Quarterly Reviews of Biophysics **34**, 3 (2001), pp. 269–324. © 2001 Cambridge University Press
DOI: 10.1017/S0033583501003699 Printed in the United Kingdom

Role of DNA sequence in nucleosome stability and dynamics

J. Widom

Department of Biochemistry, Molecular Biology, and Cell Biology, and Department of Chemistry, Northwestern University, 2153 Sheridan Road, Evanston, Illinois 60208-3500, USA

1. Introduction 270
1.1 Overview of nucleosome structure 271

2. Relative equilibrium stability (affinity) of histone–DNA interactions in nucleosomes 272
2.1 Relative affinity equals relative equilibrium stability 272
2.2 Competition assays for relative free-energy measurements 273
2.3 Technical issues in relative free-energy measurements 275
2.4 Range of affinities 278

3. Relation of nucleosome stability to nucleosome positioning 279
3.1 Translational nucleosome positioning 279
3.2 Rotational positioning 280
3.3 Unfavorable positioning 281
3.4 Experiments 281

4. Physical basis of DNA sequence preferences 282
4.1 Free-energy cost of DNA bending 283
4.2 Molecular mechanics of DNA bending and bendability 284
4.3 Bent and bendable DNA sequences 286
4.4 Parameter sets for prediction of DNA bending and bendability 288
4.5 DNA twisting 290
4.6 Energetics of nucleosomal DNA packaging 291

5. DNA sequence motifs for nucleosome packaging 292
5.1 Natural and designed nucleosomal DNAs 293
5.2 New rules and reagents from physical selection studies 294
5.3 Molecular basis of DNA sequence preferences 299
5.4 Special properties of the TA step 300
5.5 Unfavorable sequences 302
5.6 Natural genomes 303
5.7 Evolutionary approach toward an optimal sequence 305

Correspondence: J. Widom, Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, 2153 Sheridan Road, Evanston, IL 60208-3500, USA.
Tel.: 847/467-1887; Fax: 847/467-1380; E-mail: j-widom@northwestern.edu

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

5.8 Optimization by design 305

6. Dynamic nucleosome instability 308
6.1 Site-exposure equilibria 308
6.2 DNA sequence-dependence to site-exposure equilibria 312
6.3 Nucleosome translocation 315
6.4 Action of processive enzymes 319

7. Conclusions 319

8. Acknowledgements 320

9. References 320

1. Introduction

The nucleosome core particle is the fundamental repeating subunit of chromatin. It consists of two molecules each of the four 'core histone' proteins, H2A, H2B, H3 and H4, and a 147 bp stretch of DNA. The lowest level of chromatin organization consists of a repeated array of nucleosome core particles separated by variable lengths of 'linker DNA'. In many, but not all, cases, each core particle plus its linker DNA is associated with one molecule of a fifth 'linker' histone protein, H1. The complex of the core particle plus its linker DNA and H1 (when present) is called a 'nucleosome'.

Nucleosomes are of interest in part because of their importance in chromatin function. The substrates for the essential biological processes of transcription, replication, recombination, DNA repair, and cell division are chromatin (i.e. chains of nucleosomes), not naked DNA. New discoveries in nucleosome and chromatin function are emerging almost daily. Nucleosomes play both active and passive roles, in both positive and negative gene regulation. They dominate the behavior of key steps in replication, recombination and repair. Specialized nucleosomes facilitate the function of chromosome centromeres and telomeres. These aspects of nucleosome biology are reviewed elsewhere (Widom, 1998; Wolffe, 1998; Elgin & Workman, 2000).

The present review will focus on a second remarkable aspect of nucleosomes. This involves the structural organization of the nucleosomal DNA and, in particular, the interplay between DNA sequence and nucleosome stability. These topics are of particular significance because they bear on fundamental issues in DNA structure and mechanics that are broadly applicable to most or all protein–DNA complexes. Whenever proteins and DNA combine to make a protein–DNA complex, the structures of both the protein and the DNA (and their solvation) will inevitably exhibit changes from the conformations preferred as individual (solvated) molecules. Often these conformational changes are dramatic, involving substantial protein folding or unfolding or conformational rearrangements, and substantial bending, twisting, melting, or other conformational changes of the DNA. Much of the free energy of protein–DNA 'bond' formation is used up doing work to distort the DNA conformation. Individual structural and mechanical properties of differing DNA sequence motifs can greatly influence both the detailed structures that are formed, and, especially, the energetic cost of forming them, i.e. the overall stability of the protein–nucleic acid complexes. Because

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

nucleosomes package *any* arbitrary DNA sequence, yet greatly distort *all* of them, nucleosomes serve as an excellent laboratory for studies of the sequence-dependence of DNA structural and mechanical properties.

We will briefly review key earlier studies, and will then focus on recent discoveries and challenges for future work.

1.1 Overview of nucleosome structure

The structure of the nucleosome core particle has been solved by X-ray crystallography at near-atomic resolution (Luger *et al*. 1997a). The structure is reviewed in more detail elsewhere (Widom, 1997a; Luger & Richmond, 1998a; Richmond & Widom, 2000), but I will briefly review here some aspects that are particularly relevant to the relationship between DNA sequence and nucleosome stability. The structure reveals a basic tripartite assembly for the histone octamer, reflecting its organization in functional subunits comprising two H2A/H2B heterodimers and one $H3_2H4_2$ tetramer. The tetramer itself is a stable complex of two H3/H4 heterodimers, which have a 'handshake' interlocking protein fold similar in structure to that of the H2A/H2B heterodimers, now known as the histone fold motif. The histone fold domains organize the central 121 bp of DNA, with the additional 13 bp at each end organized by an N-terminal extension to the histone fold of H3. Each histone heterodimer contributes three main DNA-binding motifs, in two types (referred to as L1L2 and $\alpha 1\alpha 1$). These occur in the order L1L2, $\alpha 1\alpha 1$, L1L2, spaced $\approx$ 10 bp apart.

The histone octamer has overall pseudo-twofold symmetry. In addition each histone heterodimer has an *intra*-dimer pseudo-twofold symmetry axis (running through the $\alpha 1\alpha 1$ DNA-binding motif), and the two heterodimers on each side of the overall twofold are related by an additional *inter*-dimer pseudo-twofold symmetry axis.

A prominent feature of the histone octamer surface is a positively charged superhelical ramp, which plays an important role in the DNA organization. The DNA is wrapped around the histone octamer on this charged ramp, in $\approx$ 1.65 turns of an irregular superhelix, with the DNA helix axis at a (variable) radius of $\approx$ 42.5 Å and a (variable) pitch of $\approx$ 26.5 Å. Bending is distributed along the length of the DNA and includes bends into both the major and minor grooves; eight sites (at superhelix locations ± 1, 2, 4 and 5 in their numbering system, corresponding roughly to base pairs ± 10, 20, 40, 50 for a nucleosome centered at base pair 0) which involve bends into the DNA major grooves appear particularly deformed. Not only is the DNA dramatically bent, but its twist is also distorted away from the canonical solution value of $\approx$ 10.5 bp/turn, varying in detail with position in the particle, with an overall average value of $\approx$ 10.2 bp/turn.

DNA binding occurs primarily to the sugar phosphate backbone over the short stretches of each DNA helical turn where the minor groove (and hence the backbones of both DNA chains) faces inward toward the octamer surface. Contacts between the histones and DNA are primarily non-specific. They include extensive salt bridges and hydrogen bonds to the phosphate groups contributed by both main-chain and side-chain groups; extensive non-polar contacts with the DNA sugar; and electrostatic interactions of the positively charged N-termini of α-helices with DNA phosphates. However, there do exist a number of base-specific contacts, including non-polar contact of the 5-methyl group of thymidine in the major groove.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

Two additional modes of histone–DNA interaction are particularly striking. First, an arginine side-chain is inserted into the minor groove every time it faces inward to the histone surface. These side-chains generally make salt bridges to the phosphodiester chains on both sides of the minor groove. In most cases the arginine is held by additional bonds to protein functional groups so as to prevent it from penetrating deeply into the groove and making base-specific contacts such as hydrogen bonding to pyrimidine O2 or purine N3 groups in the minor groove. Secondly, each of the core histones has a $\approx$ 10–40 amino-acid-long, highly positively charged N-terminal domain. These are referred to as 'tail domains' because they are highly extended and mobile. The N-terminal tails of both copies of H2B and H3 pass out from the nucleosome between the gyres of DNA, threading through adjacent minor grooves and extending out from the particle. The N-terminal tails of the two copies of H2A pass over the DNA along a minor groove. Thus together the H2A, H2B and H3 tail domains bracket turns of DNA, passing over and between the DNA gyres, and seem likely to influence the energetics and dynamics of DNA wrapping. These tail domains are of further interest because they mediate interactions with numerous gene regulatory proteins, are the sites of numerous post-translational modifications that are linked to nucleosome function, and [particularly for histone H4 (Luger *et al.* 1997a)] are implicated in stabilization of higher order chromatin folding (Widom, 1989, 1998; Fletcher & Hansen, 1995).

2. Relative equilibrium stability (affinity) of histone–DNA interactions in nucleosomes

Any arbitrary sequence DNA molecule can be incorporated favorably into nucleosomes *in vitro*, consistent with the biological requirement for packaging the vast DNA sequence diversity represented in the collective genomes of all eukaryotes and euryarchaeota. The *absolute* affinity of histones for DNA in nucleosome formation has not been reported, due in part to complications from the quaternary but stepwise-ordered nature of the assembly and disassembly processes (Widom, 1989) and accompanying kinetic traps (Widom, 1999). However, differential methods are available that allow one to analyze *relative* free energies for nucleosome formation on one sequence compared to another. These experiments reveal that, while nucleosomes can form on any DNA, certain DNA sequences have relative affinities that exceed by factors of 100–1000 or more the affinity of arbitrary sequence DNA. The sequence features responsible for this higher affinity are discussed below. Here we discuss what these different affinities imply with regard to nucleosome stability, and then we focus on how these relative affinities are determined experimentally.

2.1 Relative affinity equals relative equilibrium stability

Consider any molecule A which can combine with molecule B_1 to make AB_1, or may instead combine with a closely similar molecule B_2 to make AB_2. We may choose to speak of the *affinities* of A for B_1 or B_2, as quantified by the corresponding equilibrium association constants or their reciprocals, the dissociation constants. *Affinity is equivalent to equilibrium stability*: suppose that B_2 binds to A more tightly than does B_1 (i.e. with greater association constant, or with a lower dissociation constant for the AB_2 complex). This equates to less

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

dissociation of AB_2 than of AB_1, given A, B_1, and B_2 at identical total concentrations. That is, the higher-affinity species makes the more stable complex.

This property of chemical equilibria, which applies to *all* molecules, must also apply to the case of differing DNA sequences binding to histone octamers. DNA sequences having lower or greater relative *affinities* for histone octamer must create nucleosomes having correspondingly lower or greater relative equilibrium *stabilities*.

2.2 Competition assays for relative free-energy measurement

The assays used at present derive from the work of Crothers and colleagues (Shrader & Crothers, 1989, 1990). They are based on a competitive nucleosome reconstitution procedure, in which radiolabeled tracer DNA competes with a large excess of unlabeled competitor DNA for binding to limiting amounts of histone octamer during *in vitro* nucleosome reconstitution. Nucleosomes are formed in conditions that allow for full equilibration, and then the resulting species are gradually 'frozen-in' to allow their quantification. Depending on the relative affinity of the specific radiolabeled tracer DNA molecule compared to the (invisible) competing pool of unlabeled DNA, more or less of the tracer DNA is incorporated into nucleosomes. The resulting labeled products are resolved by native gel electrophoresis or other equivalent procedure to separate free tracer DNA from tracer DNA in nucleosomes. The ratio of counts in nucleosomes to counts in naked DNA defines an equilibrium constant and a corresponding relative free energy that is valid for that competitive environment. When the procedure is carried out for different tracer DNAs in parallel, using the identical competitive mixture for each, some DNA sequences are found to compete more effectively than others. Quantitative difference-free energies between any two tracer DNAs are obtained by subtraction of the corresponding relative free energies.

Several variants of this basic approach are in use. All need to satisfy several interrelated technical requirements, including: they must actually yield native nucleosomes as products; they must overcome kinetic blocks inherent in nucleosome reassembly; and they must rigorously establish equilibrium. (A question that arises, concerning the conditions in which this equilibrium is actually established, will be discussed below.) All of these approaches take advantage of two key facts. First, sufficiently high salt concentrations suppress the affinity of histones for DNA to negligible levels [because of the polyelectrolyte properties of DNA (Manning, 1978)], thereby allowing free equilibration of the histones between differing DNAs. And secondly, subsequent appropriate decreases in the concentration of salt (and temperature) reversibly allow nucleosome reconstitution, and ultimately trap and freeze-in stable nucleosomes at sub-physiological salt concentrations.

Exchange procedures for competitive nucleosome reconstitution take advantage of the fact that native nucleosomes can be efficiently assembled on tracer DNAs simply by mixing the tracer DNA with native nucleosomes or chromatin, in conditions of elevated salt concentration and temperature. In the original method developed by Crothers, defined amounts of H1-depleted chromatin served as a histone donor, while a large excess of chicken erythrocyte DNA serves as competitor. These reagents are incubated with specific radiolabeled tracer DNA in conditions of elevated salt concentration (typically 1 M) and temperature (37 °C) for a sufficient period to establish equilibrium, and then diluted stepwise to freeze-in the products for analysis.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

Early studies (Shrader & Crothers, 1989) confirmed that an equilibrium distribution can in fact be achieved by this process. The distribution of products was found not to vary with increased incubation time, and was independent of whether the radiolabeled tracer started out in nucleosomes or as free DNA (a classic test for equilibration).

A concern with the use of exchange-based methods for measurement of difference-free energies is based on classic observations that, in 1 M NaCl and 25 or 37 °C, histones (particularly, the $H3_2H4_2$ tetramer) retain substantial affinity for DNA (van Holde, 1989). This means that the exchange procedures could be sensitive to uncontrolled variations in salt concentration, temperature, or details of the mixing during dilution. Thus while Crothers' original method demonstrably does achieve equilibrium, variants of that method may not. Indeed, with a variant of the exchange method that uses histone H1- (and H5-) depleted chromatin as *both* the histone octamer donor and the DNA competitor, evidence is described below suggesting that equilibrium may not always be achieved.

For these reasons we developed an alternative method based on salt-dialysis (Lowary & Widom, 1997, 1998). Purified histone octamer, competitor DNA and radiolabeled specific tracer are incubated together in 2 M NaCl, conditions in which all four histones are known to have negligible affinity for DNA, ensuring rapid, free exchange. [The same salt concentration quantitatively displaces the histone octamer from DNA, in the purification procedure for the histone octamer itself (Feng *et al*. 1993).] The concentration of NaCl is then slowly reduced by dialysis, thereby again reconstituting nucleosomes and ultimately freezing-in the distribution of nucleosomal and free DNA products to allow quantitative analysis. Provided that the dialysis occurs sufficiently slowly, the literal reversibility of the process guarantees as a matter of physical principle that the reconstitution achieves a true equilibrium prior to freezing it in.

The dialysis method facilitates a direct comparison of free energies for many sequences since both the competitive solution environment and the process by which the competition is ultimately 'frozen' for analysis can be made literally identical for many samples simultaneously. This is accomplished as follows. We produce a single reaction mixture containing limiting histone octamer, excess DNA competitor, and initial 2 M NaCl solution, in sufficient amount for use in all of the desired reactions at once. This mixture is aliquoted into different tubes, and tracer quantities of the differing radiolabeled DNA fragments to be analyzed are added. As an additional precaution, we thoroughly equilibrate all of the samples against the initial 2 M NaCl solution prior to dialysis down to lower NaCl concentration. We place all of the dialysis buttons inside a dialysis bag, which in turn is placed into a single vessel of dialysis buffer, so that the samples will have an identical history during reconstitution.

The results from such an experiment with two different DNAs are illustrated in Fig. 1. The reconstituted samples are resolved by native gel electrophoresis, and the relative concentrations of radiolabeled DNA molecules in nucleosomes and in free DNA for each sample are determined by Phosphorimager or Instantimager analysis. It is apparent from the raw data themselves that the tracer DNA sample in lane 3 has a significantly higher affinity than that in lane 2.

The robustness of the dialysis procedure is evident from the fact that different investigators, studying different preparations of the same DNA molecules at different times, obtain results that are in good quantitative agreement (e.g. compare the results of Lowary & Widom, 1998; Thåström *et al*. 1999; Anderson & Widom, 2000). Additional evidence suggesting that the system is behaving properly includes the following. When the tracer

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

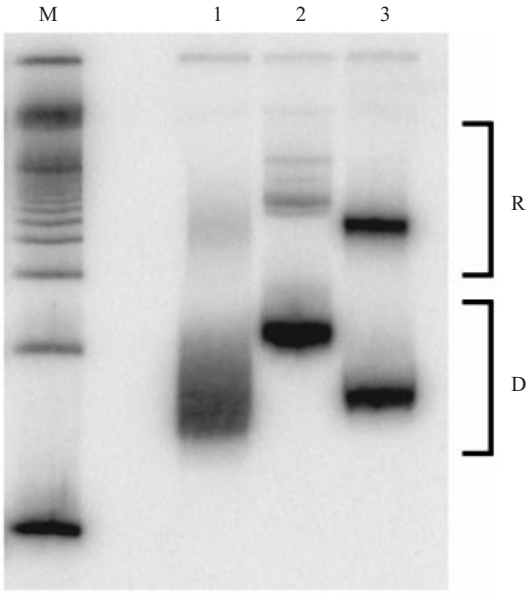


Fig. 1. Competitive nucleosome reconstitution analyzed by native gel electrophoresis. A quantitative phosphorimage is shown. Radiolabeled tracer competes with a large excess of unlabeled natural nucleosome core particle DNA for limiting quantities of histone octamer in dialysis from concentrated NaCl. Lane 1, some of the natural nucleosome core particle DNA is labeled and used as tracer. Lane 2, a derivative of the well-characterized natural nucleosome-positioning sequence from sea urchin 5S rRNA gene (Simpson & Stafford, 1983; FitzGerald & Simpson, 1985). Lane 3, high-affinity sequence 601.2 (Anderson & Widom, 2000). M, 100 bp ladder; D, mobility of naked DNA; R, reconstituted nucleosome mobilities. The diverse mobilities in the nucleosomal region of the gel represent reflect a range of positionings on the different molecules. (Also, the core particle DNA tracer is not homogenous in length). The raw data show that 601.2 competes more effectively for the limiting histone octamer than does the 5S sequence, since there is plainly a greater ratio of counts in bands 'R' *versus* counts in 'D' for 601.2 than for 5S. (From Anderson & Widom, 2000.)

DNA is identical to the unlabeled competitor, the fraction of tracer DNA reconstituted into nucleosomes is equal to the molar ratios of histone octamer and competitor DNA in the reaction. And, the actual equilibrium fraction of tracer DNA incorporated into nucleosomes can be greatly changed through the use of higher-affinity competitor DNAs, yet the difference-free energies between differing DNAs are unchanged. Results from such an experiment are illustrated in Fig. 2*a*. This approach will likely prove to be important as even higher-affinity DNA sequences are designed or selected.

2.3 Technical issues in relative free-energy measurements

Non-trivial technical issues do in fact arise during experimental measurement of relative (difference-) free energies for histone–DNA interactions in nucleosome reconstitution. One key potential problem is a failure to achieve equilibrium. Fig. 2*b* compares the results obtained using the dialysis method with those obtained using a popular variant of the exchange method. The results deviate systematically and significantly from the identity line. This implies that at least 1 of the 2 experiments is not in fact yielding equilibrium values. The discussion summarized above leads us to conclude that the problem lies with this variant of the exchange method.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

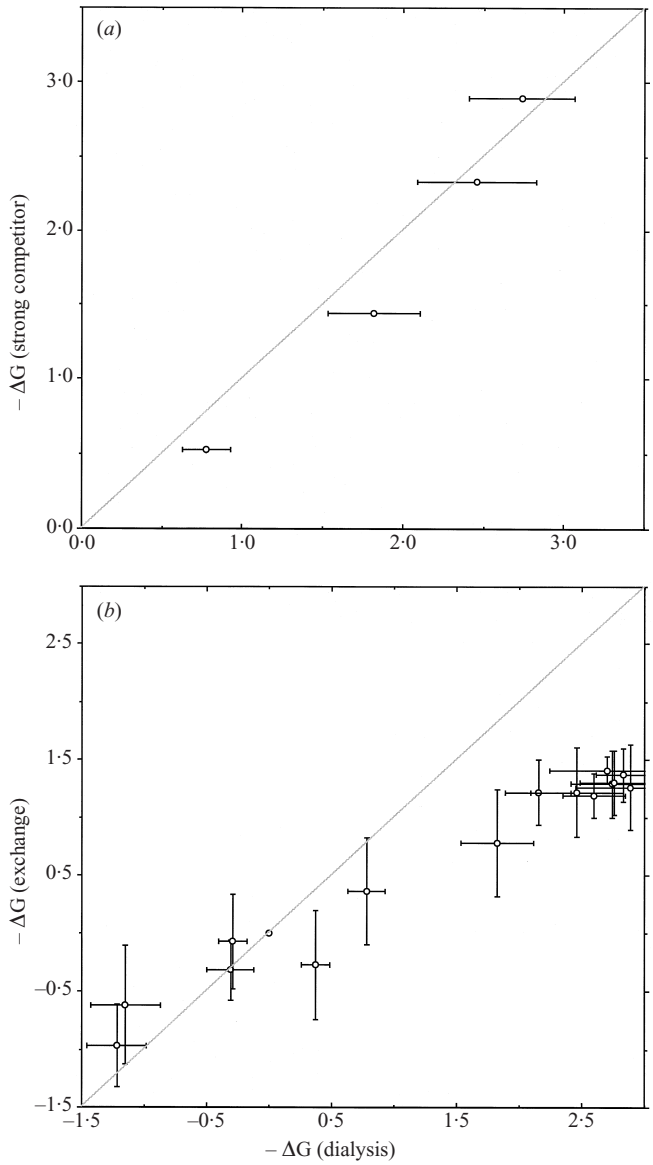


Fig. 2. (*a*) Comparison of difference-free energies ($\Delta\Delta G°$s) for histone–DNA interaction in nucleosome reconstitution measured by the equilibrium dialysis method *versus* the results obtained using a dialysis approach but with a much stronger competitor DNA. Numerical values are reported relative to the results obtained for the sea urchin 5S rRNA gene nucleosome-positioning sequence, which is used as a reference to facilitate comparison between studies. (Data from A. Thåström, presented in tabular form in Thåström *et al*. 1999.) The gray line has a slope of 1 and an intercept of 0 (an identity line), as required if the two methods give equivalent results. The results fall close to the identity line despite arising from very different absolute values for the fraction of radiolabeled tracer incorporated into nucleosomes. (*b*) Comparison of difference free energies ($\Delta\Delta G°$s) for histone–DNA interaction in nucleosome reconstitution measured by the equilibrium dialysis method *versus* the results obtained using a variant of the exchange method. (Data from A. Thåström, presented in tabular form in Thåström *et al*. 1999.) In this case the results *do not* follow the identity line. This means that at least one of the two methods as implemented in this study is failing to give correct equilibrium values. Other studies suggest that the problem lies with this variant of the exchange method. Error bars indicate ± 1 S.D. among repeated measurements.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

A separate question concerns the conditions in which equilibrium is established (Drew, 1991). This question applies whether the reconstitution reactions are carried out using the equilibrium method or with variants of the exchange method that do achieve equilibrium. As the concentration of NaCl drops below ≈ 1.5 to 1 M, the $H3_2H4_2$ tetramer develops significant and increasing affinity for DNA; and at ≈ 1 to 0.75 M NaCl the H2A/H2B heterodimers begin to develop significant affinity (van Holde, 1989). Thus at some indeterminate salt concentration, probably in the range 0.5–1 M, the histones develop sufficiently high affinity for DNA that they may no longer be in rapid exchange between differing DNA molecules.

