

Open Access Article

The authors, the publisher, and the right holders grant the right to use, reproduce, and disseminate the work in digital form to all users.



Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 27, Issue Number 6, (2010)
Current Perspectives on Nucleosome Positioning
©Adenine Press (2010)

The DNA Sequence-dependence of Nucleosome Positioning *in vivo* and *in vitro*

<http://www.jbsdonline.com>

Andrew Travers^{1,2,*}
Edwige Hiriart¹
Mark Churcher¹
Micaela Caserta^{3,4}
Ernesto Di Mauro^{3,4}

Abstract

The contribution of histone-DNA interactions to nucleosome positioning *in vivo* is currently a matter of debate. We argue here that certain nucleosome positions, often in promoter regions, in yeast may be, at least in part, specified by the DNA sequence. In contrast other positions may be poorly specified. Positioning thus has both statistical and DNA-determined components. We further argue that the relative affinity of the octamer for different DNA sequences can vary and therefore the interaction of histones with the DNA is a 'tunable' property.

Introduction

Both *in vitro* and *in vivo* the histone octamer can form nucleosomes on a wide spectrum of DNA sequences, independent of base composition. It thus lacks the base-specific sequence selectivity typical of transcription factors. Yet both *in vitro* and *in vivo* the octamer adopt a rather precise position on a given DNA sequence (1-10). Importantly *in vivo* histones are essential participants in gene regulation (11, 12). The *in vitro* data argue strongly that precise positioning is a consequence of selection by histone-DNA interactions but defining the role, if any, of these interactions *in vivo* has proved elusive. A particular problem is the finding that *in vivo* the same DNA sequence can accommodate different nucleosome spacings, and hence different positions, not only in different tissues of the same organism (13, 14) but also when DNA from one organism is transferred to another (15, 16). These observations indicate that the DNA sequence by itself need not specify a unique nucleosome array and that *in vivo* the organisation of nucleosomes may be determined by mechanisms other than direct histone-DNA interactions.

The Structure of the Nucleosome Core Particle

If histone-DNA interactions are strong determinants of positioning the nature of these determinants should be apparent in the structure of the nucleosome core particle and in the octamer itself. The detailed crystal structures of a nucleosome core particle assembled *in vitro* on a DNA palindrome derived from α -satellite DNA revealed that the DNA was wrapped around octamer in ~ 1.63 left-handed superhelical turns (17, 18), similar to the value deduced from a low-resolution structure of core particles isolated from chicken erythrocyte chromatin (19). The tight wrapping is confined to the central 12 double-helical turns in which, on average, the DNA is bent by $\sim 49^\circ$ /double helical turn. Put another way, ~ 123 bp of DNA are bent by 590° where, by comparison, the persistence length of mixed-sequence B-form DNA in solution is ~ 140 -150 bp (20-26).

¹MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK

²Fondation Pierre-Gilles de Gennes pour la Recherche, c/o LBPA, École Normale Supérieure de Cachan, 94235 Cachan France

³Fondazione Istituto Pasteur-Fondazione Cenci Bolognetti, c/o Dipartimento di Genetica e Biologia Molecolare Università 'La Sapienza', 00185 Rome, Italy

⁴Istituto Biologia e Patologia Molecolari del Consiglio Nazionale delle Ricerche Università 'La Sapienza', 00185 Rome Italy

*Phone: +33147407768
E-mail: aat@mrc-lmb.cam.ac.uk

The tight wrapping of DNA around the histone octamer is facilitated by certain physicochemical properties of the polymer. One such is anisotropic flexibility (aka anisotropic bendability or deformability), first proposed by Zhurkin (27) and Trifonov (28). This is essentially a geometric property which favours the adoption of DNA trajectory that not only, for a nucleosome, conforms to the superhelical path of the DNA on the surface of the histone octamer but also disfavours other trajectories. Such anisotropic deformability can be conferred by the occurrence of certain short DNA sequences in helical phase (29-32) or by intrinsic curvature dependent upon phased oligo(dA)(dT) tracts (33, 34). The short sequences facilitate tight bending by enabling the distortion of DNA such that the grooves close to the surface of the octamer are compressed while those facing the solvent are widened. These preferences are reflected in the periodic distributions of the AA/TT and GC dinucleotides in chicken erythrocytes core nucleosomal DNA sequences (30). In this set of sequences the AA/TT base-step occurs preferentially where the minor groove is in contact with the octamer, *i.e.*, on the inside of the DNA bend, and GC and other G/C-containing base steps occurs preferentially where the minor groove points away from the octamer. This pattern is characteristic of at least the outer 6 double-helical turns of the nucleosomal DNA and confers directional bending, *i.e.*, the pattern defines the orientation or rotational positioning of DNA on the histone octamer. Similar sequence preferences are associated with tight bending induced by CRP, FIS, Tn3 resolvase and the BPV E4 protein (35-38). However, around the midpoint of the chicken erythrocyte sequences, corresponding to a position where the minor groove points away from the octamer, the periodic phase of the AA/TT and GC dinucleotides is opposite to that in the outer six turns on each side, leading to the suggestion that this singularity constituted a translational positioning signal (39).

The bending preferences conferred by short nucleotide sequences can be directly related to their preferred conformations. A/T containing sequences, especially those containing sequences such as AA/TT and AAA/TTT can preferentially adopt a conformation with a narrow minor groove (40-43). It is this preferred morphology that determines the preferred orientation of these sequences in bent DNA. Conversely widening of the minor groove is energetically unfavourable. Similarly G/C containing sequences can, in crystal structures preferentially adopt a conformation with a wide minor groove (42-45). Because of the tight curvature of nucleosomal DNA the narrowing of the minor groove to ~ 3 Å at the contacts with the histone octamer is disfavoured by exocyclic groups, notably the purine 2-amino group, that protrude into the minor groove and likely hinder groove compression (46). The electrostatic potential of the narrow minor groove of A/T containing sequences also favours the binding of arginine residues at inward-facing minor grooves around the histone octamer (43). These observations imply that the periodic modulation of di- and tri-nucleotide frequencies results from the preferential exclusion of these sequences from disfavoured rotational orientations.

Although the sequence dependent bending preferences of DNA are likely important determinants of positional preferences, another determinant of affinity is the flexibility of DNA (47). This parameter is essentially the inverse of the persistence length and is a determinant of the deformation energy required for tight wrapping. While flexibility per se does not contain positional information it may affect positioning indirectly by distinguishing sequences with similar anisotropic bending preferences. Thus intrinsically curved DNA sequences, such as those conferring phased runs of oligo(dA)(dT), are inherently more rigid than isotropically bendable sequences of high flexibility (48), such as oligo(dAT)(dAT), and consequently occupy a smaller region of configurational space. Consequently the entropic penalty on binding to the octamer would be greater for a highly flexible, isotropically bendable sequence than it would be for an anisotropically deformable sequence. Although intrinsic curvature can undoubtedly influence the affinity of the octamer for DNA (33, 47, 49) the relative contributions of flexibility and anisotropic

deformability to octamer binding will be determined by the local thermodynamics of nucleosome organisation.

