

RESEARCH ARTICLE

Open Access

Physical properties of naked DNA influence nucleosome positioning and correlate with transcription start and termination sites in yeast

Özgen Deniz^{1†}, Oscar Flores^{1†}, Federica Battistini¹, Alberto Pérez², Montserrat Soler-López¹ and Modesto Orozco^{1,3,4*}

Abstract

Background: In eukaryotic organisms, DNA is packaged into chromatin structure, where most of DNA is wrapped into nucleosomes. DNA compaction and nucleosome positioning have clear functional implications, since they modulate the accessibility of genomic regions to regulatory proteins. Despite the intensive research effort focused in this area, the rules defining nucleosome positioning and the location of DNA regulatory regions still remain elusive.

Results: Naked (histone-free) and nucleosomal DNA from yeast were digested by micrococcal nuclease (MNase) and sequenced genome-wide. MNase cutting preferences were determined for both naked and nucleosomal DNAs. Integration of their sequencing profiles with DNA conformational descriptors derived from atomistic molecular dynamic simulations enabled us to extract the physical properties of DNA on a genomic scale and to correlate them with chromatin structure and gene regulation. The local structure of DNA around regulatory regions was found to be unusually flexible and to display a unique pattern of nucleosome positioning. *Ab initio* physical descriptors derived from molecular dynamics were used to develop a computational method that accurately predicts nucleosome enriched and depleted regions.

Conclusions: Our experimental and computational analyses jointly demonstrate a clear correlation between sequence-dependent physical properties of naked DNA and regulatory signals in the chromatin structure. These results demonstrate that nucleosome positioning around TSS (Transcription Start Site) and TTS (Transcription Termination Site) (at least in yeast) is strongly dependent on DNA physical properties, which can define a basal regulatory mechanism of gene expression.

Keywords: DNA physical properties, Molecular dynamics, MNase digestion, nucleosome positioning, gene regulation, chromatin structure

Background

Genomic studies mostly provide one-dimensional information encoded in DNA, but we cannot ignore the fact that in eukaryotic organisms, DNA is packaged into chromatin structure, where DNA folds to a global compaction of at least 10^4 [1]. Genome homeostatic histone concentration ensures most of DNA to be wrapped into

nucleosomes (~75-90%) [2], which are structural units of 145-147 base pairs (bp) long, where the interaction with regulatory proteins is severely handicapped. Nucleosomes are separated from each other by short linkers (around 20 bp long in yeast) where site-specific recognition by proteins is easier. Therefore, DNA compaction has clear functional implications, since it modulates the accessibility of genomic regions to regulatory proteins. Indeed, a close relationship was established between nucleosome positioning and important regulatory signals [3], such as proximal promoters [4,5] and splicing sites [6]. Further evidence on the connection between three-dimensional

* Correspondence: modesto.orozco@irbbarcelona.org

† Contributed equally

¹Institute for Research in Biomedicine and Barcelona Supercomputing Center Joint Research Program on Computational Biology. Baldiri i Reixac 10. Barcelona 08028. Spain

Full list of author information is available at the end of the article

chromatin structure and function was obtained from genome-wide analysis of chromatin DNase I degradation profiles, which revealed a cross-link between DNase I hypersensitive sites and regulatory regions [7-9].

DNA underlying sequence has long been considered to be an important contributor to nucleosome assembly [10-13]. Crystal structures of nucleosome core particles revealed a lack of direct readout mechanisms between histones and DNA bases (the so-called base readout) [14-17] which led to the postulate that histone-DNA direct interactions are not the major determinant of nucleosome positioning [18]. Accordingly, the DNA relative affinities for nucleosome formation (e.g. high-affinity Widom601 sequence) [19] should be based on an indirect readout mechanism, where the ability of a given DNA sequence to be deformed would account for the nucleosome assembly preferences [20-24]. Nevertheless, to which extent nucleosome positioning *in vivo* is really dictated by the DNA sequence is still an issue of strong discussion [25-27].

