Cell PRESS
Structure
Article

Structural Insight into the Sequence Dependence of Nucleosome Positioning

Bin Wu,[1,2,3] Kareem Mohideen,[1,2] Dileep Vasudevan,[1] and Curt A. Davey[1,*]
[1]Division of Structural and Computational Biology, School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551
[2]These authors contributed equally to this work
[3]Present Address: Department of Biological Chemistry and Molecular Pharmacology, Immune Disease Institute, Harvard Medical School, Center for Life Science Building, 3 Blackfan Circle, Boston, MA 02115, USA
*Correspondence: davey@ntu.edu.sg
DOI 10.1016/j.str.2010.01.015

SUMMARY

Nucleosome positioning displays sequence dependency and contributes to genomic regulation in a site-specific manner. We solved the structures of nucleosome core particle composed of strong positioning TTTAA elements flanking the nucleosome center. The positioning strength of the super flexible TA dinucleotide is consistent with its observed central location within minor groove inward regions, where it can contribute maximally to energetically challenging minor groove bending, kinking and compression. The marked preference for TTTAA and positioning power of the site 1.5 double helix turns from the nucleosome center relates to a unique histone protein motif at this location, which enforces a sustained, extremely narrow minor groove via a hydrophobic "sugar clamp." Our analysis sheds light on the basis of nucleosome positioning and indicates that the histone octamer has evolved not to fully minimize sequence discrimination in DNA binding.

INTRODUCTION

Distinctions in hydration, base pairing, and base-base stacking interactions result in sequence-dependent DNA structure, which is exploited by nuclear proteins for genomic regulation (El Hassan and Calladine, 1997; Olson et al., 1998; Travers, 2004). For instance, many transcription factors utilize indirect readout in binding their cognate sites, whereby a pre-existing conformation or distorted state predisposed by the DNA sequence is recognized. Commonly, such binding involves kinking at highly flexible TA or CA = TG dinucleotide elements, which serve as identifiable soft spots in DNA (Dickerson and Chiu, 1997; Dickerson, 1998; Olson et al., 1998). Although core histone proteins may have evolved to largely minimize base pair (bp) discrimination in binding the double helix, the tight, systematic wrapping of DNA in nucleosomes gives rise to an indirect readout of the sequence.

Eukaryotic genomes have apparently coevolved with histones in the regulatory optimization of nucleosome structure, dynamics and organization. By influencing spatial context and factor recognition, nucleosome positioning provides a platform for regulating DNA transactions (Jiang and Pugh, 2009; Segal and Widom, 2009a, 2009c; Radman-Livaja and Rando, 2010). For instance, nucleosome depletion at promoters and discrete positions occupied at flanking regions play a key role in transcription. Nucleosome positioning is strongly influenced by DNA sequence in vivo, although the genome-wide magnitude of the dependency is currently a subject of inquiry (Kaplan et al., 2009; Zhang et al., 2009; Travers et al., 2009). Thus, in addition to sequence, chromatin remodeling factors, polymerase activity and other DNA binding factors can modulate nucleosome location. Importantly however, strong sequence-based positioning signals present even at a very minor fraction of key genomic sites could make a disproportionate contribution to regulation.

From the twist of the double helix wrapped in the nucleosome, the major and minor grooves alternate between facing away from and being bent toward the histone octamer (Richmond and Davey, 2003; Ong et al., 2007). Five bp sections typically constitute major and minor groove "blocks" where each faces the histone octamer (Figure 1). As a consequence of the intrinsic anharmonicity of bending the double helix (Dickerson and Chiu, 1997; Dickerson, 1998), roll (base unstacking/compression at the groove edges) makes a substantially greater contribution toward DNA wrapping compared with tilt (base unstacking/compression at the phosphodiester edges; Richmond and Davey, 2003). Moreover, the relative unfavorableness of compression into the minor (negative roll) versus the major groove (positive roll) gives rise to at least two specialized modes of minor groove bending in the nucleosome core. Minor groove kinking has been observed when a highly flexible CA = TG bp step takes up a central position in a minor groove block, wherein pronounced negative roll at a single step can make a majorative contribution to wrapping. Minor groove blocks lacking a centrally located CA = TG step display negative roll over three to four steps accompanied by alternating shifting of bp into the major and minor grooves, which prevents steric clashing at the narrowed minor groove edge. One exception occurs with the tremendous unstacking requirements associated with DNA stretching, in which extreme minor groove kinking at a GG = CC dinucleotide has been found (Ong et al., 2007; Davey et al., 2009). Although context may influence deformability, the GG = CC step type, and purine-purine and purine-pyrimidine dinucleotides in general, are not observed to

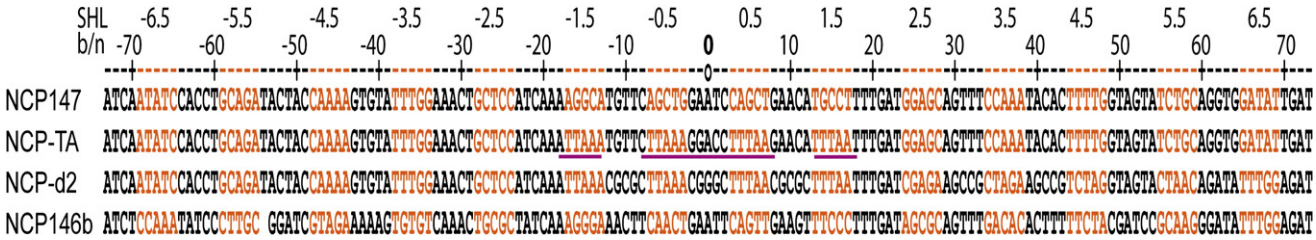

Figure 1. DNA Sequence and Histone-DNA Register of NCP Constructs

Minor and major groove blocks are orange and black, respectively. The base or nucleotide numbering scheme (b/n) is relative to NCP147 and corresponds to the 5′ (−) to 3′ (+) direction of either DNA strand in the duplex (only one strand is shown for each construct; SHL = superhelix location, turns from center). Bp steps in NCP-TA that differ with respect to NCP147 are underlined in magenta, and a gap in the sequence represents a shift in histone-bp register from DNA stretching in NCP146b. NCP-d2 is the only nonpalindromic sequence shown.

display attributes of flexibility (El Hassan and Calladine, 1997; Dickerson and Chiu, 1997; Dickerson, 1998; Olson et al., 1998; Krueger et al., 2006; Balasubramanian et al., 2009; Morozov et al., 2009).

