking the persistence length of DNA to be 50 nm and assuming that DNA in a nucleosome is uniformly bent in a circular trajectory at a radius of curvature of 4.4 nm leads to an estimated free energy cost (1) for bending DNA in a nucleosome of  $\approx 75$  kcal·mol<sup>-1</sup>. Indeed, because of previous estimates such as this, it was once considered remarkable that nucleosomes exist at all. Analogous problems exist for DNA twist.

The discovery that certain DNA sequences are naturally bent (2) suggested that eukaryotic genomic DNA sequences might be evolved or constrained to facilitate their packaging in nucleosomes (3, 4). Moreover, one might anticipate the existence of a relationship between the free energy of nucleosomal packaging and the phenomenon of nucleosome positioning (such an equation is developed in the present study), suggesting the possibility that nucleosome positioning signals may be encoded in genomic DNA. This idea is interesting in part because of the relationship between nucleosome positioning and gene regulation (5).

Results from several studies support the hypothesis that genomes are evolved to facilitate their own packaging. (i) Analyses of DNA fragments present in isolated nucleosome core particles, chromatosomes, and dinucleosomes reveal nonrandom periodic distributions of certain dinucleotides and longer sequences, with particular relative phases (see ref. 6 and

references therein). The results suggest that such sequences facilitate the bending of DNA that is necessary for nucleosome formation and that the bending may be produced by an anisotropic flexibility of the A+T-rich and G+C-rich regions. (ii) Shrader and Crothers (7, 8) designed sequences *de novo* that obeyed these rules. The designed sequences were found to bind histones (and form nucleosomes) with particularly high affinity. Moreover, they confer “rotational,” but not “translational,” positioning, consistent with the proposed explanation of their high affinity. (iii) Sets of sequences that have been found by chance to be organized in preferentially positioned nucleosomes *in vivo* have been analyzed (9–12) and are found to have nonrandom periodic distributions of certain dinucleotides, related to (although different in detail from) those discovered in the sequences of isolated nucleosomal DNA. (iv) Finally, a recent study using Fourier transformation to analyze data emerging from the *Saccharomyces cerevisiae* and *Caenorhabditis elegans* genome sequencing projects reveals the periodicities detected in those other studies of selected sequences and shows other dinucleotides to also be nonrandomly periodic at similar wavelengths (13).

While these results collectively suggest that genomic DNA sequences do indeed participate in their own nucleosome packaging and in nucleosome positioning, they leave open the question of what fraction of the genomic DNA may have this capability and how large the energetic contribution is.

The goal of the present study was to assess the extent to which bulk genomic DNA sequences contribute to their own packaging in nucleosomes and to nucleosome positioning. We measured the relative free energies for DNA binding to histone octamer for genomic DNA fragments and for randomly synthesized DNA, using a competitive nucleosome reconstitution assay, and we developed an equation that relates the free energy measured in nucleosome reconstitution assays to the fractional occupancy of specific nucleosome positions *in vitro*. The implications of our results for gene regulation *in vivo* are discussed.

## MATERIALS AND METHODS

**Random DNA.** The original pool of synthetic random DNA was the generous gift of D. Bartel and J. Szostak (14) and has been extensively characterized by them. We received 2  $\mu$ g of a 294-bp double-stranded DNA fragment constructed as follows: T7 promotor–L<sub>22</sub>–N<sub>72</sub>–StyI–N<sub>76</sub>–BanI–N<sub>72</sub>–R<sub>20</sub>, where the sequence of L<sub>22</sub> is 5'-ggaacacatccgactggcacc-3' and the sequence of R<sub>20</sub> is 3'-ggaaccagtaatcctagggc-5'. N<sub>72</sub> and N<sub>76</sub> represent segments of synthetic random DNA of 72 or 76 bp; these are joined together by two 6-bp defined restriction sequences, StyI (5'-ccaagg-3') and BanI (5'-ggcacc-3'). We used the primers L<sub>30</sub> (5'-gggagctcggaacacatccgactggcacc-3',

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Copyright © 1997 by THE NATIONAL ACADEMY OF SCIENCES OF THE USA  
0027-8424/97/941183-6\$2.00/0  
PNAS is available online at <http://www.pnas.org>.

‡To whom reprint requests should be addressed at: Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, 2153 Sheridan Road, Evanston, IL 60208-3500. e-mail: j-widom@nwu.edu.

which deletes the T7 promoter) and R<sub>20</sub> (above) to amplify molecules.

We subjected 1.6  $\mu\text{g}$  of the random sequence DNA to eight cycles of PCR in 50 mM Tris-HCl (pH 9.0 at 25°C), 20 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1.5 mM MgCl<sub>2</sub>, 0.2 mM each dNTP, 0.5 mM each primer, and 50 units/ml of *Tfl* DNA polymerase (Epicentre Technologies, Madison, WI). PCRs were pooled and purified as described (15, 16). Our final yield of DNA was 132  $\mu\text{g}$ , corresponding to  $\approx 82$  copies each of  $5 \times 10^{12}$  different molecules.

Molecules were cloned into the *HincII* site of the plasmid pGEM 3Z (Promega). When desired, individual molecules were amplified from plasmid DNA either using the PCR conditions described above, followed by labeling with T4 polynucleotide kinase, or by including [ $\alpha$ -<sup>32</sup>P]dATP in the PCRs. Labeled fragments were purified on 5% acrylamide gels.

