

# Archaeal Histone Selection of Nucleosome Positioning Sequences and the Procaryotic Origin of Histone-dependent Genome Evolution

Kathryn A. Bailey<sup>1</sup>, Suzette L. Pereira<sup>1</sup>, Jonathan Widom<sup>2</sup>  
and John N. Reeve<sup>1\*</sup>

<sup>1</sup>Department of Microbiology  
The Ohio State University  
Columbus, OH 43210-1292  
USA

<sup>2</sup>Department of Biochemistry  
Molecular Biology and Cell  
Biology, Northwestern  
University, Evanston  
IL 60208-3500, USA

Archaeal histones and the eucaryal (eucaryotic) nucleosome core histones have almost identical histone folds. Here, we show that DNA molecules selectively incorporated by rHMfB (recombinant archaeal histone B from *Methanothermobacter ferredoxinus*) into archaeal nucleosomes from a mixture of  $\sim 10^{14}$  random sequence molecules contain sequence motifs shown previously to direct eucaryal nucleosome positioning. The dinucleotides GC, AA (=TT) and TA are repeated at  $\sim 10$  bp intervals, with the GC harmonic displaced  $\sim 5$  bp from the AA and TA harmonics [(GCN<sub>3</sub>AA or TA)<sub>n</sub>]. AT and CG were not strongly selected, indicating that TA  $\neq$  AT and GC  $\neq$  CG in terms of facilitating archaeal nucleosome assembly. The selected molecules have affinities for rHMfB ranging from  $\sim 9$  to 18-fold higher than the level of affinity of the starting population, and direct the positioned assembly of archaeal nucleosomes. Fourier-transform analyses have revealed that AA dinucleotides are much enriched at  $\sim 10.1$  bp intervals, the helical repeat of DNA wrapped around a nucleosome, in the genomes of *Eucarya* and the histone-containing *Euryarchaeota*, but not in the genomes of *Bacteria* and *Crenarchaeota*, procaryotes that do not have histones. Facilitating histone packaging of genomic DNA has apparently therefore imposed constraints on genome sequence evolution, and since archaeal histones have no structure in addition to the histone fold, these constraints must result predominantly from histone fold-DNA contacts. Based on the three-domain universal phylogeny, histones and histone-dependent genome sequence evolution most likely evolved after the bacterial-archaeal divergence but before the archaeal-eucaryal divergence, and were subsequently lost in the *Crenarchaeota*. However, with lateral gene transfer, the first histone fold could alternatively have evolved after the archaeal-eucaryal divergence, early in either the euryarchaeal or eucaryal lineages.

© 2000 Academic Press

**Keywords:** Archaea; SELEX; nucleosome positioning; genome evolution; HMfB

\*Corresponding author

## Introduction

Essentially the same histones, H2A, H2B, H3 and H4 (Sullivan *et al.*, 2000), package chromosomal DNA into nucleosomes in almost all *Eucarya* (Woese *et al.*, 1990; Woese, 2000), arguing that these proteins and this system of DNA packaging must have been present before the divergence of the eucaryal lineage. With the discovery of proteins with histone-like sequences in the *Archaea*,

a potential procaryotic origin for this system was identified (Sandman *et al.*, 1990, 1998), and structural studies have since confirmed that these archaeal proteins are *bona fide* histones (Starich *et al.*, 1996; Zhu *et al.*, 1998; Decanniere *et al.*, 2000). Although they lack the N and C-terminal structures that flank the histone folds of the eucaryal nucleosome core histones, archaeal histones form dimers with quaternary structures that are almost identical to the histone-fold dimers in the eucaryal nucleosome core (Luger *et al.*, 1997; Luger & Richmond, 1998; Decanniere *et al.*, 2000). Archaeal

E-mail address of the corresponding author:  
reeve.2@osu.edu

histones similarly bind and wrap DNA into complexes designated archaeal nucleosomes (Pereira *et al.*, 1997; Bailey *et al.*, 1999; Soares *et al.*, 2000) that have many features in common with the tetramers formed by eucaryal histone (H3 + H4)<sub>2</sub> tetramers (Alilat *et al.*, 1999). Both have DNA wrapped around a histone tetramer core, overwound at ~10.1 bp per helical turn, with the DNA constrained in either a positive or negative toroidal supercoil (Musgrave *et al.*, 1991; 2000; Hamiche *et al.*, 1996; Hamiche & Richard-Foy, 1998; Alilat *et al.*, 1999). Both tetramers also similarly respond to DNA-sequence dependent positioning signals and assemble nucleosomes preferentially at specific locations *in vivo* and *in vitro* (Dong & van Holde, 1991; Hayes *et al.*, 1991; Spangenberg *et al.*, 1998; Pereira & Reeve, 1999; Sandman & Reeve, 1999).

Positioned nucleosomes play a major role in regulating eucaryal gene expression (van Holde, 1993; Wolffe, 1994; Svaren & Horz, 1996; Beato & Eisfeld, 1997), and considerable effort has been invested in identifying and characterizing sequences that are preferentially incorporated in nucleosomes and which therefore direct positioned nucleosome assembly (Satchwell *et al.*, 1986; Shrader & Crothers, 1989; 1990; Ioshikhes *et al.*, 1992, 1996; Bina, 1994; Bolshoy, 1995; Widlund *et al.*, 1997, 1999; Lowary & Widom, 1997, 1998; Widom, 1998; Thåström *et al.*, 1999). DNA molecules isolated from nucleosomes assembled *in vivo* contain an over-abundance of AA (=TT), GC and GG (=CC) dinucleotides repeated at ~10.2 bp intervals, with ~5 bp between the A/T- and G/C-rich repeats (Ioshikhes *et al.*, 1992, 1996; Bina, 1994; Bolshoy, 1995; Widlund *et al.*, 1997), and DNA molecules synthesized to conform to a [(A/T)<sub>3</sub>NN(G/C)<sub>3</sub>NN]<sub>n</sub> consensus pattern have very high levels of affinity for histone octamers *in vitro* (Shrader & Crothers, 1990; Bina, 1994). Consistent with these results, DNA molecules selected using SELEX procedures (Tuerk & Gold, 1990) for incorporation into nucleosomes by histone octamers also have sequences much enriched for AA (=TT) and TA dinucleotides separated at ~10 bp intervals (Lowary & Widom, 1998; Thåström *et al.*, 1999). Apparently, molecules that contain alternating anisotropically-flexible A/T and G/C-rich regions most readily accept the minor and major groove distortions required for tight DNA wrapping around a nucleosome core and are therefore preferentially incorporated into nucleosomes (Shrader & Crothers, 1989, 1990; Widom, 1998; Fitzgerald & Anderson, 1999; Thåström *et al.*, 1999). Analysis of the yeast and nematode genome sequences revealed an over-abundance of AA (=TT) dinucleotides at ~10.1 bp intervals, indicating that these eucaryal genomes are selectively enriched for sequences that facilitate their own assembly into nucleosomes (Widom, 1996).