Variation in Positioning Signals

Although the nature of the sequence-dependence of DNA bending is well established the optimum organization of short sequences conferring anisotropic deformability in octamer binding sites is less clear (50). Variation has been reported in the sequence and structural periodicities in nucleosomal DNA, while sequences associated with well-positioned octamers in yeast differ in character from strong positioning sequences selected by *in vitro* reconstitution of nucleosomes (3, 51-53). This apparent variation can be attributed to two factors - differences in sampling procedures that result in the analysis of differing sets of nucleosomal DNA sequences and intrinsic differences in the DNA sequences associated with the histone octamer.

Analysis of different sets of budding yeast nucleosomal DNA sequences identified by parallel sequencing and by partial micrococcal nuclease (MNase) cleavage reveal differences in sequence organization. Whereas one set of sequences identified by parallel sequencing is enriched in G/C containing sequences at the midpoint (10) another such set, together with that identified by partial MNase cleavage are enriched in A/T containing sequences at the midpoint (9, 51). Again, while nucleosome core sequences isolated from chicken erythrocytes are enriched in A/T sequences at the centre those isolated from chromatosomes from the same source do not (29, 54). Similarly, preferred octamer sequences identified by reconstitution *in vitro* also differ. The commonly-used salt dilution protocol at low octamer concentration selects sequences with a G/C rich midpoint and strong periodicities of A/T- and G/C-rich short sequences (3, 52, 53). In contrast, octamer-binding sequences selected on the ovine β -lactoglobulin gene, lack a strong A/T periodicity and instead contain GG/CC and CC/GG (defined with respect to the same DNA strand) periodicities in opposite phases (55). Since there is no reason to doubt that the collated sequences are derived from nucleosomes, the observed differences in preferred sequence organisation could likely arise from differences in sampling procedures and/or from intrinsic differences.

Different sampling procedures can potentially identify different sets of nucleosome-associated DNA sequences. Initially nucleosome positions in chromatin were mapped by a method in which partial micrococcal nuclease (MNase) digestion of chromatin was followed by location of the preferred cleavage sites by indirect end labeling (5-8). A related method uses copper-*o*-phenanthroline (56), which like MNase preferentially cleaves untwisted DNA but lacks the specificity for A-T base-pairs (57). This method assumes that MNase principally, but not necessarily exclusively, cleaves DNA in chromatin in the linker between nucleosomes. The resulting cleavage pattern is then interpreted as defining a preferred array of nucleosome positions. The major drawbacks of this method are that in general it can only be applied to a limited region (up to 7-10 kb (58)) of DNA and that the nucleosome positions are identified by the length of DNA protected and not, in general, by positive identification of the associated histones.

More recently high-throughput methods have been applied to the organization of genomic positioning patterns (for reviews see refs. 49, 59-62). All these studies first isolate core nucleosome particles by limit MNase digestion followed in some cases by immunoprecipitation of particular populations and then analyse the associated DNA either by tiling arrays (63-65) or by parallel sequencing (9, 10, 53, 66, 67) The data from tiling array analysis provides a quantitative estimate of the probability of a given DNA sequence being contained within a nucleosome. These data are then further processed by Hidden Markov Model methods to identify preferred positions. The resolution of this method depends on the density of the tiling array used. In contrast parallel sequencing directly provides the sequences of

DNA molecules associated with the isolated core nucleosomes. By mapping these sequences onto genomic DNA preferred nucleosome positions can be identified. Ideally parallel sequencing in principle provides a means for mapping the relative occupancy of all occupied nucleosome positions on genomic DNA.

The coverage of the genomic position maps obtained by the tiling array and parallel sequencing methods is comparable and for many loci there is impressive agreement between the two methods. But there are also loci where there is significant disagreement both between the high throughput methods and between one or other of these methods and the partial MNase method (50). For example, at some yeast loci different studies identify different nucleosome arrays (for example refs. 9, 10) or different average nucleosome spacings (for example: refs 9, 65, 68 at the URA3 locus). Again mismatches occur between parallel sequencing studies and partial MNase cleavage studies. While some of the mismatches observed may simply result from a lack of resolution resulting from, for example, the averaging of a relatively poorly positioned – or ‘fuzzy’ – nucleosome, this consideration is unlikely to be the explanation for different studies identifying differently phased arrays at the same locus.

The disagreements between different methods, although limited, raise the issue of how representative the identified populations of nucleosomes are in relation to the actual population in chromatin. This issue is important because estimates of absolute nucleosome occupancy at a given position require that the analysed population be entirely representative. Another measure of whether different studies isolate equivalent populations is DNA sequence organization. While tiling array studies cannot be assessed for this parameter, both the partial MNase digestion and parallel sequencing methods can. When the populations of nucleosome positions identified in different studies are compared striking differences in the occurrence of A/T containing (AA,AT,TA,TT) and G/C containing (CC,CG,GC,GG) dinucleotides at the sequence midpoint – presumed to correspond approximately to the structural dyad – become apparent (50). Whereas the yeast nucleosomes analysed by the parallel sequencing studies of Field *et al.*, (10) and Kaplan *et al.*, (53) are enriched in G/C containing dinucleotides at the midpoint those analysed by the parallel sequencing studies of Albert *et al.*, (66) and Mavrich *et al.*, (9) are both enriched in A/T containing dinucleotides at the same position (50). By the criterion of sequence organization these two sets of populations are, on average, not equivalent. The yeast nucleosomes identified by partial MNase digestion, albeit a much smaller set, are even more enriched in A/T-containing dinucleotides at the midpoint, the average A/T content at this position being 73% compared with the genomic content of 62% (51).

The differences in average A/T content at the midpoint between sequences obtained by partial and limit MNase digestion could simply result from the genomic limit MNase digestion analyses sampling a far larger set of nucleosomal DNA sequences. However, another possible explanation is that the sequence variations are a consequence of procedural differences. In particular, whereas partial MNase digestion identifies the location of presumed nucleosomes in a preferred array, all the other methods involve limit digestion to core nucleosomes. For the isolated populations to be wholly representative of *in vivo* chromatin these latter studies all require the implicit assumption that during isolation MNase cleavage either occurs only exterior to core nucleosomal DNA or, that any internal cleavage within core DNA is essentially random. However, since MNase cleavage is highly sequence selective (69-71), with preferred cutting at the TA dinucleotide (10, 71), any internal cleavage by this enzyme would potentially bias the population of nucleosomes isolated (72) and also the population of DNA molecules sequenced. Even if a core nucleosome particle containing internally cleaved DNA were isolated the associated DNA would be shorter than full-length nucleosomal DNA and its sequence would not necessarily be incorporated into a database requiring full-length reads.

