A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome


Genome Res. 2008 18: 1073-1083 originally published online June 12, 2008
Access the most recent version at doi:10.1101/gr.078261.108

Supplemental Material
http://genome.cshlp.org/content/suppl/2008/08/11/gr.078261.108.DC1.html

References
This article cites 54 articles, 20 of which can be accessed free at:
http://genome.cshlp.org/content/18/7/1073.full.html#ref-list-1

Article cited in:
http://genome.cshlp.org/content/18/7/1073.full.html#related-urls

Related Content
A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning
Anton Valouev, Jeffrey Ichikawa, Thaisan Tonthat, et al.
Genome Res. July , 2008 18: 1051-1063

Two strategies for gene regulation by promoter nucleosomes
Itay Tirosh and Naama Barkai

A high-resolution map of nucleosome positioning on a fission yeast centromere
Jun S. Song, Xingkun Liu, X. Shirley Liu, et al.
Genome Res. July , 2008 18: 1064-1072

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here

To subscribe to Genome Research go to:
http://genome.cshlp.org/subscriptions

Copyright © 2008, Cold Spring Harbor Laboratory Press
A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome


Most nucleosomes are well-organized at the 5’ ends of S. cerevisiae genes where “−1” and “+1” nucleosomes bracket a nucleosome-free promoter region (NFR). How nucleosomal organization is specified by the genome is less clear. Here we establish and inter-relate rules governing genomic nucleosome organization by sequencing DNA from more than one million immunopurified S. cerevisiae nucleosomes (displayed at http://atlas.bx.psu.edu/). Evidence is presented that the organization of nucleosomes throughout genes is largely a consequence of statistical packing principles. The genomic sequence specifies the location of the −1 and +1 nucleosomes. The +1 nucleosome forms a barrier against which nucleosomes are packed, resulting in uniform positioning, which decays at farther distances from the barrier. We present evidence for a novel 3’ NFR that is present at >95% of all genes. 3’ NFRs may be important for transcription termination and anti-sense initiation. We present a high-resolution genome-wide map of TFIIIB locations that implicates 3’ NFRs in gene looping.

[Supplemental material is available online at www.genome.org.]
nuclease (MNase), immunopurified with antibodies against tagged histones H3 and H4, and 1,206,057 individual nucleosomal DNAs were sequenced ("ChiP-seq," see Methods). 53,026 consensus nucleosome locations were identified by at least three sequencing reads of >100 bp each. The long sequencing reads afforded by the GS20 genome sequencer (454 Life Sciences) allow nucleosome identification in regions of low complexity DNA sequence (e.g., telomeric regions, transposons, tRNA genes, paralogous genes, etc.), which are problematic for other approaches that employ hybridization (microarrays) or identify short (∼25 bp) sequence tags.

A median of 14 sequence reads per nucleosome was attained (Supplemental Fig. S1; Supplemental Table S1), resulting in >93% sequencing coverage. Yeast therefore have ∼57,000 nucleosomes, which agrees with recent reports (Lee et al. 2007; Shivashwamy et al. 2008). Nucleosome locations also agreed well with other data sets (Supplemental Fig. S2) (Raisner et al. 2005; Albert et al. 2007; Lee et al. 2007; Shivashwamy et al. 2008).

The distribution of reads and nucleosome locations at genomic sites is presented in Figure 1A and Supplemental Figure S3. Data and nucleosome distributions for any locus in the genome can be visualized at http://atlas.bx.psu.edu. The data suggest that nucleosomes do not exist as distinct populations in which roughly half are well-positioned and half are delocalized (Yuan et al. 2005; Segal et al. 2006; Lee et al. 2007), although for purposes of analysis it is often useful to categorize them as such. Instead, nucleosome positions exist as a continuum from low to high delocalization (fuzziness).

Patterns of nucleosome organization across the genome

The arrangement of nucleosomes along chromosomal DNA has several physical properties that include location, occupancy level, fuzziness/phasing, spacing, and width (Fig. 1B; see Table 1 for operational and descriptive definitions). These properties take on significance in the context of chromosomal features (e.g., beginning and end of genes). Figure 1C shows a composite distribution of nucleosomes relative to the 4799 transcriptional start sites (TSS) reported elsewhere (David et al. 2006). A number of