results of using a SELEX approach to isolate DNA molecules that were preferentially incorporated into archaeal nucleosomes. DNA molecules were selected from a large population of random-sequence synthetic DNAs by their above

average affinities for recombinant (r) archaeal histone HMfB (histone B from *Methanothermobacter feroidus*; Sandman *et al.*, 1990, 1995). They have been shown to contain repetitive dinucleotide motifs established previously as eucaryal nucleosome-positioning elements and, furthermore, Fourier-transform analyses have confirmed that these motifs are enriched in DNA molecules assembled into archaeal nucleosomes *in vivo* and that they are over-represented in the genomes of *Euryarchaeota*, procaryotes that contain histones (Bult *et al.*, 1996; Klenk *et al.*, 1997; Smith *et al.*, 1997; Kawarabayashi *et al.*, 1998) but not in the genomes of *Crenarchaeota* (Kawarabayashi *et al.*, 1999; <http://niji.imb.nrc.ca/sulphome>) and *Bacteria* (Fleishmann *et al.*, 1995; Kunst *et al.*, 1997; Deckert *et al.*, 1998; Stevens *et al.*, 1998; Nelson *et al.*, 1999), procaryotes that do not contain histones. Since archaeal histones do not contain structures in addition to the histone fold (Starich *et al.*, 1996; Zhu *et al.*, 1998; Decanniere *et al.*, 2000), these results argue that histone-fold-DNA interactions are primarily responsible for nucleosome positioning, and that this system of genome packaging and the opportunity to use positioned nucleosomes to regulate gene expression most likely evolved first in the procaryotic archaeal lineage (Woese *et al.*, 1990, 2000).

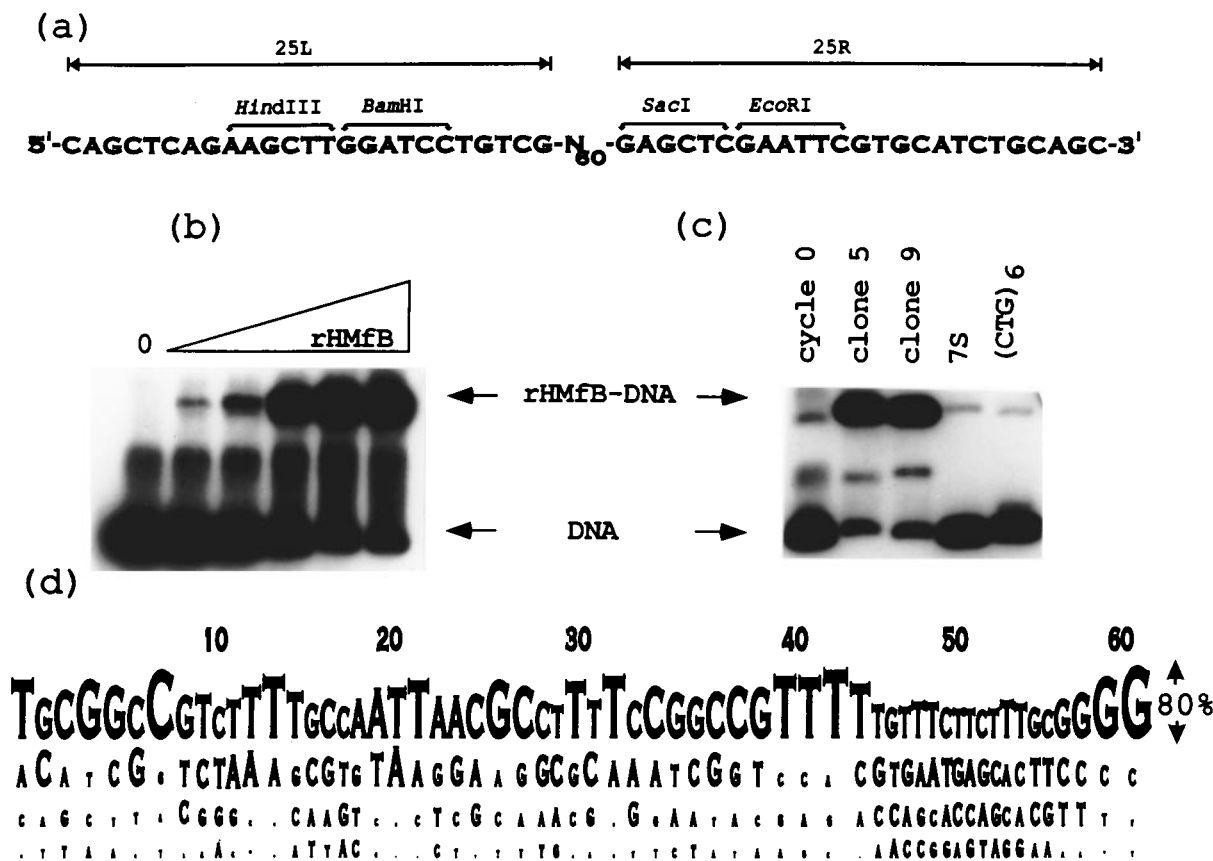
## Results

### Selection of DNA molecules by assembly into archaeal nucleosomes

A population of ~10<sup>14</sup> different 110 bp DNA molecules (cycle 0 DNA) was synthesized, each molecule having a 60 bp region of random sequence (N<sub>60</sub>) flanked by two fixed 25 bp sequences (25L and 25R) that contained restriction enzyme cleavage sites and provided priming sites for PCR amplifications (Figure 1(a)). Aliquots of this population were incubated with increasing amounts of rHMfB, and the DNA molecules incorporated into archaeal nucleosomes under conditions of ~5 % DNA incorporation (Figure 1(b)) were isolated, amplified and used as the DNA in the next round of selection. After eight rounds of selection and amplification, individual molecules from the selected population (cycle 8 DNA) were cloned and 132 were sequenced, resulting in 111 different sequences. A total of 89 had full-length (60 bp) N<sub>60</sub> regions (available at <http://www.biosci.ohio-state.edu/~microbio/Archaealhistones/DNAsubstrates.html>), and the remainder had N<sub>60</sub> regions with 57-59 bp. Ten cycle 0 molecules were also cloned and sequenced, all of which had 60 bp N<sub>60</sub> regions.

### Sequences of cycle 8 molecules

Sequencing revealed that almost all of the cycle 8 molecules had N<sub>60</sub> regions with oligo-T tracts



**Figure 1.** Structure of cycle 0 DNA molecules, rHMfB affinities and consensus of cycle 8 sequences. (a) Each cycle 0 molecule had a 60 bp region of random-sequence (N<sub>60</sub>) flanked by the 25 bp fixed sequences (25L and 25R) shown. (b) Electrophoretic separation of cycle 0 DNA from archaeal nucleosomes (rHMfB-DNA) assembled at increasing rHMfB to DNA ratios. (c) Comparison of the affinities of cycle 0 DNA, cycle 8 clones 5 and 9, *M. fervidus* 7S RNA encoding sequence (Pereira & Reeve, 1999) and (CTG)<sub>6</sub> repeat sequence (Sandman & Reeve, 1999) for rHMfB. (d) The sequences show by relative font size the frequency with which A, C, G or T occurred at each location in the 89 different 60 bp cycle 8 N<sub>60</sub> sequences. The most prevalent nucleotide at each location, relative to 25L, are shown in the sequences from top to bottom. The font size equivalent to 80% occupancy of a site by one nucleotide is indicated.