Differences in the frequency of A/T and G/C containing dinucleotides at the mid-points of sets of sequences obtained by the parallel sequencing could thus be potentially explained simply by differences in the extent of treatment with MNase during the initial isolation of core nucleosome particles. Indeed an early study of the sequences of chicken erythrocyte core nucleosomal DNA noted that the TA dinucleotide, and not other dinucleotides, was significantly depleted, relative to genomic DNA, in the sequences isolated (30). Although studies routinely control for the sequence selectivity of MNase cleavage at the ends of nucleosomal DNA (9, 10) there is no good experimental measure of the extent of internal cleavage in any population of isolated nucleosomes. A recent study comparing nucleosomal DNA sequences isolated from chromatin with different extents of MNase digestion found that some nucleosomes, notably those in the nucleosome-depleted regions of promoter regions, are particularly sensitive to internal cleavage by MNase while others are relatively resistant (73). This finding accords with the observation that the central region of DNA from promoter nucleosomes identified by partial MNase cleavage is enriched in TA, a preferred site for MNase cleavage (50). A further potential complication is that nucleosomes may change their positions during extensive MNase digestion.

However, not all described differences in the organisation of nucleosomal DNA can be ascribed to differences in sampling. First the nature of the short sequences determining the trajectory is variable. In contrast to periodic repetition of the AA/TT and GC base-steps found in chicken and yeast nucleosomal DNA (10, 30), sequences identified by reconstitution of nucleosomes on ovine DNA do not, on average, exhibit a strong periodic occurrence of AA/TT but instead exhibit periodic occurrences of GG/CC and CC/GG in opposite phases (55). A major difference between AA/TT and GG/CC is that whereas the AA/TT step in protein-DNA complexes has a small or negative roll angle, GG/CC can adopt both positive roll/negative slide and negative roll/positive slide configurations (42). These different geometries are important since when AA/TT is placed where the minor groove points in towards the octamer it will, in conjunction with GC in the opposite rotational orientation, confer a planar bend on the DNA. In contrast the preferred configuration for the roll/slide combination is a superhelix (74). This property was first noted by Tolstorukov *et al.*, (74) for the TA base-step which is preferentially located where the minor groove faces the octamer in high-affinity octamer binding sequences selected *in vitro* (52). The finding that geometrically distinct types of sequence can direct binding to the octamer raises the issue as to whether the DNA can adopt different trajectories under different assembly conditions. On present evidence this cannot be excluded. Nevertheless it should be noted that the short sequences with a preferred high slide geometry are more typically observed when sequences are selected *in vitro* using octamers from higher eukaryotes.

Another intrinsic variable is the structural periodicity of nucleosomal DNA which has been directly determined from crystal structures (17, 75) and also from the pattern of pyrimidine dimers formed from pyrimidine-pyrimidine base-steps by UV-irradiation of chromatin or isolated NCPs (76). The DNA periodicities in crystal structures of the core nucleosome containing 145, 146 and 147 bp of DNA are respectively 10.15, 10.23 and 10.3 bp (17, 75, 77). This variation is accomplished by keeping the number of double-helical turns constant while incorporating additional DNA base-pairs into 'stretched' double-helical turns containing 11 instead of 10 base-pairs. These values are close to the average value of 10.3 ± 0.1 bp for the periodicity of pyrimidine dimers.

In contrast to crystal structures the periodicity values obtained from compilations of sequences associated with both core nucleosomes and chromatosomes are average values for the whole population and do not necessarily reflect that in any particular sequence. These sequence periodicities are generally, but not always, characterised by regular modulations in the frequency of occurrence of A/T and G/C containing

di- or tri-nucleotides in opposite phases. The periodicity is usually interpreted as the surface helical repeat of the wrapped DNA (29, 30, 78) and, in principle, is equivalent to the structural repeat measured from crystal structures. For nucleosomal DNA sequences from chicken erythrocyte core particles the average sequence periodicity is ~10.2-10.25 bp (30), again in good agreement with the crystal structures and also with the DNase I digestion profile of core particles of similar provenance (79). However this average value conceals variation of periodicities for different short sequences. For example while the AAA/TTT trinucleotide associated with an inward-facing minor groove has an average periodicity of 10.31 bp that of the GGC/GCC trinucleotide has an average periodicity of 10.15 bp. The correlation of periodicity with base-composition is also observed in other compilations with A/T-containing di- and tri-nucleotides invariably exhibiting a higher periodicity than the corresponding G/C-containing dinucleotides (Table). This effect is apparently independent of the rotational orientation of the short sequences since in core particles reconstituted on ovine DNA the dinucleotides GG/CC and CC/GG have the same periodicity but opposite rotational orientations (54). In general only a small number of the 26-28 optimal rotational positions in a given natural octamer-associated sequence contain DNA-bending signals. Thus a possible explanation for a difference in AA/TT and GC periodicities would be that because A/T-rich sequences are more readily untwisted than G/C-rich sequences there would be a slightly greater frequency of 'stretched' double-helical turns between tight histone-DNA contacts. This would result in a higher periodicity. In this situation the sequence periodicities of individual sequences could differ from each other. The variation in observed periodicities also suggests that octamer-bound DNA sequences from chromatin preparations containing linker histone may have a slightly lower periodicity and consequently the DNA exit and entry trajectories could differ from those of core particles. While two analyses agree that the periodic occurrence of the AA/TT and TT/AA base-steps exhibit the same phase (80, 81), other studies, based initially on a library of nucleosomal DNA sequences from different sources and subsequently on the analysis of genomic nucleosome positions in *Caenorhabditis elegans*, find that the periodic occurrences of AA/TT and TT/AA have opposite phases with a periodicity of 10.4 bp (82, 83). This finding was coupled to the identification of a consensus sequence repeat, YYYYYRRRR, in nucleosomal DNA where the YR step is positioned, on average, where the minor groove points away from the histone octamer. Thus, in this analysis, the phases of AA/TT and TT/AA differ by $\pm 90^\circ$ from the common phase determined in other analyses of nucleosomal DNA sequences.