From both in vitro and genome-wide in vivo analysis, there is overall consensus that A|T-rich and G|C-rich sequences tend to position in the minor and major groove blocks, respectively (Satchwell et al., 1986; Widom, 2001; Thastrom et al., 2004; Kaplan et al., 2009; Zhang et al., 2009; Travers et al., 2009). This general trend may relate to the tendency of A|T elements to prefer a narrow, or compressed, minor groove, opposite to that of G|C motifs. However, the detailed energetics are not so straightforward, because poly-A:T tracts, which adopt a very narrow minor groove, are in fact nucleosome destabilizing and excluding elements (Segal and Widom, 2009b). Although there is conflicting evidence on the actual deformability of poly-A:T tracts (Olson et al., 1998; Segal and Widom, 2009b), recent comprehensive analyses of nonnucleosomal protein-DNA complexes indicate that AA = TT as well as AT dinucleotides are overall rigid or nonflexible (Balasubramanian et al., 2009; Morozov et al., 2009), which is consistent with the idea that DNA flexibility is a major factor governing positioning (Travers, 2004; Virstedt et al., 2004).

Considering it has by far the lowest bp stacking energy and displays one of the largest degrees of conformational variability of any bp step type, TA steps are apparently the most flexible (El Hassan and Calladine, 1997; Travers, 2004; Dickerson and Chiu, 1997; Dickerson, 1998; Krueger et al., 2006; Balasubramanian et al., 2009; Morozov et al., 2009). This step type stands out disproportionately as the most pronounced nucleosome positioning element from in vitro studies, in which the strongest histone octamer-binding DNA fragments are selected for (Widom, 2001; Thastrom et al., 2004). Moreover, the TA elements display a strong ~10.1 bp enrichment periodicity (Widom, 2001), corresponding to localization in minor groove blocks (Virstedt et al., 2004). Likewise, the TA-rich, A|T tract of the strongest known genomic nucleosome positioning element, TATAAA CGCC, also localizes to minor groove blocks (Widlund et al., 1997; Widlund et al., 1999).

The energetic challenge associated with minor groove bending (Dickerson and Chiu, 1997; Dickerson, 1998) may give rise to a preference for superflexible TA steps to position in minor groove blocks. However, the influence on positioning of different histone-DNA binding sites is nonuniform and is particularly strong at the location 1.5 double helix turns from the nucleosome dyad (center; Fitzgerald and Anderson, 1999; Fernandez and Anderson, 2007). This is a site of especially sharp bending in many different nucleosome sequences and coincides with the point where extreme minor groove kinking has been observed (Richmond and Davey, 2003; Ong et al., 2007; Fitzgerald and Anderson, 1999; Fernandez and Anderson, 2007; Richmond et al., 1984; Hogan et al., 1987). Moreover, this location has a distinct preference for the sequence TTTAA/TTAAA, very similar to the aforementioned genomic positioning motif (Widlund et al., 1997; Widlund et al., 1999), which appears as a consensus element from in vitro selection of the strongest histone octamer-binding DNA fragments (Thastrom et al., 2004).

We conducted a crystallographic study of nucleosome core particle (NCP) containing TTTAA elements in minor groove blocks flanking the dyad, and found sequence-dependent structure and histone-imposed distinctions at different DNA binding sites. This sheds light on the special function of TA dinucleotides and A|T-rich elements in nucleosome positioning. Moreover, we identify a conserved motif that corresponds to a translational positioning mechanism in the histone system.

RESULTS

Novel NCP Constructs and Nucleosome-Nucleosome Interface

The diffraction quality of NCP crystals is heavily dependent on the DNA length and sequence (Ong et al., 2007; Luger et al., 1997; Davey et al., 2002; Bao et al., 2006). We screened a series of different DNA fragments containing TTTAA elements at the 0.5- and 1.5-turn positions and found two that yielded well-diffracting NCP crystals (Figure 1 and Table 1). NCP-d2 is composed of a 147 bp nonpalindromic DNA, in which approximately one-half of the sequence is identical to that of human α-satellite constructs NCP147 or NCP146b that yield the best diffracting crystals to date (Davey et al., 2002). Using asymmetric DNA fragments runs the risk of obtaining a mixture of the two pseudo-symmetry-related NCP orientations in the crystal, which may generally result in poorly defined electron density. However, particles appeared to crystallize preferentially in one orientation in a previous study on a nonpalindromic DNA-containing NCP (NCP-A_{16}; Bao et al., 2006).

In spite of the favorable diffraction properties of the NCP-d2 crystals, the B-factor average is very high (142 $Å^2$) and the

Table 1. Data Collection and Refinement Statistics

	NCP-d2	NCP-TA
Data Collection[a]		
Space group	$P2_12_12_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	106.4, 109.6, 179.9	106.7,178.5,110.4
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 102.8, 90.0
Resolution (Å)	3.05-60.0 (3.05-3.21)	2.95-60.0 (2.95-3.11)
R_{merge}	5.6% (48.3%)	6.9% (47.1%)
$I / \sigma I$	24.5 (3.0)	12.4 (1.5)
Completeness (%)	95.2 (76.6)	89.7 (55.6)
Redundancy	5.7 (5.6)	3.7 (3.2)
Refinement		
Resolution (Å)	3.05-60.0	2.95-60.0
No. reflections	36,713	74,471
R_{work} / R_{free}	25.7% / 33.0%	22.9% / 30.0%
No. atoms	12,086	24,335
Protein	6,055	12,259
DNA	6,021	12,042
Ion	10	34
B-factors (Å^2)	142	97
Protein	103	73
DNA	181	122
Ion	174	126
Rmsd		
Bond lengths (Å)	0.008	0.008
Bond angles (°)	1.35	1.35

[a] Data sets are based on single crystal diffraction, and values in parentheses are for the highest-resolution shell.

electron density is overall poorly defined and does not allow discrimination between the two possible orientations, which apparently prevail at roughly equal proportion. This is consistent with similar, high crystallographic R-values obtained for refinement of an NCP-d2 model in either orientation (0.257 versus 0.267, R_{work}; 0.330 versus 0.344, R_{free}).

We applied a more conservative approach in the design of NCP-TA, which is composed of a 147 bp palindromic fragment that differs from the NCP147 DNA only in the four minor groove block sequence elements flanking the dyad and two bp in the central major groove block (Figure 1). Surprisingly, the 22 bp alteration in the central sequence is sufficient to yield a dramatically different crystal packing configuration. In contrast to the ubiquitous $P2_12_12_1$ NCP crystal symmetry, NCP-TA crystallizes in space group $P2_1$ with two particles in the asymmetric unit.

NCP composed of frog, chicken, human, or fly histones and the same DNA fragment yield crystals with very similar packing configurations (Luger et al., 1997; Harp et al., 2000; Tsunaka et al., 2005; Clapier et al., 2008). The major interaction involves the faces of two oppositely oriented particles, which overlap by about two-thirds (Figure 2A). The remaining minor face is engaged in a metal-mediated DNA-DNA contact with a third particle (Davey and Richmond, 2002). The major interface comprises the binding of the H4 N-terminal tail in an acidic cleft formed by the H2A-H2B dimer, adjacent to a divalent metal binding site provided by proximal H2A-H2B and H3 elements. This core feature is flanked on either side by DNA interactions with H2B C-terminal α-helix extensions.

