DNA structure, nucleosome placement and chromatin remodelling: a perspective

Andrew A. Travers*,†, Cédric Vaillant‡, Alain Arneodo‡ and Georgi Muskhelishvili§

*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, U.K., †Department of Biochemistry, Tennis Court Road, Cambridge CB2 1GA, U.K., ‡Laboratoire Joliot Curie, ENS de Lyon, Lyon, France, and §Jacobs University Bremen, School of Engineering and Science, Campus Ring 1, D-28759 Bremen, Germany

Abstract

A major question in chromatin biology is to what extent the sequence of DNA directly determines the genetic and chromatin organization of a eukaryotic genome? We consider two aspects to this question: the DNA sequence-specified positioning of nucleosomes and the determination of NDRs (nucleosome-depleted regions) or barriers. We argue that, in budding yeast, while DNA sequence-specified nucleosome positioning may contribute to positions flanking the regions lacking nucleosomes, DNA thermodynamic stability is a major component determinant of the genetic organization of this organism.

DNA structure and intrinsic nucleosome positioning

In a solution, the DNA polymer is stiff with a persistence length of ~50 nm corresponding to ~150 bp or 14.3 double-helical turns of B-DNA (reviewed in [1]). This is the length over which its behaviour approximates to that of a straight rod. Yet in the nucleosome core particle the same length of DNA is coiled with an average radius of curvature of 9 nm. The widths of the DNA grooves, both major and minor, are consequently considerably greater on the outside of the bent DNA than on the inside, a difference that must be accommodated by the averaged conformation of the individual base steps. A particular limiting parameter is the width of the minor groove, which on the inside of the wrapped DNA can approach 3 Å (1 Å = 0.1 nm), much lower than the average solution value of ~8 Å. This imposes strong steric constraints on the DNA, since narrowing of the minor groove is impeded by the presence of the exocyclic 2-amino group of guanine [2]. This steric constraint is consistent with the preferred sequence organization of octamer-bound sequences in which AA/TT dinucleotides repeating every 10 bp (with GC dinucleotides 5 bp out of phase) and positioned where the minor groove contacts the histone octamer in the nucleosome core particle impose a curvature favourable to nucleosome formation and stability [3,4]. The minor groove contacts are stabilized by the phased arginine residues extending into the DNA-binding ramp of the octamer [5,6]. Within the octamer-bound DNA this type of sequence organization, even if present in only a limited region (80 bp or less) [7,8], can function as a nucleosome positioning sequence and locate the octamer very precisely relative to the DNA sequence.

A major advance in the understanding of DNA sequence-specified nucleosome positioning was the direct selection in vitro of DNA sequences for binding to the histone octamer [7,9]. This experiment yielded a set of sequences with the highest determined affinity for the octamer under the conditions of selection [7]. Notable among these is the 601 sequence that positions the octamer very precisely and has become one of the workhorses of chromatin research. Importantly this sequence also possesses a high bending anisotropy with correspondingly strongly constrained degrees of bending freedom. This is apparent both by direct visualization by AFM (atomic force microscopy) and also its capacity for facile circle formation using nucleosome length DNA fragments. Like other nucleosomal DNA sequences, the 601 sequence has short A/T blocks repeating every ~10 bp interspersed with short G/C blocks in the opposite helical phase. Notably the 601 DNA sequence also has a high overall DNA stacking energy, which provides stabilizing energy for the attraction between successive base pairs in the DNA double helix [10].

The tight bending of DNA around the octamer intuitively suggests that the most deformable DNA sequences would be those most favoured by octamer binding. We have asked whether a simple sequence-dependent physical parameter, stacking energy, could be related to the position-dependent variations in nucleosome occupancy produced by the in vitro reconstitution of nucleosomes on to genomic yeast DNA under conditions similar to those used for the selection of the 601 sequence [11]. We observed, contrary to expectation, that this occupancy largely paralleled the DNA stacking energy averaged over 125 bp DNA over a short region of the genome (Figure 1). In other words, in this experiment, nucleosome occupancy appears to correlate directly with a property that would be expected to reduce the apparent flexibility of DNA.

DNA stiffness is strongly dependent on the nature of the DNA sequence, but what are the sequence parameters that determine stiffness? The fundamental observation is that, for a given DNA sequence, with the exception of those

Key words: chromatin, DNA structure, histone, nucleosome, octamer.

Abbreviations used: HMGB, High-mobility group B; NDR, nucleosome-depleted region.