Table
Reported periodicities of nucleosomal DNA

Reference	Dominant periodicity values	Material and/or region of nucleosomal DNA considered
Sequence periodicities:		
Satchwell <i>et al.</i> , 1986 (30)	10.15 (GGC) - 10.31 (AAA) bp	chicken NCP, isolation involves exposure to 0.6 M NaCl
Satchwell & Travers, 1989 (119)	9.80 (GC) - 10.15 (AA/TT) bp	chicken dinucleosomes with linker histone - low salt isolation
Muyldermans & Travers, 1994 (54)	9.90 (GGC) - 10.20 (AAA) bp	chicken chromatosomes - low salt isolation
Fraser <i>et al.</i> , 2009 (55)	9.90 (GG/CC, CC/GG) bp	salt dilution <i>in vitro</i> reconstitution on ovine DNA
Thastrom <i>et al.</i> , 2004 (52)	10.14 (TA) bp	SELEX using salt dilution <i>in vitro</i> reconstitution (periodicity value for central 7 turns)
Structural periodicities:		
Klug & Lutter, 1981 (79)	~10.3 bp	DNase I cleavage periodicity of chicken NCP
Gale <i>et al.</i> , 1986 (76)	10.3 \pm 0.1 bp	soluble chromatin - with and without linker histone (mean minor groove out periodicity)
Davey <i>et al.</i> , 2002 (75)	10.15-10.30 bp	crystal structures for NCPs containing 145-147 bp
Richmond & Davey, 2003 (18)	10.28 bp	central 7 turns from high resolution structure of NCP

Dominant periodicity values refer to the di- or tri-nucleotides with the highest amplitudes of periodic modulation. NCP, nucleosome core particle(s). If no DNA region is specified the calculated periodicity includes all the wrapped DNA.

This issue is currently a matter of lively debate with some authors suggesting that DNA sequences play a dominant role in positioning (10) while others reject this view (83). In general parallel sequencing studies identify an, albeit weak, rotational positioning signal of periodic modulations of AA/TT frequency in the population of sequenced nucleosome core DNA molecules (9, 10, 67). In addition a DNA signature characteristic of positioned nucleosomes, particularly those in promoter regions, was characterised in a small set of octamer-associated sequences identified in budding yeast by partial MNase digestion (51). This ‘signature’ is based on an analysis of sequence periodicity signals in DNA corresponding to the patterns observed in chicken erythrocyte nucleosome core particle DNA. It has the form of a low periodicity signal at the midpoint of the nucleosomal DNA, flanked on one or both sides by a higher average periodicity (Figure 1). Although initially observed in budding yeast this signature is also found in particular promoter regions of the HIV 5’ LTR (Figure 1). Equally importantly, while there is a preferred sequence organization for DNA wrapped around the histone octamer certain sequences by virtue of their low anisotropic deformability or high average stiffness may have a lower affinity for the histone octamer and may exclude octamer binding under *in vivo* conditions. Thus *in vitro* poly(dA)(dT) is a poor ligand for the octamer (85) and oligo(dA)(dT) tracts create barriers for octamer binding (86). In budding yeast such tracts are found preferentially in nucleosome-depleted regions (87, 88). Similarly, some yeast UAS regions contain potentially stiff sequences with a high average stacking energy (51). Taking both positive and negative factors into account the organization of nucleosomes in budding yeast and in *Drosophila* can be accurately predicted in terms of general location by using an algorithm that considers only the physicochemical properties of DNA (89).

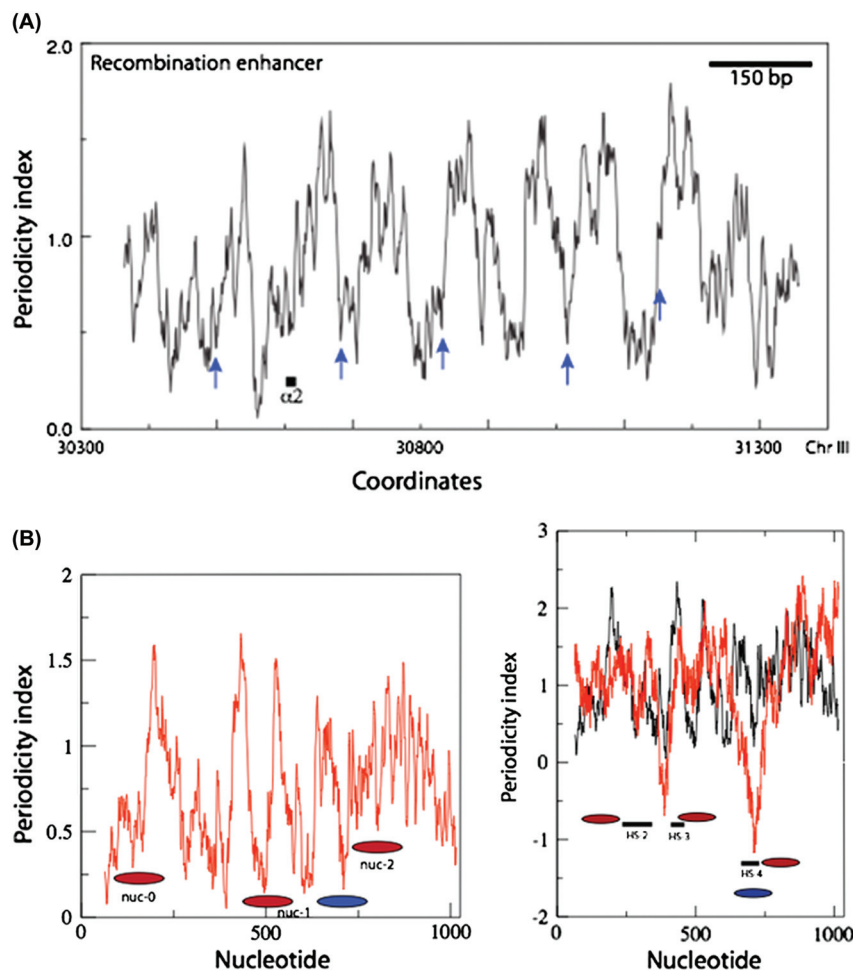


Figure 1: DNA periodicity profiles. The periodicity index is a measure of coherence of periodicity patterns of *e.g.*, AA/TT and GC over a given length of DNA, usually empirically 50 bp. A helical periodicity of ~10 bp is assumed. If direction of bending changes, the phase of the periodic sequence patterns will change and will interfere with phase of adjacent sequences. A high value of the index indicates the sequence periodicities are strong and coherent over the window chosen. (A) *Saccharomyces cerevisiae* recombination enhancer (reproduced with permission from ref. 51). Blue arrows indicate centres of positions mapped by partial MNase digestion. Coordinates refer to the nucleotide positions in the appropriate chromosome. (B) 5’ LTR of HIV. *In vivo* (red ellipses) and *in vitro* (blue ellipse) nucleosome positions for HIV are taken from ref. 120. Nucleotide indicates the numbers of nucleotides from the 5’ end of the LTR. Lh panel: periodicity profile; rh panel: stacking energy superimposed on periodicity profile. Note that regions of high stacking energy (high negative ΔG) fall within or flank nucleosome hypersensitive sites (HS).

Together these observations suggest that histone-DNA interactions are a determinant, and possibly a major one, of positioning *in vivo* but not necessarily the only one.