Figure 1. Nucleosome organization around the 5′ end of genes. (A) Browser shot showing the distribution of sequencing reads along 2.5 kb of an arbitrary gene (5′ end of YAL041W shown in blue). The read count at each coordinate is shown as a bar graph. The distribution was smoothed using a correction factor for MNase bias. Peaks correspond to assigned nucleosome locations based upon a user-defined threshold (purple tracks). Nucleosomes that are found in broad peaks or plateaus are assigned a specific location, although their actual position is delocalized (illustrated by overlapping ovals). Additional browser shots are shown in Supplemental Figure S3 and for any queried locus at http://atlas.bx.psu.edu/. (B) Illustration of the physical properties associated with nucleosome positions, which are defined in Table 1. (C) Distribution of nucleosome locations relative to transcriptional start sites (TSS) (David et al. 2006). Nucleosome distances were binned and the count divided by 100, then normalized to the number of regions (red line) present in each bin, and plotted as a smoothed distribution (black-filled plot). In an effort to represent a “pure” pattern, regions <300 bp from an adjacent TSS or TTS (transcript termination site) but also >300 bp from the reference TSS, were removed from the analysis. The same plot was generated without these filters (gray-filled plot, in background). An illustration of the statistical distribution of nucleosomes reported by Kornberg and Stryer (1988) is shown at the top of the panel. (D) Schematic illustration of nucleosome organization at telomeric regions. The arrangement of repeat elements is not representative of all telomeres, but represents a common arrangement. Positions of nucleosomes are approximate and reflect general themes, such as nucleosome-free zones, and noteworthy nucleosomes (colored blue, or bound by another protein in pink).
patterns were recognizable that have been described in previous studies (Lee et al. 2004; Sekinger et al. 2005; Yuan et al. 2005; Ioshikhes et al. 2006; Segal et al. 2006; Albert et al. 2007; Shi-vaswamy et al. 2008), and so will not be discussed here. These include (1) a depletion of nucleosomes immediately upstream of the −1 nucleosome at places where adjacent upstream genes are absent (i.e., intergenic regions), (2) a −140 bp nucleosome-free promoter region (NFR), which was found at −95% of all 5761 genes, (3) a TSS that is −13 bp inside the upstream border of the +1 nucleosome, (4) uniform 165 bp spacing of nucleosomes (18 bp linker) near the 5′ end of genes that extends to as many as nine nucleosomes, and (5) a loss of uniformity (peak-trough diminution) toward the 3′ end. Such properties represent the predominant genome-wide pattern.

The uniformity of positioning with respect to the TSS is much greater here than reported in any of the other studies. Several factors might contribute to this, including immunopurification of crosslinked nucleosomes, MNase bias correction, and methods of peak calling. Since positioning is determined with respect to experimentally measured TSSs, which have inherent experimental error, the actual uniformity of positioning could be stronger than that shown in Figure 1C.

We examined a variety of subclasses of genes as well as chromosomal elements to see whether certain types of genes or elements possessed alternative organizational arrangements (Supplemental Fig. S4). RNA polymerase II subclasses included TATA-containing, TATA-less, lowly transcribed, highly transcribed, cytosolic ribosomal protein encoding, mitochondrial ribosomal protein encoding, sporation-specific, filamentation-specific, histone H1-enriched. In most cases, the canonical nucleosome organization pattern was observed. However, there were some differences. TATA-containing genes tended to have less of a canonical pattern, which may be related to their greater dependency on chromatin remodeling complexes than most other genes (Basehoar et al. 2004). Genes encoding cytosolic ribosomal proteins were depleted of nucleosomes, and were missing a −1 nucleosome (Supplemental Fig. S4E), probably owing to their extremely high transcription rate. Genes containing histone H1 lacked a consensus organizational pattern around the −1 nucleosome (Supplemental Fig. S4I), suggesting that H1 might be contributing to gene-specific chromatin organization in the upstream regions of promoters. snRNA and snoRNA genes lacked a consensus organization pattern (Supplemental Fig. S4I), while tRNA genes were largely devoid of nucleosomes, although a consensus organization was evident for −1 kb upstream (Supplemental Fig. S4K). Replication origins (ARS) were devoid of nucleosomes, but were bracketed by positioned nucleosomes (Supplemental Fig. S4L).