It follows that the equilibrium that is frozen-in during competitive reconstitution is one that applies to the structure of the nucleosome as it exists in 0.5–1 M NaCl. Since this is close to or overlaps with the range in which histones may begin dissociating, it raises the important question of whether measurements of relative free energies for differing DNAs, as measured in competitive reconstitution experiments, are relevant to nucleosomes in physiological conditions?

Data in the literature plainly establish that there are detectable changes in *net* or *apparent* nucleosome conformation that occur over the range from roughly physiological salt concentrations up to 1 M (van Holde, 1989). For example, the sedimentation coefficient, tyrosine fluorescence anisotropy, and near-UV molar ellipticity all change slightly and gradually over this range in salt concentration. As mentioned above, nucleosomes become susceptible to DNA exchange at salt concentrations of ≈ 0.8 M or greater. Several studies indicate that the extended histone tail domains may be 'loosening' or dissociating from the nucleosomal DNA over this range in salt concentration (van Holde, 1989).

Of such observations, those concerning the tail domains plainly point to definite structural changes. Remarkably, however, these domains do not contribute to the stability of the nucleosome (Dumuis-Kervabon *et al*. 1986; Ausio *et al*. 1989; Luger *et al*. 1997b). Hence their 'loosening' from the particle need not reflect nor cause any significant structural change within the rest of the nucleosome.

The other observations may be consistent with 'two-state' behavior, in which there are no changes at all to the structure of the nucleosome itself; rather, there may simply be a gradual shift in some conformational equilibrium, such that at the higher salt concentrations the nucleosomes may spend more time (but still little, in absolute terms) in some alternative state, e.g. in which DNA is partially unwrapping off the surface. Such behavior is known to occur (see Section 6, below), and such equilibria would be expected to be strongly dependent on the salt concentration (Manning, 1978).

Several lines of evidence suggest that key aspects of the structure of the nucleosome as it exists in elevated salt concentrations are in fact relevant to the physiological state. The X-ray crystal structure of the nucleosome core particle, which is obtained in elevated Mn^{2+} is closely similar to the structure of the nucleosome *in vivo*, as probed by the distinctive pattern of position-dependent susceptibilities to cleavage of the DNA by DNAseI (Richmond *et al*. 1984). In addition, the structure of the histone octamer in high concentrations of the salt ammonium sulfate is similar to that of the octamer as seen in the high-resolution structure of the core particle (Arents *et al*. 1991). Taken together these two facts imply that there are no gross conformational changes of the histone octamer over this wide range in salt concentrations. As will be described below (Section 5), the DNA sequence rules and motifs that favor nucleosome formation as assessed from selection experiments using competitive

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

Table 1. *Quantitative free energy measurements*

DNA sequence	$\Delta\Delta G°$, kcal mol^{-1}
Lowest affinity known non-natural sequence	$+1.22 \pm 0.10$[a]
Bulk chicken genomic DNA	$+0.55 \pm 0.03$[b]
Chemically synthetic random DNA	$+0.5 \pm 0.13$[b]
Sea urchin 5S RNA gene clone	$+0.00$
Clone 601	-2.9 ± 0.14[c]
Round 15 pool of selected sequences	-2.8 ± 0.11[c]

$\Delta\Delta G°$ values reported relative to the 5S reference sequence (Simpson & Stafford, 1983). Values are expressed as mean $\pm$ standard error. $\Delta\Delta G°$s reflect intrinsic properties of the molecules independent of the particular conditions used in the competition experiments (Lowary & Widom, 1997, 1998; Shrader & Crothers, 1989, 1990; Thåström *et al*. 1999).

[a] Thåström *et al*. 1999; [b] Lowary & Widom, 1997; [c] Lowary & Widom, 1998.

Bulk chicken genomic DNA refers to 95% or more of the genome but omits the high-affinity tail ($\leqslant 5\%$) of the genome (Lowary & Widom, 1997). The round 15 pool of selected sequences refers to the average affinity of the entire pool of surviving high-affinity sequences after 15 rounds of physical selection. The pool at this stage comprises $\approx$ 30–50 distinct sequence families (Lowary & Widom, 1998).

reconstitution approaches include those that are actually enriched in natural eukaryotic genomes. This observation implies that not only basic aspects of nucleosome structure, but also the more subtle property of the *ordering* of the affinities of differing DNA sequences, as suggested by the competitive reconstitution experiments, are the same as for real nucleosomes *in vivo*. Perhaps most compelling, a direct analysis of rotational positioning of DNA in the nucleosome (see Section 3.2) shows this to be invariant over the range 0.0–0.8 M NaCl (Bashkin *et al*. 1993). Consistent with the hypothesis of a slowly salt concentration-dependent equilibrium involving partial unwrapping of the DNA ends, the *amplitudes* of the nuclease protection patterns slowly decrease with elevated salt concentration, but the *positioning* remains identical.

Taken together, the available evidence suggests that relative stabilities (affinities) of nucleosomes measured in competitive reconstitution experiments are in fact relevant to the structure of the nucleosome as it exists in physiological conditions. Nevertheless, because H2A/H2B heterodimers have limited affinity for DNA in the range 0.75–1 M NaCl, it is possible that the choice of DNA fragments during reconstitution may be dominated by the $H3_2H4_2$ tetramer alone, in which case the relative stabilities measured in these experiments may similarly be dominated by interactions of DNA with the tetramer.

2.4 Range of affinities

Sequences that have been studied to date in direct side-by-side comparisons span a 1000-fold range in relative affinity ($>$ 4 kcal mol^{-1} range in free energy) (Table 1). The lowest-affinity known sequences have 0.7 kcal mol^{-1} *lower* affinity than that of a chemically synthetic random DNA pool or of individuals chosen randomly from such a pool, while the highest-affinity known sequences have affinities that are 3.4 kcal mol^{-1} *greater* than that of random-sequence DNA.

The highest-affinity sequences were obtained in a selection experiment (described in more detail in Section 5, below) that explored an infinitesimal fraction of the full possible sequence

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

space for 147-bp-long sequences. It is certain, therefore, that even higher-affinity sequences exist and await discovery. At this time one cannot guess what the affinity of the optimal sequence may be.

3. Relation of nucleosome stability to nucleosome positioning

A DNA sequence-dependence to the free energy of histone–DNA interactions in nucleosome reconstitution constitutes a driving force that acts to bias nucleosomes to locate over certain DNA sequences in preference over others. This behavior results in the phenomenon known as nucleosome positioning. We distinguish translational and rotational nucleosome positioning. Translational positioning refers to the probability with which a histone octamer selects a particular contiguous stretch of 147 bp of DNA in preference to other stretches of the same length that are translated forwards or backwards along the DNA. Rotational positioning is a degenerate form of translational positioning in which a set of discrete translational positions, differing by integral multiples of the DNA helical repeat, are all occupied in preference to the set of other possible locations. (Rotational positioning is so-called because it maintains the rotational orientation of a given face of the DNA with respect to the underlying histone surface. Thus a site on the DNA that faces 'out', away from the histone octamer, in one member of a rotationally related set of nucleosome positions, does so for all members for that set of rotationally related positions.) As will be discussed in Sections 4 and 5 of this review, DNA sequences that are intrinsically bent, or are anisotropically bendable, may lead to rotationally positioned nucleosomes.

We consider a model (Lowary & Widom, 1997) in which a histone octamer is constrained to be bound to a DNA molecule of length L. We take 147 bp as a fixed amount of DNA in a nucleosome core particle, assuming for simplicity that locations of the octamer that extend beyond the DNA end (which would leave DNA-binding sites on the octamer unsatisfied) are negligibly populated at equilibrium. *In vitro*, for nucleosome reconstitution experiments, L is well-defined, and there are $L-146$ different possible locations for the octamer. For a further simplification, we take all possible locations of the histone octamer other than the specifically positioned one(s) to be equivalent in their free energy.

3.1 Translational nucleosome positioning

Translational positioning may be understood quantitatively in terms of the free energy change for transfer of histone octamer from unrestricted exploration of any of the $L-147$ available (and assumed equivalent) non-specific positions to fixed occupancy at a single specific location. We take the non-specifically bound state as the reference, and assign the free energy $+0$. Let ΔG_{net} be the free energy change for transfer to the specific position. Two terms contribute: $\Delta G_{\mathrm{net}} = \Delta G_{\mathrm{intr}} + \Delta G_{\mathrm{stat}}$, where ΔG_{intr} is the free energy change for transmuting a non-specific site into a specific one, and ΔG_{stat} is a statistical contribution to the free energy change, reflecting the fact that there may be many more possible non-specific locations than the single specific one. ΔG_{intr} is the same as would obtain if the process were to take place in standard state conditions; hence $\Delta G_{\mathrm{intr}} + \Delta\Delta G_{\mathrm{HO}}$, which is defined and also measured experimentally as the difference-free energy for reconstitution of histone octamer into nucleosomes, for two DNA molecules having identical lengths but where one contains

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

a single favored ('specific') position and the other contains no such sites. Transfer of histone octamer from free choice among any of $L-147$ non-specific positions to fixed occupancy at a single specific position reduces the entropy of the system by the amount $\Delta S_{\mathrm{stat}} = R\ln(1/(L-147))$, thereby contributing $\Delta G_{\mathrm{stat}} = -T\Delta S_{\mathrm{stat}} = -RT\ln(1/(L-147))$ to the net free energy change. With these definitions, the probability of occupancy of the specific position by the histone octamer (P) is given by Lowary & Widom (1997):

$$P = \frac{e^{-\Delta G_{\mathrm{net}}/RT}}{1+e^{-\Delta G_{\mathrm{net}}/RT}} = \frac{1}{1+(L-147)\,e^{\Delta\Delta G_{\mathrm{HO}}/RT}}, \tag{1}$$

where R is the gas constant and T the absolute temperature.

This equation has several important ramifications. DNA sequence-directed nucleosome positioning is inherently statistical, not 'precise': it is quantified by a probability of occupancy of the specific site that is always greater than 0 and less than 1. P depends fundamentally on two quantities: the intrinsic free-energy preference of histone octamer for the specific site compared to alternative sites, $\Delta\Delta G_{\mathrm{HO}}$, and the number of alternative non-specific sites, $L-147$. Given a particular DNA length, P is determined by $\Delta\Delta G_{\mathrm{HO}}$. In conditions in which nucleosomes translocate freely along the DNA, P will also equal the fraction of time that the 'specific position' is occupied in any particular nucleosome.

Equation (1) also provides a way to *measure* $\Delta\Delta G_{\mathrm{HO}}$. Direct experimental measurement of P, perhaps through quantitative nuclease protection studies for sites both internal and external to the positioned nucleosome, together with the known DNA length L, allows $\Delta\Delta G_{\mathrm{HO}}$ to be calculated.

3.2 Rotational positioning

Equation (1) is readily generalized to account for cases in which multiple specific translational positions are preferentially populated relative to the bulk 'non-specific' positions. Rotational positioning is an example of such a case, for which there happen to be constraints on the mutual spacings of the preferred positioning sites such that they all maintain the same relative orientation of one face of the DNA with respect to the histone octamer. For a DNA molecule of length L, let n_{R} be the number of positions occupied by rotationally positioned histone octamers, and let $\Delta\Delta G_{\mathrm{R}}$ be the intrinsic free energy for binding of histone octamer in any one of the set of preferentially occupied positions in a rotationally positioned system. (For simplicity, we let the same $\Delta\Delta G_{\mathrm{R}}$ apply to all of them.) Then, rotational positioning, too, is seen to be intrinsically statistical, quantified by a probability that any of the specific rotational positions is occupied, given by Lowary & Widom (1997):

$$P = \frac{1}{1+\dfrac{(L-146-n_{\mathrm{R}})}{n_{\mathrm{R}}}\,e^{\Delta\Delta G_{\mathrm{R}}/RT}}. \tag{2}$$

$\Delta\Delta G_{\mathrm{R}}$ can be measured as described above provided that one uses, for this purpose, a variant of the DNA molecule that is sufficiently short (e.g. $\approx$ 150 bp) so that essentially just a single one of the rotationally phased positions is accessible. Alternatively, a longer molecule may be used provided that the measured *apparent* $\Delta\Delta G_{\mathrm{R}}$ is corrected for the statistical factor $RT\ln n_{\mathrm{R}}$ prior to use in Eq. (2).

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

3.3 Unfavorable positioning

The discussions above have focussed on occupancy of the histone octamer at particular individual or sets of positions that are actively preferred in comparison to alternative sites which are taken to be representative of random sequence DNA. However there can also exist *unfavorable* forces that act to influence nucleosome positioning. This is most easily understood in connection with rotational nucleosome positioning which, as will be described further in Sections 4 and 5 below, is closely connected with bent DNA.

Consider a DNA sequence that, on average, has a fixed direction of curvature (i.e. tends to bend into a circular trajectory). We will see in Sections 4 and 5 that, other factors being equal, positioning such a sequence on a histone octamer, such that the DNA's intrinsic preferred curvature direction is maintained in the nucleosome, will be inherently favored over positioning on random-sequence DNA. But conversely, positioning that same DNA sequence on a histone octamer so as to force the *opposite* curvature direction (which corresponds to translocating the positioning by $n+0.5$ DNA helical turns for integer n) must come at a *greater* cost in free energy than for positioning on random-sequence DNA.

Similarly, we will see below that there exist other DNA sequence motifs that, for reasons that are not presently understood, are inherently unfavorable for incorporation into a nucleosome.

It follows that for any such sequences, it would not be appropriate to consider positioning as only a choice between one or more preferred location(s) *versus* positioning on random-sequence DNA; rather the existence of actively *unfavorable sequences* or *unfavorable positions* on particular sequences creates an additional driving force to 'sharpen up' the probability distribution for nucleosome positioning. All such effects are simply incorporated by separately counting the relative *intrinsic* free energy for positioning at each possible location, $\Delta\Delta G_{\mathrm{HO},i}$ as follows:

$$P = \frac{\mathrm{e}^{-\Delta G_{\mathrm{HO},i}/RT}}{\sum_{i=1}^{L-146} \mathrm{e}^{-\Delta G_{\mathrm{HO},j}/RT}}, \tag{3}$$

where P_i is the probability for positioning at location i along a DNA of overall length L.

3.4 Experiments

Studies have not yet been carried out to investigate quantitatively the relationship between the nucleosome positioning power of particular DNA sequences and the relative affinity of the sequences for histone binding in nucleosome reconstitution, as implied by Eqs. (1)–(3). However, recent studies dramatically reveal the expected qualitative relationship. An example of such a result is shown in Fig. 3. Two different DNAs, similar in length but differing by ≈ 1.0 kcal mol^{-1} in free energy (i.e. by approximately fivefold in affinity) are quantitatively reconstituted into nucleosomes (that is, sufficient histone octamer is supplied so as to allow essentially complete incorporation of the DNA into nucleosomes), and the resulting nucleosomal products are resolved by native gel electrophoresis. Nucleosomes formed at differing positions along the DNA result in differing electrophoretic mobilities (Dong *et al*. 1990; Meersseman *et al*. 1991; Pennings *et al*. 1991). The lower-affinity sequence [the well-studied sea urchin 5S rRNA gene nucleosome-positioning sequence, which is currently in use in many laboratories for physical studies of chromatin *in vitro* (Simpson &

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

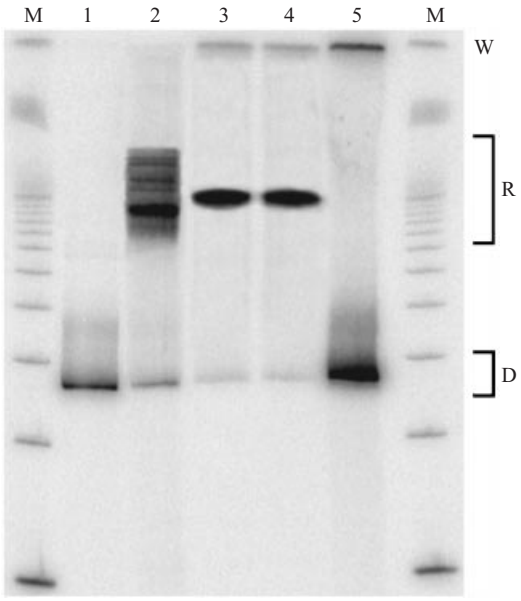


Fig. 3. Relationship between nucleosome positioning power and histone-binding affinity. A native gel electrophoresis experiment with nucleosomes reconstituted from two different DNAs is shown. Lanes M, 100 bp size standard markers. Lanes 1, 5, naked DNA samples of the relatively low-affinity sea urchin 5S rRNA gene nucleosome-positioning sequence, 256 bp in length, and the higher-affinity sequence 601.3b, 264 bp in length (Anderson & Widom, unpublished observations), respectively. Lanes 2, 3, the same two samples after reconstitution into nucleosomes. Lane 4, as in lane 3, but with an additional incubation at elevated temperature and ionic strength. Nucleosomes reconstituted at differing positions on the same DNA typically have differing electrophoretic mobilities. Note the multiplicity of bands, and the strong 'background counts' between them, in the 5S nucleosomes. These bands and the multiplicity of species having mobilities between the bands, which are manifested as 'background', reflect significant positioning at very many different positions for this sequence. In contrast, the higher-affinity sequence 601.3b yields a single strong band, implying much stronger occupancy at one preferred position. (From Anderson & Widom, unpublished observations.)

Stafford, 1983; Simpson *et al*. 1985)] gives rise to a complex mixture of separable nucleosome mobilities: the strong band reflects a single preferred position (or symmetry-related pair of positions); the set of weak bands reflect a set of particular less-preferred positions. Studies by others suggest that most of these bands correspond to rotationally positioned nucleosomes – distinct translational positions that maintain the rotational phase of the DNA. The high background between these bands may reflect significant occupancy at a large set of other positions that do not maintain even the (degenerate) rotational positioning.

In contrast, the higher-affinity sequence (601.3b) reconstitutes into a single sharp band with no detectable local background. This implies that the fractional occupancy of a single preferred position is significantly greater for this high-affinity sequence than for the 5S sequence, qualitatively consistent with the requirements of Eqs. (1)–(3).

4. Physical basis of DNA sequence preferences

How is it that differing DNA sequences can influence by factors of 1000-fold or more their affinity for interaction with histones in nucleosome reconstitution – especially given that the nucleosome structure seems to be evolved chiefly for the ability to package *any* DNA

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

sequence? The structure of the nucleosome itself suggests five different means whereby particular DNA sequence may affect the free energy for their interactions with histones. (i) Particular DNA sequences may simply make more or better 'bonds' (e.g. more H-bonds or 'salt-bridges', better H-bond geometry, better steric fit, etc. with the histone octamer and/or with solvent; an analysis must include both the solvated independent species and the solvated complex). (ii) Particular DNA sequences may possess permanent bends with magnitudes and positions appropriate for nucleosome formation. (iii) Particular DNA sequences may be more readily *bendable*, i.e. possess a lower bending force constant. (iv) Particular sequences may possess non-standard permanent helical twist with magnitudes and positions appropriate for nucleosome formation. (v) Particular sequences may be more readily *twistable*, i.e. possess a lower twisting force constant.

The rationale for the first of these five explanations is plain and will not be considered further. We focus instead on the latter four. Note that these refer to four very different properties of DNA. As with any elastically deformable material, permanent bendedness is a static, structural property of DNA that is distinct from *bendability*, which is a dynamic, mechanical property (measured by the bending force constant). That is, objects including DNA may be bent but relatively inflexible, or (on average) straight but highly flexible. (Actually, as will be discussed in more detail below, there is a subtle distinction between two *kinds* of permanent bendedness – literal static bendedness, and dynamic anisotropic flexibility.) The same ideas apply to the static structural property of the local DNA helical twist, *versus* the dynamic mechanical property of the twistability (measured by a twisting force constant).

The sequence-dependence of DNA structure and mechanics is a rich field with an extensive literature. I shall not attempt to provide a full review here but I instead refer the reader to two excellent monographs (Calladine & Drew, 1997; Bloomfield *et al*. 2000). Here I provide a brief overview, focusing in particular on questions related to nucleosome packaging.

4.1 Free-energy cost of DNA bending

We consider first the case of DNA bending. DNA is a moderately flexible polymer. It is useful to define the 'bending persistence length' for flexible polymers; this is the distance over which an initial direction of the chain tends to persist (more precisely, does persist, in the ensemble or time average). Lengths of polymer that are much shorter than the persistence length are essentially straight in the absence of exogenous forces, and are stiff, requiring large amounts of energy (work) to induce even modest bends, whereas lengths of polymer that are much longer than the persistence length are both highly bent (at any moment), and easily bendable. They behave as random coils, in which the relative orientations of widely separated segments are essentially uncorrelated.

The persistence length of arbitrary sequence DNA is $\approx$ 50 nm, or $\approx$ 150 bp (Hagerman, 1988). In this light, the nucleosome is a remarkable structure, because the 147 bp of nucleosomal DNA – roughly one persistence length of DNA, a length over which DNA would be only gently curved, on average, by equilibrium thermal fluctuations – is wrapped through almost two superhelical turns around the histone octamer. This means that much of the chemical energy of histone–DNA interactions is used up doing work against the bending force constant of DNA.

The free-energy cost of this DNA bending can be estimated using standard equations from the mechanics of elastic materials (Bloomfield *et al*. 2000). Assuming for simplicity that the

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

full length of the nucleosomal DNA is wrapped in a smooth circular trajectory, with a radius of curvature at the DNA helix axis chosen appropriately for the nucleosome structure ($\approx$ 4.4 nm), yields a remarkably high free-energy cost for DNA bending of $\approx$ 76 kcal mol^{-1}.

Evidently, if it were the case that particular DNA sequences existed that happen to favor bent, rather than straight, structures, then incorporating such sequences at appropriate positions in a nucleosomal DNA could greatly lower the energetic cost of DNA bending in nucleosome formation. *The entirety of that reduction in DNA bending energy would be manifested as increased stability* (*affinity*) *of the protein–DNA interactions in the nucleosome*.

Alternatively, there could, in principle, exist other DNA sequences that have no particular permanent bendedness, yet rather are 'softer' than usual – that is, are more easily bent, i.e. possess a lower bending force constant. This too would decrease the work required to bend the DNA into the nucleosomal conformation, with the reduction in energetic cost again translating directly into increased stability of the nucleosome.

4.2 Molecular mechanics of DNA bending and bendability

DNA bending can be described in terms of rotations about two axes that are orthogonal both to each other and to the local DNA helix axis. These two rotations are known as roll and tilt. Roll refers to a rotation of a base pair about its long axis. Such a rotation bends the local helix axis into (i.e. compresses) the major or minor groove, depending on the sign of the rotation. Bends in the direction of the major groove (compressing the major groove) are referred to as positive roll, and conversely bends that compress the minor groove are referred to as negative roll. Tilt refers to rotations about the short axis of a base pair, perpendicular to roll, and has the effect of prying the two base pairs apart at one end and pressing them together at the other end.

With these definitions, local DNA bending can be described in terms of a (roll, tilt) coordinate pair for each two consecutive base pairs (that is, a base step, if one is referring only to one strand of the duplex).

We need to distinguish several different properties: permanent (static) bendedness, isotropic (dynamic) bendability, and anisotropic bendability (which in certain respects closely resembles static bendedness). Permanent (static) bendedness is conceptually straightforward; but the distinction between bent *versus* bendable DNA is subtle and experimentally difficult. The problem is that increased bendability, even for an intrinsically straight DNA sequence, is actually manifested in solution as increased bendedness!