By comparing nucleosome positions in yeast chromatin mapped by parallel sequencing with those preferentially occupied on yeast genomic DNA using *in vitro* reconstitution Kaplan *et al.*, (53) concluded that histone-DNA interactions were a dominant determinant of positioning *in vivo*. In contrast using similar technique Zhang *et al.*, (84) came to the opposite conclusion. These conclusions relate primarily to translational positioning. The conclusions drawn from the comparisons of *in vivo* and *in vitro* positions are subject to two main caveats: the influence of MNase on the sampling of the population, as discussed above, and the assumption that the relative affinities of different DNA sequences for the histone octamer are the same under the chosen *in vitro* reconstitution conditions as *in vivo*.

The nature of the problem can be simply defined. The histone octamer, or tetramer, should be regarded as a DNA-binding entity that lacks the capacity for base-specific recognition (17). Instead the selectivity of octamer binding *in vitro*, and most likely also *in vivo*, is determined in part, not by a particular defined sequence, but by the overall physicochemical properties of the DNA sequence bound (42, 80). These physicochemical properties comprise not only the anisotropic or directional deformability (27, 28), and the flexibility, but also the torsional flexibility. In practice this implies, at least *in vitro*, that the affinity of the octamer for different DNA sequences, should spread over a wide range, as indeed is observed (for example, refs. 52, 90). Given a sufficiently large number of different DNA sequences this range of affinities would become a continuum. This phenomenon is also apparent in the DNA-binding properties of certain abundant bacterial proteins that organise the DNA in the bacterial nucleoid and which possess little, if any, capacity for direct base-specific recognition. For example, H-NS, which can repress transcription by stabilising DNA plectonemes (91, 92) and was long considered to bind non-specifically to DNA, has recently been shown, like the histone octamer (3), to bind with high affinity *in vitro* to sites that nucleate cooperative binding. Another such DNA-bending protein, FIS, with likely only limited base-specific recognition, again binds to different DNA sequences over a wide range of affinity (36).

An additional, and largely neglected, consequence of indirect recognition in DNA-protein interactions is that since the physicochemical properties of DNA are directly influenced by local changes in its environment, so also will be the interaction of DNA with, for example, the histone octamer. Thus both temperature and DNA supercoiling influence the torsional properties of DNA, and hence the elastic constants for unwinding and bending, while water activity and temperature can influence anisotropic deformability by affecting the structure of the relatively rigid oligo(dA)(dT) tracts (93-95). More pertinently contact with histone octamer will likely lower the elastic constants for DNA bending and also possibly for torsion (47). In the context of nucleosome positioning this means that the relative affinities of different DNA sequences for the octamer are not fixed but instead will depend on the precise conditions of nucleosome formation. Indeed, the sequence binding preferences of the octamer *in vitro* have been shown to be dependent on both octamer concentration and temperature (48).

A more realistic way of considering the nucleosome positioning problem might be to view the octamer/DNA interaction in terms of a tunable energy landscape that is sensitive to variables such as temperature and applied torque. In principle these variables could act on the DNA, or the octamer, or both. Discrepancies between positions mapped *in vitro* and *in vivo* on the same DNA sequences (96, 97) could thus simply arise because the energy landscapes for nucleosome formation are different in the two cases. Indeed a much closer correspondence between *in vivo* and *in vitro* positions has been achieved by reconstituting budding yeast chromatin in the presence of ATP and a crude extract from yeast cells (97). This contrasts with *in vitro*

reconstitution methods including the classic salt dilution protocol (86, 96) or the ACF remodelling complex (84). This is particularly pertinent in the context of positioning *in vivo* where positioning is maintained by chromatin remodelling complexes and RNA polymerase II activity. Removal of remodelling activities *in vivo* in yeast results in changes in nucleosome positions (98, 99) while the packaging of promoter DNA in nucleosomes can be dependent on transcriptional activity (70, 100). One remodelling complex, RSC, cradles a nucleosome core particle in a pocket (101-103), creating a topologically constrained DNA microdomain (104). The DNA translocase activity of the remodeling complex applies torque to the core particle which can be manifest as local changes in DNA superhelicity (105, 106). Such torque could affect the energy landscape of both the wrapped DNA and that in its immediate vicinity, potentially changing the preferred position of the octamer by altering the local thermodynamic equilibrium, as well as kinetically facilitating octamer sliding. In principle different remodellers could apply different amounts of torque and thereby stabilise different positions and in some cases, for example the ACF and certain other ISWI-containing complexes, could by constraining the inter-nucleosome distance relocate octamers to positions that are not necessarily energetically favoured by the physicochemical properties of DNA (106-109). In particular, in a repeated array of a strong positioning sequence positioning may be dominated by cation-dependent spacing as well as by the effects of DNA topology (110). In these examples the regular spacing generated by the remodelling complexes might stabilise a more regular nucleosome packing in the form of a compact 30 nm fibre (Figure 2).

The Organisation of Nucleosomes *in vivo*

The elucidation of genome-wide positioning maps - for the budding yeast *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans* and *Homo sapiens* - has revealed some strong similarities in nucleosome organization (9, 10, 53, 63-67, 111, 112). Close to the transcription site there is a nucleosome-depleted or nucleosome-free region flanked downstream, and often upstream, by a well-positioned nucleosome. In many cases these nucleosomes constitute the start of regular arrays in which positioning becomes less distinct or 'fuzzier' as the distance from the well-positioned nucleosome increases (9). Another complication is the arrays may not be unique in that alternative overlapping arrays on the same DNA sequence have been observed, particularly in *Caenorhabditis elegans* and at the CUP1 and HIS3 loci in budding yeast (111, 113, 114).

These patterns address the question raised by Kornberg (115, 116) as to whether nucleosome positioning *in vivo* is 'statistical' or specified entirely by the physicochemical properties of the DNA sequence. Statistical positioning postulates the presence of a 'boundary' nucleosome which specifies one end of an array which is not determined by the physicochemical properties of DNA sequence. The boundary nucleosome could be positioned either by a protein or by a strong intrinsic positioning sequence or by both. This model is strongly supported by the observed organisation patterns *in vivo* (9, 64). In budding yeast nucleosomes can be ordered by a protein - such as the budding yeast $\alpha 2$ repressor (7). Similarly *in vitro* reconstitution argues that energetically unfavourable DNA barriers can order nucleosomes (86) - again in agreement with the statistical positioning model. Additionally the 'DNA signature', based on sequence periodicity patterns, occurs preferentially in promoter-associated nucleosomes in a small set of nucleosomal DNA sequences identified by partial MNase cleavage (51). This signature is often less prominent downstream of the promoter but in some cases it is found within nucleosome arrays - for example at the recombination enhancer locus and at the RVS167/SAC7 locus (51). Because the length of the

