Population Analysis of Subsaturated 172-12 Nucleosomal Arrays by Atomic Force Microscopy Detects Nonrandom Behavior That Is Favored by Histone Acetylation and Short Repeat Length*

Ralph C. Bash‡, Jaya Yodh§, Yuri Lyubchenko¶, Neal Woodbury‡, and D. Lohr‡‡

From the ¤Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, the ¶Division of Basic Sciences, Midwestern University, Arizona College of Osteopathic Medicine, Glendale, Arizona 85308, and the ‡Department of Microbiology and Department of Biology, Arizona State University, Tempe, Arizona 85287

 Concatameric 5 S rDNA templates reconstituted in vitro into nucleosomal arrays provide very popular chromatin models for many kinds of studies. Here, atomic force microscopy is used to determine the population distributions for one such nucleosomal array, the 172-12, reconstituted to various subsaturated levels with nonacetylated or hyperacetylated HeLa histones. This array is a model for short linker length genomes and transcriptionally active and newly replicated chromatin. The analysis shows that as input histone levels increase, template occupation increases progressively as discrete population distributions. The distributions are random at low (n_av < 4) and high (n_av > 8) loadings but display specific nonrandom features, such as a deficit of molecules with one nucleosome more or less than the peak species in the distribution and enhanced distribution breadths, in the mid-range (n_av = 4–8). Thus, the mid-range of occupation on polynucleosomal arrays may be a special range for chromatin structure and/or assembly. The mid-range nonrandom features are enhanced in distributions from short repeat (172-12) arrays, particularly for unacetylated chromatin, and in distributions from hyperacetylated chromatin, particularly for long repeat (208-12) arrays. Thus, short repeat length and acetylation can affect basic chromatin properties, like population tendencies, in very similar ways and therefore may cause similar changes in chromatin structure. Some possible effects are suggested. The data also indicate that it is thermodynamically more difficult for hyperacetylated nucleosomes to assemble onto the 172-12 templates, a result having implications for in vitro chromatin assembly.

A characteristic feature of eukaryotic chromatin is their significant variability in average nucleosome repeat length, a feature that arises from variations in the average lengths of linker DNA that separate the conserved core nucleosomal particles in vitro (1, 2). Average genomic repeat lengths range from ~156 to ~250 bp (3, 4), but there are some trends. For example, in transcriptionally active Saccharomyces cerevisiae, the average nucleosome repeat is ~170 bp, whereas in the inactive, terminally differentiated chicken erythrocyte, it is 208 bp (3). The indication that transcriptionally active chromatin tends to have short nucleosome repeat lengths is strengthened by studies of active genes within metazoan genomes (5, 6). Differences in the lengths of DNA that space the core nucleosomes are likely to have pronounced effects on core nucleosome structure as well as on higher order nucleosome packing arrangements (4, 7).

Some years ago Simpson et al. (8) developed a DNA template composed of tandem sea urchin 5 S rDNA units. The basic 5 S unit can impose a nucleosome positioning preference. Concatamers of these units provide templates that allow the in vitro assembly of defined polynucleosomal arrays suitable for biophysical and biochemical studies. Concatamers of 208-bp 5 S units, e.g., the dodecameric array of units referred to as the 208-12, have been the most widely used, to study nucleosome positioning (9, 10), the in vitro nucleosome assembly process (11, 12), chromatin folding, and its relationship to acetylation, linker histone presence, and transcription factor binding (13–22). The 5 S templates were made with a choice of DNA lengths between the basic positioning sequences (8), e.g. 172-12, 190-12, and 208-12, thus creating templates that effectively differ in average nucleosome repeat length. The 172-12 provides a model for short linker length chromatin like yeast, and the 208-12 provides a model for long linker length chromatin such as chicken erythrocyte.

To analyze the fundamental features of nucleosome occupation on these multisite 5 S arrays, our labs have recently turned to atomic force microscopy (AFM). AFM is a powerful imaging technique that can be used to visualize a number of nucleic acid and nucleoprotein complexes (Refs. 23 and 24 and references therein), including chromatin (25–30). It offers particular advantages for the type of investigation described here. The ability to directly visualize individual molecules allows one to count precisely the numbers of nucleosomes present on DNA templates. By analyzing many molecules, it is possible to determine unambiguously the population distribution, in essence the statistical distribution of nucleosome occupation states for the template, as a function of nucleosome loading, in modified versus unmodified chromatin, etc. The experimental results can then be compared with theoretical models for insight on how nucleosomes choose to occupy these arrays. Subsaturated arrays, i.e. arrays in which the number of available 172- or 208-bp nucleosome-binding sites exceeds the number of nucleosomes available to bind, are particularly useful for these types of studies because of the occupancy choices available to nucleosomes loading onto such templates. Subsaturated arrays also...