**Genomic DNA and Histones.** Bulk chicken erythrocyte chromatin having a range of lengths from  $\approx 200$  bp to 10 kbp (weight average,  $\approx 1$  kbp) was prepared as described (17), and the DNA was extracted and digested with the restriction enzyme *RsaI* to a weight average length of  $\approx 400$  bp. Nucleosome core particle DNA (18) and histone octamer (19) were prepared from chicken erythrocyte nuclei as described.

**Analytical Reconstitutions.** 20  $\mu\text{g}$  of PCR-amplified random DNA (including trace amounts of radiolabeled material) plus 1.15  $\mu\text{g}$  of histone octamer were mixed in 250  $\mu\text{l}$  of TE (10 mM Tris/1 mM EDTA, pH 7.5) plus 2.0 M NaCl [all buffers contain 0.5 mM phenylmethylsulfonyl fluoride (PMSF) and 2 mM benzamidine]. The mixture was dialyzed for 2 hr into TE plus 2.0 M NaCl, followed by 2 hr each of dialysis into TE plus 1.5, 1.0, and 0.5 M NaCl, successively, and finally dialysis overnight into 0.5 $\times$  TE. Samples were loaded onto linear 5–30% sucrose gradients (in 0.5 $\times$  TE) and spun for 24 hr at 41,000 rpm at 4°C in a Beckman SW-41 rotor. Gradients were fractionated (by pumping from the bottom) into  $\approx 0.5$ -ml fractions and Cerenkov-counted.

**Competitive Reconstitutions.** Histone octamer (2  $\mu\text{g}$ ) was mixed with 30  $\mu\text{g}$  of *RsaI*-digested chicken erythrocyte DNA or with 20  $\mu\text{g}$  of chicken erythrocyte core particle DNA plus the desired radiolabeled tracer DNA in 50  $\mu\text{l}$  of TE plus 2 M NaCl (all buffers contain 0.5 mM PMSF and 1 mM benzamidine). This mixture was loaded into microdialysis buttons, which were in turn placed into a dialysis bag containing the same buffer. Samples were dialyzed at 4°C into the starting buffer for >2 hr, then into two changes of 0.5  $\times$  TE for a minimum of 12 hr each. Aliquots of the reconstituted material were run on 5% acrylamide gels containing 1/3 $\times$  TBE (30 mM Tris-borate/0.67 mM EDTA). Gels were dried and quantified by PhosphorImager (Molecular Dynamics).  $K_{\text{eq}}$  values were calculated as ratios of background subtracted counts in complex to counts in free DNA.  $\Delta G$  values were calculated from each equilibrium constant, and  $\Delta\Delta G$  values were calculated from differences between pairs of  $\Delta G$  values as indicated. A lower weight concentration of core particle DNA than *RsaI* digest was used as competitor; this choice was intended as a compromise between equal molar and weight concentrations, which cannot both be achieved simultaneously since the average lengths are different. It appears that the *RsaI*-digested DNA is sufficiently long that weight concentration may have been the more relevant variable controlling the association with histone octamer. In any case, the results obtained with the two different competitors at their different concentrations do not differ significantly.

**Computational Analysis.** Computations on the genome of the yeast *S. cerevisiae* were carried out using the data set previously described (13), which deletes telomeric regions. Sequences were analyzed for each occurrence of the motif “(A or T)<sub>3</sub>NN(G or C)<sub>3</sub>NN,” where N is any base; the number of

such motifs in each 146-bp window was counted, and the counts were histogrammed.

## RESULTS

**Competitive Reconstitution with Natural and Random DNA.** We use a modification of the competitive reconstitution assay developed by Shrader and Crothers (7, 8) to measure the relative free energies of different DNA molecules for binding to histone octamer in nucleosomes. In this approach, radio-labeled tracer DNA competes with a large molar excess of unlabeled competitor DNA for limiting amounts of histone octamer. Stepwise dialysis from 2.0 M NaCl allows the most stable population of nucleosomes to form in an equilibrium process (7, 8), eventually “freezing in” this equilibrium at low [NaCl]. This reconstitution procedure is well established and is known to yield native-like nucleosomes. The products of reconstitution are separated by native gel electrophoresis and quantified by PhosphorImager. The ratio of radiolabeled tracer DNA incorporated in nucleosomes to free tracer DNA defines an equilibrium constant and a corresponding free energy for histone binding of the tracer DNA that are valid for that specific competitive environment. Difference free energies are measured by subtraction of the free energies for differing radiolabeled tracer DNA molecules measured in parallel experiments having identical competitor DNA and competitive environments.

For the present study, the competitor is fragments of chicken erythrocyte genomic DNA prepared in either of two different ways. (i) Chicken erythrocyte nuclei were lightly digested with micrococcal nuclease to yield very long but soluble chromatin fibers; DNA was extracted and digested with a “four-base specificity” restriction enzyme, *RsaI*, to yield a diverse set of fragments that give an unbiased representation of the entire genome. (ii) Chicken erythrocyte nuclei were extensively digested with micrococcal nuclease to yield nucleosome core particles, and their DNA was isolated. The core particle DNA potentially gives an unbiased representation of the genome but also might not (e.g., it is feasible that there is a physical selection for especially stable nucleosomes, during the preparative digestion down to core particles); importantly, our results revealed little difference between these two DNA samples.