The outward faces of the two particles in the asymmetric unit of NCP-TA are engaged in the same interparticle contacts as NCP147, described above. The interaction between the two particles, however, involves an almost complete overlap of the faces (Figure 2B). This configuration is similar to the major crystal packing interface observed for the NCP composed of yeast histones (White et al., 2001). Thus, minor changes in either protein or DNA sequence can favor highly altered modes of internucleosomal interaction. Although the contact areas for the two systems are both roughly 700 Å^2 (ΔASA ≈ 1400 Å^2), the interface in the yeast NCP crystal differs by an approximate 10 Å shift in the particle face-to-face register that gives altered histone-histone contacts, and there is substantial canting of one particle with respect to the other, limiting the interactions to one side.

In NCP-TA, two divalent metal-mediated interactions constitute the center of the new interface (Figures 2B-2D). One cation binding site is formed by acidic residues from a histone fold α-helix of H3 and the H2B C-terminal α-helix extension and the other by glutamate and additional residues from opposing H2A-H2B dimers. The metal-mediated contacts are supported by flanking interactions between the H4 N-terminal tail and H3 on one side and elements from juxtaposed H2A proteins on the other. The latter interactions are in turn flanked at the periphery of the interface by van der Waals and H-bonding contacts between an H2B element and a DNA phosphodiester backbone.

Compared with the other major interface in the NCP-TA crystal, common to NCP147 etc., the new interface involves an ~20 Å translational shift in particle overlap that allows the two acidic H2A-H2B dimer and H3 elements, formally comprising a single metal binding site in NCP147, to each pair up individually with two other acidic elements from H2A and the H2B C-terminal α-helix extension (Figure 2). Thus, there is a significant degree of conservation between the two interaction modes. From the absence of direct interfacial DNA-DNA contacts, it is not obvious how the new, additional interface of NCP-TA is favored through the limited DNA sequence differences relative to NCP147. However, there are alterations in double-helix structure resulting from the sequence changes (see below) that could indirectly affect DNA-DNA packing configuration. In addition, DNA sequence and conformation likely influences interactions with the histone N-terminal tails (Widlund et al., 2000), which in turn can modulate nucleosome-nucleosome binding preferences.

NCP-TA DNA Conformational Parameters

Relative DNA disorder is especially low in crystals of the NCP147, which has allowed acquisition of a 1.9 Å resolution model of exceptional quality (Davey et al., 2002). Because the DNA sequence of NCP-TA is the same as NCP147 outside of the central region, comparison of double-helix structure parameters between the two can provide a measure of NCP-TA model quality. To obtain the most representative DNA parameters, in which the influence of differential crystal packing interactions is minimized, we adopt the same strategy utilized previously for

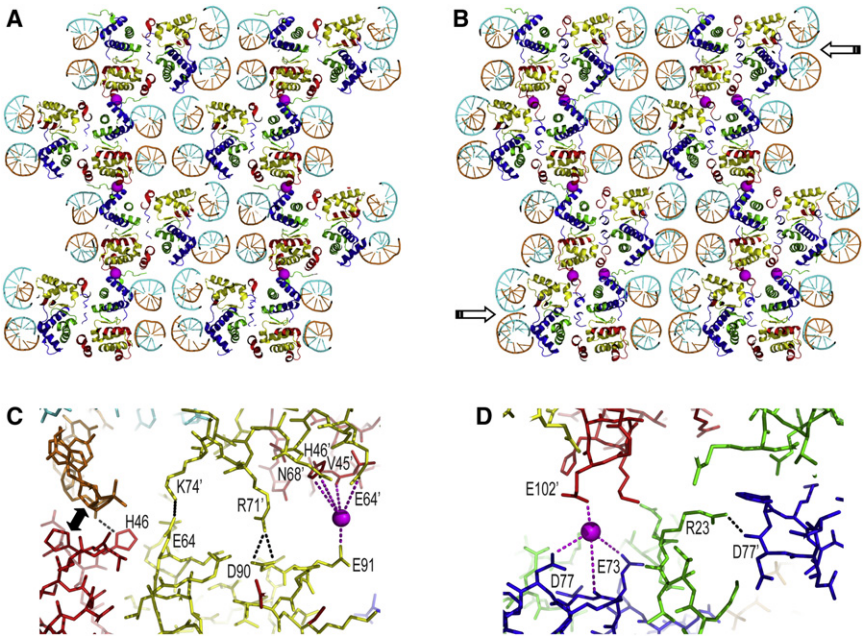

Figure 2. Major Nucleosome Interfaces in Crystals of NCP147 and NCP-TA

DNA strands are colored orange and cyan and proteins blue, H3, green, H4, yellow, H2A, and red, H2B. Mn^{2+} ions mediating histone-histone contacts appear as magenta spheres.

(A and B) Cross-section showing four rows of two particles for NCP147 (A) and NCP-TA (B). Shifting of alternate NCP-TA lattice rows relative to NCP147 (arrows, B) causes breakdown in symmetry by creating an additional interface.

(C and D) The new interface comprises two metal (hydrate)-mediated interactions (magenta dashed lines), hydrogen bonds and salt-bridges (black dashed lines) and histone-DNA van der Waals contacts (encompassing residues 44-48 of H2B; arrow, C).

analysis of NCP147 (Richmond and Davey, 2003). This entails averaging values for sequence- and pseudosymmetry-related particle halves, which from the reduced symmetry of the NCP-TA crystals involves four independent nucleosome core half-sites.

To optimize bp stacking interactions and minimize potential strain introduced into the DNA backbone, changes in roll or shift are respectively coupled to compensatory alterations in twist and slide or tilt (El Hassan and Calladine, 1997; Packer and Hunter, 1998). Principal component analysis revealed that the roll-twist-slide and shift-tilt couplings are very pronounced in the nucleosome core (Richmond and Davey, 2003), which at 50% and 30%, respectively, account for most of the variance within these five main DNA conformational parameters for NCP147 (Table 2). At 46% and 25%, the two primary couplings are only slightly weaker in NCP-TA. This is also reflected in the modest reduction in correlation magnitudes between the respective variables, indicating that the overall accuracy of the NCP-TA DNA model is high.

Graphical inspection of the NCP-TA and NCP147 bp step parameters is also indicative of model precision (Figure 3). Kinking and the associated extreme high twist and slide values at central CA = TG steps in the 3.5, 4.5, and 5.5 minor groove blocks observed previously for NCP147 are recapitulated in NCP-TA. In addition, a smooth minor groove bending mode and associated alternating shift and tilt profile is observed for the 2.5 site, which is common between the two constructs. In contrast, comparison of DNA conformation in the major groove blocks reveals a significant degree of variation between NCP147 and NCP-TA. Importantly, however, the largest differences can be seen to generally display correlated variation via the roll-twist-slide and shift-tilt couplings, as indicated by the principal component analysis. Such structural variation arises from distinctions in crystal packing interactions between NCP-TA and NCP147, and indicates that DNA in the major groove blocks has a substantially greater amount of conformational freedom compared to that within the minor groove blocks.