Our group and others have provided indirect evidence highlighting the connection between DNA physical properties and chromatin organization [28-30]. In particular, we have previously reported theoretical studies showing that human promoters display very unusual stiffness properties [31]. These might affect DNA binding of regulatory proteins, either directly by hampering or favoring complex formation, or indirectly through the modulation of the chromatin structure and hence the DNA accessibility [31]. Here, we have pursued this hypothesis by a genome-wide analysis of conformational properties across yeast naked DNA using micrococcal nuclease (MNase) degradation profiles as an experimental descriptor. We were able to characterize in detail, MNase preferences for naked DNA, extending fractional information derived from small-scale experiments. These preferences (at the tetramer level) correlate with *ab initio* physical descriptors derived from molecular dynamics (MD) simulations of short DNA oligonucleotides [32-35]. This finding confirms that MNase can signal genomic regions with unusual physical properties [36,37]. Very interestingly, MNase-hypersensitive sites in naked DNA are mainly located around TSS and TTS, which supports experimentally our suggestion that those regulatory regions are signaled by physical properties. Moreover, the correlation of genome-wide nucleosome positioning profiles with MD-derived mesoscopic calculations evinces that the main mechanism by which physical properties influence gene regulation is through nucleosome positioning. Altogether, our experimental and computational integrative analysis demonstrates a clear relationship between sequence-dependent structural properties of naked DNA, accessible from first-principles simulations, and regulatory signals in chromatin structure.