### Table 1. Physical properties of a nucleosome located in a genome

<table>
<thead>
<tr>
<th>Feature</th>
<th>General description</th>
<th>Technical description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupancy Fuzziness</td>
<td>Amount of nucleosome binding.</td>
<td>Number of sequencing reads used to assign a nucleosome location</td>
</tr>
<tr>
<td>Phasing</td>
<td>A measure of how delocalized or spread out a nucleosome position is</td>
<td>The standard deviation of all read coordinates that contribute to a nucleosome location</td>
</tr>
<tr>
<td>Width</td>
<td>A measure of whether the edges of a nucleosome are more or less protected from MNase compared to a standard 147 bp core particle</td>
<td>The number of base pairs between the consensus left and right borders of a nucleosome (measured separately on the W and C strand)</td>
</tr>
<tr>
<td>Spacing</td>
<td>Distance between nucleosomes</td>
<td>Number of base pairs between the midpoints of two adjacent nucleosomes</td>
</tr>
<tr>
<td>Feature distance</td>
<td>For example, how far a nucleosome is from a transcription start or end site</td>
<td>Number of base pairs between a nucleosome midpoint and a specified chromosomal coordinate</td>
</tr>
</tbody>
</table>

Yeast telomeric regions are composed of four classes of repetitive elements (as listed in SGD, see Fig. 1D): telomeric (~76 bp), Y′ (~6400 bp), X combination (~278 bp), and X core (~461 bp). Because of their repetitive nature, telomeric regions have been largely excluded from genome-wide mapping studies using tiled microarrays or short sequence tags, and so little is known about telomeric nucleosome organization. We found telomeric repeats to be relatively depleted of nucleosomes, although a positioned nucleosome was centered ~150 bp upstream of the beginning of these repeats (Supplemental Fig. S4M). Y′ elements contained well-positioned nucleosomes in which the +1 and +2 nucleosomes (defined relative to the start of the element) were located at +80 and +440 bp from the start (Fig. 1D; Supplemental Fig. S4N). As a result there is a ~200 bp NFR. Interestingly, the Y′ +1 nucleosome protects an extra 40 bp of DNA from MNase digestion, suggesting that it is part of a larger stable complex. Canonically spaced and positioned nucleosomes were found upstream of X core elements (designated as −3, −2, and −1 positions in Fig. 1D, see also Supplemental Fig. S4O,P). X core elements started at the downstream border of the −1 nucleosome. Inside the X core repeat, the +1 nucleosome is extremely well-positioned and normally sized (~150 bp), but spaced ~250 bp downstream from the −1 nucleosome. The +2 nucleosome, which is equivalent to the −1 nucleosome of the X combination repeats, resides another 300 bp further downstream, after which subsequent nucleosomes are normally spaced. Thus, the +1 nucleosome of the telomeric X core repeats appears to be unusually isolated and stable (blue oval in Fig. 1D).

### Nucleosome organization at the 5′ end of genes conforms to the principles of statistical positioning

Figure 1C illustrates two potential examples of statistical positioning emanating from promoters: (1) dense nucleosome packing over genes which drives uniform spacing, but decays further from the barrier (peak-trough diminution between the +1 nucleosome and +2000 bp), and (2) lower nucleosome density over intergenic regions in which uniform spacing is diminished upstream of the −1 nucleosome. The statistical positioning model further predicts that nucleosome fuzziness on individual genes will increase when moving away from a barrier position (Yuan et al. 2005), such as that presented by the +1 nucleosome. Indeed nucleosome fuzziness, as measured by the standard deviation of sequencing read locations that define each nucleosome position (Fig. 1B; Table 1), increased from the TSS to the 3′ end of the gene (Fig. 2A, general trend of the trace from the TSS to +2000 bp). The decay in phasing, as measured by the change in fuzziness over distance, was three times greater in the upstream
Figure 2. Statistical positioning emanating from barrier nucleosomes. (A) Nucleosome fuzziness and width relative to the TSS. Fuzziness (standard deviation of read locations for those nucleosomes defined by three or more reads) is plotted, against a gray backdrop of Figure 1C. Data are plotted as a moving average of 500 nucleosomes. Deviations from the canonical 147 bp nucleosomal length (in bp) are displayed as a 2000-nucleosome moving average, and represent the distance between the nucleosomal calls made separately on the W and C strand. (B) Averaged NPS correlation score and poly(dA:dT) density for genes aligned by the TSS. The NPS correlation employed an updated AA/TT nucleosomal distribution pattern (Supplemental Fig. S6C). Plot of AAA/TTTT per bp frequency (11 bp moving average) is shown. Longer poly(dA:dT) tracts gave similar results (data not shown). The gray backdrop reflects nucleosome positions from Figure 1C. NPS scores and fuzziness are not correlated (Supplemental Fig. S7), indicating that the diminishing NPS correlation at distances distal to the –1/+1 nucleosome is unlikely to cause the increased fuzziness shown in panel A. (C) NPS correlation profiles broken out by AA dinucleotide enrichment near 5′ ends vs. 10 bp periodic spacing of AA dinucleotides.