Table 2. Principal Component Analysis of DNA Conformation

Eigenvector	NCP147:1	NCP147:2	NCP-TA:1	NCP-TA:2
Eigenvalue	2.5	1.5	2.3	1.3
Variability (%)	50.1	30.3	46.0	25.3
Cumulative %	50.1	80.4	46.0	71.4
Correlations[a]				
Roll	**0.91**	0.08	**0.89**	0.02
Twist	**−0.92**	−0.15	**−0.88**	−0.15
Slide	**−0.87**	−0.11	**−0.80**	−0.21
Shift	−0.10	**0.88**	−0.19	**0.77**
Tilt	−0.26	**0.84**	−0.22	**0.78**

[a] Coefficients in bold have magnitude greater than 0.75.

Structure of TTTAA Elements in the Nucleosome Core

The TTTAA sequence elements in NCP-TA localize to the 0.5 and 1.5 minor groove blocks such that the TA steps are 5.5 and 15.5 bp from the dyad. As such, TA is the outward-positioned of the two most centrally located minor groove block steps. In spite of sequence identity, the 0.5 and 1.5 sites display notably different structures (Figure 4; see Figure S1 available online). Moreover, superposition of the two sets of four independently refined sites clearly reveals two families of distinct conformational states. In particular, the 0.5 sites display kinking with high slide at the TA step, analogous to that observed for sites with centrally located CA = TG steps (Figure 3). On the other hand, the 1.5 sites show a smooth minor groove bending profile with nominal associated step parameters.

DNA structure at the 0.5 and 1.5 sites in NCP-TA shows distinctions relative to that in NCP147, which contains respectively CAGCT and TGCCT motifs at these locations (Figure 3). These G|C-rich minor groove blocks lack a central flexible step and display smooth bending. The presence of the centrally positioned TA step may account for the occurrence of kinking at the 0.5 sites in NCP-TA. However, kinking is not observed at the 1.5 sites, which instead display the smallest roll values of any of

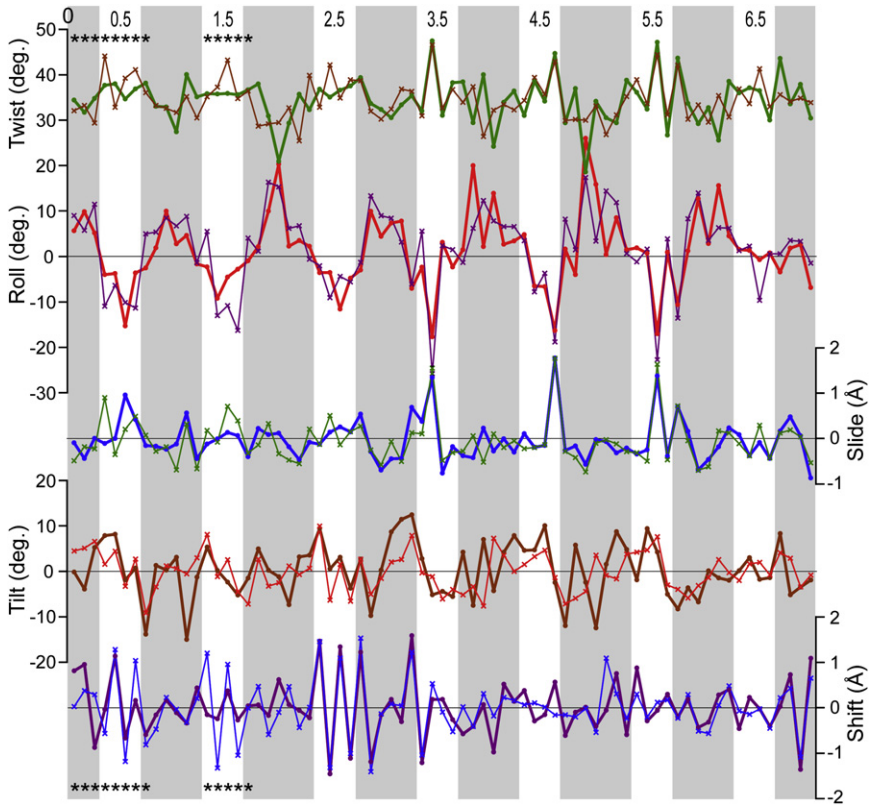

Figure 3. Nucleosomal DNA Conformational Parameters

Dinucleotide step values averaged over one particle half are shown for NCP-TA (thick lines) and NCP147 (thin lines). A "0" indicates the nucleosome center, and site numbers are given for minor groove blocks (unshaded). Bp steps in major groove blocks in addition to the flanking major-minor groove block interface steps are shaded in gray. Asterisks denote bp step sequence differences between NCP-TA and NCP147.

the minor groove blocks. Notably, this location stands out as having by far the most sustained narrow minor groove throughout the nucleosome core (Figure 5A). The narrowest point of the 1.5 site is among the minima observed in NCP147 or NCP-TA, but in particular the expanse of narrowing is substantially greater than at any other location. In NCP147, the narrowing extends over 3 bp, whereas in NCP-TA it persists for at least 6 bp.

The occurrence of extreme minor groove narrowing at the 1.5 sites, irrespective of dramatic sequence changes, indicates that the common DNA conformation is histone-imposed. In fact, this H3-H4 α1α1 helix DNA binding site contains a motif, not found elsewhere on the histone octamer, in which two hydrophobic side chains flank the minor groove at its most narrow point (Figures 5B and 5C). As such, the H3 leucine and H4 proline side chains apparently serve as a "sugar clamp" to enforce massive minor groove compression. The potential for steric clashing between the side chains and DNA backbone would necessitate groove narrowing in order for phosphate group-histone binding to occur. The presence of this unique motif appears to serve the same function in the yeast, *Drosophila*, chicken, and human nucleosomes (White et al., 2001; Clapier et al., 2008; Harp et al., 2000; Tsunaka et al., 2005).

We have previously observed that nearly one-half of the total roll angle component throughout the nucleosome core is not translated into actual bending of the double helix, for true DNA curvature, rather it gives rise to an alternation in bp tip values (Richmond and Davey, 2003). Fluctuation in tip takes the form of canting one bp with respect to its neighbor, yielding a non-zero roll value while not altering the direction of the double-helix axis. The "excess roll" so generated apparently helps accommodate tight systematic wrapping of the DNA. Considering the extreme groove narrowing at the 1.5 site, the intrinsic narrow minor groove preference of A|T tracts (Haran and Mohanty, 2009) and the small negative roll values of NCP-TA relative to NCP147 at this location suggests that negative rolling functions in part to compress the minor groove for "fitting" the histone octamer surface. As such, wide minor groove-preferring motifs, such as the 1.5 site TGCCT element in NCP147, can be forced to undergo compression via negative roll. At the same time, however, extra roll generated for groove narrowing, as opposed to helix bending, could be dissipated through changing tip. Such a mechanism is implied by the occurrence of nearly constant tip values throughout the intrinsically narrow 1.5 region in NCP-TA, which is in strong contrast to the respective profile in NCP147 (Figure 5A).