For an isotropically bendable sequence, inevitable random thermal fluctuations cause transient randomly oriented local bends *in random directions*. Thus, at any instant, isotropically flexible molecules will be bent. Averaged over time, they will have equal probabilities for bending in any particular direction, hence they will have zero net *vector* (fixed-directional) bending. Nevertheless, the time-averaged conformation will still be bent. That is, if one defines bending locally, and integrates this as a scalar quantity along the length of the DNA, the net scalar bendedness will be non-zero.

Anisotropic bendability is more subtle. In this case, random thermal fluctuations excite bending, but now in biased directions. The time-averaged conformations will be non-zero *both* in scalar bendedness *and* vector bendedness. Only experiments that are directly sensitive to the thermal *dispersion* in bendedness can distinguish permanent bendedness from

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

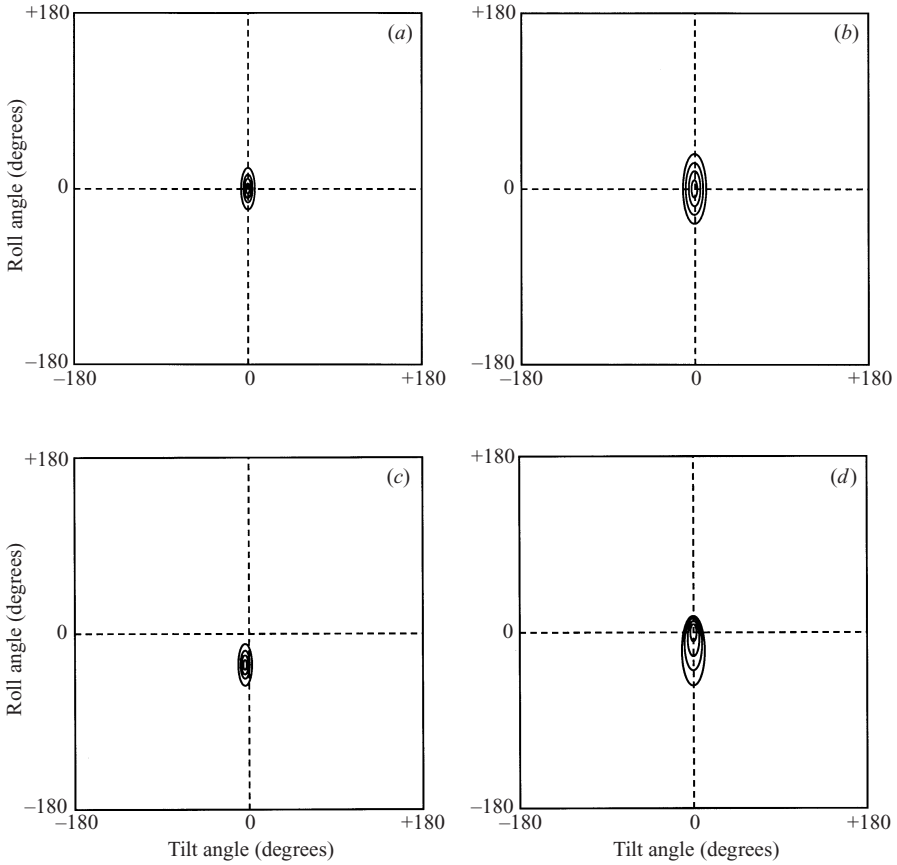


Fig. 4. Schematic illustration of DNA bendedness and bendability in the roll–tilt plane. Hypothetical contour lines of increasing free energy are shown, for combined roll and tilt rotations of an arbitrary pair of consecutive base pairs (i.e. an arbitrary base step). Available data (Olson *et al*. 1998) suggest that tilt rotations are costlier in free energy than roll rotations of comparable magnitude, hence the contours are illustrated as relatively elongated parallel to the roll axis. (*a*) A base step having little flexibility (closely spaced contours) and no net bendedness (symmetric about 0 roll and 0 tilt). (*b*) An isotropically more flexible base step. The time-averaged net vector (fixed-directional) bendedness will be 0, yet the most probable conformations at any moment are bent. Bending via base-pair roll dominates. (*c*) A base step having little flexibility but with a permanent directional bend, consisting chiefly of base-pair roll with a small contribution also from tilt. (*d*) A base step having anisotropic flexibility. The free-energy minimum is shown for an unbent conformation, but because of the anisotropic free-energy surface, the most probable conformations at non-zero temperature will have a net fixed-directional bendedness.

anisotropic dynamic flexibility. Experiments have been designed that are explicitly sensitive to flexibility itself, allowing these two distinct phenomena to be disentangled (see Section 4.3, below). But few such data are presently available.

The (roll, tilt) description of DNA bending allows for a useful representation of the distinction between bendedness and bendability, and between isotropic *versus* anisotropic bendability, in the form of free-energy contour maps (Fig. 4). Fig. 4*a* illustrates a free-energy contour map in the (roll, tilt) plane for a generic base step in DNA in which there is little flexibility and no net bendedness. The contours are centered on (roll = 0, tilt = 0), hence there is no net *static* bendedness. There is also no net dynamic *vector* bendedness because the contours rise symmetrically (for positive *versus* negative roll; and, separately, for positive

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

versus negative tilt). Nevertheless the most probable conformations will be slightly bent in random directions. The step is relatively inflexible (that is, resists bending), since the contour intervals are closely spaced – the free energy rises rapidly for small bends away from (0, 0). The figure implies an overall preference for bending via roll *versus* tilt, as the contours rise more rapidly for the tilt directions. Fig. 4*b* illustrates an isotropically more-flexible base step, corresponding to a slower rise of the free energy contours as compared to the example in Fig. 4*a*. There is no net static bendedness, and no net *vector* dynamic bendedness; but the most probable conformations at any moment are bent, and in particular are more bent than for the step illustrated in Fig. 4*a*. Bending via base-pair roll dominates. Fig. 4*c* illustrates a base step having little flexibility but possessing a permanent (static) directional bend, consisting chiefly of base-pair roll with a small contribution also from tilt. Superimposed on this static bend is additional *isotropic* (that is, isotropic for positive *versus* negative roll, and, separately, for positive *versus* negative tilt) flexibility. Finally, Fig. 4*d* illustrates a base step having anisotropic flexibility. The free-energy minimum is shown for an unbent conformation, but because of the anisotropic free-energy surface, the most probable conformations at non-zero temperature will have a net fixed-directional (i.e. *vector*) bendedness – in the direction of negative roll, for the case illustrated here.

4.3 Bent and bendable DNA sequences

DNA sequences that may be inherently straight but have finite flexibility will, at any moment, typically be bent because of thermal fluctuations. This makes it difficult to distinguish experimentally between bendedness and bendability. For example, gel electrophoretic mobility is often used as an assay for bent DNA. DNA bending reduces the net projection of the DNA's end–end vector along the direction of the electric field, thereby decreasing the electrophoretic mobility (Bloomfield *et al*. 2000). This gives rise to a mobility anomaly for bent DNA, in which the mobility is lower than that of a 'typical' DNA having the same actual length. But this behavior can occur *either* because of a net (vector) bend (whether it is a static bend or arises from anisotropic bendability), *or* because of increased isotropic flexibility.

Anisotropic bendability is particularly difficult to distinguish from static bendedness; at present little concrete information of this kind is available. Thus for the remaining discussion we will group these two phenomena together, and simply refer to them collectively, as a net, directional, 'bendedness'; and we will contrast this with 'bendability'.

Several experimental approaches have been developed to detect and measure overall flexibility ('bendability'), treated as isotropic and so distinguished from 'bendedness'; but none of these are particularly simple. The most promising include: quantitative analysis of DNA cyclization reactions (Taylor & Hagerman, 1990; Kahn *et al*. 1994; Bacolla *et al*. 1997) in conjunction with Monte Carlo simulations (Hagerman & Ramadevi, 1990; Roychoudhury *et al*. 2000); measurement of rotational relaxation times, by transient electric birefringence, for a set of DNA fragments that vary the distance (through a DNA helical turn) between a fixed bend and a test bend (Vacano & Hagerman, 1997; Zacharias & Hagerman, 1997); Fluorescence anisotropy decay (Schurr *et al*. 1992); EPR (Okonogi *et al*. 2000); and NMR on oriented samples (MacDonald *et al*. 2001).

Particular sequences that cause DNA to have net directional bends are known to exist. The best studied of these are the A-tract bending sequence motifs, originally discovered in highly

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

bent natural DNAs present in 'kinetoplasts', which are unusual DNA-containing organelles of some protists that consist of numerous small interlocking circular DNAs (Marini *et al*. 1982). The A-tracts comprise stretches of 5–6 As (in a block, along one strand), with the blocks repeating every ≈ 10 bp, in phase with the DNA helical repeat [e.g. tandem repeats of the sequence motif (CA_6TNN), for 'N' any base]. Careful studies establish that such sequences are bent because of a net *directional* bendedness rather than because of simple increased *isotropic* flexibility (Widom, 1984, 1985; Wu & Crothers, 1984; Kahn *et al*. 1994). Since the A-tracts recur in phase with the DNA helical repeat, relatively modest bends contributed by each A-tract add coherently to produce substantial overall curvature. Despite two decades of intensive studies, the molecular basis for the bending of these 'A-tract' DNAs has remained a matter of debate. However, a recent NMR structural study of A-tract bending (MacDonald *et al*. 2001), which takes advantage of a new experiment that is sensitive to vectors relative to an overall molecular orientation, may finally have settled the question. This study suggests that each repeating unit of the A-tract sequence is bent by $\approx 19°$ in the direction of the *minor* groove of the A-tract, comprising a $\approx 4°$ bend at the 5′-end of the A-tract, another $\approx 5°$ bend within the run of As itself, and a $\approx 10°$ bend at the 3′-end of the A-tract.

A later physical selection experiment designed to isolate the highly bent DNAs in an unbiased screen of a diverse random library (taking advantage of the gel mobility anomaly that arises from DNA bending) recovered sequences containing the same DNA sequence motif as in the kinetoplast DNA (Beutel & Gold, 1992). Attempts to slightly increase the stringency of the physical selection recovered *no* molecules. These findings suggest that these A-tract bending motifs may be the DNA sequence elements having the *greatest possible net bendedness*.

It is also noteworthy that the radius of curvature of the kinetoplasts, and also of non-natural A-tract-containing sequences designed to multimerize and form very small circles (Ulanovsky *et al*. 1986), is still much greater than that for DNA in the nucleosome. That is, the most-bent known sequences are much less bent as naked DNA than is DNA wrapped in nucleosomes. This fact will have important consequences in determining the dominant forces encountered when DNA is packaged in nucleosomes, and therefore in determining the DNA sequence motifs that may best facilitate nucleosomal packaging and nucleosome positioning. These questions are discussed in more detail in Section 5.

Another quite different sequence motif has also been demonstrated to contain a net (directional) bend, although one that is smaller in magnitude than the A-tract bending motif. This motif is GGGCCC, which is found to have a net bend directed toward the *major* groove, i.e. opposite to that of the A-tract (Brukner *et al*. 1991, 1993).

More recent studies suggest that the bending of the GGGCCC motif may be dependent on the presence of Mg^{2+} ions in solution (Brukner *et al*. 1994). This observation may be related to the results of high resolution X-ray crystallographic and NMR studies that detect sites of preferential occupancy by monovalent (Hud *et al*. 1999; McFail-Isom *et al*. 1999) and multivalent (Hud & Feigon, 1997; Chiu & Dickerson, 2000) cations. Such localized cations may contribute to localized DNA bending via a charge neutralization mechanism (Strauss & Maher III, 1994; Maher, 1998).

Since flexibility and bendedness are distinct properties, it follows that DNA sequence motifs responsible for greater than unusual flexibility may differ from those that cause permanent bendedness. Indeed, a number of studies have led to the identification of

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

particularly *flexible* DNA sequence motifs. Klug *et al*. (1979) were first to focus attention on special properties of the TA step in alternating TA copolymers (see also Section 5.4). More recent simulations (Lankas *et al*. 2000) and experiments (Widlund *et al*. 1999; Okonogi *et al*. 2000; Roychoudhury *et al*. 2000) show that alternating TA (i.e. also AT) copolymers exhibit relatively increased bending flexibility. Another study points to greater than average bending flexibility for repeats of CTG/CAG and CGG/GGC which are involved in triplet expansion diseases (Bacolla *et al*. 1997).

4.4 Parameter sets for prediction of DNA bending and bendability

Plainly, an important goal is to develop complete descriptions of the bending free-energy surfaces for all DNA sequences. A question that arises immediately is, How long a DNA sequence motif must be considered? The (roll, tilt) description of DNA bending implies a minimal sequence motif of a base step – that is, two consecutive base pairs. There are 10 such independent base steps. (There are 16 different dinucleotides, but 6 of these are the reverse complements of others, hence there are only 10 independent base steps.) Thus a first-order analysis of bending might focus on the base step as a fundamental structural and mechanical unit.

Since the different base steps may have different preferred conformations as well as different mechanics, it follows that for a base step comprising positions i and $i+1$ along a duplex DNA, the conformations and mechanics of the base pair at $i+1$ will be influenced to some extent by that at $i+2$, and that in turn will be influenced to some (lesser) extent by the base pair at $i+3$, and so on. Similarly the conformation of base pair i will be influenced by those at $I-1$, $I-2$, and so on. Thus higher-order analyses may need to focus on motifs that are longer than simple base steps.

There are clear examples of such behavior in the literature. Thus, for example, blocks of 5 As on one strand produce macroscopic curvature, whereas adding a sixth A but transmuting an A/T base pair in the middle of the block to a T/A base pair entirely disrupts the macroscopic curvature while at the same time eliminating the unusually long base-pair lifetimes of the A-tract as measured by thymine imino proton exchange (Leroy *et al*. 1988). Similarly, the bendedness of the GGGCCC sequence motif is not manifested in individual GG/CC or GC/GC base steps (Brukner *et al*. 1993), and base-pair lifetimes are significantly shorter in the longer (G,C)-containing blocks than in isolated G/C base pairs (Dornberger *et al*. 1999).

Given a finite number of data, one faces a choice between better defining a smaller number of parameters (say, the preferred bending and bendability of the 10 base steps), *versus* having fewer measurements contributing to each of a larger number of parameters (such as the bending and bendability of the 136 unique tetranucleotides, and so on). For these reasons, at present, studies based on experimental data focus chiefly on dinucleotide or trinucleotide (base step) parameters, while studies based on computed free-energy landscapes also focus on short motifs but extend these to slightly longer ones such as tetranucleotides, to capture the behavior of each base pair as influenced by the preceding and following base pairs, or of each base step as influenced by the preceding and following base steps.

Many such efforts have been undertaken. Here I will summarize three of the more recent of these.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

One recent approach is fully computational, based on an approximate atomic forcefield tailored for analyses of nucleic acids. This approach allows a complete analysis of the conformational free-energy surface – and hence of bending and bendability, among other properties, for the 10 dinucleotides (Packer *et al*. 2000b). A companion study examines the effects of differing flanking base pairs on the conformational dynamics of the central base step (Packer *et al*. 2000a), reporting parameters for the 136 unique tetranucleotides. Among other points, this study suggests that neighbor effects are important for some base steps and some neighbors, but not for others.

An alternative approach is based on analysis of the DNA helical parameters for the 10 differing base steps as seen in the set of all high-resolution X-ray crystal structures of protein–DNA complexes available at the time (Olson *et al*. 1998). Two key assumptions underlying this study are: (i) by focussing on protein–DNA *complexes*, concerns about the effects of crystal lattice packing on DNA conformation may be ameliorated, at least to some extent; and, (ii) by averaging over a great many differing protein–DNA complexes, any protein-specific features will be suppressed, while structural preferences inherent to the DNA will be brought out. The average of the helical parameters for each base step would then reflect the inherent structural preferences for that step, while the *dispersion* around those averages would reflect the *flexibility*.

A third approach (Anselmi *et al*. 2000) represents a hybrid between computation and experimental data. Base-step *structural* parameters are computed based on molecular force fields and are then 'tuned' to better fit experimental data on sequence-dependent gel mobility. Based on a particularly interesting assumption that *flexibility* could be related to inherent duplex *stability*, the base-step flexibility parameters are derived from intrinsic base-step thermal stabilities as measured in DNA melting studies (SantaLucia *et al*. 1996).

These disparate studies agree on certain major points. They agree that, at least for most base steps, bending via base-pair roll is greatly preferred over tilt. This is thought to be the case within the structure of the nucleosome itself, since the sharpest bends appear to occur via positive roll at particular locations within the DNA superhelix. On the other hand, the recently determined NMR solution structure of the A-tract bending motif may provide a counter-example. In this case, both roll and tilt contribute importantly to the overall bend. Moreover, while the structure at the site of greatest local bending can be described as a combination of roll and tilt, it is due more fundamentally to an abrupt reduction in inclination of the base pairs with respect to the helix axis. Additional points of major agreement between these disparate studies include: all three support the view that particular DNA sequences do exist that prefer locally more-or-less bent conformations, or that are more readily bent in one direction than another; and all three support the view that particular sequences having enhanced *bendability* do exist.

Examined in detail, however, the different analyses disagree to a significant extent for many specific cases. The causes and origins of this disagreement are not established. Some possibilities are: (i) Studies of DNA structural parameters in X-ray crystal structures of protein–DNA complexes may not be adequately free of conformational biases. For example, while all crystal structures of A-tracts either as naked DNA or as protein–DNA complexes are straight, the NMR solution structure suggests that in solution they are bent. One goal of analyzing crystals of protein–DNA complexes is to minimize contributions from crystal packing forces, yet it may be that they persist to an important extent. (ii) The possibility that DNA flexibility is *not* directly related to local melting stability as assumed in some analyses

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

(Anselmi *et al*. 2000). (iii) The detailed potentials and force fields used in mechanics computations may well contain errors that are small on absolute scales but are non-negligible in comparison to free energy differences between important DNA conformations.

Taken together, the available data support the ideas that: (i) certain sequences are, on average, bent in a particular direction, either because of a permanent (static) bendedness or an anisotropic flexibility; and (ii) certain sequences are 'softer' (more easily bendable) than others. Incorporating either kind of such sequence motifs at appropriate locations along a nucleosomal DNA could significantly reduce the energetic cost of bending the DNA into its conformation in the nucleosome, thereby increasing the stability (affinity) of the nucleosome.

Aspects of the sequence-dependence to DNA bending and bendability as they relate specifically to nucleosomal DNA packaging are discussed in Section 5, below.

4.5 DNA twisting

Just as DNA is a semi-flexible polymer with respect to its bendability, so too it is semi-flexible with respect to its twistability. Thus there exists a persistence length for DNA twisting, and a corresponding twisting force constant. (For flexible polymers, the persistence length and the force constants are closely related quantities; it happens that the literature on DNA bending usually refers to persistence length, while the literature on twisting refers to the twisting force constant or modulus.) And just as DNA in a nucleosome is significantly bent, so too it is distorted away from its canonical solution helical twist. This change in the helical twist comes at the cost of some of the chemical energy of histone–DNA bond formation, which is used to do work against the DNA twisting force constant.

The free energy of this DNA twisting can again be estimated using standard equations from the mechanics of elastic materials (Bloomfield *et al*. 2000). Such an estimate requires that we know the local DNA helical twist along the length of the nucleosomal DNA. Whereas the global bending (i.e. superhelical wrapping) of DNA in the nucleosome is a relatively low-resolution property, reasonably well-defined in the 2.8 Å X-ray crystal structure of the nucleosome, twist is a local property, less-well defined in the present structure. It may therefore be preferable at this time to use estimates for the local DNA twist in the nucleosome that are derived from analysis of hydroxyl radical footprinting studies (Hayes *et al*. 1991a, b, 1990). These studies suggest that the central ≈ 30 bp of the nucleosomal DNA have a twist of ≈ 11.0 bp/turn, whereas the remaining ≈ 60 bp on either side have helical twists of ≈ 10.0 bp/turn. This gives an overall average twist for the nucleosomal DNA of ≈ 10.2 bp/turn, consistent with the crystal structure. Estimates for the average twist of DNA free in solution vary; we adopt the value ≈ 10.5 bp/turn (Olson *et al*. 1998). Finally, estimates of the twisting force constant (twist modulus) for arbitrary sequence DNA also vary; many measurements are roughly consistent with a value for the twist modulus of 2.4×10^{-19} erg cm^{-1} (Shore & Baldwin, 1983; Roychoudhury *et al*. 2000); and in any case, for the present purposes the exact value will not matter. Using these parameters yields a free energy cost of ≈ 6.6 kcal mol^{-1} for changing the twist of DNA in the nucleosome away from its preferred conformation in solution.

This quantity is again quite large by chemical standards. It follows that, were there to exist DNA sequences that happen to favor particular helical twists that accord with the local requirements of DNA in the nucleosome, then incorporating such sequences at appropriate

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

positions in a nucleosomal DNA could in principle greatly lower the energetic cost of DNA twisting in nucleosome formation, with the entirety of that reduction in DNA twisting energy manifested instead as increased stability of the protein–DNA interactions in the nucleosome.

Alternatively, just as for DNA bending, there could in principle exist other DNA sequences that have no particularly unusual local helical twist, yet rather are 'softer' than usual – more easily twisted, i.e. possess a lower twisting force constant. This too would decrease the work required to distort the DNA into the nucleosomal conformation, with the reduction in energetic cost again translating directly into increased stability of the nucleosome.

As for the case with DNA bending, the distinction between a particular non-standard static helical twist *versus* particularly *twistable* DNA is a difficult one. Indeed, altogether less is known about both DNA twist and twistability than for bendedness and bendability, and therefore less is known about the sequence-dependences of these quantities.

In some cases DNA sequence motifs having greater than average *bending* flexibility concomitantly exhibit greater than average *twisting* flexibility (Roychoudhury *et al*. 2000), while in other cases they do not (Bacolla *et al*. 1997). And conversely one computational study describes cases in which sequences may have greater than average *twisting* flexibility with little effect on the bending flexibility (Lankas *et al*. 2000).

Perhaps the most reliable complete data are those obtained in the recent analysis of the DNA helical parameters for all base steps in the set of all high-resolution X-ray crystal structures of protein–DNA complexes (Olson *et al*. 1998). These data show that particular base steps *do* prefer larger- or smaller-than-average values for the local helical twist, and they further support the conclusion that particular sequences having enhanced *twistability* could exist.

4.6 Energetics of nucleosomal DNA packaging

We considered five differing means whereby particular DNA sequences might contribute to increased affinity for histone octamers in nucleosome reconstitution, and hence to increased nucleosome stability. The first of these (more/better bonds) is not completely independent of the other four: for example, an increased ability of a particular DNA sequence to distort in a particular way may *allow* it to make more/better bonds. The remaining four are however distinct from each other, as described above.

To rigorously quantify the contributions of the many interactions that are important for nucleosome stability will require much effort, involving many experimental cycles in which one makes mutations in the DNA *and/or in the proteins*, then determines the effects on relative free energies of histone–DNA interactions. It can also be important to determine the structures obtained with these altered DNA sequences or with mutant histones, to test whether changes to the structure can be influencing the observed energetics. Methods have not yet even been developed for quantifying the effects on nucleosome stability of changes to the histone proteins, although one can imagine analogues of the competition experiments presently used to assess the effects of changes in the DNA.

Similarly, to resolve effects that are attributable to DNA structure or mechanics more precisely into bendedness, bendability, twist, or twistability, will require detailed studies of the sequence-dependences to these properties for free DNA and for DNA in the nucleosome.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

Relatively few such data are available, hence at present we are restricted chiefly to speculation. With that disclaimer, the following views may be useful as guides for our thinking in future studies.