Received for publication, May 30, 2001, and in revised form, September 24, 2001
Published, JBC Papers in Press, October 2, 2001, DOI 10.1074/jbc.M104918200

This paper is available on line at http://www.jbc.org

48362 This paper is available on line at http://www.jbc.org
have intrinsic interest as models for specific chromosomal regulatory regions like replication origins (31, 32) and gene promoters (33, 34), which are often not fully saturated with nucleosomes.

Previously, we determined the population features for subsaturated 208-12 arrays (30). In this work, the population distributions for subsaturated 172-12 arrays are determined, and the results are compared with the 208-12 results. The 172-12 is a very important template because it provides a model for organisms with short genomic repeats and for transcriptionally active and newly assembled chromatin, which also tend to have short nucleosome repeat lengths (3, 5, 6). However, despite its obvious interest as a model for these functionally important chromatin states, the 172-12 has not been systematically analyzed except to determine that its folding (13) and nucleosome positioning characteristics (9) resemble those of the 208-12. Thus, our analysis provides novel structural studies that can serve as the basis for further analyses and applications of this important model system.

Histone acetylation is a provocative and much studied chromatin feature that is associated with both transcription and replication (35-37). Because of the link between histone acetylation and these functionally important chromatin states and because yeast histones are known to have high levels of acetylation in vivo (38, 39), we also analyzed hyperacetylated 172-12 reconstitutes.

**MATERIALS AND METHODS**

**Plasmid and Histone Isolation**—The plasmid (p5S172-12) containing the 172-12 was a generous gift of J. Hansen. DNA and acetylated or nonacetylated HeLa histones were prepared as described previously (30).

**Chromatin Reconstitution**—Nucleosomal arrays were reconstituted at various subsaturating levels following the method of Hansen and Lohr (12) with some modifications. Briefly, HeLa histone octamers were mixed with 3 μg of the 172-12 DNA template at varying molar ratios of histone octamers to DNA. DNA and histone octamers were mixed together on ice at a final concentration of 2 mM NaCl (plus TE (10 mM Tris-1 mM EDTA, pH 8.0)) and a final volume of 30 μl. Histone octamers were dialyzed in HSB buffer (2.5 mM NaCl, 50 mM phosphate, pH 8.1, 1 mM phenethylmethylsulfonyl fluoride) and added last to the reaction mix. Reaction mixtures were dialyzed (Spectra/Port 6-8 KDa) at 4°C against 1 liter of TE containing 1 mM NaCl for >2 h, 0.80 mM NaCl for 3 h, and then 0.6 mM NaCl overnight. The final dialysis step was against 1 liter of 1 mM EDTA (pH 8.0) for >2 h. Chromatin was fixed by dialyzing against 0.1% glutaraldehyde in 1 mM EDTA (pH 8.0) for 6 h. Excess glutaraldehyde was removed by a final dialysis for 24 h against 1 liter of 1 mM EDTA (pH 8.0). Fixed samples were recovered and stored on ice until further use.

To confirm that the reconstitutions were successful and to estimate their concentrations, an aliquot of each reconstitute was electrophoresed at 45 V on a native 3.5% polyacrylamide gel overnight at 4°C. Oligonucleosomes migrate as a smear between free 172-12 DNA template and the gel well. The more nucleosomes the template contains, the slower it migrates. Those templates carrying ≥8 nucleosomes only minimally enter the gel.

**Deposition of the Samples onto AP-Mica**—The samples for AFM were prepared as described (23, 24, 40, 41). Briefly, immediately prior to deposition, chromatin samples were diluted 5-100-fold with 1 mM EDTA (pH 8). Ten microliters of the sample (DNA concentration, 0.2-0.4 μg/ml in 1 mM EDTA, pH 8.0) was pipetted onto pieces of AP-mica for 2 min, rinsed with deionized water (ModuPure Plus, Continental, Inc., Santa Barbara, CA) operating in TappingMode. NanoProbe TESP tips (Digital Instruments, Inc.), conical sharp silicon tips (NCH, Nanosensor), and V-shaped silicon cantilevers from K-Tek International, Inc. were used for imaging. The typical tapping frequency was 300-340 kHz for TESP tips and 340-380 kHz for the K-Tek probes; the scanning rate was 2-3 Hz.

**Statistical Analysis**—For each reconstituted sample, 150-400 molecules were analyzed. “Countable” molecules were those that had easily discernible nucleosomes and discernible DNA ends. They could be either extended or more compacted. We also counted molecules containing closely contacting nucleosomes so long as we could be certain, by measuring widths for example, of the number of nucleosomes present. We observe a much higher incidence of closely contacting nucleosomes in the 172-12 samples than were observed previously in the 208-12 samples (30). This feature presumably reflects increased internucleosomal contact resulting from the shorter nucleosomal spacing and will be analyzed thoroughly in another publication.