The radiolabeled tracer DNAs used were: (i) the same core particle DNA used as one of the competitors; (ii) a pool of  $5 \times 10^{12}$  different chemically random sequences; and (iii) two individual cloned isolates of the chemically random DNA.

Each synthetic random DNA molecule is 282 bp long and includes 220 bp of random sequence DNA arranged in three blocks joined together with two 6-bp nonrandom sequences (see *Materials and Methods*). Two stretches of nonrandom DNA flank each end of the constructs for use in PCR amplification. An assumption inherent in this study is that these defined sequence elements should not contribute significantly to the analysis. This assumption is consistent with available information, and we show below that this is the case experimentally. These elements contain no particular unusual sequences; they are simply particular instances of arbitrary sequence. They do not contain signals previously identified as related to nucleosome packaging (see above), nor do they contain the signals that we identified in an extensive SELEX (20) analysis of DNA determinants of nucleosome stability that we have completed (unpublished work). If by chance the defined sequence elements disfavored nucleosome formation, then that is no problem because the synthetic random sequence segments are of sufficient length that the two defined sequence ends and either one of the two restriction sequence joints can be entirely excluded from the nucleosome.

We used sucrose gradient ultracentrifugation to characterize the products of reconstitution reactions in which only the synthetic random DNA is present and the histone octamer is

supplied in 0.1 mol per mol of DNA. The results of such an experiment are shown in Fig. 1. The sucrose gradients reveal that  $\approx 10\%$  of the DNA is formed into nucleosomes, as expected, and the mobility of the nucleosomes in the gradients is consistent with their being mononucleosomes, not dinucleosomes, which sediment at or near the bottom (21). Moreover, the amount of histone octamer supplied in the reaction is sufficient to turn 10% of the DNA into only mononucleosomes, not into dinucleosomes. In other studies (data not shown; unpublished work), we found that the reconstituted nucleosomes protect  $\approx 146$  bp against digestion by micrococcal nuclease, further arguing against the possible formation of closely packed dimers. We conclude that the reconstitution procedure yields mononucleosomes, as expected.

To measure the free energies and difference free energies of reconstitution, we prepare reconstitution reactions containing competitor DNA in  $\approx 10$ -fold mass excess over histone octamer, which in turn is present in large excess over the radiolabeled tracer DNA; importantly, the competitive environment is kept identical between samples within each experiment and between experiments.

The results of such a competitive reconstitution reaction analyzed by native gel electrophoresis and PhosphorImager are shown in Fig. 2. When the radiolabeled tracer is the same DNA as the competitor (lane 5),  $\approx 10\%$  of the DNA is incorporated into nucleosomes, as expected from the histone:DNA stoichiometry. Similar results are obtained when *RsaI*-digested genomic DNA is used as competitor (lane 1). This demonstrates that there is no significant difference in free energy for these two very different preparations of genomic DNA. (We do not carry out the converse experiment using *RsaI*-digested DNA as tracer, because its large dispersion of lengths prevents clean resolution of nucleosomal and naked DNA.) The gel itself further reveals that the pool of synthetic random DNA molecules (lane 2) and the two individually cloned isolates of random DNA (lanes 3 and 4) all behave similarly to the two genomic DNA samples. Evidently, there is little difference in the free energy of reconstitution between any of these DNA samples.

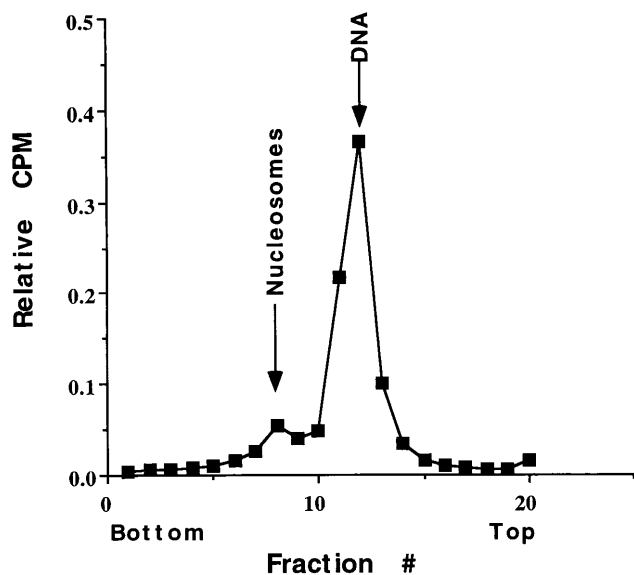


FIG. 1. Sucrose gradient analysis. Typical sucrose gradient profile of random sequence DNA selected with 0.1 mol of histone octamer per mol of DNA. The mobilities of mononucleosomes and of free DNA are indicated. The long length of the random DNA (282 bp) reduces the separation between the mononucleosomes and naked DNA compared with our earlier work with shorter reconstituted nucleosomes (15, 16). Sedimentation was from right to left.