*To whom correspondence should be addressed (email aat@mrc-lmb.cam.ac.uk).
that confer intrinsic curvature, when the average number of hydrogen bonds/bp is varied between two and three by base-analogue substitution there is a direct correlation between this quantity and the persistence length [12]. This implies that one major factor determining persistence length is the average base-pair rigidity, with G-C base pairs being more rigid than A-T base pairs. Similarly, when intrinsic curvature is conferred by phased oligo(dA–dT) tracts the presence of bifurcated hydrogen bonds between adjacent A-T base pairs may also act to enhance bending stiffness [13].

Locally the quantitative effect of the number of hydrogen bonds per base pair on DNA stiffness depends on sequence context and, in particular, the types of base-step. Of the three types of base-step, pyrimidine–purine (YR), purine–purine/pyrimidine–pyrimidine (RR/YY) and purine–pyrimidine (RY) the YR steps are the least, and the RY steps the most, thermally stable. Within each class of base-step the experimentally determined stacking and melting energies, and hence the stability, are related to base composition such that steps containing only A-T base pairs have on average both a lower stacking energy and melting energy than those containing only G-C base pairs [14]. Thus DNA stiffness, or persistence length, depends on both base composition and DNA sequence. The range of conformations that can be adopted by individual base-steps reveals a similar pattern. Analysis of base-step parameters in the crystal structures of DNA alone and of protein–DNA complexes reveals that, in DNA crystals, the YR steps in general occupy a greater conformational space than the RY and RR/YY steps [15]. Importantly, however, the extent to which the conformational space is increased on protein binding correlates approximately with the stacking/melting energy of the steps such that steps with a low stacking/melting energy (AA/TT, AT and TA) are, on average, considerably more deformable than steps with a high stacking/melting energy (AC/GT, GC and GA/TC). Thus deformability (defined here in terms of the range of conformations adopted in crystal structures and not the calculated elastic limits) can be considered to be inversely related to persistence length.

In the context of in vitro nucleosome reconstitution on yeast DNA [11] the direct correlation of nucleosome occupancy with stacking thus implies, counter-intuitively, that occupancy is inversely correlated with deformability. This is totally consistent with the sequence composition and organization of the regions of the lowest in vitro occupancy. These regions, especially those associated with the 5′ NDR (nucleosome-depleted region) are, in general, more AT-rich and frequently contain oligo(dA–dT) tracts [11]. The pattern of in vitro occupancy is thus consistent with previous observations that in vitro nucleosomes are preferentially associated with more GC-rich DNA [16,17]. To what extent do the AT-rich regions reflect the genetic organization of the yeast genome? In vivo NDRs immediately 5′ to the transcription start-site have previously been shown to be AT-rich [18,19], but the 3′ ends of genes are less well characterized. Nevertheless, both the 3′ and 5′ ends of genes correlate with regions of low DNA thermodynamic stability (Figure 1). This finding suggests that DNA stability specifies both gene organization, simply defined as the distinction between coding and non-coding regions, and chromatin structure in budding yeast and that the dominant effect is not the sequence-directed positioning of nucleosomes at specific sites [4], but instead the preferential exclusion of histone octamers from particular regions. Put another way, the DNA sequence imposes barriers modulating nucleosome occupancy. This principle has been experimentally verified in vitro for selected yeast genes [20] and is wholly consistent with theoretical physical models [21].

The correlation of nucleosome occupancy with DNA persistence length in the experiment of Kaplan et al. [11] runs counter to the apparent requirement for the tight bending of DNA around the histone octamer in the nucleosome. In direct contrast, Virstedt et al. [12] showed that in vitro the affinity of the octamer for different DNA sequences is anti-correlated with persistence length. One difference between the two sets of experiments is that reconstitution was performed at different temperatures; Kaplan et al. [11] at 4°C and Virstedt et al. [12] at 37/25°C. Indeed other experiments [10] show that the relative affinity of different DNA sequences for the histone octamer is not invariant, but depends on the experimental conditions, such that sequences with a high relative affinity at low temperatures are outcompeted by other sequences at higher temperatures. The preferred sequences for low-temperature reconstitution are those with high bending anisotropy, whereas those preferred at high temperature correspond to more deformable sequences. This switch can be understood in terms of configurational space [1]. At low temperatures, highly anisotropic sequences occupy a much more limited region of configurational space than more deformable sequences.

