Mapping in Vivo Chromatin Interactions in Yeast Suggests an Extended Chromatin Fiber with Regional Variation in Compaction*

Received for publication, August 20, 2008, and in revised form, September 29, 2008 Published, JBC Papers in Press, October 16, 2008, DOI 10.1074/jbc.M806479200

Job Dekker1

From the Program in Gene Function and Expression and Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605-0103

The higher order arrangement of nucleosomes and the level of compaction of the chromatin fiber play important roles in the control of gene expression and other genomic activities. Analysis of chromatin in vitro has suggested that under near physiological conditions chromatin fibers can become highly compact and that the level of compaction can be modulated by histone modifications. However, less is known about the organization of chromatin fibers in living cells. Here, we combine chromosome conformation capture (3C) data with distance measurements and polymer modeling to determine the in vivo mass density of a transcriptionally active 95-kb GC-rich domain on chromosome III of the yeast Saccharomyces cerevisiae. In contrast to previous reports, we find that yeast does not form a compact fiber but that chromatin is extended with a mass per unit length that is consistent with a rather loose arrangement of nucleosomes. Analysis of 3C data from a neighboring AT-rich chromosomal domain indicates that chromatin in this domain is more compact, but that mass density is still well below that of a canonical 30 nm fiber. Our approach should be widely applicable to scale 3C data to real spatial dimensions, which will facilitate the quantification of the effects of chromatin modifications and transcription on chromatin fiber organization.

The higher order organization of chromatin is thought to play critical roles in processes such as gene expression. Although the structure of nucleosomes is known in detail, much less is known about higher levels of chromatin organization. The first level of organization beyond the nucleosome may involve formation of a 30-nm-thick chromatin fiber, but the precise organization of DNA within this structure and the range of compaction levels of this fiber are poorly understood (for reviews see, e.g. Refs. 1–9).

Over the years, several models have been proposed for the arrangement of nucleosomes within chromatin fibers (reviewed in Refs. 2, 4, 8–11). Two models have been considered extensively. First, the solenoid model predicts that nucleosomes follow a one-start helical path in which consecutive nucleosomes along the DNA are also nearest neighbors in the fiber (12). Second, the crossed-linker model predicts a three-dimensional zigzag organization of nucleosomes that can give rise to a two-start helix (13–15). In this model consecutive nucleosomes are located on opposite sides, with the linker DNA crossing the fiber. This model is supported by in situ observations, nuclease sensitivity experiments, modeling, and crystallographic studies of short tetranucleosomal assemblies (10, 16–20).

Both the solenoid one-step model and the zigzag two-start helix allow formation of chromatin fibers with highly variable levels of chromatin compaction. The level of compaction is thought to affect processes such as transcription. Highly compact chromatin fibers may be less accessible for protein complexes involved in gene expression, DNA repair, and DNA replication. A key step in many chromosomal processes may therefore involve modulation of chromatin fiber compaction.

In vitro studies suggest that the level of chromatin compaction is dependent upon ionic conditions, linker length, linker histone binding, and histone modifications. At very low salt concentrations chromatin does not form 30-nm-thick fibers, and nucleosomes appear in a zigzag arrangement to form fibers with a low mass density of around 1–2 nucleosomes per 11 nm fiber (6, 12, 15, 21). Under conditions that are thought to mimic the in vivo milieu, i.e. 150 mM salt and 1–5 mM Mg2+, chromatin becomes compact to form more typical 30 nm fibers with a mass density of 6 nucleosomes per 11 nm (15, 21). The level of compaction that is observed under these conditions further depends on the length of the linker DNA between nucleosomes and by binding of linker histone (22). Further, acetylation of nucleosomes affects chromatin structure, as detected by a change in DNA linker number (23, 24), and results in a reduced ability to form compact fibers as compared with non-acetylated chromatin (6, 11, 25, 26). Strikingly, a single acetylation event at H4K16 has been shown to prevent formation of compact chromatin fibers in vitro (27, 28).

Combined these observations suggest that in vivo the level of compaction of chromatin is not uniform across the genome, but is modulated by local differences in chromatin modifications, linker length, as well as by binding of additional non-nucleosomal factors. Domain-wide differences in compaction have been observed, and these differences were correlated with...
Mass Density of the Yeast Chromatin Fiber

DECEMBER 12, 2008 • VOLUME 283 • NUMBER 50

Mass Density of the Yeast Chromatin Fiber

Regional variation in gene density, independent of gene expression level (29). More localized fluctuations in chromatin compaction have also been observed, e.g. at and around active promoters (30, 31).

Here we present a strategy that can be used to measure the average level of compaction of specific genomic segments in intact cells. In this approach, we combine spatial distance information obtained by in vivo observations, with information about the conformation of chromatin obtained by chromosome conformation capture (3C)² (32). We have applied this approach to a 95-kb GC-rich chromosomal domain on chromosome III in the yeast Saccharomyces cerevisiae. Our results show that this chromatin domain is highly flexible with a mass density of ~1.2–2.4 nucleosomes per 11 nm, which is well below the value of 6 nucleosomes per 11 nm predicted for a canonical 30 nm fiber. A neighboring AT-rich domain was found to be more compact with a mass density between 1.8 and 3.6 nucleosomes per 11 nm. Our results are consistent with a highly extended chromatin fiber, with variation in compaction at the level of sub-chromosomal domains.