Whereas the minor groove takes on a sustained, extreme narrow conformation at the 1.5 site, it displays only a single sharp point of narrowing, followed by rapid widening, at the 0.5 site (Figure 5A). The imposition of groove narrowing is in fact so drastic at the 1.5 location that there is insufficient steric clearance to accommodate protein elements (Figure 5B,C). The "minor groove inserting" arginine side chain at this location is instead situated at the mouth of the minor groove, whereas it resides within the groove at all other DNA binding sites (Figure 4D) (Davey et al., 2002). In addition to these distinctions between the 0.5 and 1.5 sites, there is another key element that promotes differential DNA conformation. Residues 39 to 43 of the H3 N-terminal tail bind to juxtaposed minor grooves formed by bp positions 6 to 10 and the nucleosome core terminus (Luger et al., 1997). In particular, a proline side chain is situated within the minor groove at the edge of the 0.5 minor groove block. This has an effect opposite to that of the sugar clamp, in which the sharp compression point of the 0.5 site must be immediately followed by minor groove widening to accommodate the proline (Figure 4D). Combined with the other DNA binding site differences, this distinct groove width modulating motif results in the 0.5 and 1.5 TTTAA elements assuming very different overall conformations.

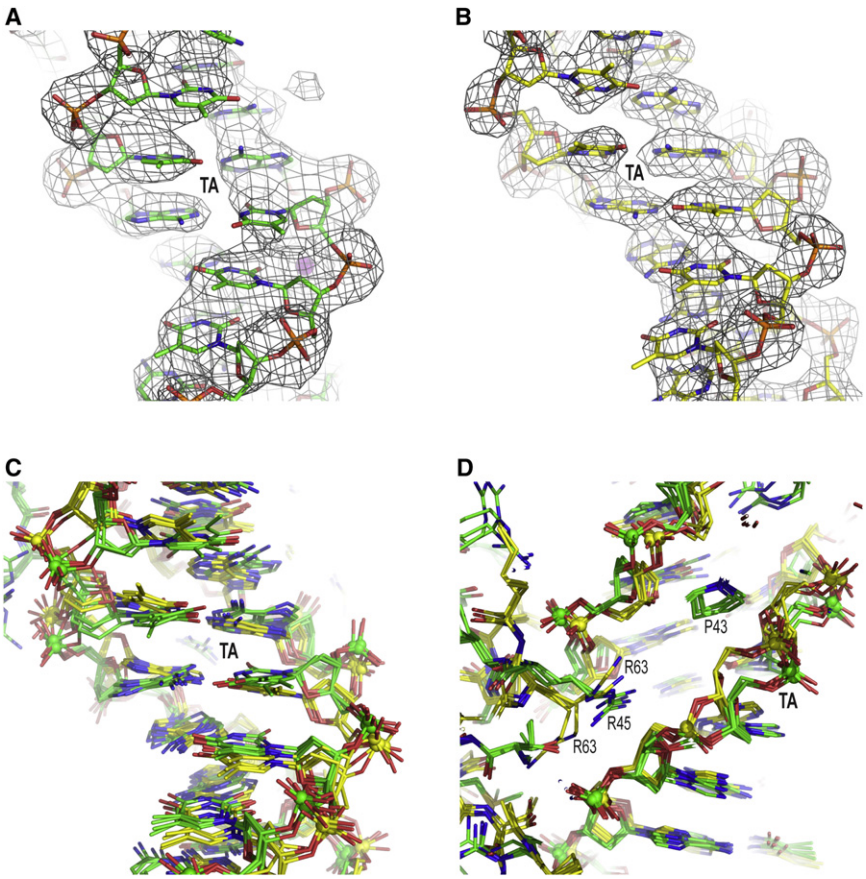

Figure 4. Structure of TTTAA Elements in NCP-TA

The central TA dinucleotide of TTTAA is indicated. (A and B) A 0.5 (A) and 1.5 site (B) with a $2F_o-F_c$ electron density map, contoured at 1.3σ (A) and 1.5σ (B), superimposed on the model.

(C and D) Least-squares superposition of all eight TTTAA elements from the four particle halves. The 0.5 and 1.5 sites are shown with green and yellow, respectively, carbon and phosphorous (spherical) atoms. The view in (C) is with the minor groove and histone binding site in the background. (D) The perspective is looking into the minor groove, with the histone binding site on the left. The side chains of H3 P43 and H4 R45 insert into the minor groove at site 0.5. The side chains of H3 R63 reside at the mouth of the 1.5 minor groove, pointing in either direction. The DNA *B*-factor values averaged for all four TTTAA elements from the 0.5 and 1.5 sites are 115 $Å^2$ and 100 $Å^2$, respectively.

DISCUSSION

Roll, twist, slide, and groove width appear to be the four main factors that govern DNA fitting on the histone octamer. The majority of bending is accomplished via roll (Richmond and Davey, 2003), which additionally acts as a groove width modulator to allow appropriate lateral spacing of the phosphodiester backbones. Twist serves to regulate the rotational strand-to-strand register important for phosphate group binding, as implied by the quantized nature of this parameter comparing stretched versus unstretched DNA conformations (Ong et al., 2007). Moreover, slide plays a key function in generating the pitch of the superhelix (Olson et al., 1998; Tolstorukov et al., 2007). However, the mechanical coupling of roll, twist, and slide and that of roll with groove width can make it difficult to distinguish a DNA wrapping "driver," if there is indeed such a single parameter. On the contrary, it appears that the system has developed to reduce sequence discrimination by accommodating the intrinsic structural tendencies of systematically bent DNA. Thus, the general positioning of flanking histone-phosphate binding sites may reflect typical dimensions for DNA bent into the major and minor grooves, which is in turn optimized for the associated reduced and elevated twist values. In addition, the anticorrelation of roll and slide ensures a left-handed, as opposed to right-handed, superhelix (Olson et al., 1998; Tolstorukov et al., 2007)— at least for the ubiquitous nucleosome (Furuyama and Henikoff, 2009).