(T<sub>3.7</sub>) at ~10 bp intervals separated by GC-rich regions. When aligned by the common 25L flanking sequence (Figure 1(d)), a pattern of alternating TT- and GC-rich regions was clearly evident between positions 0 to 42 relative to 25L, followed by a region of more random sequence (positions 43 and 56), and G nucleotides then almost exclusively filled positions 58-60 immediately preceding 25R. This pattern was not present in any of the cycle 0 molecules which had essentially random N<sub>60</sub> sequences (see <http://www.biosci.ohio-state.edu/~microbio/Archaealhistones/DNAsubstrates.html>).

#### rHMfB affinity of cycle 8 molecules

Gel-shift assays of 16 individual cycle 8 molecules, undertaken in the presence of an excess of unlabeled sonicated herring-sperm competitor DNA, revealed that these molecules had affinities for rHMfB that ranged from ~9 to 18-fold higher than the affinity of the cycle 0 population (Bailey,

2000). These cycle 8 molecules also bound rHMfB with ~six- to tenfold higher affinities than the 7S RNA encoding (Pereira & Reeve, 1999) and (CTG)<sub>6</sub>-containing sequences (Sandman & Reeve, 1999) shown previously to direct the positioned assembly of rHMfB into archaeal nucleosomes (Figure 1(c)).

#### Archaeal nucleosome positioning

A cycle 8 molecule with a single *Alu*I site in the N<sub>60</sub> region was identified and used as a representative molecule for positioning studies. Archaeal nucleosomes assembled using this DNA molecule were subjected to micrococcal nuclease (MN) digestion and, as previously observed (Pereira & Reeve, 1999), ~90 bp MN-protected fragments initially accumulated which were then replaced by ~60 bp (actually 58 bp) fragments. These molecules apparently correspond to the length of DNA that fully circumscribes the histone core, and to the length of DNA that makes direct contacts with the

histone-folds of the tetramer core, respectively (Luger *et al.*, 1997; Luger & Richmond, 1998; Bailey *et al.*, 1999; Pereira & Reeve, 1999). The 58 bp MN-protected DNA molecules were isolated, digested with *AluI*, *BamHI* and *EcoRI*, and the sizes of the digestion products determined by electrophoresis through DNA sequencing gels. None of the 58 bp molecules retained the *BamHI* or *EcoRI* cleavage sites that were originally present in 25L and 25R respectively, and *AluI* digestion generated only two discrete-sized restriction fragments, 10( $\pm$ 1) and 48( $\pm$ 1) bp in length. Therefore, virtually all of the 58 bp molecules were from the same region of the original 110 bp cycle 8 molecule identifying, as illustrated in Figure 2(b), the one preferred site on this DNA molecule at which rHMfB assembled into an archaeal nucleosome.

### Fourier-transform analyses of cycle 8 sequences

All of the different cycle 8  $N_{60}$  sequences and their reverse complements were subjected to Fourier-transform analyses which revealed enrichments for several dinucleotide steps repeated at  $\sim$ 10 bp intervals, most notably AA (=TT), TA and GC (Figure 3). Weaker, but still significant signals were also detected for TA and GC at  $\sim$ 20 bp,  $\sim$ 30 bp and  $\sim$ 40 bp intervals, demonstrating that these dinucleotides also occurred in extended harmonics. Such dinucleotide repeats were not present in the cycle 0 molecules nor in the control sequences generated by computer randomization of the cycle 8 sequences. Signals for CG (Figure 3), CC (=GG)

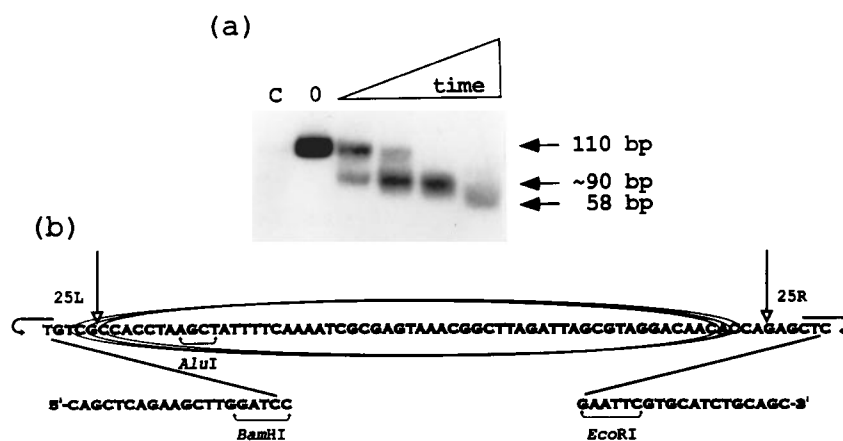
and GA repeated at  $\sim$ 10 bp intervals were also detected in the cycle 8 sequences, but at only slightly above their predicted random occurrences, and there were no signals for enrichment of AT or AC dinucleotides.

### Real-space analysis of cycle 8 sequences

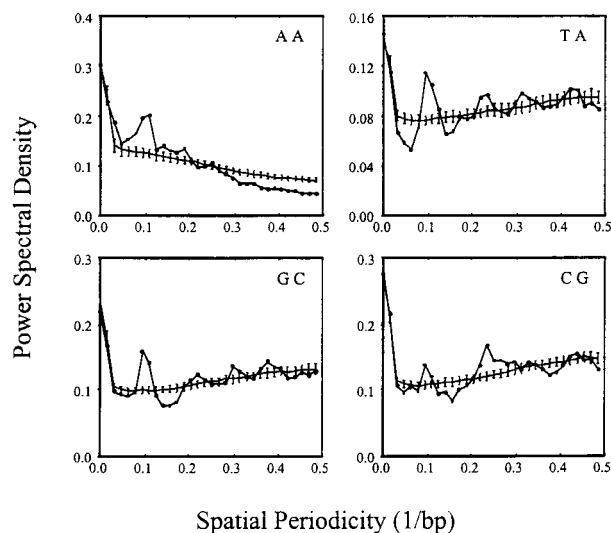
A correlation function [ $C(\lambda)$ ] was calculated as previously described (Lowary & Widom, 1998) to identify all di- and trinucleotides repeats in the cycle 8 sequences at all separations ( $\lambda$  bp) from 1 to 55 bp. Many non-random correlations were present (list available from J.W. on request), and consistent with the Fourier-transform analyses, the most enriched were AA followed by AA, AA followed by TA, and GC followed by GC at  $\sim$ 10,  $\sim$ 20 and  $\sim$ 30 bp intervals, TA followed by TA at  $\sim$ 10 bp, and GC followed by AA and TA at  $\sim$ 5 and  $\sim$ 15 bp intervals (Figure 4). When summarized, the cycle 8  $N_{60}$  sequences were therefore dominated by  $(GCN_3AA)_{1-3}$  and  $(GCN_3TA)_{1-3}$  repeats. When the same calculations were applied to the control sequences generated by randomization of the cycle 8  $N_{60}$  sequences there were no signals for the presence of non-random di- or trinucleotide spatial relationships.