DNA sequences that organize nucleosomes are concentrated at promoter regions

Repeating 10 bp patterns of AA/TT dinucleotides create DNA curvature that favors nucleosome formation and rotational positioning of the DNA (Drew and Travers 1985; Satchwell et al. 1986; Shrader and Crothers 1989; Widom 1996; Cohanim et al. 2005). Previous experimentally derived frequency distributions of AA/TT dinucleotides across nucleosomal DNA show the expected 10 bp pattern (Ioshikhes et al. 1996; Satchwell et al. 1986). However, within the same pattern there is also an enrichment of AA dinucleotides toward the 5′ end of nucleosomal DNA (see Supplemental Fig. S6A), which has not been investigated. Correlations to the experimental AA/TT pattern are strongest at the –1 and +1 nucleosome positions (Ioshikhes et al. 2006), suggesting that such patterns contribute to nucleosome positioning in the Saccharomyces genome. Whether 10 bp periodicities of AA/TT dinucleotides or 5′ enrichment of AA provides the predominant contribution to positioning is determined below.

In contrast to AA/TT patterns, contiguous poly(dA:dT) runs are rigid and exclude nucleosomes in core promoter regions (Iyer and Struhl 1995; Suter et al. 2000; Anderson and Widom 2001; Sekinger et al. 2005; Bao et al. 2006). Therefore, AA/TT NPS and poly(dA:dT) patterns are inversely related. We measured poly(dA:dT) density and AA/TT NPS correlations of genes aligned by TSS. We employed an updated AA/TT NPS pattern that took advantage of our larger collection of well-positioned nucleosomes (Albert et al. 2007). As expected from prior studies (Yuan et al. 2005; Ioshikhes et al. 2006), there was an enrichment of poly(dA:dT) and a negative NPS correlation in the core promoter region (Fig. 2B, from –150 to 0). Thus, the underlying DNA sequence of core promoters might be designed to exclude nucleosomes.

In contrast to the core promoter region, positive NPS corre-
lations aligned with the location of the −1 and +1 nucleosomes, as expected. Regions further downstream had AA/TT NPS correlations of significantly lower magnitude and lacked substantial poly(dA:dT) enrichment between nucleosomes, indicating that their direct underlying or adjacent contributions to the highly positioned +2, +3, and +4 nucleosomes (see Fig. 1C) are modest. The combination of a poly(dA:dT)-rich NFR and an NPS-directed +1 nucleosome may help create the barrier necessary for statistical positioning of downstream nucleosomes.

When analyzing the iteratively improved AA/TT NPS pattern (Supplemental Fig. S6C), we noticed that the 10 bp periodicity of AA dinucleotides had diminished, and that the slope of the AA frequency distribution had become more pronounced. To explore whether AA 10-bp periodicities or 5′ AA enrichment was the predominant parameter in the pattern, we either smoothed out the AA periodicities (using an 11 bp moving average) or eliminated the 5′ AA enrichment by symmetrizing the pattern (Supplemental Fig. S6D). As shown in Figure 2C, enrichment of AA dinucleotides toward the 5′ end of the nucleosome (and TT enrichment toward the 3′ end) was a better descriptor of positioning than 10 bp periodicities of AA/TT. Therefore, we surmise that AA enrichment toward the 5′ ends (and TT enrichment toward the 3′ ends) provides a larger contribution to positioning than 10 bp periodicities. The 10 bp AA/TT periodicity, which is rather evident in well-positioned nucleosomes (Satchwell et al. 1986; Ioshikhes et al. 2006; Albert et al. 2007), might instead contribute to rotational positioning rather than translational positioning of the −1 and +1 nucleosomes.

Nucleosomal dinucleotide patterns are evident for the −1 and +1 nucleosomes, but largely absent from genic nucleosomes

Many dinucleotides are distributed nonrandomly across yeast nucleosomal DNA (Albert et al. 2007). We determined the extent to which each of the 16 possible dinucleotides might contribute to nucleosome positioning at −1, +1, and all other genic positions. Figure 3A and Supplemental Figure S8 reveal a nonrandom distribution of certain dinucleotides across genomic regions that have been aligned at the midpoints of 38,200 nucleosomes. Plots have been separated into the −1, +1, and genomic nucleosomal groups. We first focus on the nucleosomal region located between −73 and +73 bp from the nucleosome dyad. The reciprocal 5′–3′ gradient of AA vs. TT dinucleotides was evident for the −1 and +1 nucleosomes, and was largely absent from genic nucleosomes as evident by the relatively flat line between −73 and +73 bp (Supplemental Fig. S8B).