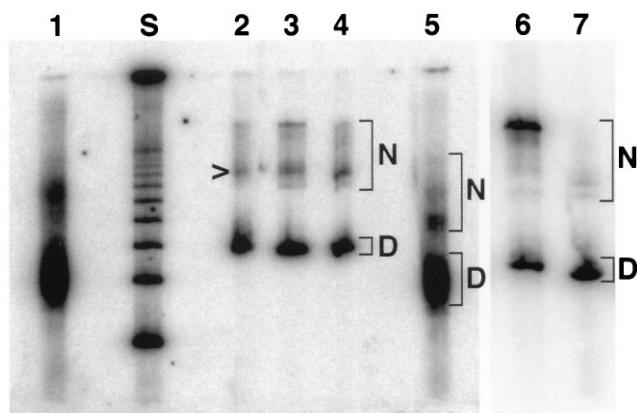


FIG. 2. Native gel analysis of competitive reconstitutions. Competitor DNAs are: lanes 1–4, *RsaI*-digested genomic fragments; lane 5, core particle DNA. Tracer DNAs are: lanes 1 and 5, core particle DNA; lane 2, pool of synthetic random DNA; lanes 3 and 4, individual cloned molecules from the synthetic random pool. S, kinased 100-bp ladder. Lanes 6 and 7, two particular cloned DNA molecules having relatively high (lane 6) or lower (lane 7) affinity for histone octamer in nucleosome reconstitution. N and D, nucleosomes and naked DNA, respectively.

These qualitative results are confirmed by the quantitative data in Table 1. The pool of synthetic random DNA and the two individual clones differ in free energy from core particle DNA by  $0.1 \pm 0.1$  kcal·mol<sup>-1</sup>. These differences are small in comparison to a benchmark of physical significance, the characteristic energies of thermal fluctuations ( $RT$ ,  $\approx 0.6$  kcal·mol<sup>-1</sup>, where  $R$  is the gas constant and  $T$  is the absolute temperature). Using core particle DNA again as the tracer but with core particle competitor DNA (instead of the *RsaI* digest) leads to a  $\Delta\Delta G$  of  $-0.2 \pm 0.2$  kcal·mol<sup>-1</sup>. This difference between the two different genomic DNAs as competitors is again small in comparison to the energies of thermal fluctuations, and moreover it may be entirely attributed to the use of a 1.5-fold lower DNA weight concentration (see *Materials and Methods*) for the core particle competitor than for the *RsaI* digest competitor; the effect of the changed weight concentration is given by  $RT \ln(20/30) \approx -0.2$  kcal·mol<sup>-1</sup>.

The small free energy differences found here represent positive results of little difference, not negative results of a failure to detect differences. Earlier studies from many laboratories establish that this methodology readily detects significant free energy differences when these in fact exist. Lanes 6 and 7 (Fig. 2) show one such example for two different individual cloned tracer DNAs, one having a high affinity for histone octamer (lane 6) and one having a more modest affinity (lane 7). The differing affinities lead to differing ratios of counts in nucleosomes to counts in naked DNA, plainly visible in the raw data (lanes 6 and 7) as well as in the

Table 1. Quantitative free energy measurements

Tracer	$\Delta\Delta G$ , kcal·mol <sup>-1</sup> ( $n$ )
Synthetic random pool	$-0.05 \pm 0.08$ (7)
Random clone 1	$0.05 \pm 0.08$ (6)
Random clone 2	$0.11 \pm 0.13$ (6)
Core particle DNA	0

$\Delta\Delta G \equiv \Delta G_{\text{sample}} - \Delta G_{\text{core particle DNA}}$ .  $\Delta G (= -RT \ln K_{\text{eq}})$  obtained using the indicated radiolabeled tracer with *RsaI*-digested genomic DNA as competitor. Values given as mean  $\pm$  SD;  $n$ , number of independent measurements. Using core particle DNA as both the tracer and the competitor yields  $\Delta\Delta G$  of  $-0.22 \pm 0.16$  kcal·mol<sup>-1</sup> ( $n = 8$ ). This difference for the two different genomic DNAs as competitors is small in comparison to  $RT$  and can be entirely attributed to the use of different DNA weight concentrations.

integrated intensities measured by PhosphorImager (data not shown).

The broad range of mobilities for the nucleosomes reconstituted with both the pool (Fig. 2, lane 2) and individual clones (lanes 3 and 4) of the synthetic random DNA is noteworthy. Mobility differences arise from variable locations of the histone octamer within the overall DNA length (22, 23), implying that on these DNA samples, many positions are represented even on individual clones (lanes 3 and 4). As will be seen below, this behavior is a corollary of the fact that both the majority of the entire random pool and representative individual clones have low free energies for nucleosome formation that moreover do not differ much for nucleosome formation in one position versus another. (Note that the high-affinity DNA molecule in lane 6 shows a very different behavior; it gives rise to a strongly biased nucleosome position with a corresponding sharp mobility on the native gel.) The core particle DNA, while somewhat disperse in length, is mostly much shorter than the 282-bp synthetic random DNA and therefore provides a more restricted range of mobilities. (The length differences do not contribute significantly to the measured free energies; similar results are obtained using the core particle DNA, which is shorter than 282 bp, and the *Rsa*I-digested DNA, which is longer.)