### Fourier-transform analyses of prokaryotic genome sequences

In a previous study (Widom, 1996), Fourier-transform analyses of the complete *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Haemophilus*



**Figure 2.** Micrococcal nuclease (MN) digestion and cycle 8 DNA positioning of an archaeal nucleosome. (a) Archaeal nucleosomes, assembled using the cycle 8 molecule with the sequence shown in panel (b), were exposed to MN for 0, 1, 5, 15 and 45 minutes. DNA fragments protected from MN digestion were separated by PAGE and visualized by autoradiography. Control (C) exposure of the DNA to MN for two minutes in the absence of rHMfB resulted in complete digestion. High-resolution electrophoresis through DNA sequencing gels demonstrated that the DNA fragments that accumulated after extended MN digestion were 58 bp. (b) None of the 58 bp MN protected DNA fragments retained the *BamHI* and *EcoRI* cleavage sites, and *AluI* digestion generated only discrete 10( $\pm$ 1) bp and 48( $\pm$ 1) bp restriction fragments. Archaeal nucleosome assembly therefore either occurred at three overlapping sites, displaced by 1 bp as indicated by the ovals, or more likely at one site with the  $\pm$ 1 bp difference reflecting MN nicking of histone-bound DNA near the entry and exit sites of the nucleosome (Hayes & Wolffe, 1993). The vertical arrows identify the sites of fusion of the  $N_{60}$  sequence with 25L and 25R.



**Figure 3.** Fourier-transform analysis of dinucleotides in the cycle 8 sequences. As shown, strong signals were present for AA (=TT), TA and GC and a weaker signal for CG at an  $\sim 0.10 \text{ bp}^{-1}$  ( $\sim 10 \text{ bp}$ ) periodicity in the non-redundant database of 111 cycle 8 sequences with subsidiary maxima in the TA and GC transforms reflecting higher harmonics of the  $\sim 0.10 \text{ bp}^{-1}$  signal. The continuous lines with (o) data points are the experimental results, and the lines with error bars are the mean and standard deviation values, respectively, of 100 identical analyses performed on computer-randomized versions of the cycle 8 sequences.

*influenzae* genome sequences revealed that AA dinucleotides were enriched at  $\sim 10.2 \text{ bp}$  intervals in both eucaryal genomes, the average helical repeat of DNA wrapped around an eucaryal nucleosome (Luger *et al.*, 1997), but that this dinucleotide repeat motif was not enriched in the bacterial genome. As this motif is also enriched in genomic DNA packaged into eucaryal nucleosomes (Ioshikhes *et al.*, 1992, 1996; Bina, 1994; Bolshoy, 1995; Widlund *et al.*, 1997; Thåström *et al.*, 1999), its over-representation in the yeast and nematode genome sequences presumably reflects a positive selection for sequences that facilitate nucleosome assembly (Widom, 1996). *H. influenzae* does not have histones (Fleischmann *et al.*, 1995), however several procaryotic genome sequences are now available from histone-containing *Euryarchaeota* (Bult *et al.*, 1996; Klenk *et al.*, 1997; Smith *et al.*, 1997; Kawarabayashi *et al.*, 1998), and Fourier-transform analyses of these genome sequences revealed that AA dinucleotides are enriched at  $\sim 10.1 \text{ bp}$  intervals (Figure 5), the helical repeat of DNA wrapped around an archaeal nucleosome (Pereira & Reeve, 1999). In contrast, this repetitive motif was not found at above random frequencies in the genomes of *Crenarchaeota* (Kawarabayashi *et al.*, 1999, <http://niji.imb.nrc.ca/sulfhome>), nor in many additional bacterial genome sequences

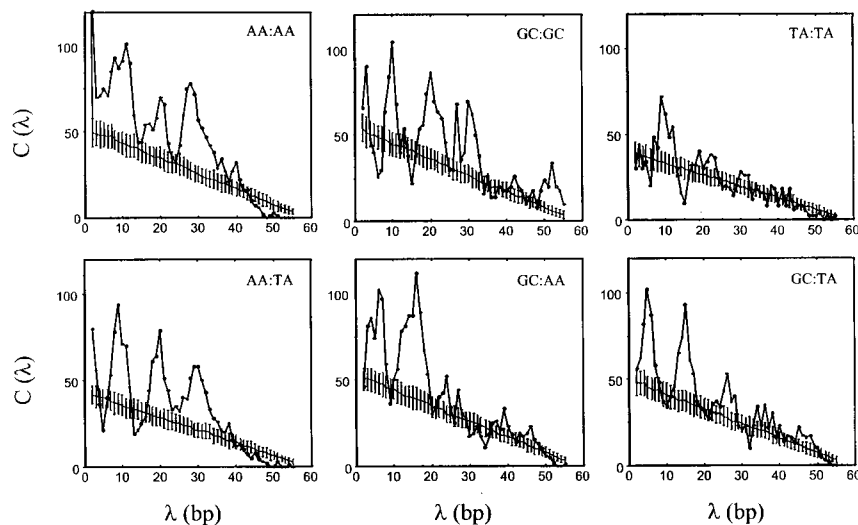
(Kunst *et al.*, 1997; Cole *et al.*, 1998; Deckert *et al.*, 1998; Stephens *et al.*, 1998; Nelson *et al.*, 1999), procaryotes that do not have histones (Figure 5). Interestingly, these procaryotic genomes do contain an AA repeat, but the spacing is  $\sim 11 \text{ bp}$  (Figure 5), which could facilitate supercoiling and/or result from sequences that encode protein  $\alpha$ -helices (Widom, 1996; Herzel *et al.*, 1999). The AA harmonic at  $\sim 11 \text{ bp}$  is compatible with the histone selection for AA at  $\sim 10.1 \text{ bp}$  intervals as both motifs are present in the yeast (Widom, 1996), *Methanococcus jannaschii* and *Methanobacterium thermoautotrophicum* genome sequences (Figure 5).

#### Fourier-transform analysis of *M. thermoautotrophicum* nucleosomal DNA

Fragments of *M. thermoautotrophicum* genomic DNA incorporated into archaeal nucleosomes *in vivo* were isolated as described by Pereira *et al.* (1997) and sequenced for comparison with the sequences of the synthetic DNA molecules selectively incorporated *in vitro* into archaeal nucleosomes. Fourier-transform analyses of 182 such sequences (58–62 bp) revealed that AA was clearly over-represented at  $\sim 10 \text{ bp}$  intervals, and weaker but still significant signals were also present for CG and TA at  $\sim 10 \text{ bp}$  intervals. These motifs were not however as pronounced as in the cycle 8 synthetic DNA molecules, consistent with the observation that natural DNAs isolated from eucaryal nucleosomes are not as enriched for positioning motifs as molecules selected *in vitro* from synthetic DNA populations by eucaryal histone octamer incorporation into nucleosomes (Lowary & Widom, 1998; Thåström *et al.*, 1999; Widlund *et al.*, 1999). It seems possible therefore that it is disadvantageous to assemble nucleosomes *in vivo* with very high intrinsic levels of stability. Five of the *M. thermoautotrophicum* nucleosomal sequences were from entirely within intergenic regions whereas the remainder contained at least part of a coding sequence. They originated from locations all around the genome (Smith *et al.*, 1997), and one contained the *M. thermoautotrophicum* homolog of the *M. feroidus* 7 S RNA encoding sequence previously shown to direct archaeal nucleosome positioning *in vivo* (Pereira & Reeve, 1999).