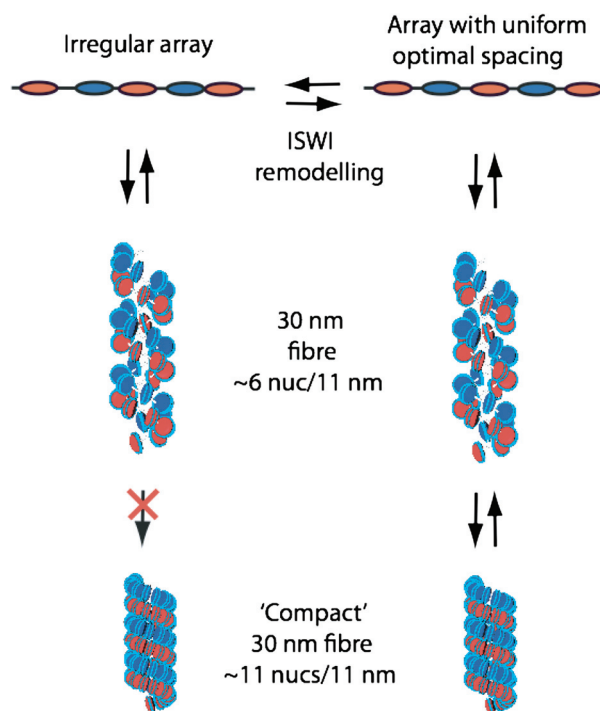


Figure 2: Coupling of nucleosome positioning and chromatin compaction. An array with irregular nucleosome spacing can, in principle, fold to form a 30 nm fibre with a packing density of ~6 nucs/11 nm. When the nucleosomes are regularly spaced at 'optimal' distances (121) tight stacking of adjacent nucleosomes can occur. With irregular spacing tight stacking is precluded because the nucleosomes would be inclined at differing orientations to the superhelical axis. The 30 nm fibres are represented as 2-start structures with the two helical stacks of nucleosomes coloured in red and blue.

DNA signature associated with a nucleosome is variable and often quite short such a motif can, and in one well characterised example is, compatible with the positioning of overlapping arrays (117). Nucleosome organization in budding yeast can thus broadly be described in terms of a statistical model but containing additional elements associated with a defined nucleosomal sequence organization. The extent to which these 'natural' positioning elements tune, or even merely fine tune, nucleosome organization remains to be established.

This research was reported by the author in part at Albany 2009: The 16th Conversation (118).

References and Footnotes

1. R. T. Simpson and D. W. Stafford. *Proc Natl Acad Sci USA* 80, 51-55 (1983).
2. B. Piña, U. Bruggemeier, and M. Beato. *Cell* 60, 719-731 (1990).
3. P. T. Lowary and J. Widom. *J Mol Biol* 276, 19-42 (1998).
4. A. Flaus, K. Luger, S. Tan, and T. J. Richmond. *Proc Natl Acad Sci USA* 93, 1370-1375 (1996).
5. A. Almer and W. Hörz. *Embo J* 5, 2689-2696 (1986).
6. H. Richard-Foy and G. L. Hager. *Embo J* 6, 2321-2328 (1987).
7. M. Shimizu, S. Y. Roth, C. Szent-Gyorgyi, and R. T. Simpson. *Embo J* 10, 3033-3041 (1991).
8. L. Verdone, G. Camilloni, E. Di Mauro, and M. Caserta. *Mol Cell Biol* 16, 1978-1988 (1996).
9. T. N. Mavrich, I. P. Ioshikhes, B. J. Venters, C. Jiang, L. P. Tomsho, J. Qi, S. C. Schuster, I. Albert, and B. F. Pugh. *Genome Res* 18, 1073-1083 (2008).
10. Y. Field, N. Kaplan, Y. Fondufe-Mittendorf, I. K. Moore, E. Sharon, Y. Lubling, J. Widom, and E. Segal. *PLoS Comp Biol* 4, 1-25 (2008).
11. M. Han and M. Grunstein. *Cell* 55, 1137-1145 (1988).
12. M. S. Schlissel and D. D. Brown. *Cell* 37, 903-913 (1984).
13. J. L. Compton, M. Bellard, and P. Chambon. *Proc Natl Acad Sci USA* 73, 4382-4386 (1976).
14. J. O. Thomas and R. J. Thompson. *Cell* 10, 633-640 (1977).
15. F. Bernardi, M. Zatchej, and F. Thoma. *Embo J* 11, 1177-1185 (1992).
16. J. McManus, P. Perry, A. T. Sumner, D. M. Wright, E. J. Thomson, R. C. Allshire, N. D. Hastie, and W. A. Bickmore. *J Cell Sci* 107, 469-486 (1994).
17. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. *Nature* 389, 251-260 (1997).
18. T. J. Richmond and C. A. Davey. *Nature* 423, 145-150 (2003).
19. T. J. Richmond, J. T. Finch, B. Rushton, D. Rhodes, and A. Klug. *Nature* 311, 532-537 (1984).
20. P. J. Hagerman. *Biopolymers* 20, 1503-1535 (1981).
21. S. D. Levene, H.-M. Wu, and D. M. Crothers. *Biochemistry* 25, 3988-3995 (1986).
22. E. S. Sobel and J. A. Harpst. *Biopolymers* 31, 1559-1564 (1991).
23. D. Pörschke. *Biophys Chem* 40, 169-179 (1991).
24. D. M. Crothers, J. Drak, J. D. Kahn, and S. D. Levene. *Methods Enzymol* 212, 3-29 (1992).
25. C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante. *Proc Natl Acad Sci USA* 94, 6185-6190 (1997).
26. M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block. *Biophys J* 72, 1335-1346 (1997).
27. V. B. Zhurkin, Y. P. Lysov, and V. Ivanov. *Nucleic Acids Res* 6, 1081-1096 (1979).
28. E. N. Trifonov. *Nucleic Acids Res* 8, 4041-4053 (1980).
29. H. R. Drew and A. A. Travers. *J Mol Biol* 186, 773-790 (1985).
30. S. C. Satchwell, H. R. Drew, and A. A. Travers. *J Mol Biol* 191, 659-675 (1986).
31. T. E. Shrader and D. M. Crothers. *Proc Natl Acad Sci USA* 86, 7418-7422 (1989).
32. E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I. K. Moore, J. Z. Wang, and J. Widom. *Nature* 442, 772-778 (2006).
33. Y. Wada-Kiyama, K. Kuwabara, Y. Sakuma, Y. Onishi, E. M. Trifonov, and R. Kiyama. *FEBS Lett* 444, 117-124 (1999).
34. A. P. Wolffe and H. R. Drew. *Proc Natl Acad Sci USA* 86, 9817-9821 (1989).
35. A. A. Travers and A. Klug. *Nature* 324, 280-281 (1987).
36. L. R. Lazarus and A. A. Travers. *Embo J* 12, 2483-2494 (1993).
37. D. G. Blake, M. R. Boocock, D. J. Sherratt, and W. M. Stark. *Curr Biol* 5, 1036-1046 (1995).
38. Y. Zhang, Z. Xi, R. S. Hegde, Z. Shakked, and D. M. Crothers. *Proc Natl Acad Sci USA* 101, 8337-8341 (2004).
39. A. A. Travers. *Trends Biochem Sci* 12, 108-112 (1987).
40. R. E. Dickerson and H. R. Drew. *J Mol Biol* 149, 761-786 (1981).
41. H. C. M. Nelson, J. T. Finch, B. F. Luisi, and A. Klug. *Nature* 330, 221-226 (1987).
42. W. K. Olson, A. A. Gorin, X. J. Lu, L. M. Hock, and V. B. Zhurkin. *Proc Natl Acad Sci USA* 95, 11163-11168 (1998).