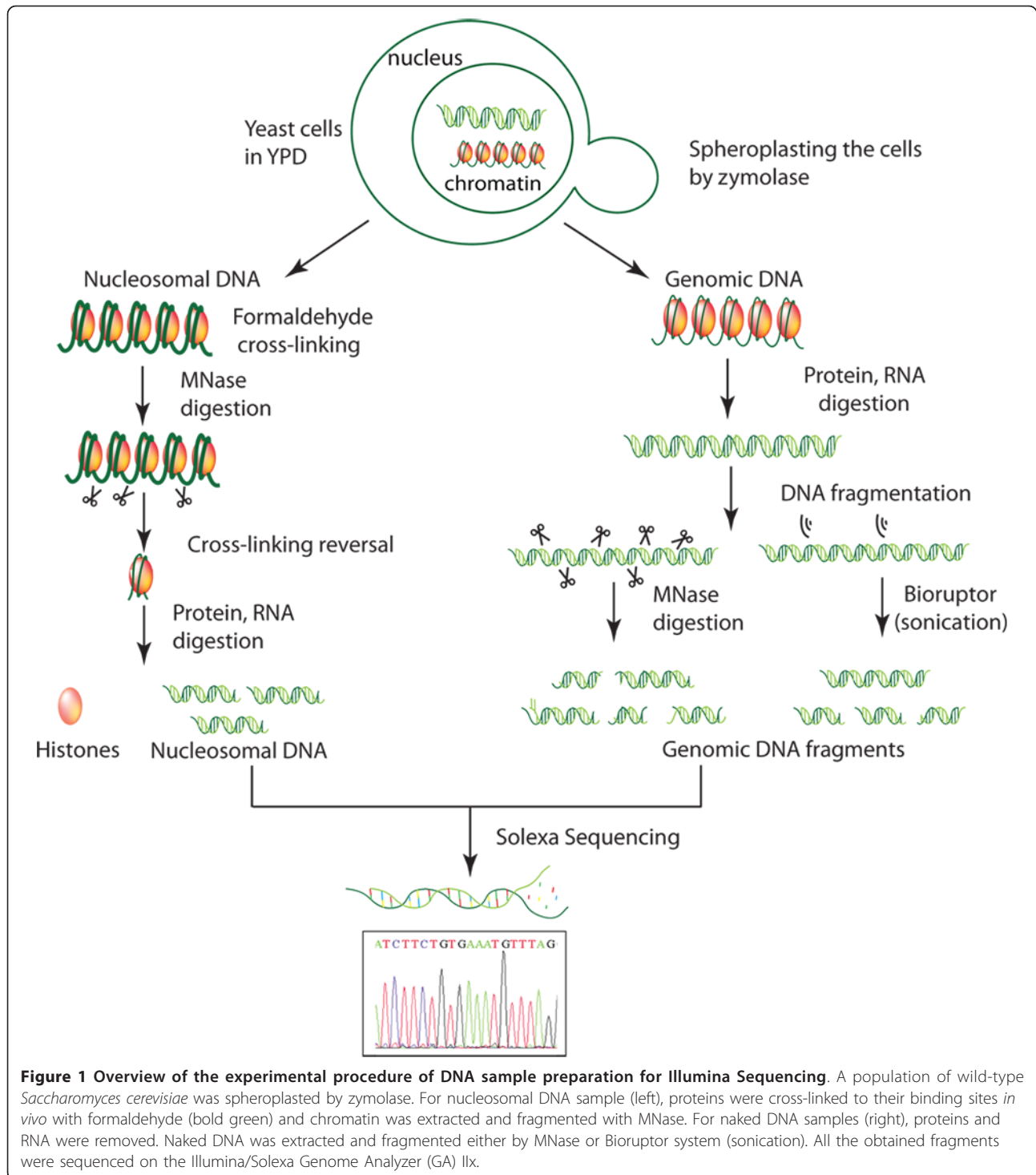
Results and Discussion

Preferential MNase cut sites

Yeast DNA fragments were prepared and sequenced following the experimental approach described in Figure 1. The analysis of our whole genome sequencing experiments, containing more than 75 million short fragments, provided a fully converged description of the MNase sequence preferences for cutting naked and nucleosomal DNAs (Table 1). As suggested from previous small-scale experiments [38], we indeed observed that in naked DNA, the enzyme preferentially cuts tetramers with a central d(A-T) step, but without the requirement of flanking dC or dG bases suggested by low-scale experiments. The high-cutting susceptibility for d(CATA)·d(TATG) tetramers found in mouse satellite DNA [39] was also detected in our massive experiments, although these tetramers were not the most predominant cutting sites. On the other hand, tetramers resistant to MNase cleavage were very diverse, except for the presence of a central purine-purine dinucleotide step (Additional File 1: Table S1). Overall, MNase displayed quite strong sequence preferences in naked DNA (up to a factor of 200) (Table 1) that could not be simply ascribed to experimental artifacts, given the fact that control experiments where DNA was fragmented by sonication did not show any marked variation in genome-wide profile (Additional File 1: Figure S1). It is noteworthy to mention that MNase resistant tetramers were different between naked and nucleosomal DNA samples, which demonstrate that the nucleosome structure protects specifically certain sequences from MNase degradation. Conversely, we found a good agreement in the preferred cutting sites between naked and nucleosomal DNAs (Table 2). This suggests that tetramer signals that are directing the first MNase cut in chromatin are intrinsic to naked DNA.

Preferential MNase degraded regions

Upon an initial endonucleotic cleavage, MNase displays an exonuclease activity that continues with the degradation of DNA [40], leading to digested areas that we identified as low coverage regions (LRs) in our sequencing experiments (see Methods). Tetramer composition along naked DNA LRs was different from the one observed in the cutting sites, suggesting that the degradation of a particular fragment does not only depend on the existence of cleavage sites in its vicinity, but also on the differential sequence preferences of endo- and exo-nuclease activities. For example, d(AAAA·TTTT) was the most abundant tetramer in naked DNA LRs, nearly four times more frequent than expected ($p < 10^{-8}$), while the same tetramer was rarely present at primary cutting sites (1/4 than expected, $p < 10^{-7}$, Additional File 1: Table S1). Moreover, the tetramer composition was very similar in both naked and



nucleosomal LR and in the common low regions (CLRs) (definitions in Additional File 1: Additional Methods) indicating that sequence susceptibility for MNase degradation in nucleosomal DNA was not exclusively dependent on the chromatin structure, but was also related to the intrinsic properties of naked DNA (Table 2).

Low coverage regions and physical properties

The MNase tetramer preferences (Tables 1 and 2) are so diverse that they cannot be explained in terms of direct DNA base reading. Analysis of MD-derived physical properties [32,41] revealed that primary cutting sites are characterized by high flexibility (affecting roll and tilt parameters)