#### Discussion

Based on the results reported, eucaryal and archaeal histone folds are not just structural homologs but also exhibit functional homology in terms of DNA sequences preferentially incorporated into nucleosomes. The molecules selectively incorporated by rHMfB into archaeal nucleosomes have sequences that conform to repeats of the 5'-(A/T)<sub>3</sub>NN(G/C)<sub>3</sub>NN-3' motif that is enriched in bulk eucaryal nucleosomal DNA, natural eucaryal nucleosome positioning sequences, and synthetic DNA molecules with high-level affinities for eucaryal histone octamers (Satchwell *et al.*, 1986;



**Figure 4.** Real-space correlation analyses of the spatial relationships between dinucleotides in cycle 8 sequences. The spatial relationships between all pairs of dinucleotides in the non-redundant database of cycle 8 sequences were determined for all distances from 1 to 55 bp (Lowary & Widom, 1998). The most-strongly correlated dinucleotide pairs, plotted as a function of the distance (in bp) of the second dinucleotide from the first, are shown. Namely, AA, TA and GC repeated at distances of ~10, 20 and 30 bp, AA followed by TA at distances of ~10, 20 and 30 bp, and GC followed by AA or TA at distances of ~5, 15 and 25 bp. Also shown as lines with error bars are the mean and standard deviations of random expectation (generated from randomized sequences as in Figure 3).

Shrader & Crothers, 1989, 1990; Widlund *et al.*, 1997; Lowary & Widom, 1997, 1998; Thåström *et al.*, 1999). It is now generally accepted that such molecules with alternating A/T and G/C-rich regions most readily accept the DNA groove distortions needed to wrap DNA tightly around a histone core (Shrader & Crothers, 1990; Widom, 1998; Fitzgerald & Anderson, 1999), however the sequences selected by rHMfB point to the presence of additional sequence parameters. AA (=TT) and TA but not AT were selected at ~10 bp intervals, and GC was selected far more strongly than CC (=GG) or CG. Purine *versus* pyrimidine differences were also not simple determinants as AA (=TT) and TA were equivalent, but TA, CA (=TG) and CG were not. Specific DNA structures, mechanics or histone residue-DNA interactions that facilitate archaeal nucleosome assembly are apparently provided by some, but not all dinucleotide sequences. Based on very similar electrophoretic mobilities, none of the cycle 8 molecules had grossly aberrant shapes (Bailey, 2000). Their selective incorporation into archaeal nucleosomes presumably therefore reflected their above-average abilities to form such structures and accommodate structural distortions rather than their being inherently highly-curved molecules.

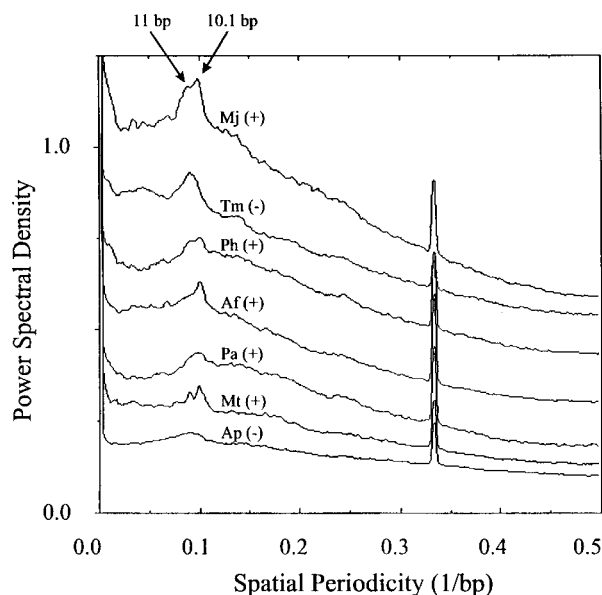
Based on the Fourier-transform analyses of genome sequences reported here (Figure 4) and by Widom (1996), there is a striking and direct correlation between the presence of histones and the over-representation of dinucleotide repeats that facilitate nucleosome assembly in genomic DNAs. It seems therefore that the use of histones, or more

precisely the histone fold, to package genomic DNA and employing nucleosome positioning to regulate gene expression has imposed constraints on genome sequence evolution. Based on only vertical gene transmission through the three-domain universal phylogenetic tree (Woese *et al.*, 1990; Woese, 2000), histones and this system genome packaging apparently evolved before the archaeal-eucaryal divergence and were then lost in the cre-narchaeal lineage. However, with lateral gene transfer, the histone fold could have evolved first early in either the euryarchaeal or eucaryal lineages. Recently, an alternative cell-cell fusion hypothesis was proposed for the origin of the *Eucarya*, based on comparative biochemistry arguments in which the eucaryal nucleus is descended from a hydrogen-consuming methanogen (Martin & M(ller, 1998) which, it seems worth noting, would have had been an *Euryarchaeon* with histones (Sandman & Reeve, 1998). In terms of DNA packaged per unit protein, the histone-fold based system of DNA wrapping is the most efficient, and it seems possible that the evolution of this system was a key event that allowed genome expansion and therefore the evolution and divergence of the *Eucarya* (Minsky *et al.*, 1997; Sandman *et al.*, 1998).

## Materials and Methods

### Preparation of rHMfB

rHMfB was synthesized, purified and quantitated from *Escherichia coli* JM105 as described by Sandman *et al.* (1995).



**Figure 5.** Power spectra for the AA dinucleotide in archaeal and bacterial genome sequences. The results show data from analyses of the complete euryarchaeal genome sequences of the histone containing (+) *Archaeoglobus fulgidus* (Af; Klenk *et al.* (1997)), *Methanococcus jannaschii* (Mj, Bult *et al.* (1996), *M. thermoautotrophicum* (Mt, Smith *et al.*, 1997), *Pyrococcus abyssi* (Pa, (<http://www.genoscope.cns.fr/Pab>)) and *Pyrococcus horikoshii* (Ph, Kawarabayashi *et al.*, 1998), and of the crenarchaeal and bacterial genomes of *Aeropyrum pernix* (Ap, Kawarabayashi *et al.*, 1999) and *Thermotoga maritima* (Tm, Nelson *et al.*, 1999), respectively, that do not (–) contain histones. Analyses of genome sequences from *Sulfolobus solfataricus* P2 (<http://niji.imb.nrc.ca/sulf-home>), and additional bacterial genome sequences (Kunst *et al.*, 1997; Cole *et al.*, 1998; Deckert *et al.*, 1998; Stevens *et al.*, 1998) confirmed the presence of a spatial periodicity peak at  $0.09 \text{ bp}^{-1}$  ( $\sim 11 \text{ bp}$  repeat), but absence of the peak at  $0.1 \text{ bp}^{-1}$  ( $\sim 10.1 \text{ bp}$  repeat) that is present in the euryarchaeal genomes. A difference in AA spatial periodicities in bacterial versus archaeal genomes was noted previously (Herzel *et al.*, 1999) but crenarchaeal genome sequences were not then available. Power spectra are shown on a relative scale with a small constant increment added to space the curves for visual clarity.

### Cycle 0 DNA

A mixture of single-stranded oligonucleotides (1.6  $\mu\text{M}$ ) with the sequences 5'-CAGCTCAGAAGCTTGGATCCTGTCG-N<sub>60</sub>-GAGCTCGAATTCGTGCATCTG-CAGC-3' was purchased from Integrated DNA Technologies (Coralville, IA). Full-length molecules (110-mers) were purified from shorter products by polyacrylamide gel electrophoresis (PAGE) and converted into double-stranded DNA by primer extension. Reaction mixtures contained 15 nM of primer r25R (5'-GCTGCAGATG-CACGAATTCGAGCTC-3'), 500  $\mu\text{M}$  dNTPs, 5 mM  $\text{MgCl}_2$ , 7.5 mM DTT, 10 mM Tris-HCl (pH 7.5) and were incubated at 95 °C for one minute, 72 °C for one minute, and then at 37 °C for three minutes. DNA polymerase I

was added (3 U of 3'  $\rightarrow$  5' exonuclease<sup>-</sup> Klenow fragment (New England Biolabs, Beverly, MA)) and incubation continued at 37 °C for 30 minutes. The 110 bp double-stranded DNA products (cycle 0 DNA) were deproteinized by phenol-chloroform extraction, purified by PAGE through an 8% T, 0.11% C polyacrylamide gel and dissolved in 80  $\mu\text{l}$  TE, 10 mM Tris-HCl, 1 mM EDTA (pH 7.5).