43. R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mannand, and B. Honig. *Nature* 461, 1248-1253 (2009).
44. M. McCall, T. Brown, and O. Kennard. *J Mol Biol* 183, 385-396 (1985).
45. D. S. Goodsell, M. L. Kopka, D. Cascio, and R. E. Dickerson. *Proc Natl Acad Sci USA* 90, 2930-2934 (1993).
46. M. Buttinelli, A. Minnock, G. Panetta, M. Waring, and A. Travers. *Proc Natl Acad Sci USA* 95, 8544-8549 (1998).
47. J. Virstedt, T. Berge, R. M. Henderson, M. J. Waring, and A. A. Travers. *J Struct Biol* 148, 66-85 (2004).
48. C. Wu and A. Travers. *Biochemistry* 44, 14329-14334 (2005).
49. C. H. Laundon and J. D. Griffith. *Cell* 52, 545-549 (1988).
50. A. Travers, M. Caserta, M. Churcher, E. Hiriart, and E. Di Mauro. *Mol Biosystems* 5, 1582-1592 (2009).
51. M. Caserta, E. Agricola, M. Churcher, E. Hiriart, L. Verdone, E. Di Mauro, and A. Travers. *Nucleic Acids Res* 37, 5309-5321 (2009).
52. A. Thåström, P. T. Lowary, H. R. Widlund, H. Cao, M. Kubista, and J. Widom. *J Mol Biol* 288, 213-229 (1999).
53. N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, Y. Field, E. M. LeProust, T. R. Hughes, J. D. Lieb, J. Widom, and E. Segal. *Nature* 458, 362-366 (2009).
54. S. Muyldermans and A. A. Travers. *J Mol Biol* 191, 855-870 (1994).
55. R. M. Fraser, D. Keszenman-Pereyra, M. W. Simmen, and J. Allan. *J Mol Biol* 390, 292-305 (2009).
56. M. Gencheva, S. Boa, R. Fraser, M. W. Simmen, C. B. A. Whitelaw, and J. Allan. *J Mol Biol* 361, 216-230 (2006).
57. F. Schaeffer, S. Rimsky, and A. Spassky. *J Mol Biol* 260, 523-539 (1996).
58. A. B. Fleming and S. Pennings. *Embo J* 20, 5219-5231 (2001).
59. C. Jiang and B. F. Pugh. *Nat Rev Genet* 10, 161-172 (2009).
60. O. J. Rando and H. Y. Chang. *Annu Rev Biochem* 78, 245-271 (2009).
61. E. Segal and J. Widom. *Nat Rev Genet* 10, 443-456 (2009).
62. A. Stein, T. E. Takasuka, and C. K. Collings. *Nucleic Acids Res*, doi:10.1093/nar/gkp1043 (2009).
63. B. Guillemette, A. R. Bataille, N. Gévry, M. Adam, M. Blanchette, F. Robert, and L. Gaudreau. *PLoS Biol* 3, e384 (2005).
64. G. Yuan, Y. Liu, M. F. Dion, M. D. Slack, L. F. Wu, S. J. Altschuler, and O. J. Rando. *Science* 309, 626-630 (2005).
65. W. Lee, D. Tillo, N. Bray, R. H. Morse, R. W. Davis, T. R. Hughes, and C. Nislow. *Nature Gen* 39, 1235-1244 (2007).
66. I. Albert, T. N. Mavrich, L. P. Tomsho, S. J. Zanton, S. C. Schuster, and B. F. Pugh. *Nature* 446, 572-578 (2007).
67. T. N. Mavrich, C. Jiang, I. P. Ioshikhes, X. Li, B. J. Venters, S. J. Zanton, L. P. Tomsho, J. Qi, R. L. Glaser, S. C. Schuster, D. S. Gilmour, I. Albert, and B. F. Pugh. *Nature* 453, 358-362 (2008).
68. S. Tanaka, M. Livingstone-Zatchej, and F. Thoma. *J Mol Biol* 257, 919-934 (1996).
69. W. Hörz and W. Altenburger. *Nucleic Acids Res* 9, 2643-2568 (1981).
70. C. Dingwall, G. P. Lomonosof, and R. A. Laskey. *Nucleic Acids Res* 9, 2659-2673 (1981).
71. J. T. Flick, J. C. Eissenberg, and S. C. R. Elgin. *J Mol Biol* 190, 619-633 (1986).
72. J. D. McGhee and G. Felsenfeld. *Cell* 32, 1205-1215 (1983).
73. A. Weiner, A. Hughes, M. Yassour, O. J. Rando, and N. Friedman. *Genome Res* 20, 90-100 (2010).
74. M. Y. Tolstorukov, A. V. Colasanti, D. M. McCandlish, W. K. Olson, and V. B. Zhurkin. *J Mol Biol* 371, 725-738 (2007).
75. C. A. Davey, D. F. Sargent, K. Luger, A. W. Mäder, and T. J. Richmond. *J Mol Biol* 319, 1097-1113 (2002).
76. J. M. Gale, K. A. Nissen, and M. J. Smerdon. *Proc Natl Acad Sci USA* 84, 6644-6648 (1987).
77. C. A. Davey and T. J. Richmond. *Proc Natl Acad Sci USA* 99, 11169-11174 (2002).
78. J. H. White, N. R. Cozzarelli, and W. R. Bauer. *Science* 241, 323-327 (1988).
79. A. Klug and L. C. Lutter. *Nucleic Acids Res* 9, 4267-4283 (1981).
80. A. A. Travers and A. Klug. *Phil Trans Roy Soc Lond B* 317, 537-561 (1987).
81. J.-P. Z. Wang and J. Widom. *Nucleic Acids Res* 22, 6743-6755 (2005).
82. F. Salih, B. Salih, and E. N. Trifonov. *J Biomol Struct Dyn* 26, 273-281 (2008).
83. I. Gadbank, D. Barash, and E. N. Trifonov. *J Biomol Struct Dyn* 26, 403-412 (2009).
84. Y. Zhang, Z. Moqtaderi, B. P. Rattner, G. Euskirchen, M. Snyder, and K. Struhl. *Nat Struct Mol Biol* 16, 847-852 (2009).
85. D. Rhodes. *Nucleic Acids Res* 6, 1805-16 (1979).
86. P. Milani, G. Chevereau, C. Vaillant, B. Audit, Z. Haftek-Terreau, M. Marilley, P. Bouvet, F. Argoul, and A. Arneodo. *Proc Natl Acad Sci USA* 108, 22257-22262 (2009).
87. V. Iyer and K. Struhl. *Embo J* 14, 2570-2579 (1995).
88. E. Segal and J. Widom. *Curr Opin Struct Biol* 19, 65-71 (2009).