Superimposed on the broad range of mobilities for the nucleosomes reconstituted with the synthetic random DNA pool (Fig. 2, lane 2) or with either of the two individually cloned isolates (lanes 3 and 4) is a discernable band at a distinct mobility (arrowhead). The existence of this band could suggest that a fraction of the nucleosomes reconstituted on each of these DNA samples may be preferentially adopting a particular position on the DNA. But for the pool of unrelated random sequences, this could only happen if some element that all molecules have in common such as the fixed sequence elements or proximity to a DNA end were contributing detectably to the reconstitution. If this effect were large, it would invalidate a key assumption underlying this study. To quantify this effect, we measured the fraction of counts in reconstituted nucleosomes due to this species having the distinct mobility. We obtained results of 15%, 23%, and 33%, for lanes 2–4, respectively. Thus the equilibrium constants for incorporating the tracer into nucleosomes is increased by 1.15-, 1.23-, and 1.33-fold, respectively by the ability to form this species. The corresponding free energies are 0.08, 0.11, and 0.16 kcal·mol<sup>-1</sup>, far lower than the benchmark of thermal energies and smaller than the likely experimental error of the differential measurements reported in Table 1. Thus effects attributable to fixed sequence elements can be seen in this assay but are quantitatively small and do not significantly affect the conclusions of this study.

Since the best (i.e., highest affinity) ≈10% of the DNA is reconstituted into nucleosomes in this procedure, these results imply that there is little difference in the free energy of reconstitution between the best 10% of the synthetic random sequences and the best 10% of the natural genomic sequences. Importantly, however, the results for the individually cloned isolates of the synthetic random DNA show that the results from the best 10% of the synthetic random sequences are representative; there is <1% chance that both individual isolates are each in the top 10th percentile of affinity in the entire pool, yet their free energy of reconstitution is no different. We conclude that there is little free energy difference between typical random synthetic DNA and the best 10% of the natural genomic DNA. Moreover, half of the best 10% of the natural genomic DNA is worse than its measured average (by definition). Hence, we conclude that at least 95% of genomic DNA has a free energy of reconstitution (i.e., affinity for histone octamer) that differs little from that of typical synthetic random DNA. The bulk of the eukaryotic genome is evidently not constrained or evolved to aid sub-

stantially in its own packing at the level of individual nucleosomes.

**Relation of Histone Binding Affinity to Nucleosome Positioning.** We consider next the relation between the free energy of nucleosome formation and the preferential occupancy of specific DNA sites by the histone octamer, also known as nucleosome positioning. We distinguish translational and rotational positioning (5). Translational positioning refers to the extent to which a histone octamer selects a particular contiguous stretch of 147 bp of DNA in preference to other stretches of the same length that are translated forwards or backwards along the DNA. Rotational positioning is a degenerate form of translational positioning in which a set of discrete translational positions, differing by integral multiples of the DNA helical repeat, are all occupied in preference to the set of other possible locations. DNA sequences that are intrinsically bent or are anisotropically bendable may lead to rotationally positioned nucleosomes.

We consider a model in which a histone octamer is constrained to be bound to a DNA molecule of length  $L$ . We take 147 bp as the amount of DNA in a nucleosome core particle, assuming for simplicity that locations of the octamer that extend beyond the DNA end (leaving DNA binding sites on the octamer unsatisfied) are negligibly populated. *In vitro*, for nucleosome reconstitution experiments,  $L$  is well defined, and there are  $L - 146$  different possible locations for the octamer. *In vivo*, there is not a unique definition of  $L$ ; one appropriate definition is the nucleosome repeat length (see ref. 24). For a further simplification, we take all possible locations of the histone octamer other than the specifically positioned one(s) to be equivalent in their free energy.

**Translational Nucleosome Positioning.** Translational positioning may be understood quantitatively in terms of the free energy change for transfer of histone octamer from unrestricted exploration of any of the  $L - 147$  available (and equivalent) nonspecific positions to fixed occupancy at a single specific location. We take the nonspecifically bound state as the reference, and assign the free energy  $\equiv 0$ . Let  $\Delta G_{\text{net}}$  be the free energy change for transfer to the specific position. Two terms contribute:  $\Delta G_{\text{net}} = \Delta G_{\text{intr}} + \Delta G_{\text{stat}}$ , where  $\Delta G_{\text{intr}}$  is the free energy change for transmuted a nonspecific site into a specific one, and  $\Delta G_{\text{stat}}$  is a statistical contribution to the free energy change, reflecting the fact that there may be many more possible nonspecific locations than the single specific one.

$\Delta G_{\text{intr}}$  is the same as would obtain if the process were to take place in standard state conditions; hence  $\Delta G_{\text{intr}} \equiv \Delta \Delta G_{\text{HO}}$ , which is defined and also measured experimentally as the difference free energy for reconstitution of histone octamer into nucleosomes, for two DNA molecules having identical lengths but where one contains a single favored ("specific") position and the other contains no such sites.

Transfer of histone octamer from free choice among any of  $L - 147$  nonspecific positions to fixed occupancy at a specific position reduces the entropy of the system by the amount  $\Delta S_{\text{stat}} = R \ln(1/(L - 147))$ , thereby contributing  $\Delta G_{\text{stat}} = -T\Delta S_{\text{stat}} = -RT \ln(1/(L - 147))$  to the net free energy change.