From these distributions, we calculated that −8% of all nucleosomal dinucleotides (i.e., those between −73 and +73), by virtue of being either present or absent at specific locations, contribute to positioning of the −1 and +1 nucleosomes, whereas only −2% contribute to positioning of the remaining genic nucleosomes (Supplemental Table S4). This translates to −12 and −3.6 dinucleotides per nucleosome for the −1/+1 and genic nucleosomes, respectively. Of these, AA/TT made the largest contributions (approximately three and one per nucleosome, respectively). These findings are in agreement with other studies demonstrating a limited role for positioning sequences genome-wide (Lowary and Widom 1997; Peckham et al. 2007; Shivashawamy et al. 2008), and compare favorably with estimates derived from a smaller data set (Bolshoy et al. 1996). The current analysis does not preclude the possibility that the contribution of dinucleo-
teins that might position nucleosomes (Fedor et al. 1988; Roth et al. 1990). To explore potential contributions from proteins, we looked for enrichment of cis-regulatory element motifs (MacIsaac et al. 2006) in regions having weak intrinsic nucleosome positioning (Ioshikhes et al. 2006). We counted −1900 conserved cis-regulatory elements that reside in the vicinity of the +1 nucleosome for genes that had weak NPS correlations in the same region, and compared it to those with strong NPS correlations. As shown in Figure 3B, genes with weak NPS correlations had 30% more conserved cis-regulatory elements than strong NPS genes ($P < 10^{-13}$). For the same collection of genes no difference in cis-regulatory elements was detected at the −1 nucleosome ($P = 0.6$). Thus, positioning of the +1 nucleosome may be endowed in part by sequence-specific DNA binding proteins in addition to intrinsic positioning sequences and adjacent exclusion sequences.

The distribution of individual classes of cis-regulatory elements is presented in Supplemental Figure S9. Several individual classes of elements stood out. Examples include binding sites for Yap7, which belongs to a class of bZIP stress response regulators (Fernandes et al. 1997), and the zinc-finger transcription factor Stp1 (Fig. 3C) (Wang et al. 1992).

**Nucleosome organization at the 3' end of genes conforms to the principles of statistical positioning**

To further test the concept of statistical positioning, we examined nucleosome positions near the ends of genes, where nucleosomes are generally fuzzy. If statistical positioning were not occurring, we would expect that in regions of predominantly fuzzy nucleosomes, the infrequent well-positioned nucleosome would represent an isolated event. As such, neighboring nucleosomes would not be influenced by the positioned nucleosome and would therefore be fuzzy.

To avoid potential influence of positioning occurring at the 5' end, short genes (<1.3 kb) were excluded from the analysis. The remaining 2027 genes were partitioned into three groups having either highly phased, moderately phased, or very fuzzy terminal nucleosomes, as assessed by their percent rank of fuzziness (standard deviation of read locations). Nucleosome spacing upstream of the terminal nucleosome (i.e., internal to the genes) was then assessed relative to the terminal nucleosome. As shown by the filled plots in Figure 4A, uniformly spaced nucleosomes (well-defined peaks at −180, −340, and −510, with well-defined intervening troughs) were found adjacent to highly phased terminal nucleosomes (filled plot in the lower panel). Fuzzy terminal nucleosomes lacked such uniformity (filled plot in the upper panel). This provides further evidence for statistical positioning as a general rule governing nucleosome organization.

We next looked at nucleosome fuzziness as a function of distance from the terminal nucleosome (Fig. 4B). Relatively low fuzziness persisted over a greater genomic distance from highly phased terminal nucleosomes than the persistence emanating from very fuzzy terminal nucleosomes. Thus, the infrequent occurrence of a positioned nucleosome far from promoter regions creates a higher occurrence of adjacent well-positioned nucleosomes, which is predicted by the statistical positioning model.