Each of these five mechanisms by which particular DNA sequences could facilitate their own packaging into nucleosomes may contribute simultaneously, and likely does so. But one or another explanation (perhaps a couple) may prove dominant.

Numerous aspects of the nucleosome structure suggest that it is evolved to *suppress* specific protein–DNA interactions (see Luger *et al*. 1997a; Widom, 1997a; Luger & Richmond, 1998a, b) for more detailed discussion of this point). This view accords also with the fact that the histone octamer core, which is highly conserved throughout evolution, needs to package a great diversity of DNA sequences, namely the genomes of all eukaryotes and euryarchaeota. Taken together, these points suggest that sequence-specific protein–DNA 'bonding' interactions within the nucleosome probably play a minor role in discriminating *between* differing DNA sequences.

It follows that some combination of mechanisms 2–5 are likely to contribute to a quantitatively significant extent to the sequence-dependence of the energetics of nucleosomal DNA packing; and one or another of those may dominate. Since it appears that the energetic cost of bending the DNA greatly exceeds that for altering the helical twist during nucleosomal packaging, it is likely that the *sequence-dependence* of the energetic cost for DNA bending may exceed the sequence-dependence of the energetic cost of twisting by a comparable factor. In that case, one would conclude that the sequence-dependence to the bendedness and/or the bendability of DNA will quantitatively dominate the sequence-dependence to the affinity of histone–DNA interactions in nucleosomes, and hence to the sequence-dependence of nucleosome stability.

A number of experimental studies are consistent with the view that some combination of enhanced bendedness or bendability, and/or enhanced twist or twistability, contribute importantly to the favorable packaging of DNA in nucleosomes. These will be discussed in more detail in the following section. Briefly, the key findings are: (i) DNA sequences that are found by chance to have preferred directions of curvature (either from static curvature or from anisotropic flexibility) maintain this same direction of curvature when incorporated into nucleosomes (Drew & Travers, 1985; Shrader & Crothers, 1989; Hayes *et al*. 1990; Widlund *et al*. 1999). (ii) Sequences that are designed to be bent do in fact bind with enhanced affinity to histone octamers in nucleosome reconstitution (Shrader & Crothers, 1989, 1990). (iii) Conversely, DNA sequences that were selected on the basis of their high affinity for histone octamers in nucleosome reconstitution exhibit gel mobility anomalies consistent with greater than average bendedness or bendability (Lowary & Widom, 1998). (iv) DNA sequence motifs that confer high affinity for histones in nucleosomes also exhibit greater than average bending flexibility and twisting flexibility (Widlund *et al*. 1999; Roychoudhury *et al*. 2000).

5. DNA sequence motifs for nucleosome packaging

We turn next to a more detailed discussion of particular DNA sequence rules and motifs that contribute significantly to increased histone–DNA affinity and nucleosome stability. We first consider studies of DNA sequences that are actually present in nucleosomes (Satchwell *et al*. 1986), then subsequent studies that tested and extended these ideas through the use of

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

designed bending sequences (Shrader & Crothers, 1989, 1990), and finally recent studies that take a physical selection approach to define key sequence features (Lowary & Widom, 1998).

5.1 Natural and designed nucleosomal DNAs

Initial studies relating DNA bendedness to nucleosome packaging were inspired by the finding that a particular DNA sequence had a net directional bendedness (Drew & Travers, 1984), and that this preferred directionality of bending was maintained as a predominant rotational positioning when the same DNA was organized in nucleosomes (Drew & Travers, 1985). Many such cases are now known, including the most popular nucleosome-positioning sequences presently in use for studies of chromatin structure *in vitro*, such as the 5S rRNA sequence (Hayes *et al*. 1990) and the designed pentamer TG family (see below) (Shrader & Crothers, 1989, 1990), and also the natural DNA sequence having the highest known affinity for histone octamers (i.e. the strongest known natural nucleosome-positioning sequence) (Widlund *et al*. 1997, 1999; Thåström *et al*. 1999).

These initial studies led to the suggestion that runs of (A,T) – that is, sequence stretches containing only As and Ts, in arbitrary order – prefer to bend via negative base roll, i.e. into the minor groove, so that the DNA *minor* groove lies on the *inside* of a curve, whereas runs of (G,C) prefer to bend via positive base roll (into the major groove), so that the *major* groove lies on the *inside* of the curve. These putative bending preferences would correspond to preferences for a narrow minor groove for the (A,T)-rich sequences (the minor groove is compressed by bending into it) and a wide minor groove for the (G,C)-rich sequences (the minor groove expands when the major groove is compressed by bending into it), consistent with the limited crystallographic data on DNA structures available at that time.

These findings inspired a study of the DNA sequences that are actually present in isolated natural nucleosomes (Satchwell *et al*. 1986). Fourier analysis of the spatial distributions of the 10 different dinucleotide steps revealed pronounced spectral power at a periodicity of $\approx$ 10.2 bp/turn (i.e. the average helical period for DNA in the nucleosome), with strongest signals at that periodicity obtained for the dinucleotides GC and AA (= TT), respectively. The corresponding phase angles for these two dinucleotides differ by $\approx 180°$, i.e. the signals are directly out of phase with each other. This means that *oppositely* directed local bends (presumed to be positive roll and negative roll for GC and AA steps, respectively) would add *in-phase*, to produce greater overall curvature. Thus these findings supported the earlier analysis of the directional bending preferences for (A,T)- and (G,C)-rich sequences. These findings have been confirmed and extended by related studies by others on natural nucleosome positioning sequences (Ioshikhes *et al*. 1992, 1996; Bina, 1994; Fitzgerald *et al*. 1994; Staffelbach *et al*. 1994; Bolshoy, 1995; Widom, 1996; Bolshoy *et al*. 1997; Herzel *et al*. 1999; Stein & Bina, 1999; Bailey *et al*. 2000).

Crothers and colleagues applied these findings to the *de novo* design of DNA sequences anticipated to favor nucleosomal incorporation by virtue of an inherent bendedness. Their initial studies (Shrader & Crothers, 1989) investigated polymers containing direct repeats of 20-bp-long oligonucleotides in which (A,T)-rich stretches alternated, 5 bp out of phase, with (G,C)-rich patches. The designed sequences behaved as predicted: they bound to histones in nucleosomes with higher affinity than any natural sequence known at that time, and they exhibited rotational positioning (and not a specific translational positioning) consistent with their designed curvature. Subsequent studies investigated the sequence and spacing

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

preferences in more detail (Fitzgerald & Anderson, 1998, 1999; Shrader & Crothers, 1990). These studies demonstrated several significant points. First, all (A,T)-containing sequence motifs are not equivalent; and similarly all (G,C)-containing motifs are not equivalent. Replacing an individual (A,T) stretch with a different (A,T)-containing one, or a (G,C) stretch with a different (G,C)-containing one maintained the rotational positioning of the DNA but was accompanied by a significant cost in free energy (representing the effects of such a change in all five copies of the oligonucleotide at once), which for some changes eliminated nearly all of the free energy by which the originally designed 20-mer polymer was preferred over bulk genomic DNA. Secondly, replacing an individual (A,T) stretch with a (G,C) one, or vice versa, in each 20-mer, maintained the same rotational positioning – i.e. forced that stretch to bend in the direction opposite to its preferred orientation – and was accompanied by a significant cost in free energy, typically ≈ 0.7–1.6 kcal mol^{-1} (again representing the effects of such a change in all five copies of the oligonucleotide at once), roughly one half the total free energy by which the originally designed polymer was preferred over bulk genomic DNA. Thirdly, analysis of the effects of changed lengths for the central oligonucleotide showed a preference for a slightly increased length (21 bp, as opposed to 20 or 22 bp) in the center position, consistent with a locally decreased helical twist (e.g. 11 bp/turn compared to 10 bp/turn for the flanking regions) deduced from hydroxyl radical footprinting studies (Hayes *et al*. 1991a, b, 1990).

5.2 New rules and reagents from physical selection studies

The fact that one can successfully design a DNA sequence for high-affinity binding and rotational nucleosome positioning implies that at least some key aspects of DNA bending are captured by these rules regarding the bending-directional preferences of (A,T)- and (G,C)-containing sequence motifs. Remarkably, however, these ideas are *not* consistent with the real base-step bending preferences as manifested in the set of X-ray crystal structures of protein–DNA complexes, nor with the dinucleotide conformational maps derived from molecular mechanics computations, nor with the notions of a relationship between flexibility and melting stability (at least, given the actual structural and thermodynamic parameter sets used) (Anselmi *et al*. 2000).

For example, as regards the parameters deduced from the X-ray crystal structures (Olson *et al*. 1998): (i) based on the X-ray structural data, *none* of the base steps actually prefer negative values for the roll angle. Rather, the preferences range from small to large positive roll angles. Yet, for uniform circular bending (a rough approximation to the nucleosomal DNA organization) and assuming that base-pair roll is the predominant bending mode, *half* of all base steps must have negative roll in the nucleosome. (ii) The base step exhibiting the preference for the largest positive roll angle is CG, while that exhibiting the preference for the *smallest* (i.e. most negative, albeit still positive) is GC, but these are both members of the (G,C)-rich set thought to favor bending into the major groove. (iii) Both AA and AT prefer smaller (less positive, more negative) values for roll than TA, suggesting that they should be preferred in the (A,T)-rich regions, whereas the *opposite* actually obtains (Lowary & Widom, 1998; Widlund *et al*. 1999).

Existing data are similarly inconsistent with the results of the molecular mechanics computations (Packer *et al*. 2000a, b). For example, the TA step is found (in these computations) to favor large positive roll, yet it is favored in nucleosomes at sites of *negative*

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

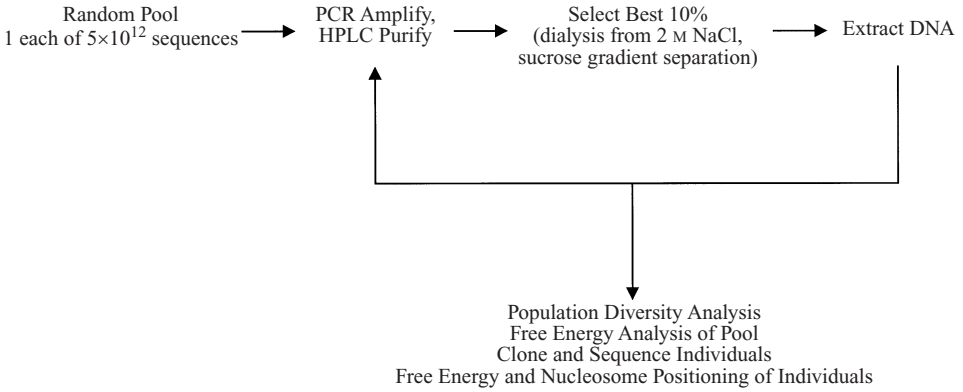


Fig. 5. Schematic illustration of the physical selection procedure used for isolation of DNA sequences having highest affinity for histone octamer in nucleosome reconstitution. (Adapted from Lowary & Widom, 1998.)

roll, and in X-ray crystal structures it prefers *small* positive roll values – that is, it is not among the steps favoring *either* extreme of roll angle. An alternative possible role of TA, namely, that it might contribute to nucleosome packaging by virtue of a particularly great *flexibility*, is also in conflict with the results of these computations, which imply that CA (= TG) has significantly greater flexibility, yet CA (and TG) are *not* among the motifs selected out in a physical selection for high-affinity binding (see below), whereas TA *is*.

Similar problems obtain with the assumptions and findings of the hybrid approach described above (Anselmi *et al*. 2000). For example, the TA step is again considered to prefer large positive roll (the largest for all 10 unique steps), yet, as mentioned above, is found in nucleosomes at sites of *negative* roll. Other issues concerning the relationship between relative melting stability of the differing base steps and their flexibility are discussed below.

At this point it simply is not clear which of the original design premises are in fact correct and responsible for the sequences' higher affinity.

For these reasons, we set out to search for DNA sequence rules governing histone–DNA interactions in a way that was independent of any preconceived notions as to what the nature of those rules may be (Lowary & Widom, 1998). We carried out a physical selection starting from a highly diverse chemically synthetic random DNA sequence library (1 molecule each of 5×10^{12} different chemically synthetic random 220-mers), and selected out those DNA sequences having the highest affinity for histones in nucleosome reconstitution. Our experimental strategy is illustrated schematically in Fig. 5. We iteratively selected the highest-affinity 10 % of the sequence population, reamplifying the winners and subjecting them to the next round of selection. The selection itself is equivalent to the competitive reconstitution procedures used for relative free energy measurement (see above). Briefly, 1 relative mol of DNA molecules compete with each other for binding to limiting quantities (0.1 relative mols) of histone octamer, during dialysis from concentrated salt solution (in which histone and DNA have negligible affinity for each other) down to subphysiological salt concentrations, forming nucleosomes and freezing them in. This strategy guarantees as a matter of physical principle that the highest-affinity sequences (the top 10 %, in this case) win the competition for the limiting histone octamers during each cycle of selection.

As the selection proceeded, the diversity of the population was progressively reduced, at close to the expected rate of tenfold per round, while the average free energy of the pool (as

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

measured by competition experiments with the entire pool labeled as tracer) progressively improved. For the final stages (rounds 12–15, approximately 200 differing sequence families), the free energy differences between individual survivor sequence families became small in comparison to thermal energies ($RT \approx 0.6$ kcal mol^{-1}), so the selection procedure was halted. At this point, the pool consisted of hundreds (at round 12) or tens (at round 15) of distinct sequence families. The (negative) free energy of the pool was 2.8 kcal mol^{-1} greater (i.e. $\approx$ 100-fold higher affinity) than that of the 5S sequence, which itself is somewhat higher than that of arbitrary sequence DNA.

Individual survivors were cloned; approximately 100 were sequenced, and the set of sequences was culled to yield a single representative from each of what proved to be 41 distinct sequence families. The free energies of several of these were determined, to confirm that the cloned molecules were representative of the pool. Finally, a variety of computational approaches were taken to ask in general ways, What is not random about the selected sequences? That is, What DNA sequences, rules or motifs are enriched by the selection for high-affinity binding? An important feature was that the computational approaches examined both local sequence features ('motifs') but also non-local ones, such as sequence periodicities, and correlations of one motif with another over long separations.

Periodic signals enriched in the selected sequences were evaluated by Fourier transformation. Fig. 6 shows the resulting power spectra obtained with four of the dinucleotide encodings – those for the GC and AA (= TT) steps (which were reported to give the strongest $\approx$ 10 bp periodic signals in the studies on natural nucleosomal DNA) and two others. The experimental signals are compared to the mean and standard deviation of random expectation.

Several aspects of these results are striking. The spectrum for AA (= TT) reveals a clear peak at a periodicity of $\approx$ 10 bp (6.64 per bp), reminiscent of the results obtained from isolated natural nucleosomes. Unlike those earlier observations, however, the selected sequences give an additional clear peak at twice the reciprocal spacing, 12.64 per bp. The simplest interpretation of this peak is that it represents the second-order harmonic of the fundamental 6.64 per bp periodicity. Its presence and strength would then be an indication of a higher degree of order in the underlying signal in our selected sequences than in the natural nucleosomal sequence sample. Analogous results although with weaker, noisier signals, are obtained for GC.

Strikingly, however, other dinucleotides give much stronger signals than any previously identified in natural DNA. The strongest of all are obtained for the dinucleotide TA: four strong peaks are found, at spacings of $\approx$ 6/64, 12.7/64, 19/64, and 25.3/64 per bp. These peaks are interpreted as the fundamental and the first three harmonics of a fundamental periodicity that can be estimated from the highest harmonic: [(25.3/4)/64] per bp, or $\approx$ 10.1 bp in real space. This implies a substantially greater degree of periodic ordering in the placements of TA dinucleotides than for AA or GC.

It is instructive also to consider the actual real-space distributions of these various dinucleotides. Such real-space distributions are usefully expressed as correlation functions, which in essence answer the following question: Given any particular sequence motif (dinucleotide in this case) at an arbitrary location, what is the probability of finding the same or any other motif at a distance, λ, for all possible distances, λ? If the two motifs are identical, these are autocorrelation functions, which are Fourier transform pairs of the corresponding power spectra. However, the two motifs may also differ (in which case these are cross-

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

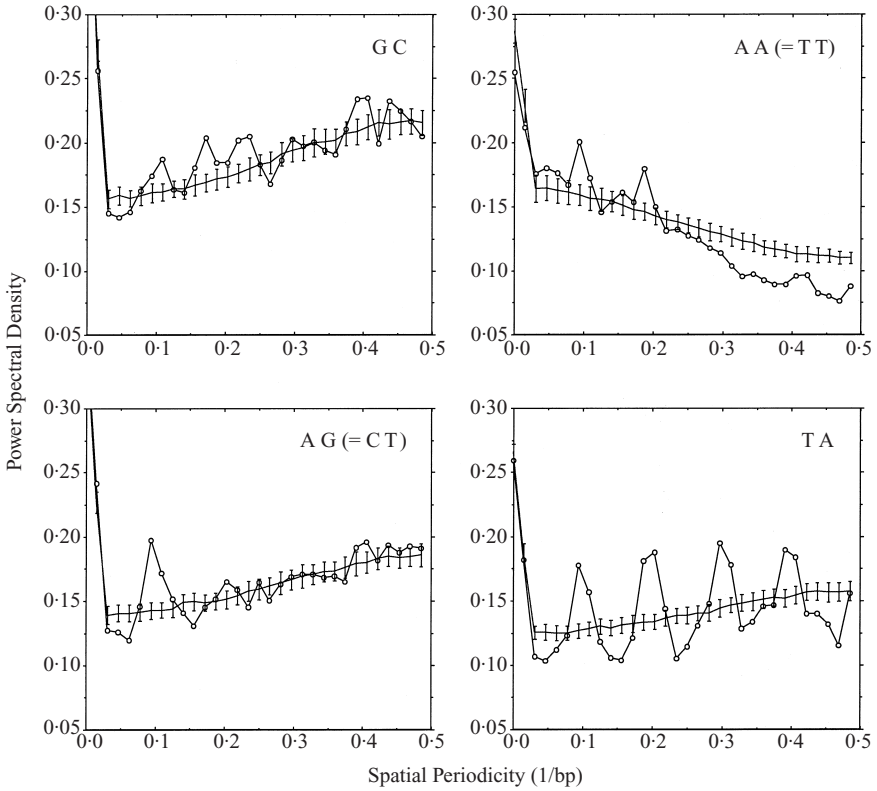


Fig. 6. Fourier transform analysis of selected sequences. Solid lines and (○) data-points: experimental transforms; fainter lines and data-points with error bars: mean values and standard deviations from transforms on randomized versions of the same sequences. *Top panels*, the spectra for the two dinucleotides reported to exhibit the strongest periodic signals in natural nucleosomes (see text); *bottom panels*, two of the much stronger periodic signals obtained in the selection experiment on chemically random DNA. Since the reverse complements of every sequence are included in the calculations, the transform for any particular encoding is identical to that for the reverse-complement encoding. (From Lowary & Widom, 1998.)

correlation functions). Cross-correlation analyses allow the relative locations of differing signals to be elucidated even in the absence of a fixed origin (reference point), e.g. even without knowing exactly where along the selected DNA sequences the optimal nucleosomal regions are located. Similar information on relative positions could in principle be obtained from Fourier analysis (in the form of the phase angle). But such signals from differing sequences could only be averaged if the sequences were aligned with respect to a common origin, which requires that accurate nucleosome mapping information be available for every sequence to be analyzed. At present such information is not available.

Auto- and cross-correlation spectra for the dinucleotides whose power spectra are illustrated in Fig. 6 are shown in Fig. 7. The graph for TA:TA (a self- or autocorrelation function) is interpreted as follows: given a TA dinucleotide at any location in a selected sequence, there is a very high relative probability of finding another TA at distances of 10 and 20 bp further along. The autocorrelation function for AA is similar but with less-statistically significant probabilities, consistent with the weaker peaks and lack of high-order harmonics in the power spectra. The autocorrelation function for GC also exhibits weak but

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

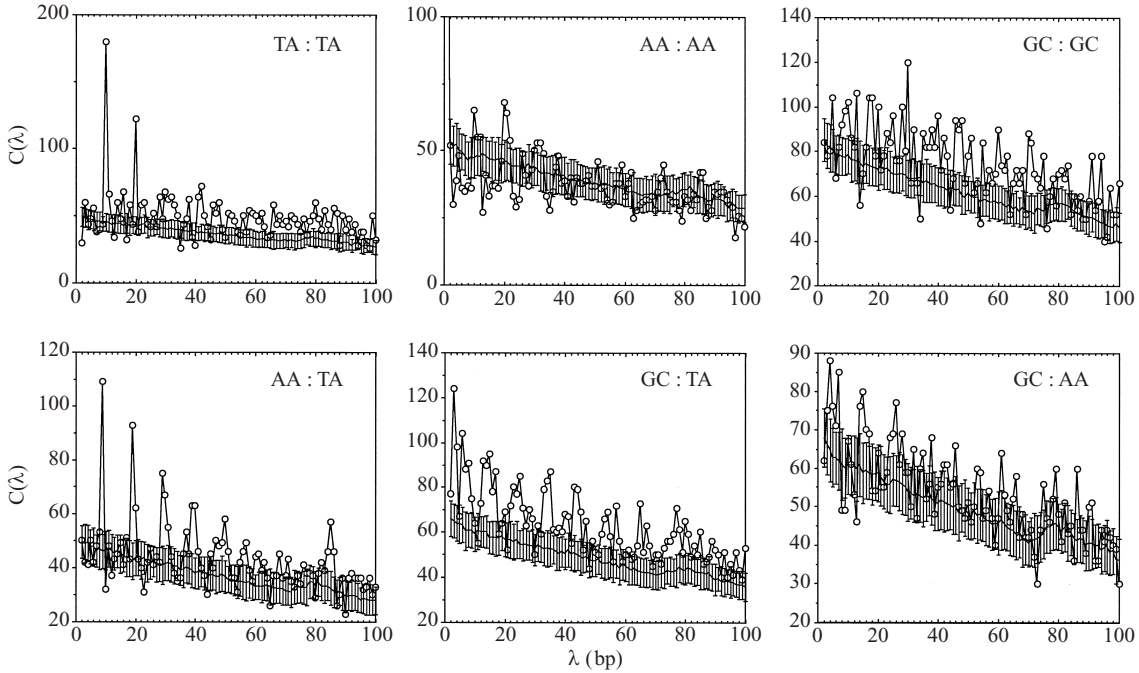


Fig. 7. Real-space correlation analyses of the spatial relationships between dinucleotides in the selected sequences. The spatial relationships between all pairs of dinucleotides in the non-redundant database of 41 high-affinity selected sequences and their reverse complements were determined, for all separations (λ) from 1 to 100 bp. Shown are auto- or cross-correlation functions corresponding to the power spectra illustrated in Fig. 6. (Note that autocorrelation functions are Fourier transform pairs of the power spectra shown in Fig. 6.) Solid lines and (○) data-points: experimental transforms; fainter lines and data-points with error bars: mean values and standard deviations from transforms on randomized versions of the same sequences. The graphs are interpreted as follows. For two dinucleotide sequence motifs *ab* and *xy*, the cross-correlation function C(λ) for *ab*:*xy* is the relative probability of finding an *xy* dinucleotide a distance λ further along the sequence, given an *ab* dinucleotide at any arbitrary location. [The apparent discontinuities in the correlation functions at λs $\approx$ 60–80 bp are due to the presence of two 6 bp fixed-sequence elements that join together the three large (72, 76, and 72 bp) random-sequence blocks from which the random library was constructed. These fixed-sequence elements are made 'invisible' for these calculations, hence no correlations can either begin or end in them. Consequently there are fewer correlations in that distance scale than would otherwise occur, since many such correlations would begin or end in one or both invisible elements.] (Unpublished data from P. T. Lowary and J. Widom; see also Lowary & Widom, 1998.)

very noisy peaks at ≈ 10 (but not 10 exactly), ≈ 20 (but not 20 exactly), 30 and 40 bp periodicities. The graph for the AA:TA cross-correlation function is interpreted as follows: given an AA dinucleotide at any location in a selected sequence, there is a very high relative probability of finding a TA dinucleotide at distances of 10, 20, 30, 40 and 50 bp further along the sequence. The remaining two cross-correlation functions show that GC is strongly correlated with TA or AA at multiple distances separated form each other by 10 bp spacing, *but these are 5 bp out of phase with the corresponding AA:XY correlations* – that is, they exhibit peaks at 5, 15, 25, 35, 45 and 55 bp spacings!