With these definitions, the probability of occupancy of the specific position by the histone octamer ( $P$ ) is given by:

$$P = \frac{e^{-\Delta G_{\text{net}}/RT}}{1 + e^{-\Delta G_{\text{net}}/RT}} = \frac{1}{1 + (L - 147)e^{\Delta \Delta G_{\text{HO}}/RT}}, \quad [1]$$

where  $R$  is the gas constant and  $T$  is the absolute temperature.

This equation has several important ramifications. Positioning is inherently statistical, not "precise": it is quantified by a probability of occupancy of the specific site that is always greater than zero and less than one.  $P$  depends fundamentally on two quantities: the intrinsic free energy preference of histone octamer for the specific site compared with alternative

sites,  $\Delta\Delta G_{\text{HO}}$ , and the number of alternative sites,  $L - 147$ . Given a particular DNA length,  $P$  is determined by  $\Delta\Delta G_{\text{HO}}$ .

Given a measured value for  $\Delta\Delta G_{\text{HO}}$  and the DNA length  $L$ , Eq. 1 allows one to calculate  $P$ . If nucleosomes prove to be freely mobile as proposed (e.g., refs.25–27),  $P$  will also equal the fraction of time that the specific position is occupied in any particular nucleosome.

Eq. 1 implies that existing “positioning sequences” have quite limited positioning power. For example, a 250-bp fragment in which occupancy at one specific position is favored by 2 kcal·mol<sup>-1</sup> [comparable in energy to the pentamer “TG” sequence element (7, 8)] yields only  $\approx 20\%$  occupancy of the specific position. This limited occupancy of the specific position is nevertheless easily detected experimentally, because it stands out against a weak background; the remaining  $\approx 80\%$  occupancy is distributed over 103 alternative positionings, each having  $< 1\%$  occupancy.

Eq. 1 also provides a new way to measure  $\Delta\Delta G_{\text{HO}}$ . Direct experimental measurement of  $P$ , perhaps through quantitative nuclease protection studies for sites both internal and external to the positioned nucleosome, together with the known DNA length  $L$ , allows  $\Delta\Delta G_{\text{HO}}$  to be calculated.

**Rotational Positioning.** Eq. 1 is readily generalized to account for cases in which multiple specific translational positions are preferentially populated relative to the bulk “nonspecific” positions. Rotational positioning is an example of such a case, for which there happen to be constraints on the mutual spacings of the preferred positioning sites.

For a DNA molecule of length  $L$ , let  $n_{\text{R}}$  be the number of positions occupied by rotationally positioned histone octamers and let  $\Delta\Delta G_{\text{R}}$  be the intrinsic free energy for binding of histone octamer in any one of the set of preferentially occupied positions in a rotationally positioned system. (For simplicity, we let the same  $\Delta\Delta G_{\text{R}}$  apply to all of them.) Then, rotational positioning, too, is seen to be intrinsically statistical, quantified by a probability that any of the specific rotational positions is occupied, given by:

$$P = \frac{1}{1 + \frac{(L - 146 - n_{\text{R}})}{n_{\text{R}}} e^{\Delta\Delta G_{\text{R}}/RT}} \quad [2]$$

$\Delta\Delta G_{\text{R}}$  can be measured as described above, provided that one uses for this purpose a variant of the DNA molecule that is sufficiently short (e.g.,  $\approx 150$  bp), such that only one rotationally-phased position is accessible. Alternatively, a longer molecule may be used provided that the measured apparent  $\Delta\Delta G_{\text{R}}$  is corrected for the statistical factor  $RT \ln n_{\text{R}}$  before use in Eq. 2.

**Protein-Directed Nucleosome Positioning.** It is possible that, *in vivo*, certain site-specific DNA binding proteins or other effects will act to influence the positioning of nucleosomes (10, 24, 28, 29). Such influences may be quantified by a coupling free energy, and Eqs. 1 and 2 may be adapted to quantify the net (but still statistical) positioning that results (J.W., unpublished work).

## DISCUSSION

**Nucleosome Packaging and Nucleosome Positioning of Genomic DNA.** The most important conclusion of this work is that the bulk of the eukaryotic genome ( $\geq 95\%$ ) is not evolved or constrained to aid substantially in its own packaging at the level of individual nucleosomes. Apparent differences which are detected (Table 1) are small in comparison to thermal energies, a benchmark of physical significance. The relation derived here between the free energy of nucleosome reconstitution and the probability of nucleosome positioning implies that the bulk of the eukaryotic genome is similarly not evolved

or constrained to substantially bias or specify nucleosome positioning at the level of individual nucleosomes. The similarity of the results obtained with two very different representations of the genomic DNA provides further evidence that each sample reliably reflects the properties of the whole genome.

The results of this study apply to systems at equilibrium. Nucleosome core particles that result from reconstitution *in vitro* are kinetically trapped; in typical solution conditions they are stable against dissociation and against exchange with labeled competitor DNA for indefinitely long periods. These are not properties of an equilibrium system. Importantly, however, Shrader and Crothers (7, 8) have demonstrated experimentally that nucleosome reconstitution *in vitro* is an equilibrium process. The procedure through which nucleosomes are reconstituted *in vitro* establishes a true equilibrium distribution and then subsequently (and reversibly) freezes this in, creating stable particles that nevertheless reflect an equilibrium distribution.

