Heavy transcription of yeast genes correlates with differential loss of histone H2B relative to H4 and queued RNA polymerases

Hope A. Cole, Josefina Ocampo, James R. Iben, Răzvan V. Chereji and David J. Clark*

Program in Genomics of Differentiation, Eunice Kennedy Shriver National Institute for Child Health and Human Development, National Institutes of Health, Bethesda 20892, MD, USA

ABSTRACT

Eukaryotic chromatin is composed of nucleosomes, which contain nearly two coils of DNA wrapped around a central histone octamer. The octamer contains an H3-H4 tetramer and two H2A-H2B dimers. Gene activation is associated with chromatin disruption: a wider nucleosome-depleted region (NDR) at the promoter and reduced nucleosome occupancy over the coding region. Here, we examine the nature of disrupted chromatin after induction, using MNase-seq to map nucleosomes and subnucleosomes, and a refined high-resolution ChIP-seq method to map H4, H2B and RNA polymerase II (Pol II) genome-wide. Over coding regions, induced genes show a differential loss of H2B relative to H4, which correlates with Pol II density and the appearance of subnucleosomes. After induction, Pol II is surprisingly low at the promoter, but accumulates on the gene and downstream of the termination site, implying that dissociation is very slow. Thus, induction-dependent chromatin disruption reflects both eviction of H2A-H2B dimers and the presence of queued Pol II elongation complexes. We propose that slow Pol II dissociation after transcription is a major factor in chromatin disruption and that it may be of critical importance in gene regulation.

INTRODUCTION

The basic subunit of chromatin is the nucleosome core, which contains ~147 bp of DNA wrapped in ~1.7 negative superhelical turns around a central core histone octamer (1). The octamer is composed of two molecules each of H3, H4, H2A and H2B. In vitro, nucleosome assembly on DNA by salt-gradient dialysis is a two-step process, involving the initial binding of an (H3-H4)2 tetramer to the DNA, followed by the binding of two H2A-H2B dimers, one on each side of the tetramer, to complete the nucleosome (1,2). The high stability and compact structure of the nucleosome, together with limited DNA accessibility due to steric hindrance by histones, result in inhibition of DNA-dependent processes such as transcription, replication and repair.

The question of how RNA polymerase II (Pol II) negotiates the nucleosome is of particular interest. In vitro, Pol II transcribes through a nucleosome causing the loss of a single H2A-H2B dimer, leaving behind an (H3-H4)2/H2A-H2B hexamer (3). Thus, the passage of polymerase causes partial disassembly of the nucleosome. This effect is exacerbated at high Pol II density, where multiple passages by Pol II result in loss of the entire nucleosome (4). In vivo, various elongation factors remove and replace histones during transcription, resulting in dynamic histone exchange from coding regions and accelerating transit of Pol II through nucleosomes (5–14). Single-gene studies suggest that complete nucleosomes are displaced from highly transcribed genes (11,12). More generally, heavy transcription in vivo correlates with chromatin disruption, both at the promoter and on the gene (15–21). At the promoter, the nucleosome-depleted region (NDR) is wider and deeper, partly due to depletion of upstream nucleosomes. On the coding region, nucleosome occupancy decreases, nucleosomes tend to pack closer together and phasing is disrupted (15–21).

Previously, we have identified a set of genes in budding yeast which exhibit a major change in chromatin structure after induction, involving a loss of nucleosome occupancy across the coding region that can extend into flanking genes (19). These genes represent a subset of the genes that are induced by 3-aminotriazole (3AT), an inhibitor of the enzyme encoded by HIS3, which elicits the starvation response mediated by translational control of the Gcn4 transcriptional activator: GCN4 mRNA is rapidly translated in response to amino acid starvation (22,23). Specific binding of Gcn4 to the promoters of many genes results in activation mediated by various cofactors, including the SWI/SNF chromatin remodeling complex and the SRB/mediator (24). Here, we examine the structural basis of this chromatin disruption by...
comparing genome-wide maps of nucleosomes and subnucleosomes (defined as histone-containing particles smaller than the core particle) prepared using MNase, with maps of histones and Pol II (prepared using sonication) in untreated and 3AT-induced cells. We show that chromatin disruption reflects a differential loss of H2A-H2B dimers relative to H3-H4 tetramers that correlates with the appearance of 3AT-induced cells. We show that chromatin disruption is proportional to Pol II density. Remarkably, the distribution of Pol II is such that it is lowest at the transcription start site (TSS) and highest on the distal coding region and at the 3'-end, where it overshoots the transcript end site (TES) and accumulates, sometimes invading the gene downstream. Thus, Pol II elongation complexes appear to queue up on the gene, waiting to complete transcription and dissociate.