These and other analyses on the selected sequences revealed many additional sequence rules and motifs that were enriched during the physical selection and thus evidently contribute to the higher affinity. These include other periodic signals, other correlated dinucleotides and trinucleotides, including ones strongly correlated at long distances, and short-range motifs such as statistical over-representation of particular *n*-mers (from $n = 2$–10).

In summary, key findings from the selection experiment include the discovery of a large number of DNA sequences having far greater affinity (Table 1, Fig. 1) and correspondingly greater nucleosome positioning power (Fig. 3) than previously known sequences. These sequences represent valuable new tools for studies of chromatin structure and function *in vitro*. A second key finding from the selection experiment is that the higher affinity and improved positioning power of the selected sequences is achieved through a set of sequence rules and motifs that includes ones that were previously known, together with a larger set of new ones, including many having much greater statistical and energetic significance.

5.3 Molecular basis of DNA sequence preferences

We reason above that some combination of unusual DNA structural and/or mechanical properties is likely to dominate the *sequence-dependence* to the energetics of nucleosomal DNA packaging. If this view proves to be correct, then this set of selected sequences could represent a valuable library of diverse DNA sequences with which to explore the relationship between sequence and DNA structure and mechanics.

Several lines of reasoning suggest that aspects of DNA structure and mechanics are in fact contributing importantly to the nature of the selected sequences. The case for unusual DNA bendedness/bendability is clearest. The cross-correlation analyses show that AA and TA occur periodically and are 'in phase' with themselves and with each other; CG and GC are similarly periodic and 'in phase' with themselves and with each other; yet AA and TA are both out of phase with CG and GC. This pattern of relative locations is strikingly consistent with the rule of thumb discussed above, that (A,T)-rich and (G,C)-rich sequences have oppositely directed bending preferences, with (A,T)-rich regions preferred at sites of bending into the minor groove via negative base-pair roll, and (G,C)-rich regions preferred at sites of bending into the major groove via positive base-pair roll. Thus the simplest interpretation is that the periodic placement of short-sequence motifs containing these elements helps the DNA bend around the histone core. Consistent with this view, native gel electrophoresis of several of the selected sequences revealed that they all exhibited anomalously slow electrophoretic mobility (in comparison to the mobility expected given their true length as determined by DNA sequencing). Decreased electrophoretic mobility in such studies is thought to reflect a decreased mean-squared end-to-end distance, consistent with a permanent bendedness or increased isotropic or anisotropic flexibility.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

The molecular basis of the sequences' increased bendedness/bendability itself is not yet known. (It is not even known which of the two broad effects – bendedness, or bendability – is chiefly affected.) But whatever the mechanism of action, it is already clear that the rules themselves are much more complicated than simple (A,T)- and (G,C)-richness. For example, the steps AA and TA are strongly periodic, whereas AT is not. It has often been supposed that 'YR' (pyrimidine followed by purine) steps are particularly 'flexible'; all YR steps exhibit greater than average flexibility as measured by the dispersion of DNA helical parameters such as twist and roll angles for these steps in the large database of solved structures (Olson *et al*. 1998). Again, however, the results of the selection suggest that all YR steps are not equivalent for histone binding and nucleosome formation. For example, the periodic signal from TA greatly exceeds that from CA (= TG) or CG, and TA dinucleotides are over-represented while CA and TG are actually *under*-represented. At present the action of the sequences discovered here cannot be understood in terms of preferred base-pair helical bending parameters as tabulated from crystal structures of protein–DNA complexes (Olson *et al*. 1998), or as computed using molecular mechanics approaches (Packer *et al*. 2000a, b), or as used in the hybrid approach based on base-step geometric preferences supplemented with thermal melting stabilities (Anselmi *et al*. 2000).

It is likely that the lengthscale over which the special structures and mechanics of these motifs becomes manifest is greater than a single base step. For example, the TA dinucleotide occurs with greater than random probability, and is strongly periodic at a ≈ 10 bp period, yet it does not occur in isolation. Rather, it is found with much greater than random probability to occur inside longer motifs, for example, in the tetranucleotide CTAG, and up to a number of particular 10-mers (which are the largest lengthscale contiguous element we examined).

Finally, it remains likely that appropriate sequence-dependence local DNA twist and increased twisting flexibility may also be contributing to the favorable free energy of the selected sequences.

Two recent studies investigated the DNA mechanics of a natural high-affinity nucleosome positioning sequence selected from the mouse genome (Widlund *et al*. 1997) (see Section 5.6, below). Consistent with the ideas outlined above, the DNA sequence motif that is responsible for high affinity in nucleosome reconstitution exhibits *both* increased bending and twisting flexibility (Widlund *et al*. 1999; Roychoudhury *et al*. 2000).

5.4 Special properties of the TA step

The TA step itself stands out as the most strongly selected feature in the selected sequences. It is the most highly over-represented dinucleotide, the most strongly periodic at any periodicity, and the most strongly correlated among all self- and cross-correlations at all lengthscales. These observations led us to re-examine the designed sequences studied by Crothers and colleagues (Shrader & Crothers, 1989, 1990). One sees that even in those sequences, over one half of their net favorable free energy can be attributed to TA steps that were present in those sequences. Moreover, those TA steps were present at what we now recognize were suboptimal spacings of 9, 11 and 20 bp, whereas the selection studies suggest that uniform spacing of 10 bp may be preferred (except possibly across the middle of the nucleosome, where 11 bp may be preferred – see below). Thus optimally positioned TA steps might potentially have contributed even more favorable free energy to those sequences.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

There were many clues that the TA step might possess unusual structural or mechanical properties, from diverse studies carried out over more than two decades. Klug and colleagues originally proposed that the TA step in alternating $(–T–A–)_n$ copolymers was inherently unstable, sacrificing relatively poor base-stacking interactions at the 5′–T–A–3′ step in order to further improve the relatively good stacking interactions at the 5′–A–T–3′ step (Klug *et al*. 1979). They noted at the time that the TA step contributes an anomalously low stability increment to the local melting temperature of DNA. [More recent studies (SantaLucia *et al*. 1996) show the TA step to have the lowest overall thermal stability, although to differ only slightly from that of AT; or to have a significantly lower thermal stability than AT as well as all other steps (Johnson *et al*. 1998)]. Klug and colleagues further noted that the thymine imino proton exchange rate is anomalously rapid, implying an anomalously high frequency of base-pair opening fluctuations. Subsequent studies revealed many additional unusual features. An internal TA step eliminates the remarkable bending properties of A-tracts, and simultaneously eliminates the unusually low base-pair opening rates characteristic of these A-tracts (Leroy *et al*. 1988). X-ray crystallographic and NMR structural studies reveal unusual structural features for the TA step in the context of the same flanking sequences enriched in the selection experiment (e.g. TCTAGA). X-ray crystallographic studies of the sequences CTCTAGAG (Hunter *et al*. 1989) and GTCTAGAC (Cervi *et al*. 1992) show them to be overall A-like, but with extensive purine–purine overlap at the central TA step [unlike that proposed for the alternating B-form (Klug *et al*. 1979)] accompanied by unusual α and γ backbone torsion angles at the A nucleotide, which produce an extended all-*trans* sugar–phosphate backbone conformation. NMR structural studies (Lefevre *et al*. 1987) of the sequence CGTACTAGTTAACTAGTACG reveal patterns of sequence-dependent twist and rise per base pair for both sets of CTAG motifs that closely resemble those seen in the two X-ray structures, suggesting that the unusual conformations of these motifs seen in the crystals may also obtain in solution. The TA step is also favored in distorted but non-contacted DNA sequence regions of some protein–DNA complexes (Lewis *et al*. 1996; Horton & Perona, 2000) and is featured in the TATA promoter element which (while being contacted by protein) is radically distorted in the TBP–DNA complex (Kim *et al*. 1993a, b). Note however that the TA steps in these latter examples are sites of large *positive* roll, whereas the (A,T)-containing (including AA and TA steps) stretches in nucleosomes appear to be favored at sites of *negative* roll. Perhaps the TA step is particularly flexible with respect to both positive and negative base-pair roll; or perhaps the TA step facilitates some other local distortion of helix (e.g. unwinding) which might allow bending via positive or negative base roll to be distributed over more base pairs while adding coherently for bending in the same direction.

At this time, the most plausible correlation for special mechanical properties of the TA step seems to be with thermal stability (Johnson *et al*. 1998). It is certainly plausible that severe DNA bending could be facilitated by unstacking a base pair, so as to essentially move obstructing atoms out of the way. Consistent with this idea, DNA containing (single-stranded) bulged loops (Kahn *et al*. 1994) or abasic sites (Ayadi *et al*. 2000) does exhibit increased bending flexibility compared to generic B-form DNA. Similarly, enhanced bending flexibility correlates with shortened base-pair lifetimes as measured by imino proton exchange (Leroy *et al*. 1988; Dornberger *et al*. 1999). And TA steps in nucleosomal DNA are also sites of hypersensitivity to permanganate, which is believed to react only with unstacked base pairs (Fitzgerald & Anderson, 1999).

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

But even this idea, that the special properties of the TA step are attributable to a particularly great tendency to unstack, has significant caveats. First, while one parameter set places the thermal stability of TA distinctly lower than that of the next higher low-stability base step (AT) (Johnson *et al*. 1998), another closely related study has the thermal stabilities of these two steps nearly identical (that of TA is lower by an insignificant amount than for AT) (SantaLucia *et al*. 1996), whereas the AT step lacks the special properties of the TA. Secondly, molecular mechanics analyses that take thermal stability into account along with other forces, do *not* point to a uniquely unstable behavior for the TA step (Packer *et al*. 2000a, b).

Taken together, the available data demonstrate that the TA step has unusual structural and/or mechanical properties, and that these properties are critically involved in biological function in multiple instances. But the nature of these unusual properties remains unclear.

5.5 Unfavorable sequences

In addition to revealing a rich set of sequence motifs that favor the incorporation of DNA into nucleosomes, the selection experiment also revealed sequence motifs that apparently *disfavor* nucleosomal packaging. The dinucleotides AT and CA (= TG) are under-represented in the selected sequences. The simplest interpretation of this finding is that these dinucleotides are actively selected against because they disfavor nucleosomal organization of the DNA.

This possibility was directly tested in a negative selection experiment on non-natural DNA. This study selected, from a large pool of non-natural DNA, those individuals that most disfavor incorporation into nucleosomes (i.e. that best resist nucleosomal incorporation) (Cao *et al*. 1998). This study revealed a set of sequence motifs that disfavor nucleosome formation. The most prevalent of these sequence motifs include repeats of TGGA, or TGA, or CA, all of which include the AT, CA, and TG steps which were under-represented in the positive selection. Taken together, these studies provide clear evidence that these dinucleotides do in fact disfavor nucleosomal incorporation of the DNA molecules that contain them.

The possible existence of another class of DNA sequences that may disfavor nucleosomal incorporation was suggested by earlier findings that stretches of poly(dA:dT) (that is, stretches with As on one strand, Ts on the other) were preferentially excluded from the central 5–6 helical turns of the nucleosomal DNA (Satchwell *et al*. 1986). Subsequent studies suggested that these sequence elements actively disfavor incorporation into nucleosomes. For example, a 40 bp stretch of poly(dA:dT) roughly centered in the nucleosome destabilizes the nucleosome by ≈ 0.8 kcal mol^{-1} (Hayes *et al*. 1991a). One study cast doubt on this conclusion (Mahloogi & Behe, 1997), but this probably reflects a failure of their system (which utilized exchange methodology for competitive nucleosome reconstitution) to reach equilibrium. In recent work we showed that incorporating an A_{16} tract into nucleosomal DNA at a position roughly midway between the DNA end and the middle of the nucleosome destabilizes the nucleosome by a small but significant amount, ≈ 0.4 kcal mol^{-1} (Anderson & Widom, 2001).

Two points concerning the unfavorable sequence elements detected in selection experiments are noteworthy. First, the unfavorable sequence motifs deduced from both the positive and negative selection experiments described above include the YR steps CA and TG, which are often considered to be particularly 'flexible' and hence would ordinarily have been thought likely to contribute *favorably* to nucleosomal packaging. Either the premise that

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

these sequence motifs are especially flexible is simply wrong, or else some other aspects of DNA mechanics are dominating the sequence-dependence to nucleosome stability.

Secondly, the actual unfavorable free energies of these under-represented or negatively selected sequences are modest. The free energy of the most unfavorable sequence isolated is only 1.2 kcal mol^{-1} worse than the free energy of the *favorable* 5S gene-positioning sequence (Table 1; Thåström *et al.* 1999). Similarly, the poly(dA:dT) elements lead to only a very modest destabilization, <1.0 kcal mol^{-1}. Such modest free energies can provide biologically significant biases to nucleosome stability and positioning – leading, for example, to a few-fold lower equilibrium probability that a particular DNA region would be found inside a nucleosome, or to a few-fold increase in the equilibrium accessibility of nucleosomal DNA target sites (see below). However, such modest free-energy penalties certainly do not suffice to keep a region of DNA literally nucleosome free.

5.6 Natural genomes

To what extent are the genomes of histone-containing organisms (eukaryotes and euryarchaeota) evolved or constrained to facilitate the packaging of their own genomic DNA into nucleosomes (Widom, 1985)? Many natural eukaryotic DNA sequences having particularly high affinity for histone octamer in nucleosome reconstitution, and correspondingly high nucleosome positioning power, have been found by chance. In addition, as summarized above, analysis of DNA sequences contained in isolated natural nucleosomes revealed a number of non-random sequence features, such as the ≈ 10-bp periodicities of AA and GC dinucleotides. These observations imply that particular natural genomic regions having properties favorable for nucleosome organization do exist, but they leave open the question of whether such regions are statistically significant in the context of entire genomes or whether they simply exist by chance; and they leave open the question of the energetic significance of the sequences.

The competitive nucleosome reconstitution experiment allows one to quantify the extent to which natural genomic DNA contributes to the favorable free energy of its nucleosomal packaging (Lowary & Widom, 1997). A pool of random genomic DNA fragments was allowed to compete against a pool of chemically synthetic random DNA, and also against randomly chosen individual random sequences. The bulk genomic DNA was preferred over random sequences by a negligible amount, at most 0.1 ± 0.1 kcal mol^{-1}. The competitive reconstitution approach used in this study resulted in the incorporation of only 10% of the DNA into nucleosomes. It follows that the interpretation of this experiment is that at least 95% of genomic DNA has a free energy of reconstitution that differs negligibly from that of typical synthetic-random DNA. Evidently, the bulk of the eukaryotic genome is *not* evolved to aid substantially in its own packing at the level of individual nucleosomes.

While the bulk of eukaryotic genomic DNA is, evidently, not evolved or constrained to facilitate its own nucleosome packaging, it remains a real possibility that a small subset of genomic sequence regions ($< 5\%$) *may* be so evolved and constrained. Fourier transform analysis of complete eukaryotic genome sequences provides a computational approach to this question. Power spectra computed from sequences encoded for the ten differing dinucleotides reveal clear peaks of spectral at a ≈ 10-bp period for several of the dinucleotides. These signals are features of eukaryotic but not of eubacterial genomes (Widom, 1996; Herzel *et al.* 1999). The strongest ≈ 10-bp periodic signals in the natural genomes are obtained for the

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

dinucleotide AA (= TT). The power spectra reveal that the distribution of AA dinucleotides differs significantly from random expectation over all lengthscales (all frequencies), and exhibits clear $\approx$ 10 bp peaks that are statistically significant both as measured by standard deviations from random expectation, and also in comparison to the fluctuations in spectral power in nearby frequency regions. The fact that these same signals are obtained in the selection experiment on (non-natural) chemically random DNA (Lowary & Widom, 1998) strongly supports their attribution to the requirements of nucleosomal packaging. Consequently these findings imply that particular regions of eukaryotic genomes are in fact evolved and constrained to facilitate their own nucleosomal packaging, or, equivalently, to bias the probabilistic positioning of histone octamers along their length.

Many additional complete genomes have subsequently been sequenced. Among the discoveries that have resulted, it is found that the kingdom of archaea comprises two subgroups: euryarchaeota, which – remarkably – contain histone-like proteins, and chrenarchaeota, which do not. The euryarchaeal histones fold into the characteristic histone dimer fold, and a pair of dimers associate with each other and with $\approx$ 60 bp of DNA to form a structure that, as probed in solution studies, appears to be closely similar to the $H3_2H4_2$ tetramer and its associated DNA from the eukaryotic nucleosome. These euryarchaeal genomes contain the same periodic genomic DNA sequence signatures attributable to nucleosome packaging as found earlier for the eukaryotic genomes (Herzel *et al*. 1999; Bailey *et al*. 2000), suggesting that the constraints on the evolution of eukaryotic genomic DNA sequences likely originated in the archaea, prior to the splitting off of the eukaryotic lineage.

A parallel experiment used physical selection approaches to isolate the DNA sequences from a synthetic random library having the highest affinity for euryarchaeal histones in the formation of euryarchaeal nucleosome-like structures (Bailey *et al*. 2000). Analysis of non-random sequence features in the selected sequences revealed a set of rules that are closely related to those discovered earlier in the selection with eukaryotic histones. These findings are consistent with a common, archaeal, origin to the presently observed constraints on eukaryotic genome sequences. Furthermore, they support the view that these sequence constraints facilitate nucleosomal DNA packaging through effects on the structure and mechanics of DNA rather than through DNA sequence-specific contacts (e.g. base-specific contacts), since the archaeal histones themselves are quite divergent in sequence from the eukaryotic.

A remaining question is: To what quantitative extent do the regions of eukaryotic genomes that are constrained in sequence, so as to facilitate their nucleosome packaging, actually contribute to the favorable free energy of histone–DNA interactions? A physical selection experiment on the mouse genome yielded a set of mouse genomic DNA fragments having the highest affinity for histones in nucleosome reconstitution, of the entire mouse genome (Widlund *et al*. 1997). The free energies of these sequences were then compared with those of the sequences isolated by selection on (non-natural) random DNA (Thåström *et al*. 1999). The single highest-affinity sequence in the entire mouse genome has a $\approx$ sixfold lower affinity for histones in nucleosome reconstitution than do the dozens of sequences selected from the random library; and the affinity of the second-highest affinity mouse sequence is $\approx$ 34-fold lower. Evidently, even the highest-affinity sequences in natural eukaryotic genomes are not evolved for highest possible affinity. Consistent with this fact, the highest-affinity natural genomic DNA regions obey only relatively weakly the rules deduced from the selection on non-natural DNA: the signal-to-noise ratio of the rules that are obeyed is lower

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

in the selected natural sequences (and also, of course, in entire eukaryotic genomic sequences) than for the selected *non*-natural sequences; and the most-strongly obeyed rules in the *natural* sequences are not the most important energetically.

Taken together, these studies reveal that eukaryotic and euryarchaeal genomic DNA sequences are indeed evolved and constrained to assist in their own nucleosome packaging and nucleosome positioning. But the fraction of the genomes so-constrained is $<5\%$ (at least, for the eukaryotic), and the free energy (i.e. histone–DNA affinity) resulting from these constraints is significantly smaller than the maximum possible.

5.7 Evolutionary approach toward an optimal sequence

The sequences isolated by physical selection from a chemically synthetic random DNA library are the highest-affinity and strongest nucleosome-positioning sequences presently known. Nevertheless, the apparent very great diversity of the library used for that selection (5×10^{12} different random 220-mers, which represents 5–6 orders of magnitude greater sequence complexity than present in the human genome) represents an infinitesimally small fraction, $\approx 10^{-120}$, of the complete space of all possible 220-mer sequences (or 10^{-74} of all possible 147-mer sequences). Therefore it is a statistical certainty that even better sequences exist and await discovery.

The question then arises how such hypothetical even-better sequences might be isolated. The largest library that can feasibly be examined may be perhaps 2–3 orders of magnitude greater diversity than that studied previously. Repeating the selection strategy used previously, but on such a larger library, would likely lead to the isolation of some of these hypothetical better sequences. However, principles of evolutionary biology suggest a far more powerful approach, based on a combination of mutagenesis and recombination. The random library studied previously facilitates recombination between three individual random segments that make up each 220-mer; and far more powerful approaches to achieving recombination over shorter lengthscales, which should be adaptable to this specific problem, have been devised (Stemmer, 1994; Crameri *et al*. 1998). Mutagenesis is readily achieved using error-prone PCR or by chemical synthesis with a controlled level of random nucleotide doping.

5.8 Optimization by design

An alternative approach to developing further-improved high-affinity binding sequences – and to elucidating the rules that govern these sequence preferences – is based on iterative cycles of sequence design and experimental test. If advances in our understanding of the sequence-dependence of DNA structure and mechanics are made in the near future, these could serve as the basis for rational design of improved sequences. In the absence of such progress, the most productive approach to sequence design may be through the creation of a 'consensus' sequence, which may be deduced through alignment of the existing set of selected high-affinity sequences, once the regions within these sequences that are actually nucleosomal are mapped. Such a consensus sequence could then serve as the starting point for cycles of refinement.

Two other interrelated criteria will likely be relevant to the design of an optimal sequence: these include symmetries and spacings of key DNA sequence motifs.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

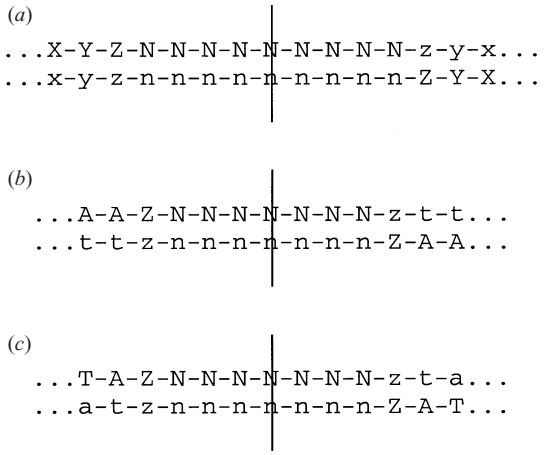


Fig. 8. (*a*) Symmetry-related sequence motifs across the nucleosomal pseudo-twofold rotational symmetry axis, indicated here by a vertical line. The overall pseudo-twofold rotational symmetry axis runs *through* a DNA base pair. Consequently, for the symmetry of the DNA sequence to match that of the histone octamer, an odd number of base pairs must separate any particular sequence motif (e.g. 5′-XYZ-3′) on one strand with another copy of itself (located across the symmetry axis) on the complementary strand. Lower-case letters imply the complements of the corresponding upper-case letters (e.g. X and x are a complementary base pair). (*b*, *c*) Examples for particular motifs implicated in nucleosome positioning; (*b*), AA (= TT); (*c*), a palindromic motif, TA. Since TA is itself palindromic, it *also* repeats along the *same* strand, but still with an odd number of base pairs of separation.