Nucleosome population distributions were obtained by plotting the fraction of molecules with a given number of nucleosomes (n) versus n. Distributions for nonacetylated and hyperacetylated samples ranged from n<sub>av</sub> < 2 to n<sub>av</sub> = 11. n<sub>av</sub> is the average number of nucleosomes present on the templates in a sample distribution. The following equation is used to calculate n<sub>av</sub>.

\[
\frac{1}{n_{av}} = \sum_{n=1}^{12} \frac{n \times (\text{no. of molecules with } n \text{ nucleosomes})}{\text{total no. of molecules counted}}
\]  
(Eq. 1)

The error bars for each point in the distribution are calculated according to the following equation.

\[
\text{error} = \sqrt{\frac{\text{no. of molecules with } n \text{ nucleosomes}}{\text{no. of molecules with } n \text{ nucleosomes}}} \times (F_r)
\]  
(Eq. 2)

where Fr is the fraction of molecules with a particular number of nucleosomes, i.e. the experimental nucleosome loading distributions from AFM analysis were compared with theoretical distributions calculated for the same average number of nucleosomes (n<sub>av</sub>), by assuming that nucleosomes bind randomly and with equal frequencies to each of the twelve 172-bp sites, regardless of the histone octamer to DNA ratio. For the theoretical (random) distributions, the fraction of DNA fragments with n nucleosomes bound (F<sub>n</sub>) is given by Equation 3.

\[
F_n = \left(\frac{n_{av}}{12}\right) \left(1 - \frac{n_{av}}{12}\right) \frac{12!}{(12-n)!!n!}
\]  
(Eq. 3)

Note that n<sub>/12</sub> is the random probability that any particular nucleosome site is occupied. Thus, \(\frac{n_{av}}{12}(1 - \frac{n_{av}}{12})\frac{12!}{(12-n)!!n!}\) is simply the total probability that any given arrangement of n nucleosomes will exist, and 12!(12-n)!!/n! is the total number of arrangements, assuming that nucleosomes are indistinguishable.

**RESULTS**

172-12 DNA templates were reconstituted with various subsaturating levels of HeLa histone octamers to obtain a series of subsaturated chromatin samples. Each sample was imaged by AFM. A representative field is shown in Fig. 1. In each field, we counted the number of nucleosomes on every molecule in which all of the nucleosomes were clearly distinguishable, including molecules containing closely contacting dimers or trimers (Fig. 1, arrows), so long as we could be certain of the number of nucleosomes present. For each reconstituted sample, the population distribution, i.e. the plot of the fraction of molecules containing n nucleosomes as a function of n, and the n<sub>av</sub>, the average number of nucleosomes per template (see “Materials and Methods”), were determined.

**General Response of Template Loading to Input Histone**

One of the powerful features inherent in the AFM approach is the ability to determine unambiguously the numbers of nucleosomes on template DNA molecules. Fig. 2 plots the relationship between the input histone level during reconstitution and the average number of nucleosomes (n<sub>av</sub>) loading onto 172-12 templates. For both acetylated and unacetylated chromatin, the n<sub>av</sub> increases smoothly with increasing input histone.

---

*R. C. Bash, J. Yodh, N. Woodbury, and D. Lohr, manuscript in preparation.
tone, showing a linear dependence at low \( n_{\text{av}} \) but becoming nonlinear at higher values. However, at all input histone levels, higher amounts of hyperacetylated than unacetylated histones are required to achieve a particular \( n_{\text{av}} \). Thus, it is apparently more difficult thermodynamically for hyperacetylated nucleosomes to assemble on these templates, at any loading level, indicating that acetylation of the N-terminal tails affects some facet of the occupation process on these templates. A study using a somewhat different experimental material, fully saturated 208-12 arrays (analyzed by restriction digestion), observed no apparent effect of acetylation on saturated loading (20). We have not analyzed saturated arrays by AFM methods because nucleosomes are too difficult to count unambiguously on such templates (30), but we do observe that loading curves for subsaturated 208-12 templates remain linear to higher \( n_{\text{av}} \) values than for 172-12 templates, which results in a higher average nucleosome occupancy per input histone at higher subsaturated levels (\( n_{\text{av}} > 9 \)) on the 208-12.\(^3\) We suggest that this difference in loading response reflects a repeat length difference between the two templates. Other repeat length-dependent differences will be described below.

Our limited data at very low input histone suggests that the plot may have a nonzero intercept (Fig. 2). One possible explanation for such behavior is a “threshold” effect, i.e., a certain level of histone must be present before nucleosomes begin to appear on the DNA templates. This would probably reflect trivial features, such as the need to saturate binding sites on the dialysis tubing, and in agreement with that suggestion, the curves for both types of 172-12 arrays (Fig. 2) and for 208-12 chromatin (not shown) would extrapolate to the same threshold value. Because AFM only needs subnanogram quantities of

\(^3\) J. Yodh, Y. Lyubchenko, L. Shlyakhtenko, N. Woodbury, and D. Lohr, unpublished results.
material, we routinely reconstitute 3 μg of DNA, and thus, at low \( n_{av} \) values, there are only nanogram quantities of histones. Reconstituting larger amounts of material (20) could probably avoid this type of behavior. A nonzero intercept would have no consequence for the major feature of Fig. 2, the consistent difference in occupation level between nonacetylated and acetylated arrays, nor for the results to be presented below, which are all based on direct visualization and counting of nucleosomes on DNA templates and thus would be unaffected by such behavior. Moreover, it is possible that the response at low input histone is actually asymptotic, which would make the curves sigmoidal in shape and characteristic of cooperativity. Analysis of nucleosome location data provides a more sensitive way to test for cooperativity and, indeed, using such an analysis, we have detected cooperativity in occupation on 208-12 templates.4 172-12 locational analyses have not been completed.2

172-12 Population Distributions

Analysis of population distributions provides more detailed information about the nucleosome occupation of these multisite arrays.