MATERIALS AND METHODS

Yeast strains

The HIS3 gene is required for 3AT induction. YDC111 (MATa ade2-1 can1-100 leu2-3,112 trpl-1 ura3-1) (25) was used for MNase digestion and Pol II experiments. For H4 experiments, YDC439 (W303 MATa ade2-1 can1-100 leu2-3,112 trpl-1 ura3-1 hht1-hhf1Δ::LEU2 URA3::HHT2-HA-HHHF2) was constructed by transformation of BUY815 (a strain carrying a single HA tag at the N-terminus of H4; the gift of Namrita Dhillon and Rohinton Kamakaka) with an EcoRI fragment containing HIS3, as described (25). For H2B experiments, YDC443 (W303 MATa ade2-1 can1-100 his3-11 leu2-3,112 lys2Δ trpl-1 ura3-1 hta2-hht2Δ::TRP1 HA-6His-HTB1::HIS3) was constructed in two steps. Firstly, wild-type strain JRY4012 (26) was transformed with a polymerase chain reaction (PCR) fragment to delete the HTA2-HTB2 locus by replacing it with TRP1, to obtain YDC417. Secondly, YDC417 was transformed with a Pci-EcoRI digest of p676 to integrate a 6-His tag followed by a single HA tag at the N-terminus of H2B, at the HTB1 locus. p676 was obtained by insertion of a 6-His and an HA tag (VSH-HHHHHGSMYPDVYPDAGSM) between the first and second codons of HTB1 in p592 (27). All inserts and tags were verified by PCR analysis and sequencing.

Paired-end sequencing of chromatin immunoprecipitates

YDC111, YDC439 and YDC443 were grown to log-phase either in synthetic complete (SC) medium (untreated) or in SC-histidine followed by induction with 10 mM 3AT (Sigma 61-82-5) for 20 min. Cells were fixed with 3% formaldehyde at room temperature and the lysate was spun for 10 min at 14 000 rpm and 4°C. The pellet was resuspended in 1 ml FA buffer and spun again. The final pellet was resuspended in 0.5 ml FA buffer and sonicated using a Diagenode Bioruptor 300 at 4°C with 40 cycles of 30 s high power and 30 s rest, with a 15-min cooling period after 20 cycles. Debris was removed by spinning for 20 min at 14 000 rpm and 4°C. The supernatant was aliquoted and stored at −80°C. DNA was quantified using the Hoechst assay. For IP, 2.5 μg DNA was adjusted to 250 μl with FA buffer, any insoluble material was removed with a brief spin and the supernatant was added to 30 μl anti-HA agarose (Sigma A2095) that had been washed three times with PBS. Binding was performed in an Eppendorf Thermomixer for 2 h at 1000 rpm and 4°C. The supernatant was removed. The HA-agarose was washed twice with FA buffer and three times with S-RIPA buffer (10 mM Tris–HCl pH 8.0, 0.14 M NaCl, 1% Na-deoxycholate, 1% Triton X100, 0.1% sodium dodecyl sulphate (SDS)) in the Thermomixer (5 min, 1000 rpm, 4°C). Bound DNA was eluted with 50 μl HA-peptide (Roche 1166975001) at 0.1 mg/ml S-RIPA buffer in the Thermomixer (30 min, 1000 rpm, 4°C). The eluate was removed, the elution was repeated, the eluates were pooled and adjusted to 150 μl and ethylenediaminetetraacetic acid (EDTA) was added to 5 mM. Cross-links were reversed by incubation at 65°C overnight, followed by protease K digestion (0.25 mg/ml, 5 h, 55°C). The digest was cooled, SDS was added to 1% and mixed, one-fifth volume of 5 M potassium acetate was added, and extracted twice with an equal volume of chloroform. DNA was precipitated with 0.7 volume of isopropanol in the presence of 5 μg glycogen, washed once with 70% ethanol and dissolved in 15 μl TE (10 mM Tris–HCl pH 8.0, 0.1 mM NaEDTA). IP-DNA was repaired as described (28), purified using a Qiagen PCR column and eluted with 50 μl TE. A 5'-phosphate and a 3'-dA were added, followed by the Illumina paired-end adaptor (#1005711) (28). Purified IP-DNA (4 μl) was subjected to PCR (21 cycles) in 50 μl with Illumina paired-end primers (InPE1.0 and InPE2.0) and an index primer at the recommended concentrations (Illumina). DNA was purified using a Qiagen PCR column and its concentration measured by A260 (10–30 μg/ml was in the linear range). Input DNA (~10 μg) was adjusted to 200 μl with FA buffer and then to 20 mM Tris–HCl pH 8.0 and 5 mM EDTA. Cross-links were reversed as described above. The purified DNA was incubated in 15 μl TE with 0.1 mg RNase/ml for 2 h at 37°C. The volume was adjusted to 50 μl with 0.2 M NaCl in TE. The DNA was eluted from a Qiagen PCR column with 50 μl TE and quantified by A260. Input DNA (0.8 μg) was prepared for paired-end sequencing as above; PCR (17 cycles) was performed according to Illumina, using ~20 ng input DNA. DNA was purified by Qia gen PCR column and its concentration measured by A260. IP-DNA from YDC111 (no tag control) was prepared for sequencing in parallel, but little final product was obtained (not sequenced). Pol II paired-end sequencing of chromatin immunoprecipitate (PESC1) was performed as described for histones, except that an Rpb3 monoclonal antibody (Neoclonal WP012) was bound to Protein A Sepharose beads (GE Healthcare 17-5280-01). For IP, 4 μg anti-Rpb3 was mixed with 30 μl beads (1:1 slurry) for 1 h at room temperature and washed three times with phosphate buffered saline (PBS) containing IgG-free BSA at 5 mg/ml before adding
2.5 (Experiment 1) or 6.25 μg (Experiment 2) DNA, adjusted to 250 μl with FA buffer. Controls without antibody were also done. The IP was performed as above, except for the elution, for which the beads were resuspended in 150 μl 0.14 M NaCl, 50 mM Tris–HCl pH 8, 5 mM Na-EDTA, 0.5% SDS. PCR amplification required ~30 cycles.