EXPERIMENTAL PROCEDURES

Chromosome Conformation Capture—3C was performed with intact cells (strain NKY2997: MATα, hоL:LYS2, lys2, ura3, nuc1::LEU2) as described for nuclei (32), with the following adaptations. Cells were grown in YPD medium (1% (w/v) Bacto-yeast extract, 2% (w/v) Bacto-peptone, 2% (w/v) glucose) until A600 = 1.0. Cells were resuspended in 200 mm Tris-Cl, pH 7.5, 1 mM dithiothreitol and incubated at room temperature for 2 min. Cells were resuspended in 0.4 M Sorbitol, 0.4 M KCl, 0.5 mM MgCl2, 40 mM sodium phosphate (pH 7.2) and incubated at 30 °C in the presence of 0.2 mg/ml Zymolase for 30 min. The resulting spheroplasts were washed three times in 0.1 M MES, 1.2 M Sorbitol, 1 mM EDTA, 0.5 mM MgCl2 (pH 6.4) and resuspended in the same buffer. Formaldehyde was added to a final concentration of 1%, and samples were incubated at room temperature for 10 min. Glycine was added to a final concentration of 125 mM, and samples were incubated for 5 min at room temperature. Chromatin was solubilized by adding SDS (0.1% final concentration) followed by incubation at 65 °C for 10 min. Triton was added to a concentration of 1%. Chromatin was digested with EcoRI at 37 °C for 18 h. Digestion efficiency was ~70% and did not vary significantly between sites (Ref. 33, and not shown). EcoRI was inactivated by addition of SDS to a final concentration of 1.6%, and samples were incubated at 65 °C for 20 min. Chromatin was diluted to 2.5 mg/ml, and Triton X-100 was added to 1%. DNA was ligated for 2 h at 16 °C using T4 ligase. The cross-links were reversed by overnight incubation at 65 °C in the presence of 5 μg/ml Proteinase K, and DNA was purified by phenol-chloroform extraction and ethanol precipitation.

A randomized control PCR template was generated by digestion and random ligation of purified yeast genomic DNA (NKY2997). The randomized control template contains all ligation products in equal molar ratios and is used to correct for differences in efficiency of PCR amplification of 3C ligation products. All templates were titrated to determine the linear range of PCR, and all quantifications of interaction frequencies were performed using template concentrations that were in the linear dynamic range of PCR. All primers were designed in a unidirectional fashion along the yeast genome to prevent detection of self-ligated partial digestion products (see Refs. 32, 34).

PCR products were quantified on 1.5% agarose gels in the presence of ethidium bromide. Interaction frequencies between loci were calculated in triplicate by determination of the ratio of the amount of ligation product detected by PCR with the 3C template divided by the amount of ligation product detected by PCR with the randomized control template (32).

Determination of Cross-linking Efficiency—We used data obtained by Nagy and co-workers (35) to determine the level of formaldehyde induced cross-linking throughout the AT- and GC-rich domains of chromosome III. Using data from their experiment dl g 067K Exp. #28 (“ORF-enrichment analysis,” see Ref. 35 for details) we determined the enrichment of open reading frames in the cross-linked fraction. The average enrichment (Log2) for genes in the AT-rich domain (n = 37) was 0.04 (S.E. ± 0.057). The enrichment (Log2) for genes in the GC-rich domain (n = 59) was 0.25 (S.E. ± 0.042). The difference in enrichment (Log2) was 0.21 (S.E. ± 0.071). This corresponds to a 1.15-fold higher cross-linking efficiency for chromatin in the GC-rich domain. When we analyzed the enrichment of intergenic regions (n = 39 for the AT-rich domain, n = 46 for the GC-rich domain) in the cross-linked fraction we found a similar fold difference (1.13-fold p < 0.01, t test).

RESULTS

The mass density of chromatin will affect the average distance between two loci. However, the spatial distance between two loci does not by itself provide sufficient information to determine the level of compaction. This is illustrated by two extreme situations in Fig. 1. The chromatin fiber can be very compact and relatively stiff (Fig. 1, A and B, left panels) or the fiber can be relatively extended and highly flexible (Fig. 1, A and B, right panels). In these examples differences in flexibility allow two loci to be separated by the same spatial distance despite significant differences in chromatin compaction. Therefore, the mass density of the chromatin fiber can only be determined when both the average distance between two loci and the flexibility of the intervening chromatin fiber are known (Fig. 2A).

We describe a strategy that combines in vivo distance data with measurements of chromatin flexibility in intact cells to determine the mass density of chromatin and applied it to analysis of a 95-kb GC-rich chromatin domain on the right arm of chromosome III of the yeast S. cerevisiae (Fig. 2A). This domain is flanked by the MATα locus on the left and the HMR locus on the right.