Nonacetylated Chromatin—A representative set of population distributions that span the range from low loading (low \( n_{av} \)) to high loading (high \( n_{av} \)) for 172-12 templates reconstituted with nonacetylated histones is shown in Fig. 3 (A–D, closed circles with solid lines). The error bars shown depend on the number of molecules counted at each value of \( n \) and are calculated as described under “Materials and Methods.” For each experimentally determined distribution in Fig. 3, we also show the distribution that would be produced by a random nucleosome loading process with the same \( n_{av} \) value as the corresponding experimental one. The nucleosome loading levels are: \( n_{av} = 1.8 \) (A); \( n_{av} = 4.4 \) (B); \( n_{av} = 6.4 \) (C); \( n_{av} = 9.1 \) (D); \( n_{av} = 4.3 \) (E); and \( n_{av} = 5.5 \) (F).

4 J. Yodh, et al., submitted for publication.
sively to a higher $n_{av}$ value but remain discrete. Thus, poly
ucleosomal occupation of these templates is multi-state in
nature, i.e. it involves stable, partially loaded intermediates
and therefore is not highly cooperative. The type of behavior
demonstrated in Fig. 3 was previously observed for the larger
repeat size 208-12 nucleosomal arrays, whether analyzed by
AFM (30) or sedimentation analysis (12).

Second, the distributions vary somewhat with the nucleo
some occupation level. At low ($n_{av}$ < 4) or high ($n_{av}$ > 8) load-
ings, the distributions are featureless and closely resemble
random distributions (Fig. 3, A and D). However, in distribu-
tions from samples in the mid-range of occupation, $n_{av}$ = 4–8
on these 12-site arrays, there are specific features that indicate
deviations from random behavior (Fig. 3, B and C). Perhaps the
most obvious of these is a deficit (compared with random ex-
pectation) in the fraction of molecules with one nucleosome
fewer (Fig. 3B) or more (Fig. 3C) than the major peak value in
the distribution, a feature that causes secondary maxima, i.e.
shoulders, to appear at ±2 nucleosomes from the major peak.
This $n_{av}$ dependence is quite repeatable. Virtually all of the
distributions at low (<4) and high (>8) $n_{av}$ closely follow a
random curve, and none of them shows a +1 or −1 deficit or a
secondary maximum, whereas all of the mid-range ($n_{av}$ = 4–8)
distributions show these features to some degree, and none
follows a random curve (not shown). Moreover, the features
appear and then disappear, respectively, in a fairly narrow
range around $n_{av}$ = 4 and $n_{av}$ = 8. Thus, the ±1 deficit appears
to be limited rather specifically to the mid-range of loading.

For a more stringent test of the generality of the mid-range
features, the data from all of the nonacetylated 172-12 chro
matin samples in the $n_{av}$ = 4–8 range was pooled (Fig. 4A).
To make this compilation, the most populated $n$ value in each
sample distribution, i.e. the $n$ value with the largest fraction
of molecules in the distribution (designated “$n_o$”), and the fractions
of molecules containing 1, 2, 3, or 4 nucleosomes fewer (−) or
more (+) than that $n$ value are tabulated. The 0, ±1, ±2, ±3,
etc. classes from each of the various sample distributions are
then summed together and displayed as a single profile (Fig.
4A). This type of analysis allows all of the various distributions
to be combined into a single data set, thus identifying features
that are generally characteristic of the entire set of mid-range
samples.

The composite profile shows a statistically significant de-
pletion of molecules containing one nucleosome fewer and one
nucleosome more than the major peak (Fig. 4A), confirming
that this specific deficit at ±1 in the distribution is a general
feature of nonacetylated 172-12 arrays in the mid-range of
loading. The appearance of shoulders at ±2 nucleosomes
from the peak is clearly due to the depletion of the ±1 classes
because the shoulders are present at approximately the ran-
don expectation level. Thus, the occupation process for these
mid-range arrays includes a component(s) that creates a
dispreference for molecules containing one nucleosome more or less than the peak value in the distribution. Note that this
is not a dispreference for odd numbers of nucleosomes on the
template (data not shown). Rather it is a dispreference for
molecules that contain one nucleosome more or less than the
peak $n$ value, whether that peak value is even or odd. Based
on the magnitude of the deviations from random, we estimate
that there are ~20% fewer molecules present in each of the ±
1 classes than a random process would be expected to produce
(Fig. 4A). Thus, the effect responsible for this nonrandom
behavior is probably only one contribution to the template
loading process. However, the existence of any such tendency
is a novel observation and is probably only detectable by an
analysis that can determine precise population distributions,
like AFM.