Regardless of the degree to which the histone octamer has evolved to minimize sequence dependency, it appears that any factor which induces pronounced DNA bending or distortion will retain at least some low level of sequence discrimination through indirect readout. In particular, because flexibility is especially sequence dependent (El Hassan and Calladine, 1997; Dickerson and Chiu, 1997; Dickerson, 1998; Olson et al., 1998; Travers, 2004; Balasubramanian et al., 2009; Morozov et al., 2009), the magnitude of the indirect sequence bias will be proportional to the degree to which the DNA is forced to deviate from its intrinsic preferences. As such, minor groove bending/compression represents the most energetically challenging DNA distortion (Dickerson and Chiu, 1997; Dickerson, 1998), suggesting it would make a disproportionate contribution to nucleosome positioning. This is consistent with the distinct sequence dependent structural modes observed for accommodating minor groove bending/compression in the nucleosome core (Richmond and Davey, 2003), in addition to the localization of superflexible TA dinucleotides in minor groove block centers as the most prominent nucleosome positioning feature (Travers, 2004; Widom, 2001; Thastrom et al., 2004; Virstedt et al., 2004; Widlund et al., 1999; Fitzgerald and Anderson, 1999; Fernandez and Anderson, 2007). The flexibility of TA is reflected in its multimodal behavior in the NCP-TA structure—supporting very different conformations and kinking as opposed to smooth bending at the 0.5 versus 1.5 sites. Moreover, bending into the major groove is relatively facile and DNA conformational freedom is elevated in major groove blocks, as observed by comparing NCP147 and NCP-TA.

Although TA steps are overall favored at minor groove block centers, the flanking sequence preferences and relative contribution toward positioning vary over different sites in the nucleosome core. In particular, the 1.5 site displays the greatest influence on positioning (Fitzgerald and Anderson, 1999).

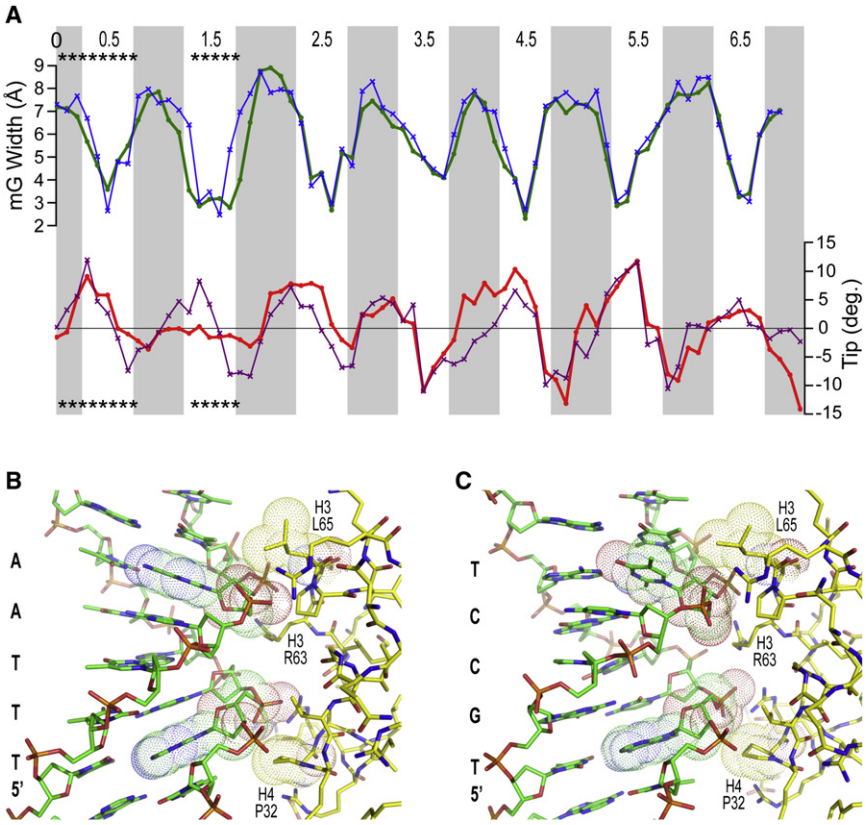

Figure 5. Groove Width Modulation in the Nucleosome Core

(A) Minor groove (mG) width and bp tip values averaged over one particle half are shown for NCP-TA (thick lines) and NCP147 (thin lines). A "0" indicates the nucleosome center, and site numbers are given for minor groove blocks (unshaded). Major groove block bp are shaded in gray. Asterisks denote bp step sequence differences between NCP-TA and NCP147.

(B and C) Histone-enforced minor groove narrowing at site 1.5 in NCP-TA (B) and NCP147 (C). Carbon atoms of protein and DNA are yellow and green, respectively. Residues H3 L65 and H4 P32 comprise the sugar clamp motif that makes extensive van der Waals contacts with the two nucleotides situated in between (space filling). Groove narrowness precludes insertion of H3 R63.

Substitution of TA at this minor groove block center with less flexible step types yields a pronounced corresponding reduction in nucleosome stability and positioning activity (Fernandez and Anderson, 2007). From in vitro selection experiments, a maximum affinity consensus sequence of TTTAA emerges for this site, which is distinct compared with flanking locations (0.5 and 2.5) that show a preference for G|C bp on either side of the central TA step (Thastrom et al., 2004). This is apparently a consequence of the sugar clamp-imposed requirement for an extremely narrow minor groove at the 1.5 location, for which a flexible, TA step-containing, intrinsically narrow poly-A|T tract (Mack et al., 2001) is ideally suited. In vivo, these properties shared by TATAAA would contribute to the positioning power of TATAAACGCC sequences (Widlund et al., 1997, 1999). On the other hand, the need for a minor groove block that can favor widening at one edge from the insertion of a proline side chain may underlie the selection for G|C-enriched sequences at the 0.5 site in vitro (Thastrom et al., 2004).

The 1.5 site makes a disproportionate contribution, but nucleosome positioning preference seemingly arises from the sequence-specific input of at least the central 12 minor and 11 major groove blocks (Figure 1). CA = TG steps at minor groove block centers over the H2A-H2B dimer binding sites (3.5, 4.5, 5.5) in NCP147 appear to drive positioning in this sequence (Richmond and Davey, 2003; Tolstorukov et al., 2007), which lacks any G|C-rich major groove blocks. These elements are common to NCP-TA and work in phase with the TA motifs. Considering the available data, namely that (a) minor groove deformation is energetically challenging and gives rise to specialized modes in the NCP, (b) DNA structure is less constrained in major groove blocks, and (c) flexible bp steps in the 1.5 and other minor groove blocks serve a unique function, indicate the sequence content in minor groove blocks makes a dominant contribution toward nucleosome positioning.