### Archaeal nucleosome assembly and selection

Before each selection cycle, a small portion of the DNA was [<sup>32</sup>P]-end-labeled and mixed with the unlabeled DNA to facilitate the quantitation of free DNA and rHMfB-DNA binding and the localization of free DNA and rHMfB-bound DNA following their separation by PAGE. An electrophoretic mobility-shift assay (Bailey *et al.*, 1999) was used to determine the rHMfB to DNA ratio that resulted in the incorporation of  $\sim 5\%$  of the input DNA into archaeal nucleosomes. Reaction mixtures (10  $\mu\text{l}$ ) that contained this ratio were then assembled in 100 mM KCl, 50 mM Tris-HCl (pH 8) and, after incubation at 25 °C for 25 minutes, 1  $\mu\text{l}$  of gel loading buffer (0.4% (w/v) bromophenol blue, 0.4% (v/v) xylene cyanol, 25% (v/v) ficoll-400) was added, and the reaction products were separated by PAGE through an 8% T, 0.11% C polyacrylamide gel run for two hours at 8 V/cm in 90 mM Tris-borate, 2 mM EDTA (TBE (pH 8)). The region of the gel that contained archaeal nucleosomes was located by autoradiography and excised. The DNA molecules were eluted from the gel by incubation overnight at 37 °C in 500 mM ammonium acetate, 0.1% SDS, 2 mM EDTA, precipitated in ethanol, and dissolved in 50  $\mu\text{l}$  TE.

### PCR amplifications

Aliquots (6  $\mu\text{l}$ ) of the selected DNA were amplified in reaction mixtures that contained 10 pmols of primers r25R and r25L (5'-CGACAGGATCCAAGCTTCTGAGCTG), 200  $\mu\text{M}$  dNTPs, 6 mM  $\text{MgCl}_2$ , 50 mM KCl, 20 mM Tris-HCl (pH 8.4) and 5 U Taq DNA polymerase (BRL, Gaithersburg, MD) by 13 cycles of incubation at 95 °C for one minute followed by 72 °C for one minute. The amplified products were purified by PAGE, pooled and used as the substrate DNA in the next round of selection.

### Cloning and sequencing

Molecules from the cycle 0 and cycle 8 populations were cloned by ligation into the pGEM-T Easy vector and transformation into *E. coli* JM109 high efficiency competent cells (Promega Corp., Madison, WI). Cloned inserts in plasmids isolated from ampicillin-resistant transformants were sequenced using ABI sequencing kits (PE Applied Biosystems, Foster City, CA).

### HMfB-DNA affinity assays

Experimental [<sup>32</sup>P]-labeled-DNA (0.1 nM) and increasing amounts of rHMfB were incubated in 100 mM KCl, 50 mM Tris-HCl (pH 8) at 25 °C for 30 minutes with either 800 ng (assays of cycle 8 molecules) or 1 ng

(assays of the cycle 0, 7S and (CTG)<sub>6</sub> molecules) of competing unlabeled sonicated herring sperm DNA (100–600 bp; Promega Corp). The reaction products were separated by PAGE (Bailey *et al.*, 1999) and the amount of DNA incorporated into archaeal nucleosomes was determined by using an Instant Imager 2024 (Packard, Meriden, CT). All assays were performed in triplicate. The cycle 0 and cycle 8 molecules were 110 bp in length (Figure 1), and the 7 S DNA from *M. fervidus* (Pereira & Reeve, 1999) and the (CTG)<sub>6</sub>-containing molecule (Sandman & Reeve, 1999) were 113 bp *KspI-SpeI* and *SnaBI-XbaI* restriction fragments, respectively.

### Archaeal nucleosome positioning

Cycle 8 molecules were amplified from plasmid clones using the PCR conditions described above in reaction mixtures supplemented with 20  $\mu$ Ci [ $\alpha$ -<sup>32</sup>P]dATP (3000 Ci/mmol). The [<sup>32</sup>P]-labeled PCR product was purified by PAGE and incubated at an ~1:10 molar ratio with rHMfB in 1 mM CaCl<sub>2</sub>, 100 mM KCl, 50 mM Tris-acetate (pH 8.8) for 25 minutes at 37°C. Micrococcal nuclease (MN; Sigma, St. Louis, MO) was added (final concentration of 0.005 U/ $\mu$ g DNA), and aliquots (10  $\mu$ l) were removed at intervals from 1 to 45 minutes and the MN-digestion stopped by addition of 2  $\mu$ l of 100 mM EDTA. The reaction tubes were placed on ice, and following phenol-chloroform extraction and ethanol precipitation, the DNA molecules remaining were dissolved in TE and separated by PAGE through 12% T, 0.16% C gels run in TBE. They were located autoradiography, isolated from the gel, subjected to restriction enzyme digestions and the digestion products were then mixed with an equal volume of formamide loading buffer (80% formamide, 10 mM EDTA (pH 8), 0.1% (v/v) xylene cyanol, 0.1% (w/v) bromophenol blue) and visualized by autoradiography after separation by PAGE at 52 V/cm through 10% T, 0.52% C sequencing gels that contained 8.3 M urea in 0.5X TBE (Pereira & Reeve, 1999).

### Isolation, cloning and sequencing of genomic DNA incorporated into archaeal nucleosomes *in vivo*

*M. thermoautotrophicum*  $\Delta$ H cells (Pihl *et al.*, 1994) were washed and resuspended in 50 mM Tris-acetate, 1 mM CaCl<sub>2</sub> (pH 8.8) and ruptured by passage through a French pressure cell (10,000 psi). The resulting lysate was incubated for one hour at 37°C with 10 U MN/g wet weight of original cell paste, SDS (0.2% w/v), EDTA (20 mM) and proteinase K (300  $\mu$ g/ml) were added and incubation continued at 37°C for three hours. DNA molecules (~60 bp) that remained after phenol-chloroform extraction, ethanol precipitation and incubation with RNase A (40  $\mu$ g/ml) for 30 minutes at 37°C were purified by excision after PAGE through an 8% T, 0.27% C gel (Pereira *et al.*, 1997). The ends of these DNA molecules were made double-stranded by incubation with the Klenow fragment of DNA polymerase I (Maniatis *et al.*, 1989) and 5'-dA extensions were added by incubation with *Taq* DNA polymerase (BRL) using the tailing procedure described by Promega (Technical Manual; Madison, WI). Individual molecules were cloned into the pGEM-T Easy vector and the inserts in plasmids isolated from ampicillin-resistant transformants of *E. coli* JM109 were sequenced and their locations on the *M. thermoautotrophicum* genome sequence (Smith *et al.*, 1997) were determined.