It is now clear that nucleosomes are mobile even in physiological ionic conditions (25–27), and our competitive reconstitution procedure sweeps slowly through physiological ionic strength before freezing in the resulting particles at subphysiological ionic strengths. Therefore we presume that the nucleosomes resulting from our competitive reconstitutions have equilibrated at physiological ionic strengths. This means that, as regards the DNA sequence determinants of nucleosome stability and positioning, the results of our study *in vitro* should reflect the same histone–DNA interactions as those that obtain *in vivo*.

We emphasize that, *in vivo*, many additional factors beyond genomic DNA sequence may act to bias the positions of nucleosomes (10, 24, 28, 29). The present results serve to define and quantify just those contributions to nucleosome positioning arising from sequence dependences to histone–DNA interactions. However, our overall conclusion—that the free energies of positioning are finite and therefore that nucleosome positioning is statistical, not precise—applies also to the total effects of all of the forces that act to govern nucleosome positioning *in vivo*.

**Relation to Previous Studies.** The present results demonstrate that the bulk of the eukaryotic genome ( $\geq 95\%$ ) is not evolved or constrained to aid substantially in its own nucleosome packaging or nucleosome positioning at the level of individual nucleosomes. Yet, at the same time, previous work by us and others demonstrates that sequence signals involved in nucleosome packaging and in nucleosome positioning are readily detectable in natural DNA. How may both of these sets of results be true at once? We imagine two limiting possibilities. (i) Perhaps signals that favor nucleosome packaging and positioning are sparsely but rather uniformly distributed along the entire genome. They can be detected by sensitive signal averaging procedures, but are quantitatively small, less than approximately  $\pm 0.2$  kcal·mol<sup>-1</sup>. They may nevertheless have biological significance. Acting individually, they yield only a small bias in nucleosome positioning. Or, they may act collectively on the positioning of an entire array of nucleosomes, perhaps through constraints on the mutual packing of nucleosomes within higher order chromatin structures (24); the collective effects may be substantial. (ii) Perhaps signals involved in nucleosome packaging and positioning are concentrated into a small subset ( $< 5\%$ ) of the genome. These are detected when enough sequence information is analyzed, while the majority of genomic sequences simply contribute incoherent “noise” to the analyses.

The second limiting possibility (above) is certainly correct to some degree; one of the best characterized “nucleosome-positioning” DNA sequences, that discovered by Simpson and Stafford (30), has a significantly higher affinity for histone

octamer in nucleosome reconstitution than typical "bulk" genomic DNA (7, 8) and is itself a natural sequence.

Both of these ideas may be true at once. An example is provided in Fig. 3, which plots the probability distribution for the number of occurrences within a nucleosome-sized window of a particular sequence motif that is correlated with favorable nucleosome packaging (7, 8), averaged over the genome of the yeast *S. cerevisiae*. There is a single peak together with a long forward tail. The main peak reflects the predominant situations, in which signals that favor nucleosome packaging are sparsely but rather uniformly distributed along the entire genome, as in idea (i). The long forward tail reflects a minority of situations in which nucleosome packaging signals are relatively concentrated in a small subset of the genome, as in the second limiting possibility presented above. (Whether these particular genomic regions actually favor nucleosome formation remains to be tested; this example provides a demonstration in principle of how real genomic sequences plausibly conform at once to both of the ideas presented above.)

It will be important to assess experimentally where the best nucleosome-packaging sequences are in genomes, and how their locations relate to other elements of gene and genome organization.

**Statistical Nature of Nucleosome Positioning.** A second important conclusion is that nucleosome positioning is inherently statistical, not precise. Briefly put, the free energies and free energy differences are finite, so probabilities are necessarily greater than zero and less than one. In this light, previous reports of precise positioning at single or multiple nearby sites must be considered as really reflecting preferential positioning at those sites. It is important to recognize this statistical property of positioning because it has substantial ramifications for proposed mechanisms of gene regulation. If positioning is not precise, then essential DNA regulatory sequences will sometimes be buried when they need to be accessible or will sometimes be accessible when they need to be repressed (buried). Mechanisms proposed for gene regulation must be robust to inevitable statistical fluctuations.