**Nucleosome sequencing and data analysis**

MNase experiments were carried out as described (28). DNA was prepared for Illumina paired-end sequencing (50 nt reads) and mapped as described (28,29). Occupancy maps were constructed using raw data smoothed using a 21-bp window moving average filter and normalized to the total number of read pairs in the untreated sample. Phasing analysis was performed using custom Perl programs (available on request (30)) combined with TSS and TES data (31). PESC data were normalized to the genome average of the sample; the IP value at each chromosomal position was then divided by the input value at the same position. Plots of PESC data for individual genes represent average values for each gene from two biological replicate experiments. Paired-end data are summarized in Supplementary Table S1.

**RESULTS**

Previously, we mapped nucleosomes genome-wide in yeast cells by digesting nuclei with MNase and subjecting purified core particle DNA to paired-end sequencing, which provides the lengths of all the DNA fragments, after alignment to the yeast genome. Induction by 3AT results in a major loss of nucleosome occupancy from 49 genes, including the well-studied ARG1 gene (19) (Figure 1A). This loss of occupancy occurs over the entire coding region and, in some cases, spreads a significant distance into a flanking gene. We sought to distinguish between three possible explanations: (i) Simple nucleosome loss. In this model, induction results in the complete removal of a large fraction of the nucleosomes on these genes, resulting in stretches of protein-free DNA that are highly accessible to MNase. (ii) Eviction of H2A-H2B dimers from nucleosomes. In this model, (H3-H4)2 tetramers and/or (H3-H4)2/H2A-H2B hexamers remain on the gene after induction, protecting significantly shorter lengths of DNA than the intact nucleosome. (iii) Nucleosomes are conformationally altered after induction, such that their DNA is more accessible and digested faster by MNase. The three models differ in their predictions concerning the effect of induction on the sizes of DNA fragments protected from MNase digestion and which histones are retained or lost. Here we test these predictions, using ARG1 as a specific example and a set of strongly induced genes to obtain a general perspective.

**Increased levels of subnucleosomal particles on induced ARG1**

Our nucleosome occupancy maps indicate the distribution and relative levels of canonical nucleosomes, as defined by protection of ~147 bp from MNase. If other particles are present, protecting different lengths of DNA, they would not be included, because the nucleosomal DNA was gel-purified prior to mapping. The models predict different degrees of MNase sensitivity: nucleosome loss would result in stretches of protein-free DNA and rapid digestion. Loss of H2A-H2B dimers would result in tetrasomes and/or hexasomes, which confer relatively weak protection of ~80 and ~110 bp, respectively (2,32). The conformational change model predicts rapid destruction of nucleosomes due to a more accessible structure.

To address this question, we subjected significantly underdigested chromatin from untreated and induced cells to paired-end sequencing without gel purification. These samples displayed a typical nucleosome ladder, with bands corresponding to mono, di and trinucleosomes (Figure 2A). For comparison, we also mapped gel-purified core particle DNA from the same digestion series. The length distributions of the core particle samples showed a major peak at
Peak was strongly reduced, and downstream of ARG1 induction resulted in a major loss of occupancy over particle data confirmed our previous observations (19). YOL057W short DNA fragments represent partially digested protein-free DNA relative to chromatin, it is unlikely that these DNA were significant amounts of DNA shorter than the core particle DNA (MNase limit digest). (A) Lengths of DNA fragments derived from the gel in (A), as determined by paired-end sequencing. The length cut-off at 50 bp reflects the sequence read length of 50 bp.

Figure 2. Length distributions of MNase-digested DNA from untreated and 3AT-induced cells. (A) Analysis of DNA extracted from nuclei digested with different amounts of MNase in an agarose gel stained with ethidium bromide. The underdigested samples chosen for analysis represent a compromise between the requirements for preservation of MNase-sensitive particles (short digestion times) and for identifying the particles by their lengths, which is improved by removing protruding linker DNA (longer digestion times). The red box shows where the gel was cut to obtain core particle DNA (MNase limit digest). (B) Lengths of DNA fragments from the gel in (A), as determined by paired-end sequencing. The length cut-off at 50 bp reflects the sequence read length of 50 bp.