An essential but currently unanswered question is whether the highest-affinity sequences will reflect one or more of the symmetries that are potentially inherent in the nucleosome structure? The organization of the nucleosome potentially allows for a true twofold rotational symmetry axis running through the middle of the DNA, plus several additional pseudo-twofold rotational symmetry axes within and between each histone heterodimer. However, symmetry is not a physical force; it could be the case that the lowest free-energy structure is asymmetric. The actual structure of the nucleosome, as solved by Richmond and colleagues, breaks the possible overall symmetry of the nucleosome, although only subtly so, thereby leaving a pseudo-twofold rotational symmetry axis in its place. However, this need not mean that an optimal structure cannot be symmetric. The DNA sequence that they analyzed was in fact a symmetric sequence, with a length of 146 bp. If the twofold symmetry axis of the nucleosome were to pass through the DNA between two adjacent base pairs, as many had imagined, then this sequence could have made twofold symmetric interactions with a twofold symmetric histone octamer, to yield a symmetric nucleosome. Surprisingly, however, the actual structure reminded us that DNA has *another pseudo*-twofold axis, that runs through the middle of a single base pair; and it turns out that this is the DNA two-fold axis that the structure has evolved to match. Thus the structure solved by Richmond and colleagues (Richmond *et al*. 1984; Luger *et al*. 1997a; Richmond & Widom, 2000) *could not* be truly symmetric. Moreover, the crystal packing imposes additional sources of asymmetry. Consequently, the fact that the nucleosome structure is nevertheless nearly symmetric suggests that an optimal structure may prove to obey overall twofold symmetry.

If it proves to be the case that the optimal structure is twofold symmetric, and if, for whatever reason (perhaps having to do with local DNA bending properties of particular sequence motifs; or perhaps some other reason) it is energetically beneficial to have some

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

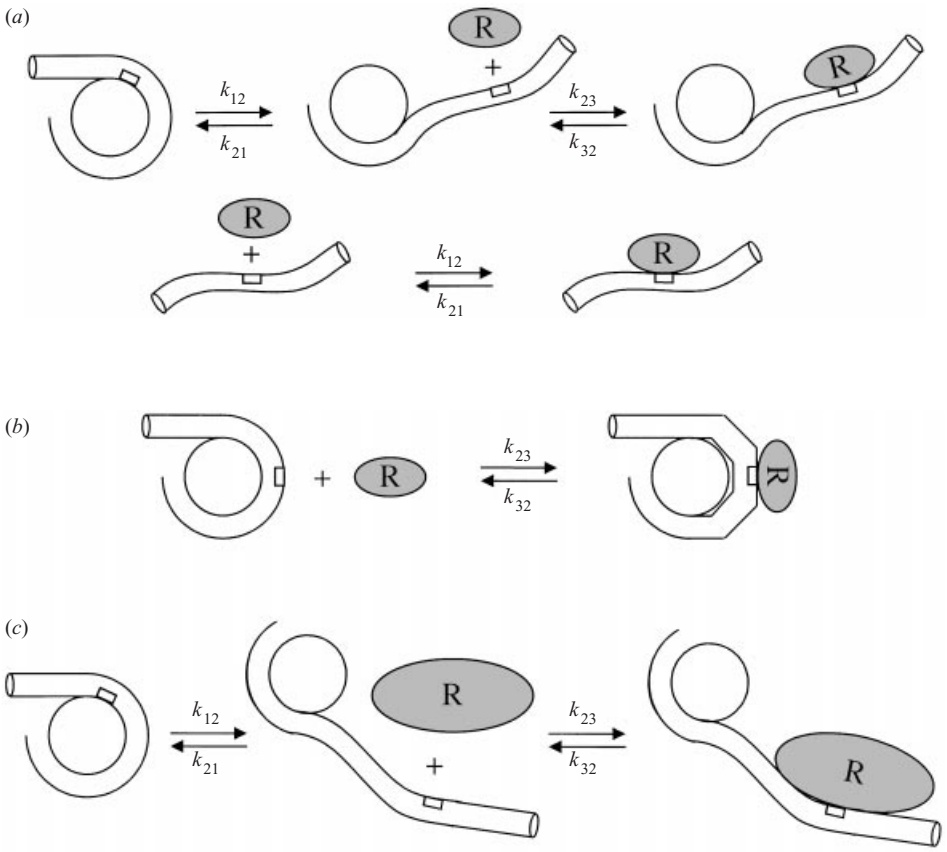


Fig. 9. (*a*) The site-exposure model illustrated for a single nucleosome (*upper panel*) compared to the same process on naked DNA (*lower panel*). The histone octamer is shown from above as a disk with the DNA coiling around it. A particular DNA target sequence (shaded) is inaccessible to the regulatory protein (R) that acts on it. k_{12} and k_{21} are position-dependent apparent rate constants for site exposure and recapture, respectively; the equilibrium constant for site exposure, $K_{\mathrm{eq}}^{\mathrm{conf}}$, $= k_{12}/k_{21}$. Exposure of sites nearer the middle of the nucleosomal DNA may occur by several successive steps of exposure of shorter segments from an end as illustrated; each smaller step would have its own microscopic rate constants. k_{23} and k_{32} are rate constants for binding and dissociation of R from its target site; for simplicity these are assumed to be identical to the rate constants for the same process on naked DNA. Real nucleosomes exist in long chains; but this need not prevent uncoiling. Experiments show that access to buried sites is affected little, or not at all, when a test nucleosome is embedded in the middle of a long chain of nucleosomes. (*b*, *c*) System-specific effects. (*b*) A hypothetical case in which the binding of a protein to a nucleosomal target site can occur with no requirement for prior site exposure, simply because the binding protein is small and can bind to its correspondingly small DNA target site with no requirement for site exposure, when the target site happens to face 'out'. A limiting case is a protein that is small enough to bind with no requirement for site exposure, and that binds to a DNA target that has the same local DNA helical parameters as the DNA in that region of the nucleosome. In this case, the affinity for the nucleosomal target site would be identical to that for naked DNA. More generally, even very small proteins are likely to require at least some distortion of their DNA target site upon binding (e.g. as indicated by the distortions in this figure). Then binding to the nucleosomal target site will be suppressed by a factor corresponding to the free energy cost of deforming the DNA away from its favored nucleosomal conformation, even though no site exposure is required for the binding to occur. (*c*) An opposite hypothetical case, in which a DNA-binding protein is particularly large, or has a particularly unfavorable shape, or is accompanied by a particularly unfavorable deformation of the DNA. In these cases, substantially more DNA than would otherwise be required must be uncoiled from the histone surface in order for the target site to be sterically accessible. Site exposure appears to proceed

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

motif (e.g. three particular consecutive base pairs, XYZ) at some particular location on the top strand in the nucleosome (e.g. as illustrated in Fig. 8; lower-case x, y, z denote the complementary bases of X, Y and Z, respectively), then it follows that the same sequence motif should occur in the same orientation on the bottom strand (that is, oriented 5′–3′ on the bottom strand, as indicated), symmetrically across the overall twofold axis of the nucleosome. That is, essential motifs will be palindromic in the sequence.

In addition, this symmetry relationship implies a particular spacing constraint: the two halves of each palindromically related motif will be separated across the overall twofold by an *odd* number of intervening base pairs (Fig. 8).

If the motif itself is palindromic (e.g. 5′–T–A–3′; Fig. 8*c*) then it also will repeat on the *same strand* at the appropriate symmetry-related location.

Even in the absence of any symmetries, the fact that the nucleosomal DNA has an overall fixed curvature implies a requirement for net bendedness every helical turn, and thus a requirement for proper helical phasing – i.e. particular separations – of any DNA sequence motifs that are used because they facilitate the required DNA distortions. This class of spacing constraints is intimately connected with the position-dependent helical twist of DNA in the nucleosome.

For all of these reasons, one anticipates that there will be a requirement for particular spacings between optimal local DNA sequence motifs, at positions along the length of the nucleosomal DNA. Existing data demonstrate that such spacing preferences exist (Shrader & Crothers, 1990; Fitzgerald & Anderson, 1998), although these studies do not yet suffice to define the molecular basis of these preferences.

Constraints based on symmetry arguments may prove to be useful in evaluating sequence alignments and also in improving consensus sequences derived from such alignments. In any case, whether the potential symmetries are manifested or not, the particular spacings between certain sequence motifs will likely prove be important determinants of histone-binding affinity, and must be considered as key parameters to be optimized during cycles of sequence design and evaluation.

6. Dynamic nucleosome instability

Recent studies reveal that the sequence-dependence of histone–DNA interaction affinity is not only an important determinant of relative nucleosome stability and positioning, but is also an important determinant of *dynamic* behaviors of nucleosomes that are potentially important in gene regulation and other aspects of nucleosome function.

6.1 Site-exposure equilibria

Nucleosomes have a remarkable dynamic property, which we refer to as ‘site exposure’ (Polach & Widom, 1995): DNA wrapped around the histone octamer is in a constant dynamic equilibrium in which stretches of DNA (less than the entire DNA length) are transiently exposed from the surface of the octamer, thereby allowing other proteins transient access to buried DNA target sites (Fig. 9*a*).

progressively inward from an end, with increasing energy penalties for increased uncoiling; thus, proteins requiring site exposure of greater lengths would have their affinity for nucleosomal target sites further suppressed. (Adapted from Polach & Widom, 1999.)

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

We make the simplifying assumption that site exposure leads to exposure of sufficient nucleosomal DNA such that the rates and equilibria for binding to an exposed nucleosomal target sequence or to a naked DNA target sequence are identical. Thus,

$$\mathrm{N} \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} \mathrm{S} + \mathrm{R} \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} \mathrm{RS} \quad \text{and} \quad \mathrm{S} + \mathrm{R} \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} \mathrm{RS} \tag{4}$$

for nucleosomes and naked DNA, respectively, where R is an arbitrary DNA-binding protein, N is a nucleosome, and S is the nucleosome after site exposure or, equivalently, naked DNA.

With this assumption, binding of a regulatory protein to a nucleosomal target sequence will occur with a net free energy change $\Delta G^{\circ}_{\mathrm{net}}$, and an apparent dissociation constant $K_{\mathrm{d}}^{\mathrm{apparent}}$, given by:

$$\Delta G^{\circ}_{\mathrm{net}} = \Delta G^{\circ}_{\mathrm{conf}} + \Delta G^{\circ}_{\mathrm{naked\ DNA}} \quad \text{and} \quad K_{\mathrm{d}}^{\mathrm{apparent}} = K_{\mathrm{d}}^{\mathrm{naked\ DNA}} / K_{\mathrm{eq}}^{\mathrm{conf}}, \tag{5}$$

where $K_{\mathrm{eq}}^{\mathrm{conf}}$ is the equilibrium constant for site exposure in the nucleosome (k_{12}/k_{21}), $K_{\mathrm{d}}^{\mathrm{naked\ DNA}}$ is the dissociation constant for binding to naked DNA (k_{32}/k_{23}), and $\Delta G^{\circ}_{\mathrm{conf}}$ and $\Delta G^{\circ}_{\mathrm{naked}}$ DNA are the corresponding free energies.

In this model, $K_{\mathrm{eq}}^{\mathrm{conf}}$ depends primarily on the translational position of the target sequence within the nucleosome (ΔG_{conf} depends on the length of DNA being uncoiled, i.e. on the length of protein–DNA interface that is disrupted). However, the effective $K_{\mathrm{eq}}^{\mathrm{conf}}$ will also depend on other factors such as the size and shape of the regulatory protein, the rotational setting of the target site around the periphery of the DNA helix, and on DNA bending induced by the protein, since these affect the amount of DNA that must be uncoiled to allow the protein to bind. These effects are discussed further, below.

This property was initially recognized and quantified in studies of the abilities of restriction endonucleases to cleave DNA at sites throughout the nucleosome (Polach & Widom, 1995, 1999). Nucleosomes were constructed with DNA containing *patches* of sites, such that sites for multiple restriction enzymes were nested together within roughly one helical turn of DNA. The rate at which buried nucleosomal sites are cleaved by the restriction enzymes, scaled by the rate of cleavage of naked DNA in identical conditions, gives the fraction of time that the nucleosomal DNA 'looks like' naked DNA, i.e. the equilibrium constant for site exposure, $K_{\mathrm{eq}}^{\mathrm{conf}}$. Measuring the apparent $K_{\mathrm{eq}}^{\mathrm{conf}}$ for the different sites within one patch allows one to discriminate between true site exposure and inadvertent construction of a site that faces 'out'.

The successes of this model are evident. When experiments reveal that proteins can in fact bind to nucleosomal target sites, this model explains how they got there, and it explains with remarkable accuracy why the observed $K_{\mathrm{d}}^{\mathrm{apparent}}$ takes on its particular value, which previously seemed to vary randomly over a 10^5-fold range. The model also clarifies a number of previously mysterious observations, in which certain proteins apparently could *not* bind to nucleosomal target sites (including especially remarkable cases where different but very close relatives of those same proteins *could* bind). In these cases, the model reveals that investigators did not carry the binding titration experiments up to the concentrations (of order $\approx K_{\mathrm{d}}^{\mathrm{apparent}}$) where binding would have been detected. Binding equilibria are controlled by concentrations of free species, and not, e.g. by ratios of total components present.

Site exposure is non-dissociative in nature – that is, buried sites are made transiently accessible without full dissociation of the DNA from the histone octamer. In a direct binding

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

assay, such as illustrated in Fig. 9*a*, histones remain present in a ternary complex comprising histones, DNA, and a bound protein 'R' (Workman & Kingston, 1992). And in a restriction enzyme cleavage assay probing sites only a short distance inside the nucleosome, nucleosomes remain intact even after their DNA has been cleaved (Polach & Widom, 1995).

The relatively progressive decrease in $K_{\mathrm{eq}}^{\mathrm{conf}}$ as one moves inward from an end into the middle of the nucleosome, the fact that binding of multiple proteins to nucleosomal target sites occurs cooperatively (see below), and the structure of the nucleosome itself, are all consistent with a picture in which DNA spontaneously and transiently uncoils inward starting from an end. As described earlier, the structure of the nucleosome (Luger *et al*. 1997a) shows the DNA wrapped on the histone surface as making contacts ('bonds') in a patch, every ≈ 10 bp, each time the phosphodiester backbone (minor groove) faces inward toward the octamer. Thus, uncoiling would naturally proceed stepwise, with an incremental increase in energetic cost (i.e. decreased equilibrium constant for site exposure) associated with each additional 10-bp-long segment uncoiled.

Experiments using coupled enzymatic assays to detect site exposure reveal that site exposure sufficient to allow access to all of the DNA in a nucleosome occurs on a timescale of seconds or faster (Protacio *et al*. 1997), and possibly *much* faster. [In these studies, as in the earlier studies using restriction enzymes as probes of accessibility (site exposure), the enzymes serve as neutral probes: they do not drive the processes that they are being used to detect.] Simple theoretical models for site exposure and recapture processes – assuming either an activated or a diffusive process for recapture – lead to the expectation that $k_{21} \geqslant 10^5\ \mathrm{s}^{-1}$ (J. Widom, unpublished observations), in which case the forward process of site exposure (k_{12}), may occur on the millisecond timescale or faster.

Real nucleosomes exist in long chains; but this need not interfere significantly with site-exposure processes. With just modest deformation of the linker DNA, a combined uncoiling coupled to a motion of the uncoiled DNA in a direction parallel to the axis of the nucleosomal disk allows uncoiling beyond the dyad (which is as far as necessary to allow binding anywhere) with no required crossings and with little motion of other nucleosomes. Experiments show that access to buried sites is affected little, or not at all, when a test nucleosome is embedded in the middle of a long chain of nucleosomes (Owen-Hughes *et al*. 1996; Owen-Hughes & Workman, 1996). Higher levels of chromosome structure may need to be disassembled prior to the site-exposure process illustrated here, but are also believed to possess only marginal stability (Widom, 1986, 1989).

Despite the generally good agreement between the results from two very different assays for nucleosomal site exposure – one based on restriction enzyme digestion kinetics, the other on simple equilibrium binding of a site-specific DNA binding protein – one may, nevertheless, anticipate that there will exist system-specific, second-order effects, which may lead to discrepancies between results from the two experiments. Similarly, one may anticipate the existence of discrepancies between results obtained using either one of these methods with two different probe molecules (restriction enzymes or binding proteins) monitoring accessibility at the same or closely spaced sites.

Fig. 9*b* illustrates a hypothetical case in which the binding of a protein to a nucleosomal target site can occur with no requirement for prior site exposure, simply because the binding protein is small and can bind its (correspondingly small) site with no requirement for site exposure when the site happens to face 'out'. (Note that since the free energy of nucleosome positioning is finite and, as will be described below, nucleosomes may be mobile in

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

physiological conditions, no site can be counted on to face 'out' all of the time. Thus, such scenarios do not represent mechanisms for guaranteeing the accessibility of binding sites *in vivo*.) If binding to this site also requires no significant deformation of the DNA, the binding affinity would closely approximate that for naked DNA, i.e. differing from the prediction of the site-exposure model by the large factor $K_{\mathrm{eq}}^{\mathrm{conf}}$. If binding required a conformational change in the DNA site, as illustrated in this figure, the binding affinity would be reduced by an amount corresponding to the energetic cost of any required conformational changes. In either case, the site-exposure model does not describe the binding event.

Fig. 9*c* illustrates an opposite hypothetical case, in which a binding protein (or restriction enzyme) is particularly large, or has a particularly unfavorable shape (or distorts the DNA into a particularly unfavorable shape), requiring that substantially more DNA be uncoiled off the histone surface in order for the target site to be sterically accessible. As described above, increased uncoiling will generally be accompanied by an increased cost in free energy, and will therefore occur with decreased equilibrium constant for site exposure, and a correspondingly suppressed binding affinity or rate of restriction enzyme digestion.

These potential exceptions do not refute this model for the general behavior of nucleosomes. Rather, they simply represent second-order corrections which are inevitably required to account for system-specific details in most physical models.

The site-exposure model has within it the potential for important novel cooperative (synergistic) interactions between multiple proteins binding simultaneously to sites within a single nucleosome (Polach & Widom, 1996). This arises essentially as a consequence of the two proteins fighting together against a common competitor, the histone octamer. A model that accounts for this behavior is illustrated in Fig. 10. Let X and Y be two arbitrarily chosen site-specific DNA binding proteins for which the target sites are contained within the same nucleosome; and let X be the protein having the binding site nearer the end of the nucleosomal DNA (the 'outermost' site). Suppose that Y binds (bottom path in the figure); then, as can be seen in the figure, if and only if binding events at the *two* sites are mechanically linked, the prior binding of Y facilitates the subsequent binding of X, because now X can bind without having to pay the full energetic penalty for site exposure (here defined as ΔG_1°) which otherwise would be required. Similarly, the ability of X to bind facilitates the subsequent binding of Y, since at least some of the final free-energy penalty for the required conformational change is already paid. These processes are linked in a thermodynamic cycle, hence the same coupling free energy and corresponding binding cooperation necessarily applies, regardless of the order in which the proteins bind. Thus X and Y act cooperatively (synergistically) even if they do not touch: the binding of one protein radically alters the binding ability of the other. No special properties are required of X or Y: they need only bind DNA for this cooperation to be manifested. X and Y may be two different proteins, or they may be two molecules of the same protein.

The amount of cooperation between X and Y (the coupling free energy δG_{XY} – the free energy by which the prior binding of X facilitates the binding of Y, or vice versa), is equal to $-\Delta G_1^\circ$, i.e. minus one times the energetic cost for exposing the outermost site. In real systems, X and Y may also have direct 'conventional' cooperative interactions, represented by ΔG_3°. Then the net cooperation will be given by the sum of $\delta G_{\mathrm{XY}} + \Delta G_3^\circ$. Importantly, however, the novel cooperative free energies inherent in this mechanism (δG_{XY}) are substantially greater than typical free energies of previously recognized 'conventional' cooperative interactions (Polach & Widom, 1996).

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

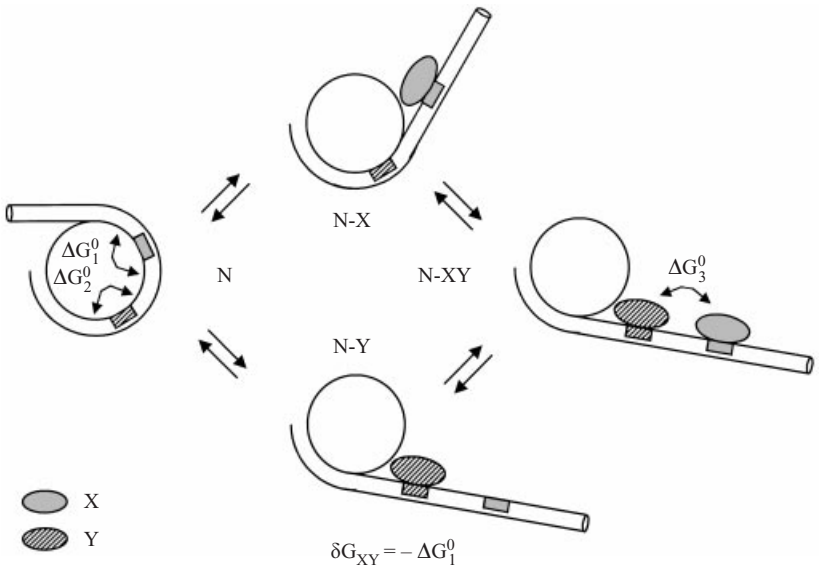


Fig. 10. Cooperativity in the binding of multiple proteins to target sites in a single nucleosome. A nucleosome is shown containing binding sites (shaded) for two proteins (X and Y). X and Y may be two unrelated proteins, or two molecules of the same protein. X is defined as the protein binding to the outermost site, and Y as binding to the innermost site. ΔG_1° is the free energy cost for uncoiling enough DNA to expose the site for X. ΔG_2° is the additional free energy cost for uncoiling sufficient additional DNA so as to expose the site for Y. In some cases, X and Y may have 'conventional' cooperative interactions, also detectable in their binding to naked DNA (e.g. from favorable protein-protein contacts between X and Y); these are represented as ΔG_3°. Even in the absence of conventional cooperative interactions, the prior binding of X facilitates the binding of Y, and the prior binding of Y facilitates the binding of X, both increasing the favorable free energy for binding of the second protein by the same amount, given by $\delta G_{XY} = -\Delta G_1^\circ$. If either X or Y happens to be a restriction enzyme, then binding of the other protein to near-saturation will increase the time-averaged accessibility of the restriction enzyme's target site – and hence the rate of digestion – by the factor $\exp(-\delta G_{XY}/RT)$. (Adapted from Polach & Widom, 1996).

This model, *with no adjustable parameters*, accounts quantitatively for a diverse set of experimental results on cooperative binding of various proteins to nucleosomal target sites (Adams & Workman, 1995; Polach & Widom, 1996). Moreover, there is good agreement between the predictions of this model and experimental data even using ΔG_{conf} obtained from the restriction enzyme digestion kinetic measurements instead of the δG_{XY} obtained directly from the primary cooperative binding data. These two experiments are completely unrelated except through the site-exposure model, and therefore the agreement between the two provides strong evidence for the applicability of the site-exposure model to the behavior of real nucleosomes.

An important question for further investigation is: Do these cooperative interactions persist across the nucleosomal dyad?

6.2 DNA sequence-dependence to site-exposure equilibria

In the site-exposure model, equilibrium constants for site exposure are related to sums of free energies for all of the histone–DNA contacts that need to be broken in order to expose the site. This implies that DNA sequence could be a major determinant of site-exposure

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

equilibrium constants, and that sequences having especially high affinity for the histone octamer [or equivalently, strong nucleosome-positioning sequences (Lowary & Widom, 1997)] should exhibit significantly reduced equilibrium site accessibility.