Determination of the Average Spatial Distance between MAT and HMR—The spatial distance between MATα and HMR has been analyzed in a cells (36). These authors introduced Lac repressor binding sites at the MATα locus and Tet repressor binding sites at the HMR locus in a strain expressing GFP-LacI and GFP-TetR fusion proteins. Binding of these fusion proteins to their binding sites resulted in the formation of fluorescent spots that marked the positions of HMR and MATα in living cells (see Ref. 36 for details). Fluorescence microscopy and

² The abbreviations used are: 3C, chromosome conformation capture; MES, 4-morpholineethanesulfonic acid; GFP, green fluorescent protein.
We used 3C to determine the conformation and flexibility of the chromatin fiber between HMR and MATa. Previously we have analyzed this chromosomal region using 3C with nuclei isolated from α-factor-arrested cells. Because the three-dimensional distance data were obtained using non-arrested intact cells, we wished to perform 3C using intact yeast cells as well, which required minor adaptations of our 3C protocol (see Refs. 32, 34 and “Experimental Procedures” above). Using intact cells we determined 30 interaction frequencies between 19 restriction fragments located in between MATa and HMR (see Table 1 for 3C primer sequences and Table 2 for 3C data). Fig. 2C shows the interaction frequencies plotted against the site separation of the interacting loci (in kilobases).

Next we fitted the 3C data to a polymer model to determine chromatin flexibility. Previously we successfully used a model that combines the Kraty-Porod worm-like chain model and the freely jointed chain model to describe cross-linking frequencies of loci along chromatin fibers in isolated nuclei (32, 46, 47). The facts that the region we analyze here appears relatively unconstrained in its movement (48) and that end-to-end distances appear to display a Gaussian distribution (36, 37) support the use of a polymer model to describe the domain. The model is given in Equations 1 and 2 and describes the relationship between the cross-linking frequency X(s) of two sites, determined by 3C, and the genomic distance, or site separation, between those sites (see Ref. 32). It is important to emphasize the assumptions that underlie application of this polymer model to chromatin. First, the model treats chromatin as a flexible and homogeneous polymer, even though native chromatin most likely is not entirely homogeneous and regional variation in compaction and flexibility will occur. For instance, histone acetylation levels are different along the length of a gene. Thus, our method only allows determination of regional averages in apparent flexibility and mass density. Second, the model is only valid when chromatin does not engage in specific or preferred long or short range looping interactions. Indeed, the 3C data do not indicate the presence of such interactions, which would appear as local peaks in the interaction profile (40). Third, the model can only be applied to regions that are sufficiently large, i.e. up to at least 10–20 times the persistence length of chromatin (for discussion see Ref. 47).

\[
X(s) = k \times 0.53 \times \beta^{-3/2} \times \exp\left(-\frac{2}{\beta^2}\right) \times l^{-3} \text{nm}^3\text{mol/liter}
\]  
(Eq. 1)

with

\[
\beta = \frac{5}{2} \times \left(1 - \frac{s}{c}\right)
\]  
(Eq. 2)

In Equations 1 and 2, s is the site separation between two loci (in kilobases). The parameter l is the length of the statistical segment in nanometers. The statistical segment is twice the length of the persistence length of the chromatin fiber and is a measure for its flexibility (45, 49). S is the length of a segment in kilobases, c is the apparent circle size of the fiber (in kilobases; c will be infinitely large in the case of an unconstrained fiber).
Finally, $k$ is the efficiency of cross-linking and can vary somewhat between experiments (32).

Equations 1 and 2 do not include a separate parameter for the compaction, or mass density of the chromatin fiber. We introduce the mass density parameter $L$, which is the length of 1 kb of chromatin in nanometers. $L$ is directly related to the parameters $l$ and $s$, as given by Equation 3.

$$I = L \times S \quad (\text{Eq. 3})$$

Substitution of Equation 3 in Equations 1 and 2 yields Equation 4.

$$X(s) = \frac{(k \times L^{-3}) \times 0.53 \times \beta^{-3/2}}{s^3} \times \exp\left(-\frac{2}{\beta^2}\right) \times \text{nucleosomes/11nm} \quad (\text{Eq. 4})$$

The 3C data in Fig. 2C were fitted to Equation 4 yielding $S = 4.7 \pm 0.45 \text{ kb}$, $c = 130 \pm 11 \text{ kb}$ for and $[k \times L^{-3}] = 1002 \pm 229 \text{ M}^{-1} \text{ nm}^{-3} \text{ kb}^3$ ($R^2 = 0.76$). Values for the individual parameters $k$ and $L$ cannot be directly obtained from this analysis, and thus no value for the mass density (or the length of a statistical segment in nanometers) is obtained from directly fitting 3C data, which are in arbitrary units. As described below, values for these parameters are determined by combining the fitted 3C data with real average distance data measured in vivo.