The positioning power of the 1.5 site apparently arises, at least in part, from this location having the most stringent requirements for minor groove compression. Additionally, however, extreme minor groove kinking can occur at this location in the context of DNA stretching around the two-turn region (Ong et al., 2007). The tremendous negative rolling contributes both to the translational component of stretching and to the minor groove bending/narrowing at this site. The occurrence of the stretching-induced distortion apparently results from DNA positioning preferences at flanking regions, because it can be brought about by minute sequence changes at distant locations and is observed in both the crystalline and solution states (Ong et al., 2007; Davey et al., 2009). With consideration of the distortion mapping data (Fitzgerald and Anderson, 1999; Fernandez and Anderson, 2007), TA elements at the 1.5 site may also be capable of extreme kinking with local or distant DNA sequence contexts that differ relative to NCP-TA. This would help explain why a superflexible step appears as a central consensus element at this location. Unfortunately, a 145 bp version of NCP-TA, designed to test this premise, yielded very poorly diffracting crystals. However, previous studies have shown that pronounced, if not maximal, distortion generally occurs at the 1.5 site in genomic and many synthetic DNA sequences (Richmond and Davey, 2003; Ong et al., 2007; Fitzgerald and Anderson, 1999; Fernandez and Anderson, 2007; Richmond et al., 1984; Hogan et al., 1987). Therefore, it is likely that this average feature arises from most sequences, such as the G|C-rich 1.5 element in NCP147, requiring substantial negative rolling for minor groove narrowing in combination with the occurrence of stretching-induced extreme kinking in particular nucleosomal sequences or contexts.

Structural analysis of the NCP composed of different sequences has helped to illuminate key features important for nucleosome positioning. In particular, the unique histone-DNA binding site at position 1.5 and its geometric relationship with flanking sites that can promote stretching (Richmond and Davey, 2003) appears to underlie the distinct distorting and positioning potential of this location. The H3-H4 $\alpha1\alpha1$ helix motif at the 1.5 site and the neighboring histone regions could presumably have evolved to be the same or more similar to the less discriminating symmetry-related 4.5 and flanking sites on the H2A-H2B dimer. However, the histone octamer has emerged to possess such a translational positioning signal, and thus has apparently not absolutely minimized sequence bias, which in turn suggests a "preconceived" function for DNA sequence in nucleosome organization and activity.

Given the current debate on the global sequence dependence of nucleosome organization (Kaplan et al., 2009; Zhang et al., 2009; Travers et al., 2009 ; Stein et al., 2010), a positioning motif built into the histone system raises the possibility that stringent signals in the genomic DNA may only arise where it is necessary to have a defined position. Thus, the default could be mobile nucleosomes, with reservation for well positioned ones established through DNA sequence signals when warranted. As such, the uncommon TATA and other strong positioning elements found in mammalian genomes (Widlund et al., 1997) may be a reflection of an infrequent necessity for highly defined localization. This would mean that the sequence dependence of positioning apparent from genome-wide studies could be moderate if many DNA locations were composed of relatively random "mixed signals" (e.g., out of phase TA elements).

EXPERIMENTAL PROCEDURES

Nucleosome Core Particle Preparation

NCP was prepared from recombinant *Xenopus laevis* histones and 147 bp DNA fragments using established methodologies (Luger et al., 1999). NCP-d2 and NCP-TA DNA expression constructs were generated by inserting multiple repeats into the EcoRV site of the pUC57 plasmid (EZBiolab, Carmel, IN, USA). The NCP-d2 insert consisted of tandem full length repeats connected by EcoRV restriction sites, whereas that of the palindromic NCP-TA comprised inverted half-site repeats, as described before (Luger et al., 1999). The half-site fragments were excised with EcoRV and reassembled via a central AvaII restriction site to generate the full-length DNA.

Crystals of NCP-TA were prepared as described previously (Davey et al., 2002). Crystallization of NCP-d2 was carried out by a similar approach, whereby 2-methyl-2,4-pentanediol (MPD), Hg, spermine, and β-octyl glucoside were additionally present in the crystallization buffer, which allowed acquisition of larger crystals with reduced incidence of twinning. Crystals were grown in droplets containing 4 mg/ml NCP-d2, 85 mM $MnCl_2$, 60 mM KCl, 20 mM K-cacodylate (pH 6.0), 12% (v/v) MPD, 2 mM $HgCl_2$, 2 mM spermine-Cl_4, and 0.5% (w/v) β-octyl glucoside and equilibrated against a 42.5 mM $MnCl_2$, 30 mM KCl, 10 mM K-cacodylate (pH 6.0), 12% (v/v) MPD, 1 mM $HgCl_2$, 1 mM spermine-Cl_4, and 0.25% (w/v) β-octyl glucoside buffer. Crystals were stabilized in a harvest buffer composed of 37 mM $MnCl_2$, 40 mM KCl, 20 mM K-cacodylate (pH 6.0), 24% (v/v) MPD, 2 mM $HgCl_2$, 2 mM spermine-Cl_4, 0.5% (w/v) β-octyl glucoside, and 2% (w/v) trehalose.

Structure Solution and Analysis

X-ray diffraction data were recorded as described previously (Ong et al., 2007) at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) using the PILATUS detector on beamline X06SA (NCP-d2, λ = 1.00 Å) and a Mar225 CCD detector on beamline X06DA (NCP-TA, λ = 1.00 Å). Data were processed using MOSFLM (Leslie, 2006) and SCALA from the CCP4 suite (CCP4, 1994). The histone-DNA model of NCP147 (Protein Data Bank code 1KX5; Davey et al., 2002) was used for structure solution by molecular replacement. Structural refinement and model building were carried out with routines from the CCP4 suite (CCP4, 1994). DNA conformational analysis was conducted using CURVES (Lavery and Sklenar, 1988). Graphic figures were prepared with PyMOL (DeLano Scientific LLC, San Carlos, CA, USA). Principal component analysis and plot renderings were done with Microsoft Excel and the XLSTAT routine (Microsoft Corporation).

ACCESSION NUMBERS

Atomic coordinates and diffraction data for NCP-TA have been deposited in the RCSB Protein Data Bank under accession code 3LEL.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure and can be found with this article online at doi:10.1016/j.str.2010.01.015.

ACKNOWLEDGMENTS

We are grateful to C. Schulze-Briese, M. Wang, V. Olieric, T. Tomizaki, R. Bingel-Erlenmeyer, M. Fuchs, and A. Pauluhn at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland), and R. Muhammad, who ensured success with data collection. We thank T. Richmond for comments on the earlier design of NCP DNA constructs and G.E. Davey for critical input on the manuscript. This work was supported by A*STAR Biomedical Research Council grant 06/1/22/19/452 and Academic Research Council grant 19/08 from the Ministry of Education (Singapore). B.W. conducted the crystallographic work on NCP-TA; K.M. carried out experiments with NCP-d2; D.V. grew crystals of NCP-TA; C.A.D. designed constructs, conducted analysis, and wrote the manuscript.

Received: November 10, 2009
Revised: January 19, 2010
Accepted: January 28, 2010
Published: April 13, 2010

REFERENCES

Balasubramanian, S., Xu, F., and Olson, W.K. (2009). DNA sequence-directed organization of chromatin: structure-based computational analysis of nucleosome-binding sequences. Biophys. J. *96*, 2245–2260.

Bao, Y., White, C.L., and Luger, K. (2006). Nucleosome core particles containing a poly(dA.dT) sequence element exhibit a locally distorted DNA structure. J. Mol. Biol. *361*, 617–624.