Table 1 Frequency of MNase-preferred tetramers at the cutting sites

Naked DNA	ratio	p-val	Nucleosomal DNA	ratio	p-val
TATA.TATA	13.28	$< 10^{-18}$	CTAG.CTAG	4.07	$< 10^{-18}$
ATAG.CTAT	8.45	$< 10^{-18}$	ATAG.CTAT	3.93	$< 10^{-18}$
CTAA.TTAG	7.90	$< 10^{-18}$	CAAG.CTTG	3.57	$< 10^{-18}$
CTAG.CTAG	6.80	$< 10^{-18}$	CTTA.TAAG	3.52	$< 10^{-18}$
ATTA.TAAT	5.74	$< 10^{-18}$	CATG.CATG	3.42	3.01×10^{-4}
CATA.TATG	5.62	$< 10^{-18}$	CATA.TATG	3.11	$< 10^{-18}$
ATAA.TTAT	5.14	$< 10^{-18}$	CTAA.TTAG	3.00	$< 10^{-18}$
CTTA.TAAG	4.92	$< 10^{-18}$	CTAC.GTAG	2.98	$< 10^{-18}$
TTAA.TTAA	4.64	$< 10^{-18}$	ATTG.CAAT	2.96	$< 10^{-18}$
ATAT.ATAT	4.52	$< 10^{-18}$	AAAG.CTTT	2.82	$< 10^{-18}$
TAAA.TTTA	3.48	$< 10^{-18}$	CTTC.GAAG	2.79	$< 10^{-18}$
ATTG.CAAT	3.25	$< 10^{-18}$	AATG.CATT	2.50	$< 10^{-18}$
GTAA.TTAC	2.64	1.01×10^{-4}	CATC.GATG	2.24	6.03×10^{-4}
ATAC.GTAT	2.39	2.01×10^{-4}	CAAC.GTTG	2.19	10^{-3}
			CAAA.TTTG	2.17	$< 10^{-18}$

Experimentally detected and expected frequency ratios of MNase-preferred tetramers at the cutting sites for naked (left) and nucleosomal (right) DNAs. Displayed tetramers are observed in at least two-fold higher frequency than expected, with a statistically significant difference ($p < 10^{-3}$) (Supplementary Methods). Ratios for d(TAAA)-d(TTTA) and d(GTAA)-d(TTAC) tetramers in nucleosomal DNA are 1.39 ($p < 0.08$) and 1.6 ($p < 0.03$) respectively.

and wide opening in the major groove (high roll values) at the equilibrium geometry (Additional File 1: Figure S2). Furthermore, the total dinucleotide-based stiffness parameter k_{total} (see Methods for definition) unveiled that LRs (in both naked and nucleosomal DNA) are located in regions with large variations in flexibility, where an extremely flexible site is surrounded by stiff motifs (Figure 2 and Additional File 1: Figure S3). Remarkably, the same results were obtained when we considered the parameters fitted to the tetramer level by the ABC consortium [42] confirming the robustness of our conclusions. Dinucleotide and tetranucleotide data (see below and Additional File 1: Additional Methods) are available upon request and are incorporated in our DNAlive webserver (<http://mmb.pcb.ub.es/DNAlive>),

Nucleosome positioning and gene structure

As previously suggested by other groups ([10,11,43-50]) MNase resistant regions in nucleosomal DNA are mainly concentrated at the beginning of transcribed regions (Figure 3). Whereas very sensitive regions (i.e. LRs) were mostly identified at regulatory regions, either upstream of transcription start sites (TSSs) (Figure 3A) or downstream of transcription termination sites (TTSs) (Figure 3B). Additional differential regions, such as MNase resistant areas upstream of TTSs or downstream of TSSs, were less certain than the major signals mentioned above (Figure 3). Considering that MNase degradation profiles were only dependent on nucleosome positioning [51-55], we could locate more than 33,000 “well positioned” and around 48,000 “fuzzy” nucleosomes along the yeast genome (see

Table 2 Frequency of tetramers in MNase-digested LRs and CLRs

Naked DNA	ratio	p-val	Nucleosomal DNA	ratio	p-val	Common low regions (CLR)	ratio	p-val
AAAA.TTTT	3.87	$< 10^{-18}$	TATA.TATA	4.06	$< 10^{-18}$	AAAA.TTTT	4.48	$< 10^{-18}$
TAAA.TTTA	2.38	$< 10^{-18}$	ATAT.ATAT	3.09	$< 10^{-18}$	TATA.TATA	3.18	$< 10^{-18}$
TATA.TATA	2.38	9.05×10^{-4}	AAAA.TTTT	2.91	$< 10^{-18}$	TAAA.TTTA	2.67	$< 10^{-18}$
AAAT.ATTT	2.16	$< 10^{-18}$	ATAA.TTAT	2.21	$< 10^{-18}$	ATAA.TTAT	2.62	$< 10^{-18}$
ATAA.TTAT	2.13	$< 10^{-18}$	AATA.TATT	2.08	$< 10^{-18}$	ATAT.ATAT	2.57	$< 10^{-18}$
TTAA.TTAA	2.10	7.54×10^{-3}	ATTA.TAAT	1.99	10^{-4}	AATA.TATT	2.43	$< 10^{-18}$
AATA.TATT	2.02	$< 10^{-18}$	TAAA.TTTA	1.84	7.04×10^{-4}	TTAA.TTAA	2.29	3.22×10^{-3}
ATAT.ATAT	2.00	4.62×10^{-3}	AAAT.ATTT	1.62	4.22×10^{-3}	AAAT.ATTT	2.27	$< 10^{-18}$
AATT.AATT	1.84	5.53×10^{-3}				ATTA.TAAT	2.15	$< 10^{-18}$
ATTA.TAAT	1.79	3.62×10^{-3}				AATT.AATT	1.81	1.30×10^{-2}
GAAA.TTTC	1.45	3.44×10^{-2}						