The values for chromatin flexibility ($S$) and apparent circularity ($c$) determined here in intact cells are comparable but not identical to the values we previously obtained for this chromatin domain with nuclei isolated from α-factor-arrested cells (32, 33). The value of $S = 4.7 \pm 0.45 \text{ kb}$ corresponds to a persistence length of 2.4 kb. The value of $c = 130 \text{ kb}$ indicates that this 95-kb domain displays apparent circular behavior, as we described previously (32). This is apparent in Fig. 2C by the fact that for site separations larger than 40 kb, the cross-linking frequencies decrease more slowly with increasing site separation. Possible explanations for this apparent circular behavior include the presence of some intrinsic features of this domain, or tethering to the envelope, although imaging suggests this domain is relatively unconstrained (48). The apparent circle size determined here in intact cells is smaller than previously observed in nuclei from α-factor-arrested cells (130 ± 11 kb versus 171 ± 19 kb see Ref. 32). Possibly, some chromosome structural features are lost during nuclei isolation or they may have changed as a result of α-factor arrest.

Determination of the Mass Density of Yeast Chromatin—Now that a statistical description of the conformation of this domain is obtained from 3C data and the average spatial distance between two loci located at the borders of this domain has been established by in vivo measurements, the mass density can be determined.

First we derived a general equation for the mass density in terms of spatial distance between two loci and the conformation of the intervening chromatin. The cross-linking frequency $X(s)$ of two loci is proportional to the local concentration $J_{hist}(s)$, which is the concentration of one locus of the chromosome in proximity of the other locus (32, 47).

Mass Density of the Yeast Chromatin Fiber
Mass Density of the Yeast Chromatin Fiber

TABLE 1
3C primer sequences

| Primer number | Primer sequence |
|---------------|----------------|
| 82 | GGTGACGGCTTTGCTTTTATAG |
| 83 | GCTGCACTTTTGCTGCTAGTAG |
| 84 | ATCTTCACTTTAATACACCTTACAG |
| 85 | GTCTCACTTTAATACACCTTACAG |
| 123 | TCTTTGCTTTGCTGCTAGTAG |
| 124 | GGTGACGGCTTTGCTTTTATAG |
| 125 | GCTGCACTTTTGCTGCTAGTAG |
| 126 | ATCTTCACTTTAATACACCTTACAG |
| 127 | TCTTTGCTTTGCTGCTAGTAG |
| 145 | GGTGACGGCTTTGCTTTTATAG |
| 146 | ATCTTCACTTTAATACACCTTACAG |
| 147 | TCTTTGCTTTGCTGCTAGTAG |
| 148 | ATCTTCACTTTAATACACCTTACAG |
| 149 | GGTGACGGCTTTGCTTTTATAG |
| 150 | ATCTTCACTTTAATACACCTTACAG |
| 151 | TCTTTGCTTTGCTGCTAGTAG |
| 152 | ATCTTCACTTTAATACACCTTACAG |
| 153 | TCTTTGCTTTGCTGCTAGTAG |
| 154 | ATCTTCACTTTAATACACCTTACAG |
| 155 | TCTTTGCTTTGCTGCTAGTAG |
| 156 | ATCTTCACTTTAATACACCTTACAG |
| 182 | CCGCCGACGGTACAACTAGCATG |
| 273 | GGTGACGGCTTTGCTTTTATAG |
| 79 | CCGCCGACGGTACAACTAGCATG |

TABLE 2
3C interaction frequencies (arbitrary units)

| Primer 1 | Primer 2 | Site separation | Interaction frequency | S.E. |
|----------|----------|-----------------|-----------------------|------|
| 83 | 145 | 7 | 1.15 | 0.061 |
| 82 | 123 | 10 | 1.25 | 0.070 |
| 83 | 125 | 11 | 0.96 | 0.020 |
| 145 | 148 | 11 | 1.35 | 0.059 |
| 148 | 150 | 12 | 0.84 | 0.070 |
| 83 | 146 | 13 | 0.65 | 0.028 |
| 83 | 147 | 15 | 0.72 | 0.049 |
| 83 | 148 | 18 | 0.42 | 0.041 |
| 125 | 150 | 19 | 0.68 | 0.050 |
| 83 | 126 | 21 | 0.59 | 0.052 |
| 82 | 124 | 22 | 0.90 | 0.036 |
| 83 | 149 | 24 | 0.40 | 0.040 |
| 148 | 151 | 28 | 0.66 | 0.054 |
| 83 | 150 | 30 | 0.42 | 0.052 |
| 82 | 83 | 30 | 0.39 | 0.033 |
| 83 | 182 | 37 | 0.34 | 0.028 |
| 150 | 273 | 39 | 0.46 | 0.050 |
| 146 | 127 | 43 | 0.50 | 0.036 |
| 85 | 86 | 46 | 0.41 | 0.055 |
| 82 | 126 | 51 | 0.30 | 0.025 |
| 273 | 147 | 54 | 0.34 | 0.011 |
| 123 | 84 | 57 | 0.43 | 0.022 |
| 82 | 150 | 60 | 0.19 | 0.009 |
| 124 | 152 | 62 | 0.38 | 0.026 |
| 273 | 83 | 69 | 0.30 | 0.010 |
| 123 | 152 | 74 | 0.37 | 0.062 |
| 123 | 127 | 75 | 0.27 | 0.033 |
| 82 | 152 | 84 | 0.24 | 0.044 |
| 82 | 127 | 86 | 0.18 | 0.024 |
| 123 | 273 | 95 | 0.31 | 0.014 |