**Distribution Broadening**—The distributions for samples in
the mid-range of loading are also broader than random expec-
tation. In the composite profile, the fractions of molecules found
toward the fringes, for example containing three and four nu-
cleosomes more or less than the peak value (±3/±4), are higher
than random, whereas the fractions of molecules found near
the center, for example the peak and ±1 nucleosome from the
peak (0/±1), fall at or below random levels (Fig. 4A). In other
words, the population of molecules around the center of the
profile is relatively depleted, and the population at the fringes
is enhanced compared with a random profile. Broadening can
also be detected in the individual distributions (Fig. 3, B and
C). To quantitate it, we calculated the fraction of molecules at
the center of the distribution, within a span of ±2 nucleosomes
from the peak $n$ value, for each reconstituted sample (Fig. 5).
Although the data show significant scatter, the fractions of
molecules found within ±2 of the peak consistently fall below
random expectation for samples in the mid-range ($n_{av}$ = 4–8),
again indicating an enhanced breadth compared with that ex-
pected for a random process.

At low and high $n_{av}$, the $f(±2)$ values are closer to random
(Fig. 5). However, there may be some broadening even at these
occupation levels. The peak value in the individual distribu-
tions often falls below random expectation, and the fractions
of molecules toward the fringes of the distribution can sometimes
exceed random levels (Fig. 3, A and D). Thus, the breadth
enhancement is largest in the mid-range (see above) but is
probably present, at a lower extent, at low and high occupation
levels. From a purely statistical viewpoint, the mid-range
should provide the best opportunity to detect nonrandom be-
havior because that range provides the most equal mix of
nucleosomes and open sites. Such a statistical effect would act
at all loading ranges but simply be most detectable in the
mid-range, i.e. just as the distribution broadening behaves.
In contrast, the ±1 dispreference is only seen in the mid-range
and thus appears to be specific to that occupation range.

**Hyperacetylated Chromatin**—The experimental distribu-
tions for subsaturated 172-12 templates reconstituted with
hyperacetylated histones are qualitatively similar to the dis-
tributions for unacetylated samples, i.e. random at low and
high occupation levels (data not shown) but containing nonran-
dom features, such as a deficit at ±1 nucleosome from the
distribution peak, for samples in the $n_{av}$ = 4–8 range (Fig. 3, E
and F). However, the nonrandomness seems to be more prom-
inent in distributions from (mid-range) acetylated samples
(Fig. 3, B and C versus E and F). This is confirmed in their
composite profile (Fig. 4B), which shows a marked depletion of
molecules that contain one nucleosome less than the major
peak value. This deficit is more pronounced (values are further
enhanced, whereas the fractions of molecules at the center
(0/±1 nucleosomes from the peak) is depleted compared with
random values (Fig. 4B), but in the acetylated profile the fringe versus center difference is slightly greater than in the nonacetylated profile (Table I). Individual distributions show the same behavior; $f(\pm 2)$ values are well below the random line for the mid-range acetylated reconstitutes and tend to fall below the comparable values for nonacetylated samples (Fig. 5). For example, a best fit line for acetylated samples falls below the best fit line for nonacetylated samples (not shown). Thus, both mid-range nonrandom features, the $\pm 1$ dispreference and distribution broadening, are more pronounced in distributions from acetylated compared with unacetylated 172-12 nucleosomal arrays.

Comparison with 208-12

In an analysis of 208-12 subsaturated nucleosomal array populations, we observed in the mid-range of loading some of the same types of nonrandom features noted here for 172-12 arrays (30). However, the only composite profile analyzed in that work was a combination of data from all of the mid-range 208-12 samples, acetylated + nonacetylated + fixed + unfixed, in a single profile. The acetylation-dependent differences noted in the 172-12 analysis prompted us to reevaluate the 208-12 data. In Fig. 4 (C and D), we present composite profiles for 208-12 nonacetylated (fixed) and acetylated (fixed) samples in the mid-range of nucleosome occupation, i.e. data comparable...
AFM Analysis of Subsaturated 172-12 Nucleosomal Arrays

FIG. 5. The fraction of molecules within two nucleosomes of the peak n value. The sum of the fractions of molecules having the peak n value plus one and two nucleosomes more or less (±1) than that value, denoted as f(±2), is shown as a function of n_{av} for individual reconstituted samples. Because of the variability among individual distributions we summed the 0, ±1, and ±2 classes instead of just the 0 and ±1 classes as in Table I. Using the latter sum gives the same trends (but more scatter) for this analysis. The open circles are 172-12 hyperacetylated reconstitutes; the filled circles are 172-12 nonacetylated reconstitutes. The solid line gives the behavior expected for a random process.

with the 172-12 data in Fig. 4 (A and B), thus allowing a direct comparison of the population features of these differing repeat length templates.