Experimentally detected and expected frequency ratios of different tetramers in MNase-digested LRs for naked (left) and nucleosomal (center) DNAs, and in CLRs (right). Displayed tetramers show a significant enrichment ($p < 0.05$) respect to genome average (Additional File 1: Additional Methods).

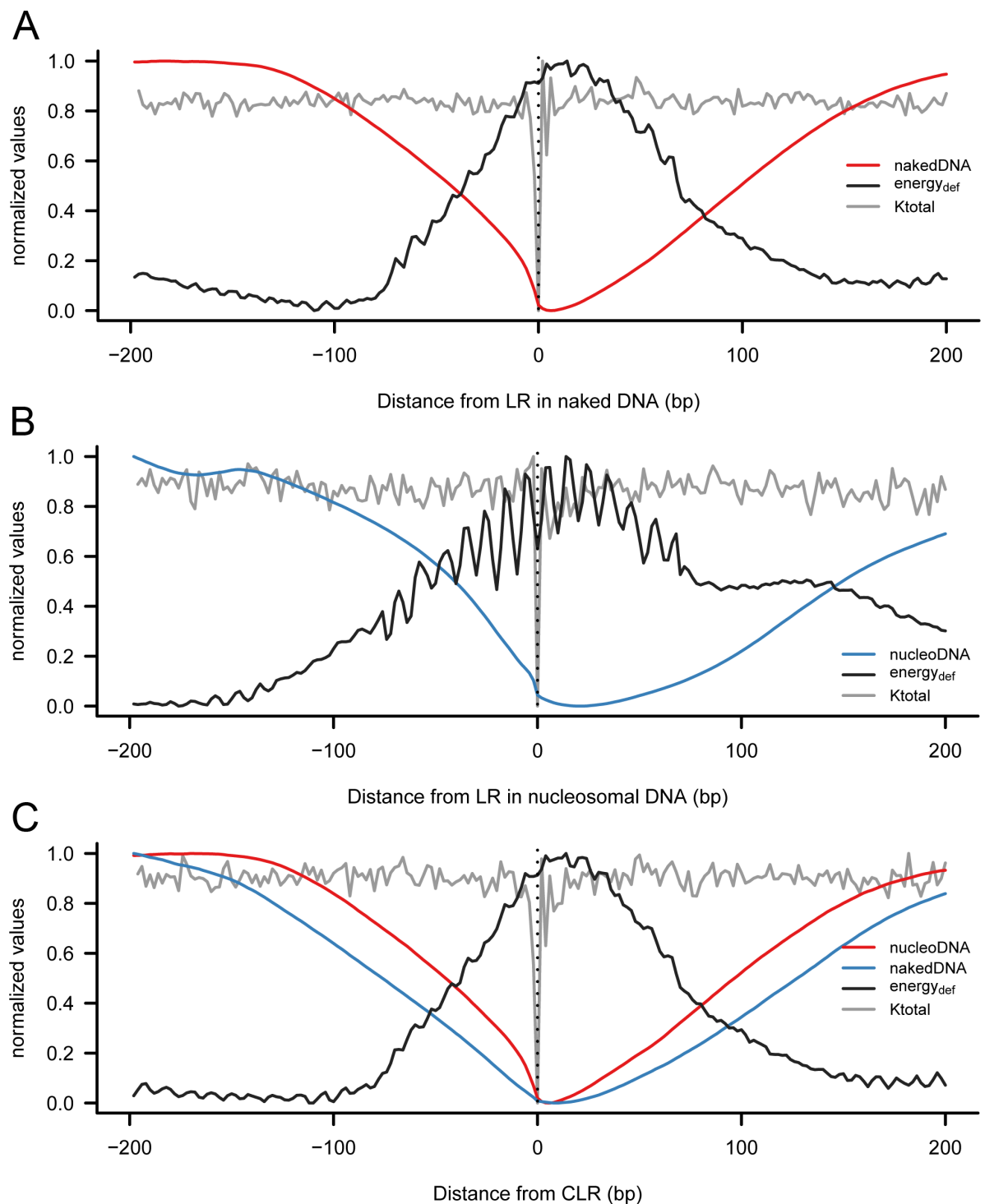
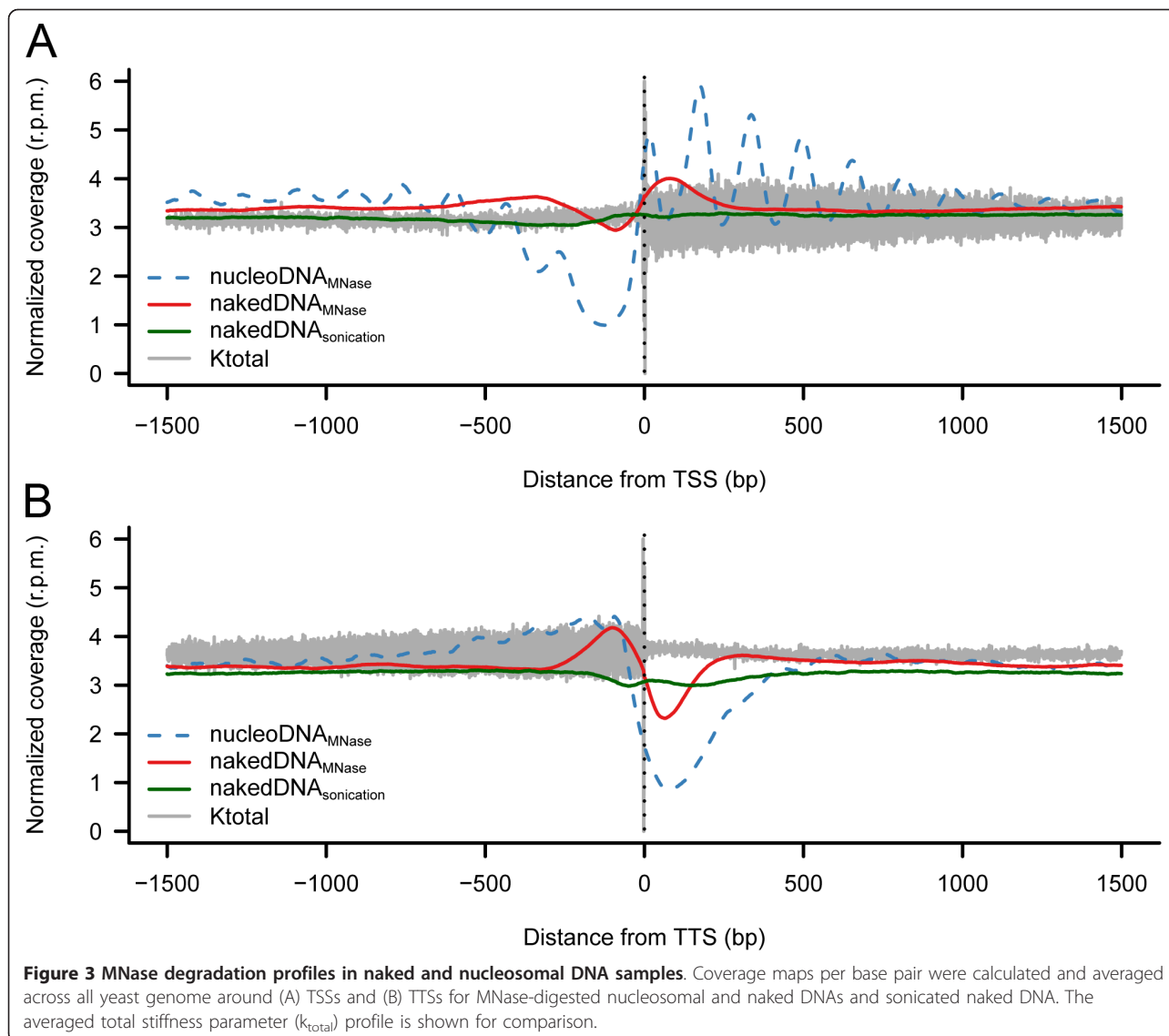


Figure 2 Stiffness, deformation energy and coverage profiles in low coverage regions. Total stiffness parameter (k_{total}), deformation energy and coverage maps were calculated and averaged across all yeast genome, around (A) LRs in naked DNA, (B) LRs in nucleosomal DNA and (C) CLRs in nucleosomal and naked DNA. Deformation energy describes the energetic cost of wrapping a 147 bp DNA fragment into the nucleosome conformation (higher the energy, more the nucleosome location is disfavored). All values are normalized (scaled in the range 0-1) to facilitate analysis and comparisons.