For a freely jointed chain the average spatial distance between two loci, \( r(s) \), is directly related to the local concentration.

\[ X(s) = k \times J_M(s) \]  
\[ (\text{Eq. 5}) \]

\[ \sqrt{\langle r(s)^2 \rangle} = 0.82 \times (J_M(s))^{-1/3} \]  
\[ (\text{Eq. 6}) \]

Equation 6 is based on Equations 3 and 4 (in Ref. 45) including a conversion to moles/liter. Combining Equations 5 and 6 yields the following expression for \( k \),

\[ k = X(s) \times (\sqrt{\langle r(s)^2 \rangle}/0.82)^3 \]  
\[ (\text{Eq. 7}) \]

and thus,

\[ [k \times L^{-3}] = X(s) \times (\sqrt{\langle r(s)^2 \rangle}/0.82)^3 \times L^{-3} \]  
\[ (\text{Eq. 8}) \]

which can be rewritten as Equation 9,

\[ L = [k \times L^{-3}]^{-1/3} \times X(s)^{1/3} \times \langle r(s)^2 \rangle/0.82 \]  
\[ (\text{Eq. 9}) \]

which describes the linear relationship between the mass density and the average spatial distance between two loci for a given chromosome conformation. This equation can be used to scale 3C data, which are in arbitrary units, so that absolute interaction frequencies can be derived and the real spatial dimension of the chromatin fiber can be calculated. Note that the slope, \( [k \times L^{-3}]^{-1/3} \times X(s)^{1/3} \times 1/0.82 \), does not vary between experiments, despite the presence of the experiment-specific parameters \( k \) and \( X(s) \). Further, it should be noted that Equations 6–9 are only valid for loci separated by >3 or 4 statistical segments, when the fiber approaches a freely jointed chain (45). Given a statistical segment length of 4.7 kb for the GC-rich chromatin (above), Equations 6–9 are valid for sites separated by >14–19 kb.

Equation 9 was used to determine the average mass density per unit length for the GC-rich domain on the right arm of chromosome III. The observed 3C interaction frequency between \( HMR \) and \( MATa \) was 0.31, whereas the predicted value according to the fit to Equation 4 was 0.38. Fig. 2D shows the relationship between \( L \) and \( r(s) \) for \( X(HMR,MATa) = 0.38 \) and \( [k \times L^{-3}] = 1002 \) M\(^{-1}\) nm\(^{-3}\) kb\(^{3}\). The average spatial distance \( r(s) \) between \( HMR \) and \( MATa \) is calculated based on the SK1 genome sequence.
and MATa is 463 nm, which yields $L = 41$ nm/kb. When we used the experimental value of $X_{(HMR,MATa)} = 0.31$, a value of $L = 38$ nm/kb was obtained. Given a nucleosome repeat length in yeast of 165 bp (3), these values correspond with 1.6 and 1.8 nucleosomes per 11 nm fiber, respectively.

The 95% confidence interval for the value of $L$ was obtained by determining the range of values for the mass density by varying $[k \times L^{-3}]$ and $r_{(HMR,MATa)}$ by plus and minus two times their standard deviations. Values ranging from 1.2 to 2.1 (for $X_{(HMR,MATa)} = 0.38$) and 1.2 to 2.4 (for $X_{(HMR,MATa)} = 0.31$) nucleosomes per 11 nm were found. These values correspond to $L = 31–57$ nm/kb and $L = 28–53$ nm/kb, respectively, which indicates that the yeast chromatin fiber is much more extended than a canonical 30 nm fiber, which has a mass density of 11 nm/kb.

Now that the mass density has been determined, the persistence length can be expressed in nanometers. Fitting interaction frequencies to Equation 4 yielded $S = 4.7$ kb, which corresponds to 131–268 nm (for $L = 28–57$ nm/kb) and a persistence length of 66–134 nm. This value is within the range determined with other methodologies (45).

A separate GFP-based analysis has recently been performed to determine the spatial distance between HMR and MATa (48). These authors report an average spatial distance between HMR and MATa of ~600–650 nm, which is significantly larger than the distance we determined based on data described by Simon and co-workers (36). Because very similar approaches were used, it is not immediately clear why such different distances were obtained. It is possible that the authors overestimated the distance by leaving out of the analysis cells in which MATa and HMR were too closely co-localized to optically separate the two corresponding GFP spots. A distance of ~600–650 nm would correspond to an average mass density of ~1.0–1.5 nucleosomes per 11 nm fiber (see Fig. 2D).