148 or 149 bp, corresponding to core particles, and a minor peak at 128 bp, corresponding to digestion within the nucleosome (Figure 2B). The underdigested samples gave a strong mononucleosome peak composed of subpeaks at 150, 159 and 166 bp, corresponding to core particles and mononucleosomes with some protruding linker DNA, still present because the relatively slow exonuclease activity of MNase did not have time to remove it. In addition, there were significant amounts of DNA shorter than the core particle, although there were no clearly defined peaks at ∼80 or ∼110 bp perhaps because, like mononucleosomes, tetrasomes and hexasomes have variable amounts of protruding DNA. Given the rapid rate at which MNase digests protein-free DNA relative to chromatin, it is unlikely that these short DNA fragments represent partially digested protein-free DNA.

Nucleosome maps of ARGI constructed using the core particle data confirmed our previous observations (19): induction resulted in a major loss of occupancy over the entire ARGI locus, extending both upstream of the NDR at the ARGI promoter, such that the -1 nucleosome peak was strongly reduced, and downstream of ARGI, into YOL057W, which has no known functional connection to ARGI (Figure 1A). However, the equivalent occupancy map for underdigested ARGI obtained using DNA molecules of 50–180 bp showed little loss of occupancy in induced cells (Figure 1B), indicating that the DNA fragments missing from the induced core particle map (Figure 1A) were present in the underdigested sample (Figure 1B). To identify these DNA fragments, we sorted the underdigested samples into mononucleosomes (140–180 bp) and subnucleosomes (80–120 bp, reflecting the expected size range for tetrasomes and hexasomes) in silico. Mononucleosomes gave the same result as core particles (Figure 1C). However, subnucleosomes increased over the ARGI coding region after induction (Figure 1D). Thus, the loss of canonical nucleosomes from ARGI after induction can be accounted for by an increase in subnucleosome-sized particles, which might derive from hexasomes and tetrasomes, or from partially digested conformationally altered nucleosomes. It is important to note that particles containing 80–120 bp DNA could be subnucleosomes (tetrasomes and hexasomes), or large non-histone complexes (see below).

The maps also revealed that the NDRs at the ARGI and YOL057W promoters were occupied in the underdigested samples (Figure 1B). Both NDRs were apparent when only mononucleosomes were considered (Figure 1C), but there were peaks in subnucleosomal DNA fragments (Figure 1D), indicating that particles of subnucleosome size are present at the NDRs, as observed previously (33,34).

**Differential loss of H2B relative to H4 from induced genes**

The three models make different predictions concerning the histones bound to induced genes: nucleosome removal would result in loss of all four histones, dimer loss predicts loss of H2A and H2B but not of H3 and H4, whereas a conformational change in the nucleosome predicts retention of all four histones. To determine which histones are bound, we performed a modified ChIP-seq (‘Paired-end sequencing of chromatin immunoprecipitates (PESCIs)’) experiment using yeast strains with HA-tagged H2B or HA-tagged H4. Untreated and induced cells were fixed with formaldehyde in the usual way and sonicated to obtain short DNA fragments. After immunoprecipitation (IP), the DNA was subjected to paired-end sequencing. Histone occupancy maps can be constructed for direct comparison with those of MNase-digested samples. Furthermore, the resolution of the IP is higher for shorter DNA fragments and can be improved by selection in silico of just the shortest DNA fragments to construct the map. The length distributions of the PESCIs inputs and IPs showed that the inputs had peaks at ∼90 bp, with a wide range of DNA fragments from 50 to 500 bp, whereas the IP samples were longer on average, with peaks at ∼125 bp (Supplementary Figure S1). The reason for this difference in average length is unclear, although no difference was observed for the Rpb3 subunit of Pol II (see below). The partially cross-linked nucleosome might be somewhat protected against sonication, although the peak fragment length is well below the size of a nucleosome. There is also uncertainty over whether uncross-linked histones are removed from the DNA in the sonication buffer, which contains a mixture of detergents.

PESCIs maps for H4 and H2B at ARGI constructed using only 50–150 bp fragments showed that there was a major loss of both histones from ARGI after induction, with a significantly greater loss of H2B than of H4 over the cod-
Figure 3. Induction of ARG1 by 3AT results in some loss of H4 and a major loss of H2B over the coding region. PESCI ChIP for H4 and H2B. Occupancy maps constructed using 50–150 bp fragments (normalized to the total aligned read-pair count for untreated cells and smoothed with a 21-bp moving average). Untreated (black line) and induced cells (green line). Inputs are shown in separate panels for clarity.