These ideas apply also to the nucleosome-dependent cooperation, since the *amount* of that cooperation (δG_{XY}) is directly linked to the free energy of histone–DNA interactions for the outer DNA regions, ΔG_1°.

We developed a new DNA sequence for use in an experimental test of these ideas (Anderson & Widom, 2000). The sequence is derived from a non-natural sequence that was selected for especially high-affinity binding to histone octamers in nucleosome reconstitution (Lowary & Widom, 1998). This is useful for creating a homogeneous sample; and because of its much higher affinity, the sequence should dramatically exhibit any DNA sequence-dependent effects on equilibrium constants for site exposure that might obtain. We made specific base changes to the DNA sequence to introduce sites for restriction enzymes at locations throughout the nucleosome, to allow measurement of the corresponding position-dependent equilibrium constants for site exposure at sites throughout the nucleosome. The base changes so introduced decreased the affinity of the sequence somewhat (relative to the original selected sequence) but the resulting affinity was still approximately tenfold greater than that of the 5S sequence derivatives used in our initial studies of site-exposure equilibria.

Values of K_{eq}^{conf} obtained from these studies are shown in Fig. 11, tied to their corresponding nucleosomal location on the abscissa (bottom). One interesting aspect of this study is that the DNA used allows us to probe sites across the full length of the DNA, whereas our earlier studies had analyzed accessibility on only a single side of the nucleosome. The new data support the hypothesis that access to internal sites is reached by progressive uncoiling from either end, since the equilibrium accessibilities decrease progressively from both ends inward toward the center of the nucleosomal DNA.

The results with the high-affinity sequence are shown superimposed on the earlier results obtained using the modified 5S sequences (tied to a separate abscissa, top), to facilitate a comparison. The results show that, in agreement with the predictions of the site-exposure model, equilibrium constants for site accessibility are strongly suppressed in the higher-affinity sequence. Evidently, DNA sequence is a major determinant of equilibrium accessibility of nucleosomal DNA target sites.

These predicted dependences of site-exposure equilibria on histone–DNA interaction free energy are seen also in nucleosomal DNAs containing poly(dA:dT) elements. As described earlier, we found that introducing an A_{16} element midway between the ends and the middle of the nucleosomal DNA, destabilized the DNA wrapping by ≈ 0.4 kcal mol^{-1} compared to random-sequence DNA. In that same study we monitored the effects of this destabilization on site-exposure equilibria. We found that inclusion of the A_{16} element led to a $\approx 1.7 \pm 0.1$-fold increase in position-averaged equilibrium accessibility of the nucleosomal DNA.

Thus large effects on histone–DNA interaction affinity give rise to large effects on equilibrium constants for site exposure, and small (but significant) changes in affinity result in small (but still significant) changes.

Because of the relationship between a DNA sequence's affinity for histones in nucleosome reconstitution, and its nucleosome-positioning power, these results collectively imply *two* new ways in which the choice of a particular genomic DNA sequence could be a quantitatively important determinant of the accessibility – and hence the occupancy – of regulatory target sites in genomic DNA.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

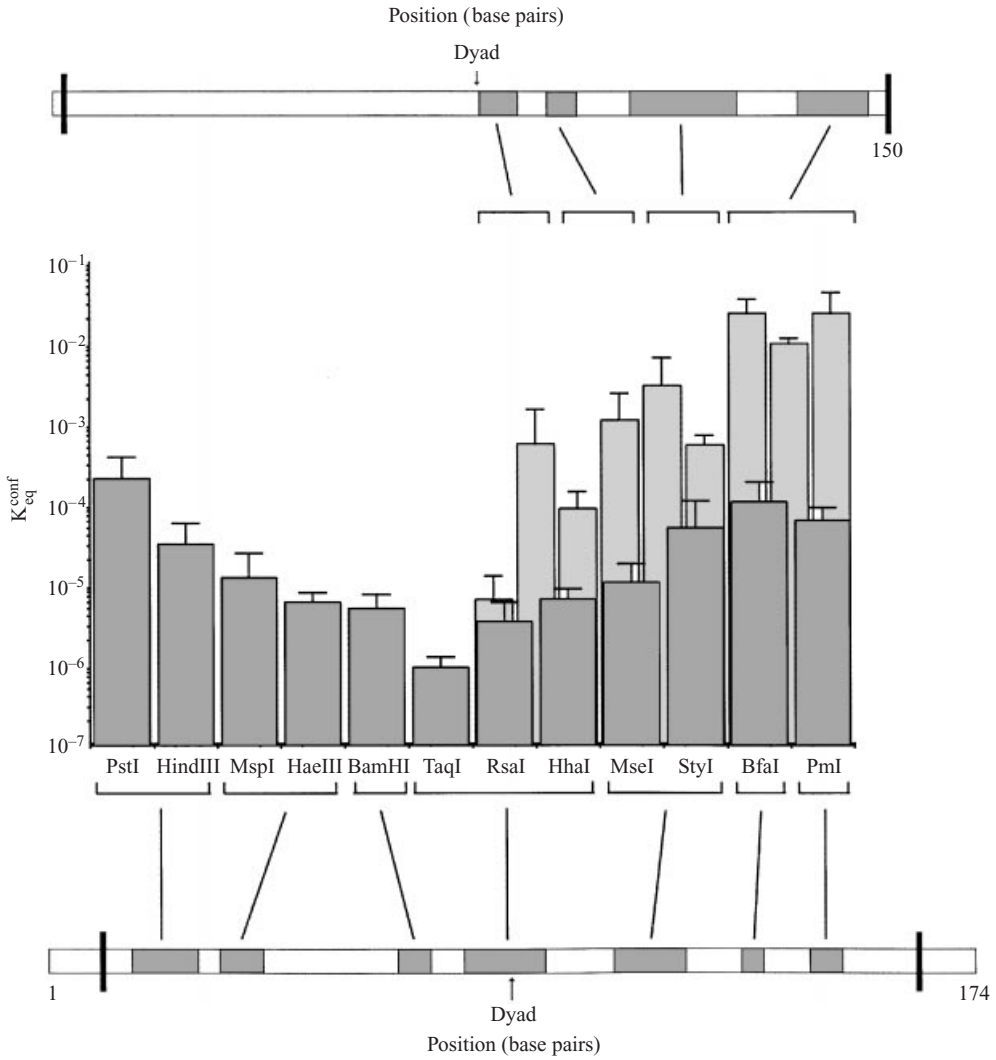


Fig. 11. DNA sequence-dependence of site-exposure equilibrium constants. Summary of measured equilibrium constants for site exposure (K_{eq}^{conf}). Means and standard deviations are plotted as a function of location in the nucleosome, for the high-affinity sequence 601.2 (Anderson & Widom, 2000) (dark shading, tied to the corresponding nucleosomal locations below the plot) and for the reference 5S sequences (Polach & Widom, 1995) (light shading, tied to their corresponding nucleosomal locations above the plot). The DNA sequence *greatly* influences the equilibrium accessibility of buried target sites. (From Anderson & Widom, 2000.)

First, the results of Fig. 11 establish that, for a given biased nucleosome positioning, the choice of genomic DNA sequence directly affects the equilibrium accessibility of nucleosomal target sites, over a range of several orders of magnitude.

Secondly, our earlier results establish (and these newer results reconfirm) that the accessibilities of sites within a nucleosome are strongly dependent on the location of the site within the nucleosome [decreasing by several orders of magnitude as one looks from one end of the nucleosomal DNA in toward the middle (Polach & Widom, 1995; Anderson &

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

Widom, 2000)] and that the preferred positioning of nucleosomes is linked to the positioning-dependent free energies of histone–DNA interactions (Lowary & Widom, 1997), with many natural sequences biasing this positioning by 100-fold or more (Thåström *et al*. 1999). From these facts it follows that the choice of genomic sequence, through biases in the preferred positioning of nucleosomes, can bias by factors of 100-fold or more whether the equilibrium constant for site exposure of a particular target site is potentially ~ 1 or greater, corresponding to preferential positioning of the target site in linker regions, or 10^{-2}–10^{-3}, corresponding to preferential positioning of the target site near the end of the nucleosomal DNA, or as low as $\sim 10^{-4}$–10^{-6}, corresponding to preferential positioning of the target site near the nucleosomal dyad.

These two mechanisms by which the choice of genomic DNA sequence can strongly influence the equilibrium accessibility of nucleosomal DNA target sites are linked by the site-exposure model: a subset of the histone–DNA contacts in the nucleosome that contribute to biasing nucleosome positioning (and thus influence site accessibility indirectly) also contribute to the energy penalty that must be paid to expose the site (thereby influencing site accessibility directly).

If it proves to be the case that nucleosomes are mobile *in vivo* and that this mobility occurs rapidly in comparison to timescales of regulatory decisions, then equilibrium occupancy of regulatory proteins may need to be considered as a set of coupled equilibria. These equilibria may include the forces (free energies) of nucleosome positioning [which depends on DNA sequence (Lowary & Widom, 1997) and many other factors (Widom, 1992; Yao *et al*. 1993)], site exposure (which again is sequence-dependent), and regulatory protein binding to DNA.

6.3 Nucleosome translocation

While histone octamers do not exchange freely between differing DNA molecules in physiological solution conditions, they have a surprising ability to translocate along DNA, thereby yielding nucleosomes in new positions (Weischet, 1979; Pennings *et al*. 1991; Meersseman *et al*. 1992). This process is often referred to as nucleosome mobility or sliding, although the latter seems certain to be a misnomer. Nucleosome mobility is readily detected using a 2D gel assay that takes advantage of the dependence of the electrophoretic mobility on the position of the octamer between the ends and the middle of a DNA fragment (Pennings *et al*. 1991; Meersseman *et al*. 1992). The rate of nucleosome mobility is enhanced at elevated temperatures and ionic strengths, but mobility can also be detected in lower temperatures and approximately physiological ionic strengths. Mobility is suppressed by histone H1 (Pennings *et al*. 1994).

The word 'sliding' suggests a mechanism for nucleosome mobility in which the DNA acts as though it is simply pulled along the histone octamer, just as if one pulled on a rope that was coiled around a smooth spool. The structure of the nucleosome suggests that such a picture cannot be correct: the histone octamer and DNA interact through myriad bonds and with good steric fits between rough molecular surfaces (Luger *et al*. 1997a). Such interactions would not allow smooth sliding; sliding intermediates would be expected to have prohibitively high energies. Could one imagine the histone octamer instead as a helical gear, with teeth that mesh with DNA base pairs? Sliding would then correspond to a rotation of the helical gear, each step releasing a gear tooth at one end and engaging a new tooth at the other end. This idea cannot apply because there *are no* '*teeth*' at the other end! Strong bonds

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

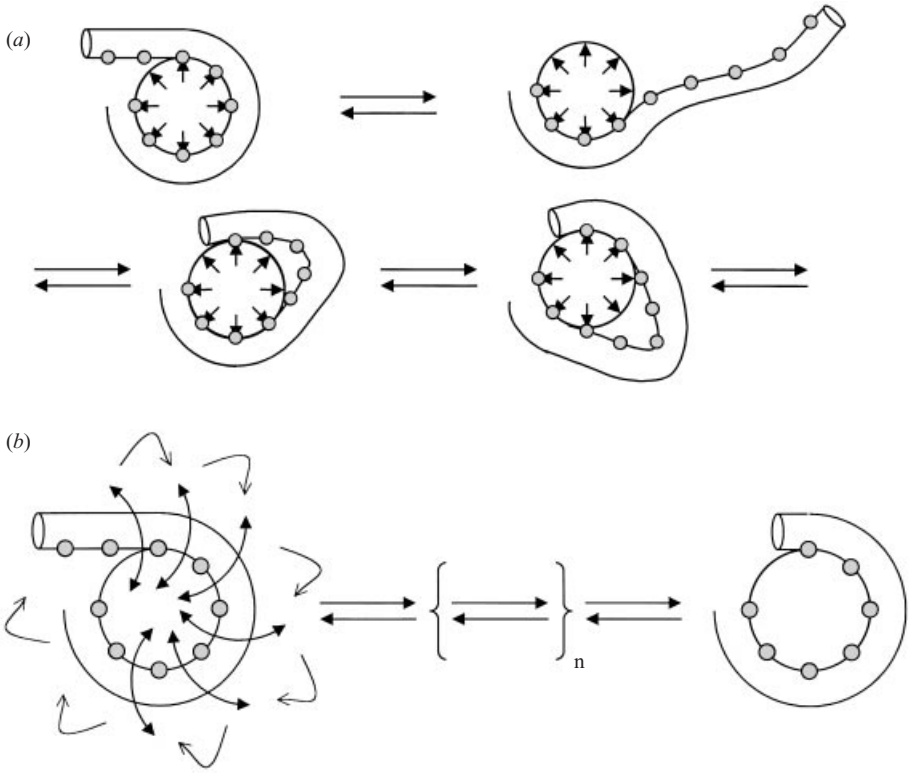


Fig. 12. Mechanisms for spontaneous nucleosome translocation. (*a*) Diffusion of a bulged loop. DNA unpeels from the nucleosome starting at an entry/exit site (that is, at one end of the nucleosomal DNA) (*upper panels*) and rebinds at another site on the DNA beyond the end of the original nucleosomal DNA (*lower left panel*). Small circles on the DNA illustrate locations at which the minor groove faces inward against the histone octamer, where it interacts with DNA-binding motifs on the histone octamer surface (arrows). DNA twisting is assumed to occur negligibly, hence all of the contact points lie along one face of the DNA, one per DNA helical turn. DNA rebinding forms a bulged loop on the nucleosome, which can migrate around the entire nucleosomal DNA by breaking bonds at a site on one side of the bulge and reforming them on the other side (*lower right panel*). If such a bulge defect diffuses off the other end of the nucleosome, the result will be that the nucleosome has translocated. Looping without twist change would lead to translocation in steps of ≈ 10 bp. (Adapted from Schiessel *et al*. 2001.) (*b*) Twist diffusion. Torsional fluctuations of the DNA (illustrated as a double-headed arrow) can jump (single-headed arrows) across the outermost binding site (histone–DNA contact) into the nucleosome, and thus be trapped as a twist defect (equivalently, a length defect), most probably a defect of ± 1 bp. If such a defect diffuses all the way through the nucleosome and out of the other end, the result will be that the nucleosome has translocated by a ± 1 bp step. Remarkably, the structure of the nucleosome itself shows that such twist defects (of -1 bp) are readily accommodated inside the nucleosome (Luger *et al*. 1997a).

that are lost at one end with each gear-step could be replaced only with inequivalent or non-existent bonds at the other end. [This situation differs from the special case of a flat (planar) gear, such as the crankshaft gear of a bicycle. The planar gear is degenerate; a tooth that rotates off the chain will eventually rotate around and re-engage with the chain. Finite length helical (i.e. non-planar) gears cannot do this.]

Two models seem more plausible as mechanisms for nucleosome translocation. One model (Schiessel *et al*. 2001) (Fig. 12*a*) is suggested by the inherent nucleosomal property of site

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

exposure itself. Suppose that a stretch of DNA unwraps from the histone surface, starting at one end of the nucleosomal DNA, and some time later is 'recaptured', but at a point along the DNA located some distance beyond the previous end of the nucleosomal DNA, thereby creating a bulged loop. Such a situation would represent a relatively high energy state (involving increased DNA curvature and some loss of histone–DNA contacts); yet it could still be much better (much lower in energy) than a slipping mechanism that required breaking *all* of the protein–DNA bonds while still keeping the DNA bent. If such a bulge were to diffuse from the other end of the nucleosome, the result would be that the nucleosome has translocated. Looping without twist change (which one presumes to minimize the energetic cost) would lead to translocation in steps of ≈ 10 bp.

In this model, specific DNA sequences that decreased the *bending* free energy (i.e. sequences that are more readily bendable) would allow for more rapid nucleosome translocation, as a consequence of lowering the free energy of the looped DNA in the transition state structures.

The slow kinetics of nucleosome translocation, together with the fact that site exposure can occur rapidly (Protacio *et al*. 1997) are consistent with the possibility that site exposure could be a mandatory initial step in the mechanism for translocation; and they argue against the opposite possibility, namely, that translocation could be the mechanism for site exposure. However, since quantitative measurements of translocation 'velocities' or diffusion coefficients have not yet been made, solid conclusions cannot yet be drawn.

An alternative model (Richmond & Widom, 2000) (Fig. 12*b*) is based on the diffusion of DNA twist defects. Suppose that a thermal fluctuation suddenly allows an absence or excess of 1 bp of DNA length to enter the outermost helical turn of the nucleosomal DNA. Such a fluctuation might require uncoiling of the outermost turn (or at least breaking all of the contacts that pin it down) as in the site-exposure model, followed by a twist fluctuation in the released DNA and then rebinding/reformation of the outermost set of contacts; or it could be that a twist fluctuation could literally diffuse right 'through' a contact patch by, for example, distorting or breaking individual atomic contacts one at a time. In either case, at this point the fluctuation is trapped as a twist defect (equivalently, a length defect), most probably a defect of ± 1 bp. If such a defect were to continue diffusing all the way through the nucleosome and out the other end, the result would be that the nucleosome has translocated by a 1 bp step.

While such a mechanism might seem far-fetched, a remarkable fact is that the structure of the nucleosome itself shows that such twist defects are readily accommodated inside the nucleosome (Luger *et al*. 1997a). As described in Section 5.8, the DNA sequence used in the initial crystal structure determination actually had a 72 bp side and a 73 bp side. Amazingly, the 'short' side made up for the missing base pair within a space of only 1–2 DNA helical turns, by overtwisting and taking a shorter path to the next contact point, thereby allowing the remaining stretches of DNA on the short half to be nearly identical in structure and trajectory to that on the longer half. That is, the nucleosome evidently stabilized precisely the kind of twist defect proposed in the twist diffusion model for nucleosome translocation. The fact that the nucleosomes crystallized without difficulty suggests that the resulting structures must not be greatly disfavored.

In this model, specific DNA sequences that decreased the *twisting* free energy (i.e. sequences that are more readily twistable) could allow for more rapid nucleosome translocation, as a consequence of lowering the free energy cost of any overwinding or underwinding required in the transition state structures.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

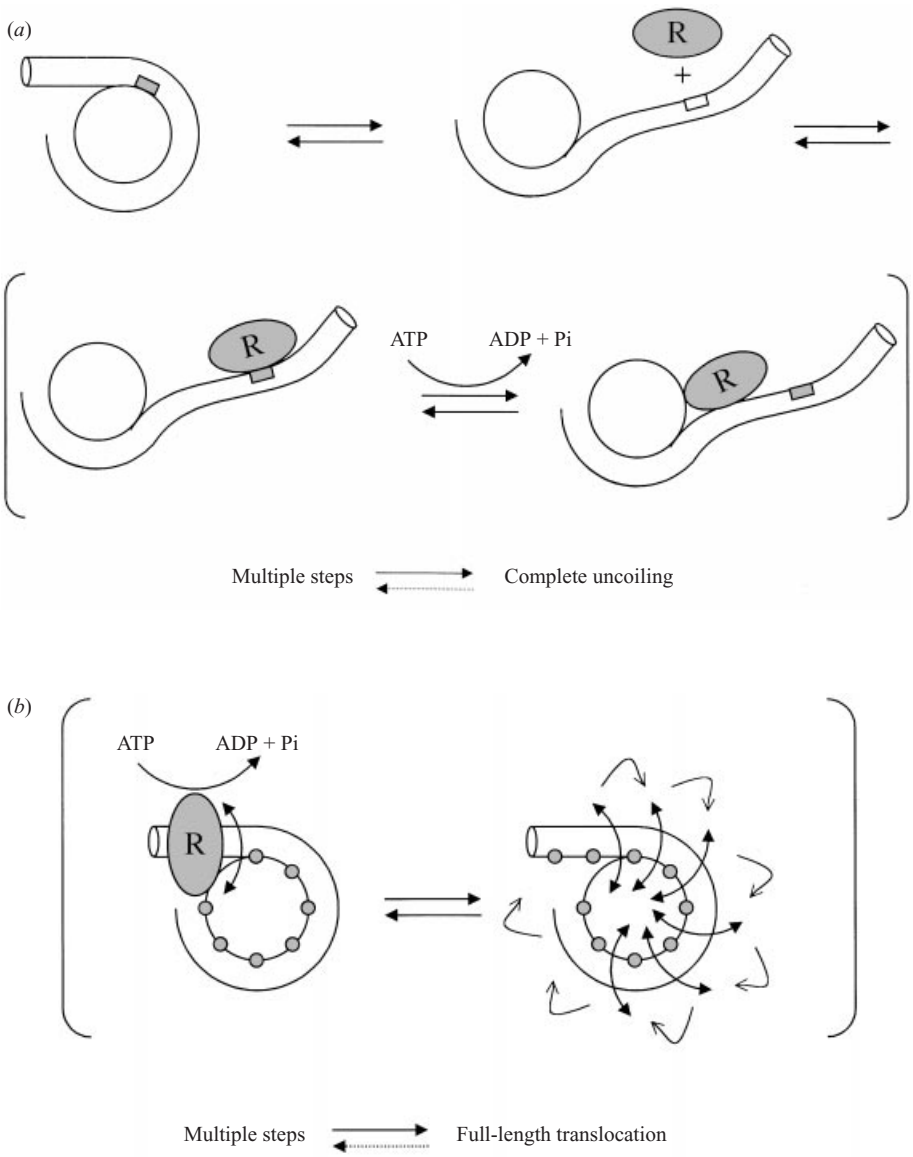


Fig. 13. Putative mechanisms of polymerases and ATP-driven chromatin remodeling enzymes. (*a*) An enzyme utilizes the free energy of ATP hydrolysis (for chromatin remodeling) or of NTP or dNTP incorporation (for RNA and DNA polymerases, respectively) to take a processive walk. In this scenario, the enzymes need not significantly increase the rate of nucleosome site-exposure fluctuations, rather, they may act as Brownian ratchets to trap such events as they occur, eventually allowing the enzyme to progress along the entire nucleosomal DNA length. This progression may lead to complete displacement of the histone octamer (illustrated). Alternatively, an initially displaced stretch of DNA may rebind behind the enzyme, forming a bulged loop as illustrated in Fig. 11*a*, leading ultimately to a net translocation of the histone octamer (i.e. nucleosome) backwards along the DNA (see text). (*b*) An enzyme acts as a Brownian ratchet, utilizing the free energy of ATP hydrolysis to rectify spontaneous twist fluctuations as they enter or leave a nucleosome, so that only fluctuations of a single sign are transferred in net into or out of the nucleosome. Alternatively, an enzyme could actively drive twist defects into the nucleosome. In either case, once they have entered the nucleosome, the twist defects diffuse along the full nucleosomal length, leading to a net translocation of 1 bp. Many such events allow nucleosomes to translocate over their full length and beyond.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

6.4 Action of processive enzymes

Analogs of the models for spontaneous nucleosome translocation can be created for enzyme-catalyzed nucleosome translocation or nucleosome displacement. Such models may plausibly account for the abilities of RNA and DNA polymerases (Widom, 1997b) [and other enzymes (Protacio *et al*. 1997)] to progress through their nucleosomal DNA templates, and may in addition account for known classes of chromatin 'remodeling' activities.

Fig. 13*a* illustrates a model in which successive repetition of two steps – partial DNA uncoiling (as in the site-exposure model), followed by enzyme translocation up to the next point of steric hindrance – allows the enzyme access to the full nucleosomal DNA length. The result could be complete displacement of the histone octamer, or [via a looping mechanism (Studitsky *et al*. 1994, 1997)], net transfer of the histone octamer backwards along the DNA. The number of such uncoiling steps required for elongation through a complete nucleosome in this model depends on the length of DNA released in each step. (Site exposure occurs non-dissociatively, so the entire DNA chain is not released in just one step.) Most likely, as suggested by the structure of the nucleosome, DNA may be exposed in successive $\approx$ 10-bp-long steps. In the simplest mechanisms of the type illustrated here, the enzymes do not actively displace DNA off the histone surface, rather they act as Brownian ratchets, trapping pre-existing spontaneous uncoiling fluctuations.