### Fourier-transform and real-space correlation function analyses

The cycle 0 and cycle 8 sequences and *M. thermoautotrophicum* genome fragments were subjected to Fourier-transform and power spectra analyses as described (Widom, 1996; Lowary & Widom, 1998; Thåström *et al.*, 1999). For each of the ten distinct dinucleotide steps, sequences were encoded "1" at positions corresponding to the first nucleotide of that step and "0" elsewhere. Power spectra were calculated using the program SPCTRM (Press *et al.*, 1986) in consecutive half-overlapping 64 bp segments, and were summed for all segments from the encoded sequence and its reverse-complement for all sequences in the non-redundant dataset. Hence, results for any particular dinucleotide are also identical for the reverse-complement of that dinucleotide. To evaluate the statistical significance of the transforms, power spectra were calculated as above in 100 separate runs over randomized versions of the non-redundant dataset, and the mean and standard deviation in the power spectral density from the 100 random runs were evaluated at each reciprocal lattice point. To create the randomized sequences, the number of occurrences of a particular mononucleotide or dinucleotide within each real sequence was counted, and then an equal number were placed at random locations chosen by the random number generator RAN1 (Press *et al.*, 1986), with proper handling of permitted and forbidden adjacencies (Widom, 1996; relevant only for the dinucleotides). As for the real sequences, both the forward and reverse-complement of each randomized sequence were included in the calculations. Calculations and checks establishing the validity of these computer programs have been discussed (Widom, 1996; Lowary & Widom, 1998; Thåström *et al.*, 1999).

Archaeal and bacterial genome sequences available in public databases (Bult *et al.*, 1996; Klenk *et al.*, 1997; Kunst *et al.*, 1997; Smith *et al.*, 1997; Cole *et al.*, 1998; Kawarabayashi *et al.*, 1998, 1999; Deckert *et al.*, 1998; Stevens *et al.*, 1998; Nelson *et al.*, 1999) were also subjected to Fourier-transform and power spectra analyses as described (Widom, 1996), using half-overlapping 1024 bp segments to provide an appropriate balance between signal to noise ratios and the spatial scope and resolution of the analyses. Calculations again included both the forward and reverse-complement sequences. Statistical significance was assessed, as described above, except that only ten random trials were used for each sequence because of the much larger datasets.

Correlation functions were evaluated for the selected sequences as described by Lowary & Widom (1998). For each dinucleotide pair (*ab*, *xy*) or trinucleotide pair (*abc*, *xyz*), the number of times that *ab* (or *abc*) was found  $\lambda$  bp ahead of *xy* (or *xyz*), was determined for all values of  $\lambda$  from 1 to 55 bp. Numbers of occurrences (for each  $\lambda$ ) were summed over all sequences and their reverse complements in the dataset. Calculations were performed for all 256 dinucleotide and 4096 trinucleotide pairs, yielding 256 dinucleotide correlation functions and 4096 trinucleotide correlation functions, each defined over 55 points ( $\lambda$  values). The statistical significance of each measurement was assessed using a method similar to that described above for the statistical evaluation of the Fourier-transforms. Each calculation (for all pairs, for all  $\lambda$  values) was carried out 100 times on randomized versions of the dataset (again with proper handling of permitted and forbidden adjacencies), and the mean and standard deviations ( $\sigma$ ) over the 100 random trials were evaluated.

## Acknowledgments

This work was supported by NIH research grants GM53185 (J.N.R.), GM54692 and GM58617 (J.W.), and DOE grant DE-FG02-8713731 (J.N.R.).

## References

- Alilat, M., Sivolob, A., Révet, B. & Prunell, A. (1999). Nucleosome dynamics IV. Protein and DNA contributions in the chiral transition of the tetrasome, the histone (H3-H4)<sub>2</sub> tetramer-DNA particle. *J. Mol. Biol.* **291**, 815-841.
- Bailey, K. A. (2000). DNA interactions of the archaeal histone HMf from the hyperthermophilic methanogen *Methanothermus fervidus*. Ph. D. Thesis, The Ohio State University, Columbus, USA, Ohio.
- Bailey, K. A., Chow, C. S. & Reeve, J. N. (1999). Histone stoichiometry and DNA circularization in archaeal nucleosomes. *Nucl. Acids Res.* **27**, 532-536.
- Beato, M. & Einfeld, K. (1997). Transcription factor access to chromatin. *Nucl. Acids Res.* **25**, 3559-3563.
- Bina, M. (1994). Periodicity of dinucleotides in nucleosomes derived from simian virus 40 chromatin. *J. Mol. Biol.* **235**, 198-208.
- Bolshoy, A. (1995). CC dinucleotides contribute to the bending of DNA in chromatin. *Nature Struct. Biol.* **2**, 446-448.
- Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D. *et al.* (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058-1073.
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C. *et al.* (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393**, 537-544.
- Decanniere, K., Babu, A. M., Sandman, K., Reeve, J. N. & Heinemann, U. (2000). Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. *J. Mol. Biol.* **303**, 35-47.
- Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L. *et al.* (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature*, **392**, 353-358.
- Dong, F. & van Holde, K. E. (1991). Nucleosome positioning is determined by the (H3-H4)<sub>2</sub> tetramer. *Proc. Natl Acad. Sci. USA*, **88**, 10596-10600.
- Fitzgerald, D. J. & Anderson, J. N. (1999). DNA distortion as a factor in nucleosome positioning. *J. Mol. Biol.* **293**, 477-491.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F. *et al.* (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496-512.
- Hamiche, A., Carot, V., Alilat, M., De Lucia, F., O'Donohue, M.-F., Révet, B. & Prunell, A. (1996). Interaction of the histone (H3-H4)<sub>2</sub> tetramer of the nucleosome with positively supercoiled DNA minicircles: potential flipping of the protein from a left- to a right-handed superhelical form. *Proc. Natl Acad. Sci. USA*, **93**, 7588-7593.
- Hamiche, A. & Richard-Foy, H. (1998). The switch in helical handedness of the histone (H3 + H4)<sub>2</sub> tetramer within a nucleoprotein particle requires a reorientation of the H3-H3 interface. *J. Biol. Chem.* **273**, 9261-9269.
- Hayes, J. J. & Wolffe, A. P. (1993). Preferential and asymmetric interaction of linker histones with 5 S DNA in the nucleosome. *Proc. Natl Acad. Sci. USA*, **90**, 6415-6419.
- Hayes, J. J., Clark, D. J. & Wolffe, A. P. (1991). Histone contributions to the structure of DNA in the nucleosome. *Proc. Natl Acad. Sci. USA*, **88**, 6829-6833.
- Herzel, H., Weiss, O. & Trifonov, E. N. (1999). 10-11 bp periodicities in complete genomes reflect protein structure and DNA folding. *Bioinformatics*, **15**, 187-193.
- Ioshikhes, I., Bolshoy, A. & Trifonov, E. N. (1992). Preferred positions of AA and TT dinucleotides in aligned nucleosomal DNA sequences. *J. Biomol. Struct. Dynam.* **9**, 1111-1117.
- Ioshikhes, I., Bolshoy, A., Derenshteyn, K., Borodovsky, M. & Trifonov, E. N. (1996). Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. *J. Mol. Biol.* **262**, 129-139.
- Kawarabayashi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y. *et al.* (1998). Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**, 55-76.
- Kawarabayashi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y. *et al.* (1999). Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.* **6**, 83-101.
- Klenk, H. P., Clayton, R. A., Tomb, J.-F., White, O., Nelson, K. E. *et al.* (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature*, **390**, 364-370.
- Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G. *et al.* (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature*, **390**, 249-256.
- Lowary, P. T. & Widom, J. (1997). Nucleosome packaging and nucleosome positioning of genomic DNA. *Proc. Natl Acad. Sci. USA*, **94**, 1183-1188.
- Lowary, P. T. & Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19-42.
- Luger, K. & Richmond, T. J. (1998). DNA binding within the nucleosome core. *Curr. Opin. Struct. Biol.* **8**, 33-40.
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997). Crystal structure of the nucleosome core at 2.8 Å resolution. *Nature*, **389**, 251-260.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Martin, W. & Müller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature*, **392**, 37-41.
- Minsky, A., Ghirlando, R. & Reich, Z. (1997). Nucleosomes: a solution to a crowded intracellular environment? *J. Theor. Biol.* **188**, 379-385.
- Musgrave, D. R., Sandman, K. M. & Reeve, J. N. (1991). DNA binding by the archaeal histone HMf results in positive supercoiling. *Proc. Natl Acad. Sci. USA*, **88**, 10397-10401.
- Musgrave, D. R., Forterre, P. & Slesarev, A. (2000). Negative constrained DNA supercoiling in archaeal nucleosomes. *Mol. Microbiol.* **35**, 341-349.