In conclusion, our histone PESCI experiments indicate that H2B is differentially lost from induced genes, suggesting that many nucleosomes have lost one or both of their H2A-H2B dimers. The fact that a few genes are also depleted of H4 indicates that some nucleosomes are completely removed from these genes. Thus, some complete nucleosomes (H4 loss) and more H2A-H2B dimers are lost compared with untreated cells. Only the three most disrupted genes as measured by the MNase assay were obviously depleted of H4 after induction (YPR036W-A, ICY2 and ARG1) (Figure 4A). The other genes were either unaffected or perhaps slightly depleted of H4. In contrast, most of the 49 genes were substantially depleted of H2B in induced cells (Figure 4B). Differential loss of H2B relative to H4 after induction was clearly demonstrated in plots of H2B occupancy against H4 occupancy for untreated and induced cells (Figure 4C and D). Moreover, H2B occupancy correlated well with the extent of chromatin disruption measured by MNase (Supplementary Figure S2).

In conclusion, our histone PESCI experiments indicate that H2B is differentially lost from induced genes, suggesting that many nucleosomes have lost one or both of their H2A-H2B dimers. The fact that a few genes are also depleted of H4 indicates that some nucleosomes are completely removed from these genes. Thus, some complete nucleosomes (H4 loss) and more H2A-H2B dimers are lost from ARG1, ICY2 and YPR036W-A after induction, but most induced genes show significant loss of H2B only.

Figure 4. Differential loss of H2B relative to H4 from highly induced genes. Analysis of 49 genes exhibiting major chromatin disruption (19). Average H4 and H2B occupancies were calculated for the coding region of each gene. The average value for each gene obtained from two biological replicate experiments was plotted. (A) H4 occupancy in untreated and induced cells. (B) H2B occupancy in untreated and induced cells. (C) H2B versus H4 occupancy in untreated cells. (D) H2B versus H4 occupancy in induced cells. Dashed line indicates no change in occupancy. Note: LYS2 is low in the HA-H2B strain because part of the gene is deleted (lys2A).
waiting to dissociate from the DNA and that dissociation is rate-limiting for transcription of ARGI and HIS4.

To determine the generality of our observations for ARGI and HIS4, we identified strongly induced genes by comparing the average Pol II occupancy on the coding regions of all genes in untreated and induced cells (Figure 5B). In induced cells, ARGI was actually the most heavily transcribed gene, and HIS4 was also very heavily transcribed. Of the 49 genes that show reduced nucleosome occupancy after induction (>25% loss) mentioned above (19), 35 have associated TSS/TES data (31); 28 of these 35 (80%) were induced >2-fold (defined by increased Rpb3 occupancy between the TSS and the TES). We analyzed the Pol II distribution on the 244 genes that were induced >2-fold (above the upper dashed line in Figure 5B) with coding regions >500 bp (Figure 5C). Relative to the TSS, the level of Pol II at the promoter was low and increased to a high plateau over distal coding regions in induced cells. Relative to the TES, the level of Pol II was high over the distal coding region, there was a slight decrease near the TES itself and an overshoot peak ~140 bp downstream of the TES in induced cells. The implication is that Pol II elongation complexes upstream of the TES are queued, waiting for the terminating Pol II to dissociate.

High Pol II density correlates with differential loss of H2B

The relationship between high levels of transcription and loss of H2A-H2B dimers was addressed by correlating the occupancy levels of H4 and H2B with those of Pol II on induced genes. There was no correlation between the levels of H4 and Pol II in untreated or induced cells, with little loss of H4 from any induced genes except ARGI (Figure 6A). This was also true for H2B in untreated cells, because induced genes are not heavily transcribed in the absence of 3AT. In contrast, there was a strong inverse correlation between H2B and Pol II levels in induced cells (Figure 6A). Thus, Pol II density correlates with loss of H2A-H2B dimers.

Analysis of nucleosome phasing on induced genes revealed substantial 3AT-dependent disruption on both sides of the TSS, with reduced core particle signal consistent with loss of nucleosomes from these genes (Figure 6B), as observed previously (20). Increased levels of 80–120 bp particles were present in the NDR and upstream of the NDR (Figure 6C). If inverted, the shape of the 80–120 bp particle peak is an approximate match for the shape of the NDR (Figure 6B), implying that these particles fill the NDR. A similar analysis of H4 occupancy revealed H4 phasing in untreated cells (Figure 6D), providing confirmation of nucleosome phasing independent of MNase. H4 phasing was weaker in induced cells, with a slight reduction in the +1 nucleosome and a slight expansion of the NDR upstream, as observed for core particles (Figure 6B). However, unlike core particles, there was little loss of H4 over the coding region. In contrast, H2B occupancy was significantly reduced over the coding region (P-value < 0.001, Welch’s t-test) and upstream in induced cells (Figure 6E). Thus, H2B occupancy correlates more precisely with core particle occupancy than H4, as expected if there is a differential loss of H2A-H2B dimers, resulting in disassembly of core particles. In untreated cells, the much weaker phasing of H2B relative

(Supplementary Figure S3A). Induction resulted in a huge increase in the amount of Pol II bound to the ARGI coding region (Figure 5A). In contrast, there was relatively little Pol II at the ARGI TSS. Furthermore, Pol II was very high even downstream of the ARGI TES, suggesting that Pol II over- shot the TES and invaded YOL057W. This overshoot was not due to long DNA fragments in the IP, because the Pol II peak extends up to ~350 bp downstream of the ARGI TES and the map is composed of 50–150 bp fragments.