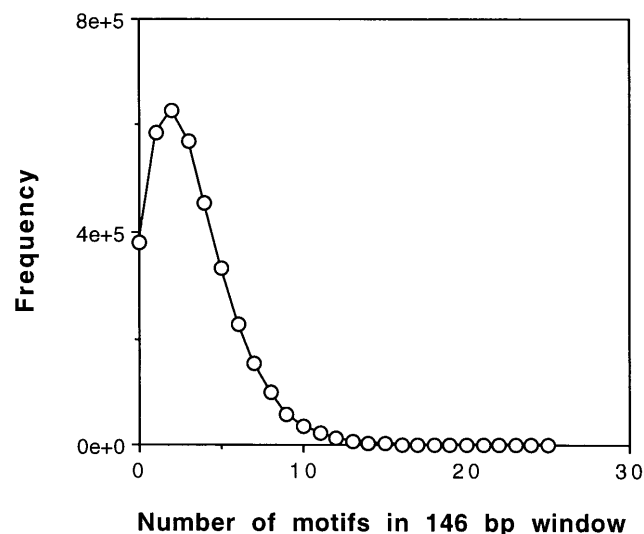


FIG. 3. Probability distribution for the number of occurrences within a nucleosome-sized window of the motif "(A or T)<sub>3</sub>NN(G or C)<sub>3</sub>NN," where N is any base, averaged over the yeast genome. The plot is represented on a linear probability scale for ease of interpretation; a logarithmic plot (data not shown) better represents detail in the forward tail and again shows there to be a single peak at 2 on the abscissa. Only nonzero values are plotted; the last nonzero datapoint is at 25 on the abscissa. Many other motifs related to nucleosome packaging or positioning were also analyzed and yielded comparable results.

It may be helpful in this context to consider the case of prokaryotic repressor proteins as gene regulatory factors. Repressor proteins do indeed bind to and act at precise sites, but it has not proven helpful to use this language nor to consider the problem in this way. Rather, it is recognized that repressor proteins obey the laws of thermodynamics; they have a finite free energy difference for binding to the specific site versus binding to a large set of alternative nonspecific sites, and one equates their activities in gene regulation to the equilibrium fractional occupancy (i.e., the probability of occupancy) of the specific site (31).

Finally, we emphasize that DNA sequence-encoded biases in (statistical) positioning do exist and may have important biological consequences. In the context of our site exposure model for the mechanism of regulatory protein binding to nucleosomal target sites (15, 32), biases in average positioning of mobile nucleosomes substantially affect the average concentration of the regulatable (accessible) state of nucleosomes. Biases in positioning may also contribute to the stability of higher order chromatin folding (24), potentially at multiple hierarchical levels in the structure.

We acknowledge with gratitude the kind gift of the random synthetic DNA by Drs. David Bartel and Jack Szostak, and we thank Kevin Polach for helpful discussions. Research in our laboratory is supported by a grant from the National Institutes of Health.

- Hagerman, P. J. (1990) *Annu. Rev. Biochem.* **59**, 755–781.
- Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7664–7668.
- Widom, J. (1985) *BioEssays* **2**, 11–14.
- Travers, A. A. & Klug, A. (1987) *Philos. Trans. R. Soc. London B* **317**, 537–561.
- Simpson, R. T. (1991) *Prog. Nucleic Acid Res. Mol. Biol.* **40**, 143–183.
- Travers, A. A. & Muyldermans, S. V. (1996) *J. Mol. Biol.* **257**, 486–491.
- Shrader, T. E. & Crothers, D. M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7418–7422.
- Shrader, T. E. & Crothers, D. M. (1990) *J. Mol. Biol.* **216**, 69–84.
- Bina, M. (1994) *J. Mol. Biol.* **235**, 198–208.
- Staffelbach, H., Koller, T. & Burks, C. (1994) *J. Biomol. Struct. Dyn.* **12**, 301–325.
- Ioshikes, I., Bolshoy, A., Derenshteyn, K., Barodovsky, M. & Trifonov, E. N. (1966) *J. Mol. Biol.* **262**, 129–139.
- Bolshoy, A. (1995) *Nat. Struct. Biol.* **2**, 446–448.
- Widom, J. (1996) *J. Mol. Biol.* **259**, 579–588.
- Bartel, D. P. & Szostak, J. W. (1993) *Science* **261**, 1411–1418.
- Polach, K. J. & Widom, J. (1995) *J. Mol. Biol.* **254**, 130–149.
- Protacio, R. U. & Widom, J. (1996) *J. Mol. Biol.* **256**, 458–472.
- Widom, J. (1986) *J. Mol. Biol.* **190**, 411–424.
- Simpson, R. T. (1978) *Biochemistry* **17**, 5524–5531.
- Feng, H.-P., Scherl, D. S. & Widom, J. (1993) *Biochemistry* **32**, 7824–7831.
- Irvine, D., Tuerk, C. & Gold, L. (1991) *J. Mol. Biol.* **222**, 739–761.
- Yao, J., Lowary, P. T. & Widom, J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7603–7607.
- Pennings, S., Meersseman, G. & Bradbury, M. E. (1991) *J. Mol. Biol.* **220**, 101–110.
- Dong, F., Hansen, J. C. & van Holde, K. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5724–5728.
- Widom, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1095–1099.
- Meersseman, G., Pennings, S. & Bradbury, E. M. (1992) *EMBO J.* **11**, 2951–2959.
- Varga-Weisz, P. D., Blank, T. A. & Becker, P. B. (1995) *EMBO J.* **14**, 2209–2216.
- Ura, K., Hayes, J. J. & Wolffe, A. P. (1995) *EMBO J.* **14**, 3752–3765.
- Kornberg, R. D. & Stryer, L. (1988) *Nucleic Acids Res.* **16**, 6677–6690.
- Yao, J., Lowary, P. T. & Widom, J. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 9364–9368.
- Simpson, R. T. & Stafford, D. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 51–55.
- von Hippel, P. H. (1994) *Science* **263**, 769–770.
- Polach, K. J. & Widom, J. (1996) *J. Mol. Biol.* **258**, 800–812.