**Nucleosome positioning signals at the end of genes are antagonized by cleavage and polyadenylation signals**

To assess whether the phased terminal and adjacent upstream nucleosomes were positioned by the underlying DNA sequence, we generated AA/TT NPS correlation profiles of genic regions that were aligned by their terminal nucleosome. Highly phased terminal nucleosomes possessed stronger NPS correlations (solid trace in lower panel of Fig. 4A) than the very fuzzy terminal nucleosomes (solid trace in upper panel of Fig. 4A). NPS correlations were of similar magnitude to those found at the −1 and +1 nucleosomes (Fig. 2B), indicating that AA/TT positioning sequences account for well-positioned terminal nucleosomes. However, AA/TT positioning sequences were relatively flat in adjacent upstream genic regions, indicating that they contribute little toward positioning of adjacent phased nucleosomes, which further implicates statistical positioning.
The NPS correlation peak was not exactly co-incident with the measured location of the well-positioned terminal nucleosome (solid trace in lower panel of Fig. 4A). We investigated whether this discrepancy was due to the presence of cleavage and polyadenylation signals (AATAAA and related sequences) that populate this region. Due to their similarity to poly(dA:dT) tracts, they might be antagonistic to nucleosome positioning (Peckham et al. 2007). When AATAAA and related sequences were masked (see Methods), the NPS correlation peak not only shifted closer to the expected position (dotted traces), but also became stronger (see Methods), the NPS correlation peak not only shifted closer to the measured location of the well-positioned terminal nucleosome.

Most Saccharomyces genes contain a novel NFR at the end of their genes

While much attention has focused on nucleosome organization at the beginning of genes, little is known about their organization at the end of genes. Interestingly, nucleosome occupancy dropped precipitously at the 3′ end of genes, essentially where the cleavage and polyadenylation site is found (Fig. 5A). A similar observation was recently reported elsewhere (Shivaswamy et al. 2008). The median length of these 3′ terminal NFRs was 173 bp (distance between flanking nucleosome borders, with the downstream nucleosome typically being a part of the next downstream gene or promoter) (Supplemental Fig. S10). 522T of 527 genes (95% of all genes) had an identifiable 3′ terminal NFR of >50 bp. Both the 5′ and 3′ NFR increased in size with highly transcribed genes (Supplemental Fig. S11), indicating that they are dynamic and linked to transcription. The 3′ NFR might be a site for disassembling the terminating RNA polymerase, much as the 5′ NFR is the site for assembly. An additional possibility is that the 3′ NFR is where anti-sense pre-initiation complexes assemble. Consistent with the latter possibility, we identified 71 anti-sense transcripts (Perocchi et al. 2007) that initiated within 300 bp of a 3′ NFR midpoint (Fig. 5B). Thus, for some genes the 3′ NFR might represent the promoter region for anti-sense transcription.

Selective enrichment of TFII B at 3′ NFRs supports a gene looping model

Recent studies suggest that a terminating RNA polymerase II might recycle directly back to its promoter via DNA looping that involves general transcription factor TFII B (Singh and Hampsey 2007). Looping genes contain TFII B at their 5′ and 3′ ends. However, other initiation factors, such as TBP, reside only at the promoter. To explore whether the 3′ NFRs identified here might accommodate such a model, we conducted high-resolution genome-wide location analysis (ChIP-chip) on TFII B and TBP, using microarray probes spaced every 5 bp throughout the S. cerevisiae genome. We searched for genes that met three criteria: (1) The 3′ NFR resided in an intergenic region that either lacked a promoter or was far from a promoter, so as to avoid complications associated with TFII B bound to nearby promoters; (2) TFII B was present at the promoter, which is a requirement of the looping model (Singh and Hampsey 2007); and (3) TFII B was present at the 3′ NFR, but TBP was absent. The absence of TBP selects against 3′ NFRs that might be sites of anti-sense initiation.

Remarkably, 120 genes met the looping prediction criteria ($P < 10^{-5}$; Fig. 6A; Supplemental Table S6). At these genes, TFII B was found just downstream from the cleavage and polyadenylation site (Fig. 6B), which is where RNA polymerase II is likely to terminate transcription. These data meet one prediction of the gene looping model. If reflective of gene looping, the results suggest that the coordination of initiation and termination via gene looping between the 5′ and 3′ NFRs may be common in yeast, although not necessarily universal. The physiological importance of looping remains to be determined. We found nothing extraordinary about the transcription frequency or gene ontologies of these genes (data not shown). Taken together, our results suggest that the 3′ NFR may be as “busy” a place as the 5′ NFR during the transcription cycle.

Discussion

We propose that the genomic organization of nucleosomes in Saccharomyces largely originates from the 5′ end of genes, where a combination of poly(A:T) exclusion sequences in the core promoter, AA positioning sequences near the 5′ end of nucleosomal DNA (TT near the 3′ end), AT and TA dinucleotides near the nucleosome borders, and transcription factor binding sites...
define the location of the −1/+1 nucleosomes and the nucleosome-free promoter region (Fig. 7). The +1 nucleosome establishes a barrier that largely relies on the principles of statistical positioning to position densely packed downstream nucleosomes for as far as −1 kb, with little contribution from the underlying DNA sequence. Most yeast genes are −1 kb or less. Beyond 1 kb, genic nucleosomes lack positional relationships with the TSS and tend to be less phased than the +1 nucleosome. The DNA entering and exiting the downstream nucleosomes tends to be more susceptible to MNase digestion, suggesting that the DNA is more loosely associated with these nucleosomes. Conceivably, positioning sequences near the 5' end of genes, or histone modifications/variants that are enriched at the 5' end of genes, could contribute to a “stronger” wrap of the DNA helix on the histone surface.