An alternative explanation for the Pol II peak at the YOL057W promoter is that it represents Pol II waiting to transcribe YOL057W, but this is unlikely because the peak was not present in untreated cells. Another interpretation is that the overshoot peak represents paused Pol II ready to transcribe the ARGI anti-sense transcript, which is initiated in the 3’-UTR (36). This possibility is more difficult to rule out, but the same observations were true of HIS4, another gene strongly induced by 3AT (Supplementary Figure S3B), which has no anti-sense transcript (37). Instead, we propose that the Pol II overshoot peak represents Pol II...
Heavily transcribed genes also show differential loss of H2B and queued polymerases

We have shown that, for induced genes, Pol II density correlates with loss of H2B (Figure 6). Is this effect peculiar to genes induced by 3AT, or is it a general property of heavily transcribed genes? We constructed scatter plots to compare the average H2B occupancy of a gene with its average H4 occupancy in untreated cells (Figure 7A) and in induced cells (Figure 7B). Most, but not all, of the genes had occupancies close to the normalized average value of 1. The histone occupancy values of the 244 inducible genes in untreated cells were superimposed on the scatter plot for untreated cells (compare Figure 7C with Figure 7A): these genes show a distribution of values similar to that for all genes, as expected, because most of them are poorly transcribed in the absence of 3AT. For comparison, the histone occupancy values of the 244 induced genes in untreated cells were superimposed on the scatter plot for treated cells (compare Figure 7D with Figure 7B): many induced genes shift to lower H2B occupancy and a few also show lower H4 occupancy. Some genes show little shift (as in Figure 6A), most likely because a high level of transcription does not necessarily result in a high level of transcription: a gene can be strongly induced but still only transcribed at moderate levels. Consequently, we examined heavily transcribed genes in both untreated and induced cells. We identified the 250 genes with the highest Pol II density in untreated cells and in induced cells; 123 genes appeared in both lists (49%), indicating that they are heavily transcribed independently of 3AT. Their histone occupancy values were superimposed on the scatter plots for untreated and induced cells (compare Figure 7E with 7A and Figure 7F with 7B). In both cases, it is clear to H4 might reflect the fact that the two H2B molecules in a nucleosome are farther apart than the two H4 molecules, resulting in a broader signal (clear phasing of H2B was observed using dyad phasing analysis, which has higher resolution; see below). Finally, both H4 and H2B (Figure 6D and E) showed a trough at the NDR similar to that observed for core particles (Figure 6B), with no peak corresponding to the subnucleosomal particles (Figure 6C), indicating that the 80–120 bp particles in these NDRs do not contain H4 or H2B and are therefore non-histone complexes.

<table>
<thead>
<tr>
<th>Table</th>
<th>Untreated cells</th>
<th>Inducible genes</th>
<th>Heavily transcribed genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>All genes</td>
<td>Inducible genes</td>
<td>Heavily transcribed genes</td>
</tr>
<tr>
<td>B</td>
<td>Untreated</td>
<td>Inducible</td>
<td>Heavily transcribed</td>
</tr>
<tr>
<td>C</td>
<td>Untreated</td>
<td>Inducible</td>
<td>Heavily transcribed</td>
</tr>
<tr>
<td>D</td>
<td>3AT-induced</td>
<td>Inducible</td>
<td>Heavily transcribed</td>
</tr>
<tr>
<td>E</td>
<td>Heavily transcribed</td>
<td>Heavily transcribed</td>
<td>Heavily transcribed</td>
</tr>
</tbody>
</table>

Figure 7. Heavily transcribed genes show reduced levels of H2B relative to H4. Scatter plots comparing H2B occupancy with H4 occupancy for all genes longer than 500 bp: (A) Untreated cells; (B) 3AT-induced cells; (C) As in (A), but with the 244 inducible genes (black circles) superimposed on the plot for all genes (gray circles); (D) As in (B), but with the 244 3AT-induced genes superimposed (green circles); (E) As in (A), but with the 250 most heavily transcribed genes in untreated cells superimposed (red circles); (F) As in (B), but with the 250 most heavily transcribed genes in 3AT-induced cells superimposed (red circles).