Figure 1 | Correlation of DNA stacking energy (blue) and in vitro nucleosome occupancy (red) on yeast DNA
The scale on the ordinate is the log2 of the experimental profile centred around the mean value. Nucleosome occupancy data are taken from [10].
The chromatin structure of genes

In the case of budding yeast, the same principle of DNA stability determining the genetic organization of the genome can also be applied to the fine structure of genes. In an array of nucleosomes covering a gene, not all nucleosomes are equivalent in properties. Depending on their position relative to the start of the gene, different nucleosomes carry different histone modifications and are targeted by different chromatin remodelling complexes [26,27]. Notably nucleosomes immediately adjacent to NDRs containing transcription factor–binding sites are often the first to be remodelled on gene activation [28,29]. But are differently positioned nucleosomes in a gene distinguishable on the basis of DNA sequence? An intrinsic property of isolated nucleosomes in solution is the transient and rapid unwrapping and rewinding of one or other end of the bound DNA. This property, much studied by the late Jon Widom and his co-workers, has the potential to expose target sites for proteins such as restriction enzymes and transcription factors [30–32]. Although it is not known whether unwrapping and rewinding are sequence-dependent, thermodynamic considerations suggest that unwrapping should be facilitated by sequences of low DNA stability, since such sequences would, relative to more stable sequences, favour the entropic driving of unwrapping. Two isolated examples of nucleosomes immediately downstream of an NDR suggest this possibility. In both the +1 nucleosomes of ADH2 (antidiuretic hormone 2) [29] and RNR3 (ribonucleotide reductase 3) [8], a (dA)20·(dT)20 tract is located immediately inside the 5′ end of the nucleosome. Although such sequences are not iminical to nucleosome formation [29,33], they are believed to antagonize tight bending and thus lower the affinity of a histone octamer for DNA [34]. They would thus be expected to facilitate local unwrapping. More generally an analysis of the sequence organization of positioned nucleosomes in yeast reveals, on average, an asymmetric distribution of A/T base-pair frequency and G/C base-pair frequency in +1 nucleosomes such that the A/T-rich segment immediately abuts the NDR [18]. This asymmetry is not found in nucleosomes positioned elsewhere in the gene. Again this base-composition distribution is consistent with notion that, in general, +1 nucleosomes have a greater potential to be preferentially unwrapped from the 5′ end. Although distal nucleosomes do not exhibit this asymmetry, they are, on average, flanked by short locally A/T-enriched DNA sequences [18]. Once again it is more stable DNA that is associated with the histone octamer.

Although in vitro the transient wrapping and unwrapping of nucleosomal DNA is spontaneous [30], it seems unlikely that in vivo such a process with potentially significant biological consequences would be left to the whim of thermal fluctuations. Indeed, there is evidence that unwrapping and wrapping are modulated by ancilliary proteins, including the abundant HMGB (high-mobility group B) proteins [35]. In vitro an HMGB protein, HMGB-D, facilitates asymmetric unwrapping while conversely the linker histone, H1, greatly reduces the exposure of the ends of the bound DNA [36]. This antagonistic action may be influenced by direct contacts between the two proteins [37]. The facilitation of unwrapping by HMGB-D requires the acidic C-terminal tail, which in HMGB1 interacts with the N-terminal tail of histone H3 [38]. Consistent with such a role for HMGB proteins are the observations that HMGB1 potentiates the sliding of nucleosomes in vitro catalysed by the Drosophila ACF complex [39] and the HMGB proteins of Drosophila interact genetically with both brm and osa, encoding subunits of the BAF (Brg1-associated factor) and PBAF (polybromo/Brg1-associated factor) remodelling complexes [40]. The antagonistic effects of HMGB proteins and linker histones on DNA wrapping around the histone octamer have important implications for the regulation of transcription and the maintenance of chromatin structure. HMGB proteins
Facilitation of transient DNA unwrapping by HMGB proteins

The cartoon illustrates a mechanism whereby a bound transcription factor (TF) recruits an HMGB protein, which can then act to facilitate local unwrapping of a nucleosome. The HMGB protein could equally be carried directly to its site of action by the transcription factor. The preferential positioning of DNA sequences of low stacking energy on the +1 nucleosome is indicated by a bar.