Regional Differences in Chromatin Compaction—Previously, we have identified structural differences between functionally distinct AT- and GC-rich domains along yeast chromosome III (32). One of those domains is the 95-kb GC-rich domain studied here. Our previous analysis of the adjacent AT-rich domain (position 100–190 kb on chromosome III) in isolated nuclei from $\alpha$-factor-arrested cells showed that the chromatin fiber was less flexible and that the value for $[k \times L^{-3}]$ was 2- to 2.2-fold higher compared with the GC-rich domain (32, 33). We re-analyzed this AT-rich domain using 3C with intact cells. We determined 27 interaction frequencies between 15 restriction fragments located throughout the genome. By determining the enrichment of genomic fragments in either the non-cross-linked or cross-linked fraction the efficiency of cross-linking could be determined throughout the genome. We used their data to estimate the relative enrichment of open reading frames located in the AT-rich domain or the GC-rich domain of chromosome III. We found that the average level of cross-linking in the GC-rich domain was 15% higher than in the AT-rich domain. Given the large number of open reading frames that was analyzed (37 in the AT-rich domain, and 59 in the GC-rich domain), this difference is statistically significant ($p < 0.01$, $t$ test). Similarly, when we used the data from Nagy and co-workers (35) these authors directly measured the level of formaldehyde-induced cross-linking throughout the yeast genome by selectively fractionating chromatin based on the level of cross-linking. By determining the enrichment of genomic fragments in either the non-cross-linked or cross-linked fraction the efficiency of cross-linking could be determined throughout the genome. We used their data to estimate the relative enrichment of open reading frames located in the AT-rich domain or the GC-rich domain of chromosome III. We found that the average level of cross-linking in the GC-rich domain was 15% higher than in the AT-rich domain.

As $[k \times L^{-3}]$ for both the AT-rich and GC-rich domains was determined in the same experiment using the same 3C material as PCR template, it is reasonable to assume that the value of $k$ (the cross-linking efficiency) is comparable for both domains. Consistently, our previous analysis in nuclei of cross-linking efficiency using a restriction enzyme accessibility assay revealed that cross-linking in the GC-rich domain was approximately only 10% more efficient, although this difference was not statistically significant (33). To further investigate potential differences in the value of $k$, we also estimated cross-linking efficiency in the AT- and GC-rich domains using data obtained by Nagy and co-workers (35). These authors directly measured the level of formaldehyde-induced cross-linking throughout the yeast genome by selectively fractionating chromatin based on the level of cross-linking. By determining the enrichment of genomic fragments in either the non-cross-linked or cross-linked fraction the efficiency of cross-linking could be determined throughout the genome. We used their data to estimate the relative enrichment of open reading frames located in the AT-rich domain or the GC-rich domain of chromosome III. We found that the average level of cross-linking in the GC-rich domain was 15% higher than in the AT-rich domain.

When we assume that $k$ is directly related to the cross-linking efficiency as measured by Nagy $et$ al. the value of $k$ for the AT-rich domain is ~1.15-fold lower than for the GC-rich
Mass Density of the Yeast Chromatin Fiber

domain. In that case, a 3.1-fold higher value for \([k \times L^{-3}]\) would correspond to a 1.5-fold lower value for \(L\), and thus correspond to a more compact chromatin fiber in the AT-rich domain. Given that the mass density of the GC-rich domains is 1.6 nucleosomes per 11 nm (95% confidence interval: 1.2–2.4 nucleosomes per 11 nm), the mass density of the AT-rich domain is around 2.4 nucleosomes per 11 nm (28 nm/kb; 95% confidence interval: 1.8–3.6 nucleosomes per 11 nm, corresponding to 18–37 nm/kb). The persistence length is 58–118 nm.

Predicting the Average Spatial Distance between Loci Using Their 3C Interaction Frequency—Now that the value of \(L\) has been determined, Equation 9 can be used to determine the value of \(k\) for a specific 3C experiment. This is important, because this experiment-specific parameter directly relates the 3C interaction frequency of any two loci to the relative concentration of one of the loci near the other (Equation 5). Once the value of \(k\) is known for a given 3C experiment, one can estimate for any pair of loci for which the 3C interaction frequency is determined, their local concentration and their average spatial distance (Equation 6). For instance, for the 3C analysis of the GC-rich domain \([k \times L^{-3}] = 1002 \text{ m}^{-1} \text{ nm}^{-3} \text{ kb}^3\) and \(L = 38 \text{ nm/kb}\) (for \(X_{\text{HMR-MATa}} = 0.31\)). From these values we determine that \(k = 5.5 \times 10^{7} \text{ M}^{-1}\). For the AT-rich domain \(k\) would be 15% lower, i.e. \(4.8 \times 10^{7} \text{ M}^{-1}\).

Using these values we can now use 3C interaction frequencies to estimate the average spatial distance between two loci using Equations 5 and 6. For instance a pair of loci in the GC-rich domain that are separated by 45 kb \((X(45) = 0.48)\) and 47 kb \((X(47) = 0.38)\). We used \([k \times L^{-3}] = 1002 \text{ m}^{-1} \text{ nm}^{-3} \text{ kb}^3\) for loci in the GC-rich domain (solid line) and \([k \times L^{-3}] = 3104 \text{ m}^{-1} \text{ nm}^{-3} \text{ kb}^3\) for loci located in the AT-rich domain (dotted line). The values for \(L\) are indicated that correspond to \(r_{\text{geo}} = 373 \text{ nm}\).