- Nelson, K., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J. *et al.* (1999). Evidence for lateral gene transfer between *Archaea* and *Bacteria* from genome sequence of *Thermotoga maritima*. *Nature*, **399**, 323-329.
- Pereira, S. L., Grayling, R. A., Lurz, R. & Reeve, J. N. (1997). Archaeal nucleosomes. *Proc. Natl Acad. Sci. USA*, **94**, 12633-12637.
- Pereira, S. L. & Reeve, J. N. (1999). Archaeal nucleosome positioning sequence from *Methanothermobacter ferredoxigenes*. *J. Mol. Biol.* **289**, 675-681.
- Pihl, T. D., Sharma, S. & Reeve, J. N. (1994). Growth phase-dependent transcription of the genes that encode the two methyl coenzyme M reductase isoenzymes and N<sup>5</sup>-methyltetrahydromethanopterin-coenzyme M methyltransferase in *Methanobacterium thermoautotrophicum*  $\Delta$ H. *J. Bacteriol.* **176**, 6384-6391.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986). *Numerical Recipes*, Cambridge Press, Cambridge, UK.
- Sandman, K. & Reeve, J. N. (1998). Origin of the eukaryotic nucleus. *Science*, **280**, 501-503.
- Sandman, K. & Reeve, J. N. (1999). Archaeal nucleosome positioning by CTG repeats. *J. Bacteriol.* **181**, 1035-1038.
- Sandman, K., Krzycki, J. A., Dobrinski, B., Lurz, R. & Reeve, J. N. (1990). HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermobacter ferredoxigenes*, is most closely related to histones. *Proc. Natl Acad. Sci. USA*, **87**, 5788-5791.
- Sandman, K. S., Grayling, R. A. & Reeve, J. N. (1995). Improved N-terminal processing of recombinant proteins synthesized in *Escherichia coli*. *Bio/Technol.* **13**, 504-506.
- Sandman, K., Pereira, S. L. & Reeve, J. N. (1998). Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cell Mol. Life Sci.* **54**, 1350-1364.
- Satchwell, S. C., Drew, H. R. & Travers, A. A. (1986). Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191**, 659-675.
- Shrader, T. E. & Crothers, D. M. (1989). Artificial nucleosome positioning sequences. *Proc. Natl Acad. Sci. USA*, **86**, 7418-7422.
- Shrader, T. E. & Crothers, D. M. (1990). Effects of DNA sequence and histone-histone interactions on nucleosome placement. *J. Mol. Biol.* **216**, 69-84.
- Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J. *et al.* (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum*  $\Delta$ H: functional analysis and comparative genetics. *J. Bacteriol.* **179**, 7135-7155.
- Soares, D. J., Sandman, K. & Reeve, J. N. (2000). Mutational analysis of archaeal histone-DNA interactions. *J. Mol. Biol.* **297**, 39-47.
- Spangenberg, C., Eisfeld, K., Stünkel, W., Luger, K., Flaus, A., Richmond, T. J., Tuss, M. & Beato, M. (1998). The mouse mammary tumour virus promoter positioned on a tetramer of histones H3 and H4 binds nuclear factor 1 and OTF1. *J. Mol. Biol.* **278**, 725-739.
- Starich, M. R., Sandman, K., Reeve, J. N. & Summers, M. F. (1996). NMR Structure of HMfB from the hyperthermophile, *Methanothermobacter ferredoxigenes*, confirms that this archaeal protein is a histone. *J. Mol. Biol.* **255**, 187-203.
- Stephens, R. S., Kalman, S., Lammel, C., Fan, J., Marathe, R. *et al.* (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science*, **282**, 754-759.
- Sullivan, S. A., Aravind, L., Makalowska, I., Baxevanis, A. D. & Landsman, D. (2000). The histone database: a comprehensive WWW resource for histones and histone fold-containing proteins. *Nucl. Acids Res.* **28**, 320-322.
- Svaren, J. & Horz, W. (1996). Regulation of gene expression by nucleosomes. *Curr. Opin. Genet. Dev.* **6**, 164-170.
- Thåström, A., Lowary, P. T., Widlund, H. R., Cao, H., Kubista, M. & Widom, J. (1999). Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. Mol. Biol.* **288**, 213-229.
- Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.
- van Holde, K. (1993). Transcription: the omnipotent nucleosome. *Nature*, **362**, 111-112.
- Widlund, H. R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P. E., Kahn, J. D., Crothers, D. M. & Kubista, M. (1997). Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.* **267**, 807-817.
- Widlund, H. R., Kuduvalli, P. N., Bengtsson, M., Cao, H., Tullius, T. D. & Kubista, M. (1999). Nucleosome structural features and intrinsic properties of the TATAACGCC repeat sequence. *J. Biol. Chem.* **274**, 31847-31852.
- Widom, J. (1996). Short-range order in two eukaryotic genomes: relation to chromosome structure. *J. Mol. Biol.* **259**, 579-588.
- Widom, J. (1998). Structure, dynamics, and function of chromatin *in vitro*. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 285-327.
- Woese, C. R. (2000). Interpreting the universal phylogenetic tree. *Proc. Natl Acad. Sci. USA*, **97**, 8392-8396.
- Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria* and *Eucarya*. *Proc. Natl Acad. Sci. USA*, **87**, 4576-4579.
- Wolffe, A. P. (1994). Nucleosome positioning and modification: chromatin structures that potentiate transcription. *Trends Biochem. Sci.* **19**, 240-244.
- Zhu, W., Sandman, K., Lee, G. E., Reeve, J. N. & Summers, M. F. (1998). NMR structure and comparison of the archaeal histone HFoB from the mesophile *Methanobacterium formicicum* with HMfB from the hyperthermophile *Methanothermobacter ferredoxigenes*. *Biochemistry*, **37**, 10563-10580.

Edited by T. Richmond

(Received 15 May 2000; received in revised form 17 August 2000; accepted 23 August 2000)