Local order within nucleosome arrays

In all organisms so far studied the nucleosomal arrays on individual genes are characterized by a strongly positioned nucleosome in the +1 position, immediately adjacent to the NDR with the positioning of subsequent distal nucleosomes becoming increasingly less well defined [18,19]. In some, but not all [41], organisms, the −1 nucleosome upstream of the NDR is also well-positioned. With the notable exception of centromeric nucleosomes [42], positioning is, in general, not conserved between individuals, but instead the organization of nucleosomes in a population may be described as a set of overlapping, mutually exclusive arrays [29,42].

Although regions of DNA instability preferentially excluding nucleosome formation in vitro accurately correlate with in vivo NDRs between genes, in vitro reconstitution with just histones and DNA fails to reproduce the in vivo pattern of organization within genes, particularly the gradation of positioning from the 5′ to the 3′ ends of arrays. Indeed accurate in vitro positioning of the nucleosomes flanking the 5′ NDRs of yeast requires a crude extract and ATP [43,44]. One necessary, but not sufficient, component is the RSC (remodelling the structure of chromatin) remodelling complex. A subsidiary question is then whether intrinsic sequence-dependent positioning [3,4] contributes to the in vivo nucleosome organization. This issue is controversial [11,45], but recent evidence suggests that the NDR flanking nucleosomes, but not, on average, more distal nucleosomes, are intrinsically specified [10,43]. In contrast, the regular spacing of downstream nucleosomes in vivo is dependent on both Iswi-containing remodelers [46], such as Mit-1 in Schizosaccharomyces pombe [41], and Chd1 closer to the 3′ end of an array [46].

The dependence of intragenic nucleosome organization on chromatin remodelers in yeast implies that this organization is actively maintained [41] and that specific remodelers are responsible for maintaining different, but possibly overlapping, sections of an array. This selectivity could be directed by specific histone modifications, such as acetylation at the 5′ end, or by particular DNA sequences as postulated by Rippe et al. [47], or, possibly, by a combination of these signals.

These results bear on the fundamental question as to whether chromatin organization, and in particular nucleosome positioning, is directed largely by DNA sequence [4] or whether nucleosome positioning is largely statistical, that is specified by barriers which essentially determine the packing of nucleosomes within an array [48]. We have argued that, in yeast, high DNA deformability is a major determinant of barriers between genes, while intrinsic sequence determinants, at least, in part, influence the positioning of the −1 and +1 nucleosomes. The positions of the more distal nucleosomes in an array are then, in general, much less dependent on DNA sequence, as predicted by the statistical model. However, whether the active maintenance of packing is wholly consistent with the statistical model is debatable [41].

The active maintenance of nucleosome arrays is also consistent with the ubiquitous distribution of linker histones and HMGB proteins. One possible reason for this distribution is that these proteins facilitate genome maintenance and are available to correct random nucleosome shuffling or, more likely, to participate in DNA repair processes [49].

Biological order and chromatin structure

The organization of chromatin in the eukaryotic nucleus and the bacterial nucleoid epitomizes the role of structural order in genetic function. In both types of chromatin, the DNA is negatively supercoiled, but this superhelicity is established by different mechanisms utilizing the energy from ATP in completely different ways [50]. In bacteria, the primary mechanism is the ATP-dependent production of negative supercoils by the topoisomerase DNA gyrase. This organizes the DNA into a dynamic pattern of coils, which can be stabilized and ordered as plectonemes or toroids by abundant DNA-binding NAPs (nucleoid-associated proteins) (Figure 3). In contrast, in eukaryotes, the initial step in chromatin packaging is the sequence-specific wrapping of DNA around the histone octamer to form nucleosomes, a process that is largely dependent on the binding energy intrinsic to the interaction. However,
a necessary condition for higher-order packaging is that spacing of the nucleosomes on a DNA molecule be regular. In principle, such spacing could result from a regular occurrence of nucleosome-positioning signals, but this is likely to confer strong constraints on the other coding functions of DNA. Instead regular spacing results from the ATP-dependent action of dedicated chromatin remodelling assemblies acting on pre-existing nucleosomes (Figure 3). These molecular motors thus act to confer structural order on chromatin and concomitantly facilitate its folding into supercoiled higher-order structures.

In his famous essay ‘What is life’, Schrödinger proposed that the behaviour of living matter was based to a large extent on an ‘order-from-order’ principle, constituted from ‘dynamical’ laws governing the interaction of single atoms and molecules [51]. We note that, in the context of chromatin structure, the specification of genetic organization of the yeast genome by DNA is in accord with this principle. Thus the structural organization encoded in a single DNA molecule is likely to be sufficient to determine the major features of the genetic organization of each chromosome.

References