The composite profile for hyperacetylated 208-12 arrays in the mid-range of nucleosome occupation shows the same types of nonrandom features observed in mid-range profiles from 172-12 arrays, an underrepresentation of molecules containing one nucleosome more or less than the peak value and the presence of shoulders at ±2 nucleosomes from the peak (Fig. 4D). Also, the profile is broader than random by the criteria used above, enhanced population in the ±3/±4 nucleosome classes and depletion of molecules in the 0/±1 classes (Table I). However, the mid-range profile from nonacetylated 208-12 samples (Fig. 4C) does not show the nonrandomness found in the other three mid-range profiles but instead shows features not seen in those profiles. For example, the ±1 dispreference is weak to nonexistent, and the profile actually appears to be slightly narrower than random; the fractions of molecules at the fringes, i.e. the ±3/±4 nucleosome classes, are at or below random expectation, whereas the fractions of molecules at the center, the 0/±1 nucleosome classes, are equal to or greater than random (Fig. 4C and Table I). In fact, the most striking feature of this profile is a significant enhancement in the fraction of molecules having exactly the peak n value, a feature that is unique to this profile. Again, the mix of distributions used to make the 208-12 acetylated and nonacetylated composite are similar; the weighted average of the n_{av} is 5.7 for the nonacetylated and 6.4 for the acetylated. We note that all four types of chromatins show random distributions at low (n_{av} < 4) and high (n_{av} > 8) loading levels (see above and Ref. 30).

The data above show that the mid-range nonrandom features are more pronounced in profiles from acetylated samples, for both repeat length templates. However, the nonrandom features also depend on template repeat size because both the ±1 dispreference and the breadth enhancements are more pronounced in the 172-12 than in the corresponding 208-12 profile, for both nonacetylated and hyperacetylated arrays (Fig. 4, A versus C and B versus D, and Table I). Thus, short repeat length can also enhance the nonrandom features in mid-range population profiles. The repeat length-induced enhancements seem to be greater in unacetylated chromatin samples; the effects induced by acetylation appear to be greater for the longer repeat (208-12) arrays (Table I). To sum up, as judged by the criteria of Table I, the nonrandom features, i.e. the amount of deviation below random for the ±1 dispreference and the degree of profile broadening, appear to be greatest for acetylated 172-12 arrays, weak to nonexistent for the unacetylated 208-12, and roughly similar for the 208-12 acetylated and the 172-12 unacetylated arrays.

**TABLE I**

| Dispreference | Breadth |
|---------------|---------|
|               | -1      | +1     | ±3/4  | 0/±1  |
| 172-12 (Non)  | 0.03    | 0.03   | 0.19  | 0.55  |
| 172-12 (Ac)   | 0.05    | 0.00   | 0.22  | 0.53  |
| 208-12 (Non)  | 0       | 0      | 0.10  | 0.68  |
| 208-12 (Ac)   | 0.04    | 0.02   | 0.17  | 0.55  |

**DISCUSSION**

We have used AFM to characterize the population features of subsaturated 172-12 5 S rDNA nucleosomal arrays. These short repeat arrays are of great intrinsic interest as *in vitro* models for chromatin from short linker length genomes and for transcriptionally active or newly replicated chromatin. However, they have not been extensively studied or utilized. Subsaturated arrays facilitate the study of template population issues because of the choices they provide for polynucleosomal assembly, and they are also models for regulatory chromosomal regions like gene promoters and replication origins. AFM allows the visualization of individual molecules and the determination of precise numerical population distributions for chromatin templates as a function of nucleosome occupation level, template repeat size, and histone acetylation state. From such data, novel insights on how nucleosomes populate these multistate arrays and how repeat length or acetylation affect the process can be obtained.

The results show that with increasing input histone, nucleosome occupation levels on both hyperacetylated and unacetylated 172-12 arrays increase smoothly (Fig. 2), as discrete population distributions that shift progressively to higher n_{av} values (Fig. 9). Thus, the 172-12 template can be reconstituted to any desired average degree of subsaturation; at all subsaturated levels, >70% of the molecules contain no more than ±2 nucleosomes from the peak (most populated) value in the distribution (Fig. 5). Similar results were observed with the larger repeat size 208-12 subsaturated arrays in sedimentation (12) or AFM (30) studies.

The 172-12 population distributions vary with nucleosome occupation level. At low (n_{av} < 4) or high (n_{av} > 8) loadings, they are virtually identical to the distributions expected for a random process. However, at intermediate levels (n_{av} = 4–8), they contain specific nonrandom features, namely a depletion...
of molecules with one nucleosome more or less than the peak value in the distribution and enhanced distribution breadths. Thus, the mid-range occupation of these arrays involves a nonrandom component. This could be associated with the polynucleosomal loading process itself, with intratemplate chromatin organization processes, or with both. That the population profile and thus the occupation process can vary with nucleosome occupation level is not unexpected because features that affect the process, for example electrostatic effects and the potential for internucleosomal interactions, will depend on the occupation level. Subsaturated 208-12 arrays show qualitatively similar population behavior (30). However, the mid-range nonrandom features are more pronounced in profiles from short repeat (172-12) arrays, particularly for unacetylated chromatin. Acetylation also enhances the nonrandom behavior, especially in the longer repeat (208-12) arrays.