Relocation of the +1 nucleosome by chromatin remodeling complexes would therefore also relocate an array of downstream nucleosomes in the same direction, creating a mechanism for long distance control through a single nucleosome. However, this would need to be tested directly, perhaps by using positioning sequences to reposition a nucleosome and examining its impact on adjacent nucleosomes. Recent studies have shown that the ISW2 complex repositions nucleosome arrays in the yeast genome (Whitehouse et al. 2007). If the principles of statistical positioning are in play, ISW2 needs to act on only one nucleosome in each array in order to reposition the entire array. Nevertheless nucleosome mobility in general may be facilitated by global chromatin remodeling complexes such as RSC (Parnell et al. 2008). A second observation that appears to be a manifestation of statistical positioning is the repositioning of downstream nucleosomes when a paused RNA polymerase II contacts and repositions the +1 nucleosome in *Drosophila* (Mavrich et al. 2008).

While the entire length of nucleosomal DNA contributes to its overall affinity for the histone core, by our calculations only a few dinucleotides contribute to positioning. Of these, AA/TT contribute the most, with −3 AA or TT dinucleotides helping position the +1 nucleosome. This contribution seems rather small to impart intrinsic curvature over the entire length of nucleosomal DNA. Thus, intrinsic DNA curvature, if present, might be limited to −50 bp regions internal to where the DNA enters and exits the nucleosome. This would allow more stable packaging of the −1 and +1 nucleosome.

Toward the 3' ends of genes, nucleosomes progressively acquire more random locations relative to the TSS, and are less phased. Nucleosome occupancy drops precipitously at the 3' end of genes, forming a 3' NFR. The 3' NFR may participate in transcription termination, anti-sense initiation, and recycling of RNA polymerase to the promoter by DNA looping. Thus, as much as the −1 and +1 nucleosomes may be important regulators of transcription initiation, the terminal nucleosomes that flank the 3' NFR may be important regulators of transcription termination. AATAAA and related elements play a clear role in transcription termination by promoting cleavage and polyadenylation of the nascent mRNA transcript. These elements might also facilitate termination by antagonizing nucleosome formation over the termination region. Whether and how the terminal nucleosome of a gene is linked to termination remains to be determined.

**Methods**

**Nucleosome mapping in vivo**

Three independent nucleosome preparations were made from BY4741 strains containing C-terminal TAP tags on histone H3 or H4. Maps produced from H3 were indistinguishable from H4 maps, and so the data were combined. Details of formaldehyde crosslinking, MNase digestion, TAP immunopurification, gel purification of mono-nucleosome sized DNA, DNA sequencing using the Roche GS20 (454 Life Sciences), mapping to genomic...
coordinates from www.yeastgenome.org, correction for MNase bias, and making calls on nucleosome locations are described elsewhere (Albert et al. 2007), with the following modifications: 0.05% sodium dodecyl sulfate was included in the ChIP. Soluble chromatin was prepared from 3 L of cells, precleared with Sepharose 4B for 15 min at 4°C, then immunoprecipitated with 600 μL of bed volume IgG Sepharose Fast Flow. Nucleosome core particles were eluted with 3 KU TEV Protease at 25°C for 2 h. Raw sequencing reads can be accessed through NCBI Trace Archives T1 SRA010003, Sequencing Center “CCGB.” Bulk downloads or specific queries of nucleosome positions can be accessed from http://atlas.bx.psu.edu or Supplemental Table S1.

**Nucleosome distributions**

For every “feature” start or end coordinate, defined by the *Saccharomyces* Genome Database (as of May 13, 2007), a search region was defined using the following criteria. First, minimum and maximum search distances were set at ±300 bp and 3000 bp, respectively, relative to the feature coordinate. Second, between the minimum and maximum, the actual range was set to be 300 bp from the next feature. Examples of “the next feature” include start or end coordinates of ARS elements, retrotransposons, telomeres, snRNAs, tRNAs, and other ORFs or mRNA transcripts (Supplemental Table S2) (David et al. 2006). A 300 bp border from the next feature was chosen so as to minimize inclusion of potential −1/NFR/+1 nucleosomal regions, which tend to be within 300 bp of an ORF start site or TSS. Third, the nucleosome count along the chromosome coordinate (bin size = 20 bp) in the allowed region for each feature was determined and summed across all bins having the same distance from the feature. Fourth, the nucleosome count was normalized to the number of regions in each bin (displayed in Figs. 1C, 5A, and Supplemental Fig. S4), then multiplied by 100. Line plots of three bin moving averages were then drawn representing the normalized bin counts at varying distances from the feature coordinate. Distances were measured to the center of each bin.