CCP4 (Collaborative Computational Project, Number 4). (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Clapier, C.R., Chakravarthy, S., Petosa, C., Fernandez-Tornero, C., Luger, K., and Muller, C.W. (2008). Structure of the Drosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer. Proteins *71*, 1–7.

Davey, C.A., and Richmond, T.J. (2002). DNA-dependent divalent cation binding in the nucleosome core particle. Proc. Natl. Acad. Sci. USA *99*, 11169–11174.

Davey, C.A., Sargent, D.F., Luger, K., Mäder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. J. Mol. Biol. *319*, 1097–1113.

Davey, G.E., Wu, B., Dong, Y., Surana, U., and Davey, C.A. (2009). DNA stretching in the nucleosome facilitates alkylation by an intercalating antitumour agent. Nucleic Acids Res. Published online December 21, 2009. 10.1093/nar/gkp1174.

Dickerson, R.E. (1998). DNA bending: the prevalence of kinkiness and the virtues of normality. Nucleic Acids Res. *26*, 1906–1926.

Dickerson, R.E., and Chiu, T.K. (1997). Helix bending as a factor in protein/DNA recognition. Biopolymers *44*, 361–403.

El Hassan, M.A., and Calladine, C.R. (1997). Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. Phil. Trans. R. Soc. A *355*, 43–100.

Fernandez, A.G., and Anderson, J.N. (2007). Nucleosome positioning determinants. J. Mol. Biol. *371*, 649–668.

Fitzgerald, D.J., and Anderson, J.N. (1999). DNA distortion as a factor in nucleosome positioning. J. Mol. Biol. *293*, 477–491.

Furuyama, T., and Henikoff, S. (2009). Centromeric nucleosomes induce positive DNA supercoils. Cell *138*, 104–113.

Haran, T.E., and Mohanty, U. (2009). The unique structure of A-tracts and intrinsic DNA bending. Q. Rev. Biophys. *42*, 41–81.

Harp, J.M., Hanson, B.L., Timm, D.E., and Bunick, G.J. (2000). Asymmetries in the nucleosome core particle at 2.5 Å resolution. Acta Crystallogr. D Biol. Crystallogr. *56*, 1513–1534.

Hogan, M.E., Rooney, T.F., and Austin, R.H. (1987). Evidence for kinks in DNA folding in the nucleosome. Nature *328*, 554–557.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. *10*, 161–172.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362–366.

Krueger, A., Protozanova, E., and Frank-Kamenetskii, M.D. (2006). Sequence-dependent basepair opening in DNA double helix. Biophys. J. *90*, 3091–3099.

Lavery, R., and Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. J. Biomol. Struct. Dyn. *6*, 63–91.

Leslie, A.G. (2006). The integration of macromolecular diffraction data. Acta Crystallogr. D Biol. Crystallogr. *62*, 48–57.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251–260.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Preparation of nucleosome core particle from recombinant histones. Methods Enzymol. *304*, 3–19.

Mack, D.R., Chiu, T.K., and Dickerson, R.E. (2001). Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: a comparative analysis of T-A and A-T steps within A-tracts. J. Mol. Biol. *312*, 1037–1049.

Morozov, A.V., Fortney, K., Gaykalova, D.A., Studitsky, V.M., Widom, J., and Siggia, E.D. (2009). Using DNA mechanics to predict in vitro nucleosome positions and formation energies. Nucleic Acids Res. *37*, 4707–4722.

Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M., and Zhurkin, V.B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc. Natl. Acad. Sci. USA *95*, 11163–11168.

Ong, M.S., Richmond, T.J., and Davey, C.A. (2007). DNA stretching and extreme kinking in the nucleosome core. J. Mol. Biol. *368*, 1067–1074.

Packer, M.J., and Hunter, C.A. (1998). Sequence-dependent DNA structure: the role of the sugar-phosphate backbone. J. Mol. Biol. *280*, 407–420.

Radman-Livaja, M., and Rando, O.J. (2010). Nucleosome positioning: How is it established, and why does it matter? Dev. Biol. *339*, 258–266.

Richmond, T.J., and Davey, C.A. (2003). The structure of DNA in the nucleosome core. Nature *423*, 145–150.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7 Å resolution. Nature *311*, 532–537.

Satchwell, S.C., Drew, H.R., and Travers, A.A. (1986). Sequence periodicities in chicken nucleosome core DNA. J. Mol. Biol. *191*, 659–675.

Segal, E., and Widom, J. (2009a). From DNA sequence to transcriptional behaviour: a quantitative approach. Nat. Rev. Genet. *10*, 443–456.

Segal, E., and Widom, J. (2009b). Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr. Opin. Struct. Biol. *19*, 65–71.

Segal, E., and Widom, J. (2009c). What controls nucleosome positions? Trends Genet. *25*, 335–343.

Stein, A., Takasuka, T.E., and Collings, C.K. (2010). Are nucleosome positions in vivo primarily determined by histone-DNA sequence preferences? Nucleic Acids Res. *38*, 709–719.

Thastrom, A., Bingham, L.M., and Widom, J. (2004). Nucleosomal locations of dominant DNA sequence motifs for histone-DNA interactions and nucleosome positioning. J. Mol. Biol. *338*, 695–709.

Tolstorukov, M.Y., Colasanti, A.V., McCandlish, D.M., Olson, W.K., and Zhurkin, V.B. (2007). A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. J. Mol. Biol. *371*, 725–738.

Travers, A., Caserta, M., Churcher, M., Hiriart, E., and Di Mauro, E. (2009). Nucleosome positioning-what do we really know? Mol. Biosyst. *5*, 1582–1592.

Travers, A.A. (2004). The structural basis of DNA flexibility. Philos. Transact. A Math. Phys. Eng. Sci. *362*, 1423–1438.

Tsunaka, Y., Kajimura, N., Tate, S., and Morikawa, K. (2005). Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. Nucleic Acids Res. *33*, 3424–3434.

Virstedt, J., Berge, T., Henderson, R.M., Waring, M.J., and Travers, A.A. (2004). The influence of DNA stiffness upon nucleosome formation. J. Struct. Biol. *148*, 66–85.

White, C.L., Suto, R.K., and Luger, K. (2001). Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. EMBO J. *20*, 5207–5218.

Widlund, H.R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P.E., Kahn, J.D., Crothers, D.M., and 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., and Kubista, M. (1999). Nucleosome structural features and intrinsic properties of the TATAAACGCC repeat sequence. J. Biol. Chem. *274*, 31847–31852.

Widlund, H.R., Vitolo, J.M., Thiriet, C., and Hayes, J.J. (2000). DNA sequence-dependent contributions of core histone tails to nucleosome stability: differential effects of acetylation and proteolytic tail removal. Biochemistry *39*, 3835–3841.

Widom, J. (2001). Role of DNA sequence in nucleosome stability and dynamics. Q. Rev. Biophys. *34*, 269–324.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nat. Struct. Mol. Biol. *16*, 847–852.