The deficit at \( \pm 1 \) nucleosome from the peak, i.e. the \( \pm 1 \) dispreference, is a general feature found in the composite data sets (>3000 molecules) from several types of samples. It is only observed in the mid-range (\( n_{av} = 4-8 \)), not at low or high occupation levels, and by various AFM analysts, so it is not likely to be due to a systematic error in the analysis. We conclude that it is a \textit{bona fide} feature of at least these, and perhaps all, subsaturated polynucleosomal arrays. As noted previously (30), the presence of this feature results in the occurrence of secondary maxima at \( \pm 2 \) nucleosomes from the peak and thus a \textit{relative} pairwise preference in the profile, behavior that is consistent with \textit{in vivo} and model building data suggesting that nucleosome pairs may be a fundamental unit of chromatin structure (42–45). However, the distributions do not show an absolute preference for pairs, i.e. peak values only at even numbers of nucleosomes. Perhaps the effect(s) responsible for this behavior is not strong enough to establish an absolute preference but only strong enough to create a relative pairwise preference, by disfavoring molecules with one nucleosome more or less than the peak value, whether that peak value is odd or even. This suggests some sort of intratemplate correlation effect. Internucleosomal interactions, which might seem to be the most obvious explanation for the \( \pm 1 \) dispreference, are not likely to be the cause because the dispreference is more pronounced in acetylated chromatin, where such interactions are thought to be less strong (15, 20). Our best candidate explanation is a DNA torsional effect, perhaps involving internucleosomal orientations (30).

The \( \pm 1 \) dispreference is a subtle feature and one that is probably only detectable by a precise and quantitative approach like that provided by single molecule AFM analysis. Its magnitude, \(-20–25\%\) (Fig. 4), indicates that the effect responsible for it is only one contribution to the template occupation process. However, the existence of any such nonrandom effect is unexpected and, indeed, rather remarkable in such a simple \textit{in vitro} system. It must reflect a fundamental aspect of (mid-range) polynucleosomal array occupation. It appears to be a small effect but might nevertheless be important because chromatin structure is probably a balance among a number of small effects; for example, histone N-terminal tail-mediated internucleosomal interactions and many nucleosome positioning preferences involve \( \pm 1 \) Kcal/mol energy differences (46, 47). Moreover, \textit{in vitro} profiles might underestimate the effect because \textit{in vitro} population behavior is probably dominated by the basic histone-DNA interaction, thus masking secondary effects like this one. Also, the effect may be stronger under other conditions (ionic strength, polyamine presence, topologically constrained DNA) or if nucleosome loading is mediated by assembly complexes. Salt reconstitution conditions are generally considered to produce nucleosomal template distributions that reflect equilibrium properties (48, 49). Thus, the \( \pm 1 \) dispreference and distribution broadening reflect inherent thermodynamic characteristics of the mid-range of occupation for these nucleosomal arrays and are not the result of kinetic trapping processes. Therefore, these \textit{in vitro} features are likely to be relevant for \textit{in vivo} processes.

These population results have some specific implications for chromatin structure and chromatin assembly. For example, the limitation of the \( \pm 1 \) dispreference to the mid-range of occupation indicates that there may be unique constraints associated with chromatin organization and/or the nucleosome loading process in that occupation range. Both of the traits that favor the \( \pm 1 \) dispreference, acetylation and short repeat length, are traits associated with assembling chromatin (3, 37). \textit{In vivo} chromatin assembly involves both old and newly synthesized histones (3, 50, 51), but regardless of how this two-component assembly is coordinated, loading of daughter chromatin must at some point pass through the mid-range of nucleosome occupation and thus be subject to the nonrandom effect(s) that appear to be associated with that loading range. Indeed, preferential assembly involving either histone class could bring daughter chromatin to half-saturation and in the middle of the range in which nonrandom behavior is observed. On the other hand, the more limited nucleosomal array reassembly associated with such processes as DNA repair or transcriptional repression may operate solely in the random (\( n_{av} > 8 \)) range. For example, transcription activation of the yeast \( GAL1 \) gene causes the loss of only the two promoter region nucleosomes, corresponding to \(-20–25\%\) of the nucleosomes in the entire \( GAL1 \) gene array (7, 52). Reloading of these nucleosomes, which occurs rapidly upon transcriptional inactivation (52), would thus operate in the random range (25\% loss corresponds to \( n_{av} = 9 \) in the 172-12 system) and not in the mid-range where nonrandom effects occur. The localized nature of DNA repair suggests that associated nucleosome reassembly would also operate in the random (\( n_{av} > 8 \)) range.