89. C. Miele, Y. Vaillant, C. d'Aubenton-Carafa, C. Thermes, and T. Grange. *Nucleic Acids Res* 36, 3746-3756 (2008).
90. H. R. Widlund, H. Cao, S. Simonsson, E. Magnusson, T. Simonsson, P. E. Nielsen, J. D. Kahn, D. M. Crothers, and M. Kubista. *J Mol Biol* 267, 807-817 (1997).
91. E. Bouffartigues, M. Buckle, C. Badaut, A. Travers, and S. Rimsky, *Nat Struct Mol Biol* 14, 441-448 (2007).
92. B. Lang, N. Blot, E. Bouffartigues, M. Buckle, M. Geertz, C. O. Gualerzi, R. Mavathur, G. Muskhelishvili, C. L. Pon, S. Rimsky, S. Stella, M. M. Babu, and A. Travers. *Nucleic Acids Res* 35, 6330-6337 (2007).
93. S. Diekmann. *Nucleic Acids Res* 15, 247-265 (1987).
94. H. S. Koo and D. M. Crothers. *Biochemistry* 26, 3745-3748 (1987).
95. H. R. Drew and A. A. Travers. *Cell* 37, 491-502 (1984).
96. C. J. Wippo, B. S. Krstulovic, F. Ertel, S. Musladin, D. Blaschke, S. Stürzl, G. C. Yuan, W. Hörz, P. Korber, and S. Barbaric. *Mol Cell Biol* 29, 2960-2981 (2009).
97. P. Korber and W. Hörz. *J Biol Chem* 279, 35113-35120 (2004).
98. B. Xella, C. Goding, E. Agricola, E. Di Mauro, and M. Caserta. *Mol Microbiol* 59, 1531-1541 (2006).
99. I. Whitehouse and T. Tsukiyama. *Nat Struct Mol Biol* 13, 633-640 (2006).
100. B. J. Venters and B. F. Pugh. *Genome Res* 19, 360-371 (2009).
101. F. J. Asturias, W. H. Chung, R. D. Kornberg, and Lorch Y. *Proc Natl Acad Sci USA* 99, 13477-13480 (2002).
102. Y. Chaban, C. Ezeokonkwo, W. H. Chung, F. Zhang, R. D. Kornberg, B. Maier-Davis, Y. Lorch, and F. J. Asturias. *Nat Struct Mol Biol* 15, 1272-1277 (2008).
103. A. E. Leschziner, A. Saha, J. Wittmeyer, Y. Zhang, C. Bustamante, B. R. Cairns, and E. Nogales. *Proc Natl Acad Sci USA* 104, 4913-4918 (2007).
104. G. Muskhelishvili and A. Travers. *Nucleic Acids Mol Biol* 11, 179-190 (1997).
105. G. Lia, E. Praly, H. Ferreira, C. Stockdale, Y. C. Tse-Dinh, D. Dunlap, V. Croquette, D. Bensimon, and T. Owen-Hughes. *Mol Cell* 21, 417-425 (2006).
106. Y. Zhang, C. L. Smith, A. Saha, S. W. Grill, S. Mihardja, S. B. Smith, B. R. Cairns, C. L. Peterson, and C. Bustamante. *Mol Cell* 24, 559-68 (2006).
107. D. V. Fyodorov, M. D. Blower, G. H. Karpen, and J. T. Kadonaga. *Genes Dev* 18, 170-183 (2004).
108. J. G. Yang, T. S. Madrid, E. Sevastopoulos, and G. J. Narlikar. *Nat Struct Mol Biol* 13, 1078-1083 (2006).
109. L. R. Racki, J. G. Yang, N. Naber, P. D. Partensky, A. Acevodo, T. J. Purcell, R. Cooke, Y. Cheng, and G. J. Narlikar. *Nature* 460, 1016-1021 (2009).
110. T. Blank and P. B. Becker. *J Mol Biol* 260, 1-6 (1995).
111. A. Valouev, J. Ichikawa, T. Tonthat, J. Stuart, S. Ranade, H. Peckham, K. Zeng, J. A. Malek, G. Costa, K. McKernan, A. Sidow, A. Fire, and S. M. Johnson. *Genome Res* 18, 1051-1063 (2008).
112. D. E. Schones, K. Cui, S. Cuddapah, T. Y. Roh, A. Barski, Z. Wang, G. Wei, and K. Zhao. *Cell* 132, 887-898 (2008).
113. C. H. Shen and D. J. Clark. *J Biol Chem* 276, 35209-35216 (2001).
114. Y. Kim, N. McLaughlin, K. Lindstrom, T. Tsukiyama, and D. J. Clark. *Mol Cell Biol* 26, 8607-8622 (2006).
115. R. Kornberg. *Nature* 292, 279-280 (1981).
116. R. D. Kornberg and L. Stryer. *Nucleic Acids Res* 16, 6677-6690 (1988).
117. G. Costanzo, E. di Mauro, R. Negri, G. Pereira, and C. Hollenberg. *J Biol Chem* 270, 11091-11097 (1995).
118. Abstracts of Albany 2009: 16th Conversation. June 16-20, Albany, New York, USA; M. Caserta, E. Agricola, M. Churcher, E. Hiriart, L. Verdone, E. Di Mauro and A. Travers. A Translational Signature for Nucleosome Positioning *In vivo*, Abstract #179. *J Biomol Struct Dyn* 26, 787-927 (2009).
119. S. C. Satchwell and A. A. Travers. *Embo J* 8, 229-238 (1989).
120. M. J. Pazin, P. L. Sheridan, K. Cannon, Z. Cao, J. G. Keck, J. T. Kadonaga, and K. A. Jones. *Genes Dev* 10, 37-49 (1996).
121. J. Widom. *Proc Natl Acad Sci USA* 89, 1095-1099 (1992).

Date Received: December 10, 2009

Communicated by the Editor Ramaswamy H. Sarma