**NPS correlation plots**

NPS correlation profiles were calculated using a derivative of an earlier AA/TT pattern (Supplemental Fig. S6A) (Ioshikhes et al. 1996). The updated pattern was derived by first acquiring 2285 of the most highly positioned H2A.Z-containing nucleosomes identified elsewhere (Supplemental Table S3) (Albert et al. 2007) and −200 other well-defined nucleosomes (Ioshikhes et al. 1996). The sequence of both DNA strands was filtered to retain a subset that positively correlated with the original AA/TT patterns (Ioshikhes et al. 1996, 2006). From that subset a new empirical distribution pattern of AA and TT dinucleotides was determined (Supplemental Fig. S6B). This new pattern was used to refilter the subset for positively correlating sequences, and a newer AA/TT pattern derived. After six iterations, new patterns originating from the two nucleosome data sets became virtually indistinguishable. The new pattern (Supplemental Fig. S6C) was then used to calculate the NPS correlation profiles as described elsewhere (Ioshikhes et al. 2006). Distances represent the distance from the feature to the center of the NPS correlation pattern. When masking sequences related to AATAAA (at the 3’ end of genes), trinucleotides AAT, TAA, and AAA located downstream from the terminal nucleosome were replaced by NNT, TNN, and NNA, respectively, and the NPS correlation was calculated.

**Definition of 3’ NFR**

A 3’ NFR is defined here as the closest linker DNA border to the 3’ end of the ORF. For purposes of “closeness” we masked those linkers that were <50 bp in length. If no linker >50 bp long was within 500 bp of the ORF end point, then the closest linker that was <50 bp was used as the 3’ NFR. This strategy therefore preferentially looks for NFRs that are >50 bp, but if it does not find one, it will allow one that is <50 bp in length.

**Genome-wide survey for gene looping using TFIIB and TBP ChIP-Seq**

TFIIB (Sua7-TAP) and TBP ChIP material was prepared by standard methods (Zanton and Pugh 2006), and hybridized to Affymetrix *S. cerevisiae* Tiling 1.0R Arrays. Signal analysis, interval analysis, and peak calling were performed using Model-based Analysis of Tiling-arrays software (Johnson et al. 2006b). The bandwidth and max gap parameters had a value of 75 bp, and peaks were called using a significance threshold of P < 10−2. To minimize false positives due to confounding contributions from nearby promoters, genes were filtered using the following criteria: (1) only the 5082 genes with experimentally determined transcription termination sites (TTS) were used (David et al. 2006), (2) genes whose TTS was within 300 bp of a downstream ORF site were removed, and (3) genes containing a TBP peak...
within ± 150 bp of the TTS were removed. 2679 TFIIB peaks were mapped to the promoter and/or TTS for the 4151 genes that met the above criteria. Promoter regions were arbitrarily defined as the 300 bp window immediately upstream of the ORF start site, and the TTS window was 150 bp upstream of, and downstream from the TTS. TFIIB occupied both the promoter region and the TTS of 120 genes (Supplemental Table S6). The CHiTST test in Excel was used to determine the probability that TFIIB would fall within the promoter region and the TTS of the same gene by chance. The distance from the nearest ORF end site was calculated for the 224 Sua7 peaks that were found within 150 bp of the TTS. The composite distribution profile in Figure 6B was generated by binning these distance values in 50 bp intervals. Microarray data are deposited in ArrayExpress under accession no. E-MEXP-1584.

Acknowledgments

This work was supported by a grant from NIH (HG004160). We thank S. Zanton for help in generating figures. We thank J. Reese, D. Gilmour, S. Tan, and T. Koerber for many helpful comments.

References


Aloise, A., Rheinlaender, F., Merlo, C., and Trifonov, E.N. 1997. The composite distribution profile in Figure 6B was generated by binning these distance values in 50 bp intervals. Microarray data are deposited in ArrayExpress under accession no. E-MEXP-1584.

 element that stimulates transcription via its intrinsic DNA structure. EMBO J. 14: 2570–2579.


Nucleosome organization in yeast


Received March 9, 2008; accepted in revised form April 15, 2008.