The population profile most relevant to replicating chromatin is the profile for short repeat, hyperacetylated arrays (Fig. 4B). That profile suggests that the following equilibrium tendencies operate in assembling (replicating) chromatin in the mid-range of loading. For most (>70\%; Fig. 5) of the assembling chromatin, nucleosome occupancy will be near the average loading level (within \( \pm 2 \) nucleosomes); the relative pairwise preference created by the \( \pm 1 \) dispreference will favor loading of nucleosomes in pairs (especially at two nucleosomes less than the peak value). However, stretches of chromatin that contain several nucleosomes more or less (\( \pm 3, \pm 4 \)) than the peak value, \textit{i.e.} chromatin that is significantly underloaded or overloaded relative to the average occupancy, is thermodynamically most favored under these conditions (mid-range, acetylated, short repeat; Fig. 4), although the preference is still small. Thus, these conditions favor clustering (overloaded) and open patches (underloaded chromatin) more than any other conditions, \textit{i.e.} they provide the best thermodynamic opportunity for significant under- or overloads at mid-range subsaturated nucleosome occupancy levels. Interestingly, although parental nucleosomes segregate randomly, they do show a tendency to cluster (50, 51). Indeed, current models envision parental histones assembling in multinucleosome patches, distinct from the stretches where newly synthesized histones assemble (51), behavior that is consistent with the basic (but small) thermodynamic tendencies that we observe. These basic tendencies could be amplified \textit{in vivo} via specific DNA sequences or protein factors and would undoubtedly be useful to help establish the diversity of chromatin structures needed for the various types of chromosomal regions and gene organizations found in eukaryotic genomes. We also find that polynucleosomal loading is inherently (thermodynamically) more
difficult, at all loading levels, for the type of array characteristic of assembling chromatin (hyperacetylated, short repeat arrays). Perhaps by making assembly more difficult and thus disfavoring spontaneous (non-factor-mediated) assembly, acetylation helps ensure that specific complexes carry out the in vivo assembly of replicating chromatin.

The nonrandom features also provide insights on aspects of basic chromatin structure. For example, the observation that the +1 dispreferrence and distribution broadening are similarly enhanced in short repeat length or in acetylated arrays indicates that those chromatin traits could affect at least some aspects of chromatin structure in similar ways. What might these common effects be? At the individual nucleosome level, acetylation causes histones to bind DNA less tightly and DNA to be less strongly associated with the nucleosome, particularly near the termini, thus altering the core structure (15, 53, 54). Short repeat lengths constrain the wrapping of DNA near nucleosome termini. For example, in yeast nuclear chromatin (average repeat –170 bp), nucleosomes are unable to wrap more than 126 bp of DNA on the core under any circumstance (7, 55). However, the same assays show that in chicken erythrocyte nuclear chromatin (average repeat of –208 bp), nucleosomes are able to wrap 126–168 bp of DNA, depending on conditions. Thus, both acetylation and short repeat length can alter the nucleosome–DNA interaction near the core termini. Both features probably affect higher order chromatin structure as well. Acetylation is known to inhibit nucleosomal folding (15); short repeat length–induced constraints in the wrapping of DNA around individual nucleosomes can also constrain higher order organization (4). Thus, known structural effects of short repeat length and acetylation are consistent with the suggestions from this work that those traits can cause similar effects on chromatin structure.

The powerful role of acetylation has long been recognized (35–37, 53, 54, 56, 57), but the effect of (short) repeat length is less appreciated. Nor has a structural commonality between the two traits been previously suggested, so far as we are aware. Indeed, the two traits may reinforce one another. For example, we find that acetylated, short repeat (172-12) chromatin, the type of nucleosome structure most closely associated with transcriptionally active or newly replicated chromatin, shows the most anomalous behavior from a population viewpoint. On the other hand, long repeat (208-12), nonacetylated chromatin demonstrates different features, an enhanced frequency of molecules that contain exactly the peak value of nucleosomes in the distribution and the narrowest composite profile. From a population viewpoint, these features make this chromatin template the most regular of the four studied, an interesting property for a chromatin organization considered to be structured for gene inactivity. Moreover, acetylation appears to alter the population features of the long repeat chromatin more strongly than it affects short repeat chromatin. These results indicate clearly that the most basic features of chromatin, such as nucleosome repeat length and acetylation state, can cause significant effects on basic chromatin properties. Thus, even the simplest chromatin fiber is probably capable of a diversity of behavior, depending on the precise mix of basic features present. The amplification of this basic diversity by histone modifications (57) and nonhistone proteins will produce a structure with the complexity needed to mediate the many kinds of processes required of the genetic material in eukaryotes.

Acknowledgments—We thank Luda Shlyakhtenko for technical advice, particularly in the early stages of this work, and Stuart Lindsay for critical reading of the manuscript and several very helpful suggestions.