### DISCUSSION

We describe a general strategy to determine mass density of specific chromatin domains using distance measurements combined with 3C analysis of the conformation of the intervening chromatin fiber. Our results indicate that, for the actively transcribed GC-rich chromatin domain analyzed here, the chromatin fiber is significantly more extended (1.2–2.4 nucleosomes per 11 nm) than the canonical 30 nm fiber, which has a mass density of 6 or more nucleosomes per 11 nm (3, 9, 21). The AT-rich domain is somewhat more compact with a mass density of 1.8–3.6 nucleosomes per 11 nm. We conclude that active yeast chromatin is not as compact as a canonical 30 nm fiber. Instead we propose a more extended conformation. Analysis of an additional distance between loci separated by 46 kb confirmed the relatively low mass density of yeast chromatin.

The mass density that we report here for an active chromatin domain is significantly lower than that reported previously in yeast. Bystricky and co-workers used fluorescence in situ hybridization (FISH) as well as GFP-based in vivo measurements to determine the average spatial distance between a set of loci and fitted the data to a polymer model to determine flexibility and mass density of the yeast chromatin fiber (37). They found a mass density of around 7–10 nucleosomes per 11 nm fiber (7–9 nm/kb), which is much lower than the mass density we report here for an active chromatin domain.
more compact than reported here. There could be several explanations for this difference. First, different methodologies were used (fluorescence in situ hybridization versus 3C) and the data were fitted to slightly different polymer models. In addition, Bystricky and co-workers used the mode of the distances (i.e. the most frequently observed distance) for their modeling. The mode is smaller than the average, especially for smaller site separations. Thus, the spatial separation of loci may have been underestimated, especially so for loci separated by smaller genomic distances, which may have contributed to the fact that a higher mass density was found. Another reason for the observed discrepancy could be related to the fact that Bystricky and co-workers did not measure a single contiguous region but measured distances between loci scattered along several chromosomes, so that distances were measured for regions with potentially different levels of compaction. The smallest distance they measured was for a pair of loci located near the right telomere of chromosome V. Sub-telomeric and transcriptionally silenced regions may be more condensed than transcriptionally active chromatin, e.g. as a result of Sir protein loading and the presence of sub-telomeric repeats and Ty elements (50, 51). It is also possible that the highly GC-rich domain we studied here may be significantly more decondensed than most other regions in the yeast genome, although the average spatial distance of the two loci on chromosome 14 separated by 46 kb is also consistent with an extended conformation (Fig. 4). The low mass density may also be related to the observation that GC-rich chromatin in yeast is on average more transcriptionally active and displays higher levels of histone acetylation (33, 52) than AT-rich chromatin. Consistently, the AT-rich domain analyzed here has a 1.5-fold higher level of compaction.

The relatively low mass density we report here is consistent with a number of independent observations. First, the majority of nucleosomes in yeast do not contain linker histone-like proteins (53, 54). The capacity to form compact fibers is influenced by the presence of linker histones (e.g. Refs. 12, 22, 55). The lack of linker histone on many nucleosomes may result in the highly extended conformation of yeast chromatin. Second, yeast chromatin is highly acetylated, especially in GC-rich regions (33, 56). Genomic regions that contain highly acetylated nucleosomes are significantly more extended than hypo-acetylated regions (57). In vitro studies have also shown that nucleosomal arrays reconstituted with acetylated histones have a reduced capacity to become compact in the presence of Mg$^{2+}$ (25, 26). In yeast, a large fraction of nucleosomes, with the exception of nucleosomes flanking promoters, are acetylated at lysine 16 of H4 (58). Acetylation of this residue prevents formation of compact chromatin fibers (27, 28). Further, it has been shown that the length of the linker DNA affects the architecture and compaction of chromatin fibers (22). In vitro reconstitution experiments under physiological conditions showed that, in the absence of linker histones and with a nucleosomal repeat length of 167 bp, chromatin forms relatively extended fibers with 3–4 nucleosomes per 11 nm (22). This reconstituted chromatin resembles yeast chromatin that mostly lacks H1 and has a short nucleosomal repeat length of around 165 bp, although it is somewhat more compact than our in vivo observations indicate. Finally, recent molecular dynamics simulations of crossed-linker fibers with short yeast-like linker lengths and in the absence of linker histone yielded values for mass density (3.1 nucleosomes per 11 nm) and flexibility (persistence length of 2.4 kb) that are in good agreement with the values we report here for in vivo chromatin (1.2–3.6 nucleosomes per 11 nm) (59). We conclude that the absence of H1, the high levels of histone acetylation (especially at H4K16), and the short linker length may all contribute to the formation of the relatively extended conformation of yeast chromatin in vivo.