Figure 6. Pol II density correlates with differential loss of H2B from induced genes. (A) Strong inverse correlation between Pol II and H2B levels on 244 genes induced by 3AT (R2 > 0.075) and longer than 500 bp (ICY2 and YPR056W-A are shorter than 500 bp and are therefore not included). Occupancy phasing analysis on the induced genes relative to the TSS: (B) Nucleosome core particles (140–160 bp). (C) 80–120 bp particles in under-digested chromatin. (D) H4 (50–150 bp fragments). (E) H2B (50–150 bp fragments).
that heavily transcribed genes tend to be depleted of H2B and a few genes also show some loss of H4. This differential loss of H2B from heavily transcribed genes in both untreated and induced cells was confirmed by plots of average histone occupancy against average Pol II density (Supplementary Figure S4; compare with Figure 6A). Thus, like 3AT-induced genes, heavily transcribed genes exhibit a differential loss of H2B relative to H4, reinforcing the correlation between Pol II density and H2B loss.

We also asked whether the two surprising characteristics of the Pol II distribution—low levels at the promoter and high levels post-TES—are properties of heavily transcribed genes, or whether they are peculiar to 3AT-induced genes. For both untreated and induced cells, all genes were sorted according to their Pol II density and then divided into quintiles, such that the first quintile contained the most transcribed genes and the last quintile contained the least transcribed genes. The distribution of Pol II relative to the TSS or the TES was then calculated as an IP/input ratio (Supplementary Figure S5). As observed for induced genes (Figure 5C), the most heavily transcribed genes (quintile 1) in both untreated and induced cells, showed the lowest Pol II levels at promoters (TSS plots) and the highest levels on both sides of the TES, with significant overshoot (TES plots). The amount of Pol II on the genes in the other quintiles is much lower than for quintile 1, although the plots are in the order expected from the sort. The disproportionate genomic distribution of Pol II has been commented on previously (38). We conclude that our observations for induced genes are generally true for heavily transcribed genes.

In addition, we asked whether the chromatin structure of TATA box-containing genes (regulated by the SAGA complex and mostly inducible) is different from that of genes lacking TATA boxes (regulated by TFII D and mostly constitutively expressed) (39). On average, TATA-containing genes displayed the characteristics of highly active genes: high Pol II over the coding region and around the TES, a wider NDR and weaker phasing, and a slight decrease in H2B relative to H4 (Supplementary Figure S6). These observations can be explained by the fact that TATA-containing genes are enriched for highly transcribed genes relative to TATA-less genes (23% of the TATA-containing genes are in the most highly transcribed gene quintile (see Supplementary Figure S5); only 10% of the TATA-less genes are present in this quintile). Consequently, the difference in chromatin structure between TATA-containing and TATA-less genes can be accounted for by the fact that TATA-containing genes have a higher proportion of heavily transcribed genes.

**DISCUSSION**

**Induced genes show differential loss of H2B from coding regions**

Our initial aim was to understand the structural basis of the loss of nucleosome occupancy from genes induced by 3AT. We considered three models: loss of complete nucleosomes, loss of H2A-H2B dimers or a conformational change. At induced ARG1, we observed an increase in particles of subnucleosomal size, reduced H4 and a larger reduction in H2B (Figure 8A). These observations are consistent with a limited loss of complete nucleosomes (reduced H4 and H2B) and loss of H2A-H2B from some of the remaining nucleosomes (greater loss of H2B), resulting in tetrasomes and/or hexasomes. Thus, our data support a combination of the first two models for ARGI and a few other very heavily transcribed genes. However, genes with somewhat lower Pol II density exhibit a major loss of H2B but not of H4, indicating that H2A-H2B dimer loss is predominant. The disturbance of the chromatin structure of flanking genes can be explained by expansion of the NDR upstream of the induced gene and by invasion of the downstream gene by Pol II prior to dissociation.

Increased levels of subnucleosome-sized particles over the coding regions of induced genes are consistent with tetrasomes and/or hexasomes but when considered alone, do not prove that they are present. This is because they confer weaker protection from MNase than complete nucleosomes and so must be detected in underdigested samples, in which protruding linker DNA has not yet been removed from the particles by MNase, making it difficult to be
certain that they protect the characteristic lengths of DNA (∼80 and ∼110 bp). Indeed, alternative interpretations are that the shorter protected DNA fragments derive from internal digestion of intact nucleosomes, or from stable non-histone complexes. However, our histone-PESC data show that there are higher levels of H4 than of H2B on coding regions, leading to the conclusion that some particles must lack one or both of their H2A-H2B dimers. It should be noted that the tetrasomes and hexasomes are not necessarily compact; they might be unfolded. In addition, we cannot distinguish between the conventional tetrasome, which contains negatively wrapped DNA, and the reverseosome (40), which contains positively wrapped DNA, because both are compact and should protect similar lengths of DNA from digestion.

Our data are consistent with a genome-wide tiling-array study (13) showing a similar, limited loss of H4 from coding regions that correlates with transcription level (neither H2A nor H2B was measured in this study). ChIP studies of the active yeast GAL genes indicate that both H2B and H4 are reduced to similar levels on the coding regions, suggesting that complete nucleosome displacement occurs (6,11,12). However, we observe a greater loss of H2B relative to H4. The fact that this differential loss is not observed at the GAL genes might reflect their extreme activity in galactose, such that complete nucleosome loss is predominant. In conclusion, we have demonstrated that loss of H2A-H2B dimers, not complete nucleosome loss, is the predominant consequence of transcription.