In these models, the kinetics or equilibria of DNA release, and hence the kinetics of enzyme progression, will be strongly dependent on the affinity of successive stretches of DNA for the histone octamer, i.e. on the particular local DNA sequence.

Fig. 13*b* illustrates an analog of the twist diffusion model for nucleosome translocation, except here an enzyme (again acting as a Brownian ratchet) is utilizing ATP energy to rectify spontaneous twist fluctuations, so that only fluctuations of a single sign are transferred in net into or out of the nucleosome. Once they have entered the nucleosome, the twist defects diffuse along the full nucleosomal length, leading to a net translocation of 1 bp. Many such events allow nucleosomes to translocate over their full length and beyond.

In these models, as for the uncatalyzed case (Fig. 12*b*), the net rate of nucleosome translocation is likely to be sensitive to the ability of the nucleosomal DNA to accommodate the needed twist defects.

7. Conclusions

Histone octamers have high affinity for *any* DNA sequence in nucleosome formation, yet differing DNA sequences differ in their relative affinity by 1000-fold or more. Differences in affinity arising as a consequence of differing DNA sequences translate directly into differences in relative equilibrium nucleosome stability, nucleosome positioning, and the dynamic equilibrium accessibility of buried nucleosomal DNA target sites. Dynamic equilibrium accessibility, in turn, is likely linked to rates of spontaneous nucleosome translocation and catalyzed chromatin remodeling.

Because DNA in a nucleosome is so tightly bent, and has altered helical twists over most of its length, nucleosomes also represent an attractive laboratory in which to explore the properties of DNA itself. DNA sequences that have more appropriately located bends, or are more easily bendable, or have a more appropriate distribution of helical twist, or are more easily twistable, would be expected to contribute to increased affinity of a DNA for binding

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

to histones in nucleosome reconstitution. Studies of one natural nucleosomal DNA that has particularly high affinity show it to be unusually 'soft' with respect to both bending and twisting distortions. Future studies of other DNA sequence motifs implicated in high-affinity binding in nucleosome formation will likely lead to advances in our understanding of the sequence-dependence to the structures and mechanics of DNA itself.

8. Acknowledgements

Research in the author's laboratory is supported by a grant from the NIH. The author thanks Peggy Lowary and the current members of his group for valuable discussions.

9. References

Adams, C. C. & Workman, J. L. (1995). Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Molec. cell. Biol.* **15**, 1405–1421.

Anderson, J. D. & Widom, J. (2000). Sequence- and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J. molec. Biol.* **296**, 979–987.

Anderson, J. D. & Widom, J. (2001). Poly(dA, dT) Promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Molec. cell. Biol.* **21**, 3830–3839.

Anselmi, C., Bocchinfuso, G., De Santis, P., Savino, M. & Scipioni, A. (2000). A theoretical model for the prediction of sequence-dependent nucleosome thermodynamic stability. *Biophys. J.* **79**, 601–613.

Arents, G., Burlingame, R. W., Wang, B.-C., Love, W. E. & Moudrianakis, E. N. (1991). The nucleosomal core histone octamer at 3.1 Å resolution. A tripartite protein assembly and a left-handed superhelix. *Proc. natn. Acad. Sci. USA* **88**, 10148–10152.

Ausio, J., Dong, F. & van Holde, K. E. (1989). Use of selectively trypsinized nucleosome core particles to analyze the role of the histone 'tails' in the stabilization of the nucleosome. *J. molec. Biol.* **20**, 51–463.

Ayadi, L., Coulombeau, C. & Lavery, R. (2000). The impact of abasic sites on DNA flexibility. *Biomolec. struct. Dynam.* **1**, 45–653.

Bacolla, A., Gellibolian, R., Shimizu, M., Amirhaeri, S., Kang, S., Ohshima, K., Larson, J. E., Harvey, S. C., Stollar, B. D. & Wells, R. D. (1997). Flexible DNA, genetically unstable CTG-CAG and CGG-CCG from human hereditary neuromuscular disease genes. *J. biol. Chem.* 276783-16792.

Bailey, K. A., Pereira, S. L., Widom, J. & Reeve, J. N. (2000). Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution. *J. molec. Biol.* **303**, 25–34.

Bashkin, J., Hayes, J. J., Tullius, T. D. & Wolffe, A. P. (1993). The structure of DNA in nucleosome core at high salt concentration and at high temperature. *Biochemistry* **32**, 1895–1898.

Beutel, B. A. & Gold, L. (1992). In vitro evolution of intrinsically bent DNA. *J. molec. Biol.* **228**, 803–812.

Bina, M. (1994). Periodicity of dinucleotides in nucleosomes derived from Simian Virus 40 chromatin. *J. molec. Biol.* **235**, 198–208.

Bloomfield, V. A., Crothers, D. M. & Tinoco, I. Jr. (2000). *Nucleic Acids, Structures, Properties, and Functions*. Sausalito, CA: University Science Books.

Bolshoy, A. (1995). CC dinucleotides contribute to the bending of DNA in chromatin. *Nature struct. Biol.* **2**, 446–448.

Bolshoy, A., Shapiro, K., Trifonov, E. N. & Ioshikhes, I. (1997). Enhancement of the nucleosomal pattern in sequences of lower complexity. *Nucl. Acids. Res.* **25**, 3248–3254.

Brukner, I., Dlakic, M., Savic, A., Susic, S., Pongor, S. & Suck, D. (1993). Evidence for opposite groove-directed curvature of GGGCCC and AAAAA sequence elements. *Nucl. Acid Res.* **21**, 1025–1029.

Brukner, I., Jurukovski, V., Konstantinovic, M. & Savic, A. (1991). Curved DNA without AA/TT dinucleotide step. *Nucl. Acid Res.* **19**, 3549–3551.

Brukner, I., Susic, S., Dlakic, M., Savic, A. & Pongor, S. (1994). Physiological concentration of magnesium ions induces a strong macroscopic curvature in GGGCCC-containing DNA. *J. molec. Biol.* **236**, 26–32.

Calladine, C. R. & Drew, H. R. (1997). *Understanding DNA: The Molecule and How it Works*, 2nd edn. San Diego, CA: Academic Press.

Cao, H., Widlund, H. R., Simonsson, T. & Kubista, M. (1998). TGGA-repeats impair nucleosome formation. *J. molec. Biol.* **281**, 253–260.

Cervi, A. R., D'Estaintot, B. L. & Hunter, W. N. (1992). Crystal and molecular structure of d(GTCTAGAC). *Acta Crystallogr.* **B48**, 714–719.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

Chiu, T. K. & Dickerson, R. E. (2000). 1 Å crystal structures of BDNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium. *J. molec. Biol.* **301**, 915–45.

Crameri, A., Raillard, S.-A., Bermudez, E. & Stemmer, W. P. C. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291.

Dong, F., Hansen, J. C. & van Holde, K. E. (1990). DNA and protein determinants of nucleosome positioning on sea urchin 5S rRNA gene sequences *in vitro*. *Proc. natn. Acad. Sci. USA* **87**, 5724–5728.

Dornberger, U., Leijon, M. & Fritzsche, H. (1999). High base pair opening rates in tracts of GC base pairs. *J. biol. Chem.* **274**, 6957–6962.

Drew, H. R. (1991). Can one measure the free energy of binding of the histone octamer to different DNA sequences by salt-dependent reconstitution? *J. molec. Biol.* **219**, 391–392.

Drew, H. R. & Travers, A. A. (1984). DNA structural variations in the *E. coli tyrT* promoter. *Cell* **37**, 491–502.

Drew, H. R. & Travers, A. A. (1985). DNA bending and its relation to nucleosome positioning. *J. molec. Biol.* **186**, 773–790.

Dumuis-Kervabon, A., Encontre, I., Etienne, G., Jauregui-Adell, J., Mery, J., Mesnier, D. & Parello, J. (1986). A chromatin core particle obtained by selective cleavage of histones by clostripain. *EMBO J.* **5**, 1735–1742.

Elgin, S. C. R. & Workman, J. L. (2000). *Chromatin Structure and Gene Expression*, pp. 328. Oxford, UK: Oxford University Press.

Feng, H.-P., Scherl, D. S. & Widom, J. (1993). Lifetime of the histone octamer studied by continuous-flow quasielastic light scattering: test of a model for nucleosome transcription. *Biochemistry* **32**, 7824–7831.

Fitzgerald, D. J. & Anderson, J. N. (1998). Unique translational positioning of nucleosomes on synthetic DNAs. *Nucl. Acids Res.* **26**, 2526–2535.

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

Fitzgerald, D. J., Dryden, G. L., Bronson, E. C., Williams, J. S. & Anderson, J. N. (1994). Conserved patterns of bending in satellite and nucleosome positioning DNA. *J. biol. Chem.* **269**, 21303–21314.

FitzGerald, P. C. & Simpson, R. T. (1985). Effects of sequence alterations in a DNA segment containing the 5S RNA gene from *Lytechinus variegatus* on positioning of a nucleosome core particle *in vitro*. *J. biol. Chem.* **260**, 15318–15324.

Fletcher, T. M. & Hansen, J. C. (1995). Core histone tail domains mediate oligonucleosome folding and nucleosomal DNA organization through distinct molecular mechanisms. *J. biol. Chem.* **270**, 25359–25362.

Hagerman, P. J. (1988). Flexibility of DNA. *Annu. Rev. Biophys. biophys. Chem.* **17**, 265–286.

Hagerman, P. J. & Ramadevi, V. A. (1990). Application of the method of Phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure, I. Computational analysis. *J. molec. Biol.* **212**, 351–362.

Hayes, J., Bashkin, J., Tullius, T. D. & Wolffe, A. P. (1991a). The histone core exerts a dominant constraint on the structure of DNA in a nucleosome. *Biochemistry* **30**, 8434–8440.

Hayes, J. J., Clark, D. J. & Wolffe, A. P. (1991b). Histone contributions to the structure of DNA in the nucleosome. *Proc. natn. Acad. Sci. USA* **88**, 6829–6833.

Hayes, J. J., Tullius, T. D. & Wolffe, A. P. (1990). The structure of DNA in a nucleosome. *Proc. natn. Acad. Sci. USA* **87**, 7405–7409.

Herzel, H., Weiss, O. & Trifonov, E. N. (1999). 10–11 bp periodicities in complete genomes reflect protein structure and DNA folding. *Bioinformatics* **15**, 187–193.

Horton, N. C. & Perona, J. J. (2000). Crystallographic snapshots along aprotein-induced DNA-bending pathway. *Proc. natn. Acad. Sci. USA* **97**, 5729–5734.

Hud, N. V. & Feigon, J. (1997). Localization of divalent metal ions in the minor groove of DNA A-tracts. *J. Am. Chem. Soc.* **119**, 5756–5757.

Hud, N. V., Sklenár, V. & Feigon, J. (1999). Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending. *J. molec. Biol.* **286**, 651–660.

Hunter, W. N., D'Estaintot, B. L. & Kennard, O. (1989). Structural variation in d(CTCTAGAG). Implications for protein–DNA interactions. *Biochemistry* **28**, 2444–2451.

Ioshikhes, I., Bolshoy, A., Derenshteyn, K., Borodovsky, M. & Trifonov, E. N. (1996). Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. *J. molec. Biol.* **262**, 129–139.

Ioshikhes, I., Bolshoy, A. & Trifonov, E. N. (1992). Preferred positions of AA and TT dinucleotides in aligned nucleosomal DNA sequences. *J. biomolec. struct. Dynam.* **9**, 1111–1117.

Johnson, T., Zhu, J. & Wartell, R. M. (1998). Differences between DNA base pair stacking energies are conserved over a wide range of ionic conditions. *Biochemistry* **37**, 12343–12350.

Kahn, J. D., Yun, E. & Crothers, D. M. (1994). Detection of localized DNA flexibility. *Nature* **368**, 163–166.

Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993a). Co-crystal structure of T BP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993b). Crystal structure of a yeast T BP/TATA-box complex. *Nature* **365**, 512520.

Klug, A., Jack, A., Viswamitra, M. A., Kennard, O., Shakked, Z. & Steitz, T. A. (1979). A hypothesis on a specific sequence-dependent conformation of DNA and its relation to the binding of the *lac*-repressor protein. *J. molec. Biol.* **131**, 669–680.

Lankas, F., Sponer, J., Hobza, P. & Langowski, J. (2000). Sequence-dependent elastic properties of DNA. *J. molec. Biol.* **299**, 695–709.

Lefevre, J., Lane, A. N. & Jardetzky, O. (1987). Solution structure of the Trp operator of *Escherichia coli* determined by NMR. *Biochemistry* **26**, 5076–5090.

Leroy, J. L., Charretier, E., Kochoyan, M. & Guéron, M. (1988). Evidence from base-pair kinetics for two types of adenine tract structures in solution, their relation to DNA curvature. *Biochemistry* **27**, 8894–8898.

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G. & Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247–1254.

Lowary, P. T. & Widom, J. (1997). Nucleosome packaging and nucleosome positioning of genomic DNA. *Proc. natn. Acad. Sci. USA* **94**, 1183–1188.

Lowary, P. T. & Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. molec. Biol.* **276**, 19–42.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997a). Structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260.

Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M. Y. & Richmond, T. J. (1997b). Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. molec. Biol.* **272**, 301–311.

Luger, K. & Richmond, T. J. (1998a). DNA binding within the nucleosome core. *Curr. Opin. struct. Biol.* **8**, 33–40.

Luger, K. & Richmond, T. J. (1998b). The histone tails of the nucleosome. *Curr. Opin. genet. Dev.* **8**, 140–146.

MacDonald, D., Herbert, K., Zhang, X., Polgruto, T. & Lu, P. (2001). Solution structure of an A-tract DNA bend. *J. molec. Biol.* **306**, 1081–98.

Maher, J. L. (1998). Mechanisms of DNA bending. *Curr. Opin. chem. Biol.* **2**, 688–694.

Mahloogi, H. & Behe, M. J. (1997). Oligoadenosine tracts favor nucleosome formation. *Biochem. biophys. Res. Comm.* **235**, 663–668.

Manning, G. S. (1978). The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **11**, 179–246.

Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. (1982). Bent helical structure in kinetoplast DNA. *Proc. natn. Acad. Sci. USA* **79**, 7664–7668.

McFail-Isom, L., Sines, C. C. & Williams, L. D. (1999). DNA structure, cations charge? *Curr. Opin. struct. Biol.* **9**, 298–304.

Meersseman, G., Pennings, S. & Bradbury, E. M. (1992). Mobile nucleosomes – a general behavior. *EMBO J.* **11**, 2951–2959.

Meersseman, G., Pennings, S. & Bradbury, M. E. (1991). Chromatosome positioning on assembled long chromatin: linker histones affect nucleosome placement on 5S rDNA. *J. molec. Biol.* **220**, 89–100.

Okonogi, T. M., Alley, S. C., Reese, A. W., Hopkins, P. B. & Robinson, B. H. (2000). Sequence-dependent dynamics in duplex DNA. *Biophys. J.* **78**, 2560–2571.

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

Owen-Hughes, T., Utley, R. T., Cote, J., Peterson, C. L. & Workman, J. L. (1996). Persistent site-specific remodeling of a nucleosome array by transient action of the SWI/SNF complex. *Science* **273**, 513–516.

Owen-Hughes, T. & Workman, J. L. (1996). Remodeling the chromatin structure of a nucleosome array by transcription factor-targeted trans-displacement of histones. *EMBO J.* **15**, 4702–4712.

Packer, M. J., Dauncey, M. P. & Hunter, C. A. (2000a). Sequence-dependent DNA structure, tetranucleotide conformational maps. *J. molec. Biol.* **295**, 85–103.

Packer, M. J., Dauncey, M. P. & Hunter, C. A. (2000b). Sequence-dependent DNA structure, dinucleotide conformational maps. *J. molec. Biol.* **295**, 71–83.

Pennings, S., Meersseman, G. & Bradbury, E. M. (1994). Linker histones H1 and H5 prevent the mobility of positioned nucleosomes. *Proc. natn. Acad. Sci. USA* **91**, 10275–10279.

Pennings, S., Meersseman, G. & Bradbury, M. E. (1991). Mobility of positioned nucleosomes on 5S rDNA. *J. molec. Biol.* **220**, 101–110.

Polach, K. J. & Widom, J. (1995). Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J. molec. Biol.* **254**, 130–149.

Polach, K. J. & Widom, J. (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J. molec. Biol.* **258**, 800–812.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

Polach, K. J. & Widom, J. (1999). Restriction enzymes as probes of nucleosome stability. *Meth. Enzymol.* **304**, 278–298.

Protacio, R. U., Polach, K. J. & Widom, J. (1997). Coupled enzymatic assays for the rate and mechanism of DNA site-exposure in a nucleosome. *J. molec. Biol.* **274**, 708–721.

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

Richmond, T. J. & Widom, J. (2000). Nucleosome and chromatin structure. In *Chromatin Structure and Gene Expression*, *Frontiers in Molecular Biology* (eds. S. Elgin, & J. L. Workman). Oxford, UK: Oxford University Press.

Roychoudhury, M., Sitlani, A., Lapham, J. & Crothers, D. M. (2000). Global structure and mechanical properties of a 10-bp nucleosome positioning motif. *Proc. natn. Acad. Sci. USA* **97**, 13608–13.

SantaLucia, J. Jr., Allawi, H. T. & Seneviratne, P. A. (1996). Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **35**, 3555–3562.

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

Schiessel, H., Widom, J., Bruinsma, R. F. & Gelbart, W. M. (2001). Polymer reptation and nucleosome repositioning. *Phys. Rev. Lett.* **86**, 4414–4417.

Schurr, J. M., Fujimoto, B. S., Wu, P. & Song, L. (1992). Fluorescence studies of nucleic acids, dynamics, rigidities, and structures. In *Topics in Fluorescence Spectroscopy* (ed. J. R. Lakowicz), vol. 3, p. 390. New York: Plenum Press.

Shore, D. & Baldwin, R. L. (1983). Energetics of DNA twisting. *J. molec. Biol.* **170**, 957–981.

Shrader, T. E. & Crothers, D. M. (1989). Artificial nucleosome positioning sequences. *Proc. natn. Acad. Sci. USA* **86**, 7418–7422.

Shrader, T. E. & Crothers, D. M. (1990). Effects of DNA sequence and histone-histone interactions on nucleosome placement. *J. molec. Biol.* **216**, 69–84.

Simpson, R. T. & Stafford, D. W. (1983). Structural features of a phased nucleosome core particle. *Proc. natn. Acad. Sci. USA* **80**, 51–55.

Simpson, R. T., Thoma, F. & Brubaker, J. M. (1985). Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: a model system for study of higher order structure. *Cell* **42**, 799–808.

Staffelbach, H., Koller, T. & Burks, C. (1994). DNA structure patterns and nucleosome positioning. *J. biomolec. struct. Dynam.* **12**, 301–325.

Stein, A. & Bina, M. (1999). A signal encoded in vertebrate DNA that influences nucleosome positioning and alignment. *Nucl. Acids Res.* **27**, 848–853.

Stemmer, W. P. C. (1994). Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391.

Strauss, J. K. & Maher III, L. J. (1994). DNA bending by asymmetric phosphate neutralization. *Science* **266**, 1829–1834.

Studitsky, V. M., Clark, D. J. & Felsenfeld, G. (1994). A histone octamer can step around a transcribing polymerase without leaving the template. *Cell* **76**, 371–382.

Studitsky, V. M., Kassavetis, G. A., Geiduschek, E. P. & Felsenfeld, G. (1997). Mechanism of transcription through the nucleosome by eukaryotic RNA polymerase. *Science* **278**, 1960–1963.

Taylor, W. H. & Hagerman, P. J. (1990). Application of the method of T4 phage DNA ligase-catalyzed ring-closure to the study of DNA structure, II. NaCl-dependence of DNA flexibility and helical repeat. *J. molec. Biol.* **212**, 363–376.

Thåström, A., Lowary, P. T., Widlund, H. R., Cao, H., Kubista, M. & Widom, J. (1999). Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. molec. Biol.* **288**, 213–229.

Ulanovsky, L., Bodner, M., Trifonov, E. N. & Choder, M. (1986). Curved DNA, design, synthesis, and circularization. *Proc. natn. Acad. Sci. USA* **83**, 862–866.

Vacano, E. & Hagerman, P. J. (1997). Analysis of birefringence decay profiles for nucleic acid helices possessing bends: the τ-ratio approach. *Biophys. J.* **73**, 306–317.

van Holde, K. E. (1989). *Chromatin*. New York: Springer-Verlag.

Weischet, W. O. (1979). On the *de novo* formation of compact oligonucleosomes at high ionic strength. Evidence for nucleosomal sliding in high salt. *Nucl. Acids Res.* **7**, 291–304.

Widlund, H. R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P. E., Kahn, J. D., Crothers, D. M. & Kubista, M. (1997). Identification and characterization of genomic nucleosome-positioning sequences. *J. molec. Biol.* **267**, 807–817.

Widlund, H. R., Kuduvalli, P. N., Bengtsson, M., Cao, H., Tullius, T. D. & Kubista, M. (1999). Nucleosome structural features and intrinsic properties of the TATAAACGCC repeat sequence. *J. biol. Chem.* **274**, 31847–31852.

Widom, J. (1984). DNA bending and kinking. *Nature* **309**, 312–313.

Widom, J. (1985). Bent DNA for gene regulation and DNA packaging. *BioEssays* **2**, 11–14.

Widom, J. (1986). Physicochemical studies of the folding of the 100 Å nucleosome filament into the 300 Å filament. *J. molec. Biol.* **190**, 411–424.

Widom, J. (1989). Toward a unified model of chromatin folding. *Annu. Rev. Biophys. biophys. Chem.* **18**, 365–395.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press

Widom, J. (1992). A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. *Proc. natn. Acad. Sci. USA* **89**, 1095–1099.

Widom, J. (1996). Short-range order in two eukaryotic genomes: relation to chromosome structure. *J. molec. Biol.* **259**, 579–588.

Widom, J. (1997a). Chromatin, the nucleosome unwrapped. *Curr. Biol.* **7**, R653–655.

Widom, J. (1997b). Transcription through nucleosomes. *Science* **278**, 1899–1901.

Widom, J. (1998). Structure, dynamics, and function of chromatin *in vitro*. *Annu. Rev. Biophys. biomol. Struct.* **27**, 285–327.

Widom, J. (1999). Equilibrium and dynamic nucleosome stability. *Meth. molec. Biol.* **119**, 61–77.

Wolffe, A. (1998). *Chromatin Structure and Function*, 3rd edn. San Diego: Academic Press.

Workman, J. L. & Kingston, R. E. (1992). Nucleosome core displacement in vitro via a metastable transcription factor-nucleosome complex. *Science* **258**, 1780–1784.

Wu, H.-M. & Crothers, D. M. (1984). Identification of the locus of sequence-directed and protein-induced DNA bending. *Nature* **308**, 509–513.

Yao, J., Lowary, P. T. & Widom, J. (1993). Twist constraints on linker DNA in the 30 nm chromatin fiber: implications for nucleosome phasing. *Proc. natn. Acad. Sci. USA* **90**, 9364–9368.

Zacharias, M. & Hagerman, P. J. (1997). Influence of static and dynamic bends on the birefringence decay profile of RNA helices: Brownian dynamics simulations. *Biophys. J.* **73**, 318–332.

https://doi.org/10.1017/S0033583501003699 Published online by Cambridge University Press