AT-rich and GC-rich chromatin domains in yeast are functionally distinct and display lower and higher average levels of gene expression and meiotic recombination levels, respectively (33, 52, 60, 61). Combined with our observation here of differences in chromatin fiber compaction in these domains, these data may indicate that the biological activity of large chromosomal domains is correlated at least in part to fluctuations in the level of compaction of the chromatin fiber. The difference in compaction could be related to a difference in histone acetylation levels. Consistently, we previously found that GC-rich chromatin in yeast is significantly more acetylated than AT-rich chromatin throughout the genome (33). We also have described a role for the histone deacetylase Rpd3p in modulating the compaction of the GC-rich domain on the right arm of chromosome 3. Interestingly, analysis of chromatin fiber compaction status in the human genome also revealed that gene-rich (and typically GC-rich) regions are less compact than gene poor (and generally more AT-rich) domains (29, 62), suggesting that domain-wide differences in chromatin compaction may be a general feature of chromosomes.

Our approach relies on a number of assumptions, as outlined above. In particular we assume that specific preferred short and long-range looping and fold-back structures do not occur and that the distance distribution is Gaussian. Future experiments, perhaps with improved imaging technologies, may directly assess the validity of these assumptions.

The strategy we have presented here to determine in vivo chromatin compaction may help in studying domain-wide control of chromatin structure by allowing the analysis of in vivo compaction of a specific domain under a variety of experimental conditions. The chromatin domain must be sufficiently large, i.e. up to at least 10 times the persistence length to allow use of the polymer model and Equations 6–9 described here. In addition, both 3C and GFP-based methods to detect specific loci in intact cells can be applied to similar analyses in higher eukaryotes. Further, our approach provides a general strategy to scale 3C data, which are typically in arbitrary units, to real spatial distances between specific loci. This is important because it enables the use of 3C data to generate three-dimensional models of chromosomal domains in accurate spatial dimensions.

Acknowledgments—M. Walhout and members of the Dekker laboratory are acknowledged for fruitful discussions and critical reading of the manuscript.

REFERENCES
1. Belmont, A. S., Dietzel, S., Nye, A. C., Strukov, Y. G., and Tumbar, T. (1999) Curr. Opin. Cell Biol. 11, 307–311
Mass Density of the Yeast Chromatin Fiber

1. Widom, J. (1989) *Annu. Rev. Biophys. Biomol. Struct.* 18, 365–395
2. van Holde, K. E. (1989) *Chromatin*, Springer, Heidelberg
3. van Holde, K., and Zlatanova, J. (1995) *J. Biol. Chem.* 270, 8373–8376
4. Woodcock, C. L., and Dimitrov, S. (2001) *Curr. Opin. Genet. Dev.* 11, 130–135
5. Hansen, J. C. (2002) *Annu. Rev. Biophys. Biomol. Struct.* 31, 361–392
6. Swedlow, J. R., and Hirano, T. (2003) *Mol. Cell* 11, 557–569
7. Wu, C., Bassett, A., and Travers, A. (2007) *EMBO Rep.* 8, 1129–1134
8. Treemethick, D. (2007) *Cell* 128, 651–664
9. Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002)
10. Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N. P., and Bick, U. S. A.
11. Robinson, P. J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R. G.,
12. Shogren-Knaak, M., Ishii, H., Sun, J. M., Pazin, M. J., Davie, J. R., and Bradbury, E. M.
13. Norton, V. G., Imai, B. S., Yau, P., and Bradbury, E. M.
14. Norton, V. G., Marvin, K. W., Yau, P., and Bradbury, E. M.
15. Athey, B. D., Smith, M. F., Rankert, D. A., Williams, S. P., and Langmore, W. A.
16. Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., and Woodcock, C. L.
17. Rydberg, B., Holley, W. R., Mian, I. S., and Chatterjee, A.
18. Wedeman, G., and Langowski, J. (2002) *Biophys. J.* 82, 2847–2859
19. Horowitz, R. A., Agard, D. A., Sedat, J. W., and Woodcock, C. L.
20. Schalch, T., Duda, S., Sargent, D. F., and Richmond, T. J.
21. Gerchman, S. E., and Ramakrishnan, V.
22. Routh, A., Sandin, S., and Rhodes, D.
23. Norton, V. G., Imai, B. S., Yau, P., and Bradbury, E. M.
24. Norton, V. G., Marvin, K. W., Yau, P., and Bradbury, E. M.
25. Garcia-Ramirez, M., Rocchini, C., and Ausio, J.
26. Tse, C., Sera, T., Wolfe, A. P., and Hansen, J. C.
27. Shogren-Knaak, M., Ishii, H., Sun, J. M., Pazin, M. J., Davie, J. R., and Peterson, C. L.
28. Robinson, P. J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R. G., and Rhodes, D.
29. Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N. P., and Bickmore, W. A.
30. Sabo, P. J., Hawrylycz, M., Wallace, J. C., Humbert, R., Yu, M., Shafer, A., Kawamoto, J., Hall, R., Mack, J., Dorschner, M. O., McNicholl, A., and Stamatoyannopoulos, J. A.
31. Gheldof, N., Tabuchi, T. M., and Dekker, J.
32. Dekker, J., Rippe, K., Dekker, M., and Kleckner, N.
33. Dekker, J. (2007) *Genome Biol.* 8, R116