The disruption of the coding region chromatin that we observe on heavily transcribed genes can be accounted for by in vitro experiments showing that a single passage through a nucleosome by Pol II results in displacement of one H2A-H2B dimer, whereas multiple passages, expected at high Pol II density, result in complete nucleosome loss (3,4). Consistent with this, exogenous RNA polymerase II preferentially displaces H2A-H2B dimers from nucleosome core particles derived from active genes (41). Our observation that genes with lower Pol II density show proportionately reduced H2B loss and only marginal loss of H4 is consistent with this scenario. In vivo, exchange of all four histones occurs at highly transcribed genes (7–9), indicating that nucleosomes are not just disassembled on active genes but are also re-assembled.

Histone exchange occurs most often at promoters but also happens on coding regions, facilitated by elongation factors, histone chaperones and histone acetylation (5–7,10,11,42). Exchange of H2A and H2B is much more rapid than that of H3 and H4 (43,44), consistent with our observation that much more H2B is displaced than H4. The fact that we can detect subnucleosomes on active coding regions indicates that re-assembly of H2A-H2B dimers into nucleosomes is relatively slow. The molecular mechanism by which dimers are replaced in vivo is unclear, but is likely to involve histone chaperones, Spt6 and/or the FACT complex (reviewed in ref. 21).

Slow dissociation after transcription results in queued polymerases

Transcription involves several steps: recruitment of Pol II to form a ‘closed’ preinitiation complex (PIC), melting of promoter DNA to form the ‘open’ initiation complex, initiation, a conformational change to form a stable elongation complex, elongation and perhaps pausing, termination and dissociation. The paradigm for transcriptional regulation has been that initiation is the slowest step and therefore the natural target for regulation (45). More recently, it was discovered that many genes in higher eukaryotes are associated with a Pol II elongation complex paused in the vicinity of the +1 nucleosome (46). Single-gene studies in higher eukaryotes show that, on some genes, Pol II overshoots the TES and pauses several hundred base pairs downstream, where it binds mRNA processing factors prior to termination (47). Genome-wide, the distribution of Pol II on higher eukaryotic genes is consistent with these studies, showing a large Pol II peak around the TSS and relatively low Pol II levels along the gene (48,49). A small subset of genes exhibits a strong peak in the termination region, but there are no queued polymerases (48,49).

In yeast, relatively low-resolution ChIP-on-chip studies have demonstrated the expected correlation between the amount of Pol II on coding regions and mRNA levels (50). More recently, high resolution microarray studies have shown that some highly transcribed genes accumulate Pol II downstream of the TES, which could reflect either slow dissociation or anti-sense transcription (38,51,52). Our high-resolution Pol II PESC data extend and clarify these findings, by examining the effect of induction on Pol II distribution, rather than by correlation with expression data. Surprisingly, we find that the lowest level of Pol II is at the TSS, suggesting that initiation is not rate-limiting. However, if Pol II promoter binding is very slow, rapid initiation and promoter clearance might be important, making the promoter immediately available to bind another Pol II. The fact that Pol II is not in direct contact with DNA in the initial form of the PIC (53) should also be considered, since it might result in inefficient cross-linking of Pol II to DNA and underestimate the amount of Pol II bound at the promoter. The highest levels of Pol II occur just upstream of the TES and ∼140 bp downstream of the TES, indicating that Pol II overshoots the TES and is slow to dissociate from the DNA. It will be important to examine whether Pol II binding downstream of the TES is functional, perhaps by demonstrating the presence of the melted DNA indicative of open complex formation.

In summary, we propose that Pol II dissociation is the slowest step in transcription of heavily transcribed genes. The delay in dissociation results in a queue of polymerases on the gene, waiting to complete transcription and then dissociate in turn. It is likely that this polymerase queue contributes to chromatin disruption, because Pol II is a large complex and physically occupies more DNA than a typical linker in yeast: either each polymerase is paused within a disrupted nucleosome, or some nucleosomes must be partially or completely removed in order to accommodate it (Figure 8B).

ACCESSION NUMBERS
GEO database GSE54524.
SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

We thank Gary Felsenfeld, Alan Hinnebusch, Vasily Studitsky and Carl Wu for helpful comments on the manuscript. We thank the NHLBI Core Facility (Yoshi Wakabayashi, Yan Luo, Ting Ni and Jun Zhu) for paired-end sequencing, Namrita Dhillon and Rohinton Kamakaka for yeast strains with HA-tagged H4 and H2B, and Bruce Howard and Nguyen Dang for help with bioinformatic analysis.

FUNDING

Intramural Research Program of the National Institutes of Health (NICHD). Funding for open access charge: Intramural Research Program of the National Institutes of Health (NICHD).

Conflict of interest statement. None declared.

REFERENCES