Methods and Additional File 1: Additional Methods for details). Notwithstanding, the surprising similarity observed between nucleosomal and naked MNase profiles (not detected in the sonication profiles; Figure 3) indicate that nucleosomal degradation profiles might not only reflect nucleosome positioning, but also the intrinsic susceptibility of naked DNA to MNase digestion [56]. This is clearly illustrated in the reduction of nucleosome positioning signals in Additional File 1: Figure S4, when nucleosomal MNase degradation maps are corrected with the naked DNA ones (see Methods). However, Figure 3 clearly demonstrates that strong nucleosome depletion or “well positioned” nucleosome signals, such as upstream of TSS and downstream of TTS, are not affected by the correction of intrinsic MNase susceptibility biases. These observations thus support most of the claims in previously

reported nucleosome positioning studies about the connection between nucleosome organization and gene regulation [10,11,43-50] and toned down some recent criticisms about the neglect of the MNase bias.

Physical properties, nucleosome positioning and regulatory regions

The analysis of MD-derived descriptors of naked DNA showed that key genomic regions, such as at TSSs and TTSs, were marked by unusual flexibility properties (Additional File 1: Figure S5) [29]. Since those regions are strongly nucleosome depleted, we hypothesized that unusual physical properties might control nucleosome positioning in those regions, which in turn would affect the DNA accessibility to regulatory proteins and ultimately impact gene regulation. To verify this hypothesis,

we computed the deformation energy required to wrap a DNA sequence around a histone octamer by using a simple elastic energy function based on the MD-derived physical descriptors (see Methods). Figure 2 clearly shows that CLRs, which are nucleosome depleted, correlate with high deformation energy confirming that in these regions it is more difficult to wrap DNA around a nucleosome core. It is interesting to note (Figure 2) that, often, 147 bps regions with high deformation energy contain a high flexible (4 mer) step, indicating that global concepts about the impact of point flexibility on chromatin organization needs to be considered with caution. Overall, results in Figure 2 strongly suggest that the properties that make a DNA segment a good substrate for MNase are also those that avoid DNA wrapping around a nucleosome. In fact, very encouragingly, deformation energies for wrapping a DNA around a nucleosome core particle can accurately predict *in vivo* nucleosome distribution around TSSs and TTSs in yeast (Additional File 1: Figure S6). These results suggest that, without dismissing the importance of cellular mechanisms for controlling chromatin structure, very important details of the nucleosome organization around TSS and TTS can be rationalized considering physical properties of the naked DNA sequence.

Conclusions

The molecular mechanisms that regulate gene expression in eukaryotic organisms are very diverse and complex. Considering the large amount of basal gene expression in cells, it is difficult to believe that regulation is entirely modulated by specific direct-readout mechanisms, where regulatory proteins would directly interact with DNA through hydrogen bonds in the major/minor grooves and compete with histones [57]. Thus, a combination of direct and indirect readout mechanisms is required to achieve the correct interaction affinity and specificity [58]. Direct mechanism can be very specific, but has implicitly a large energetic cost. Indirect mechanism is obviously less precise, but implies no energy cost for the cell and might be useful in cases where no specific regulation of the gene is needed.

Genome-wide sequencing of MNase treated nucleosomal DNA shows that key regulatory regions such as the start and the end of transcribed sites, which have been traditionally interpreted as nucleosome depletion sites, are actually signaled by a differential pattern of MNase susceptibility in naked DNA. This observation, which could initially raise some concerns, does not contradict previously reported nucleosome maps where MNase degradation was supposed to only reflect nucleosome positioning [10,11,43-50,59-61]. Indeed, nucleosomal degradation profiles corrected with naked DNA data

maintained major nucleosome positioning signals, such as nucleosome depletion upstream of TSS or downstream of TTS, thereby supporting previous MNase based nucleosome positioning conclusions [62,63]. Nevertheless, our experiments with nucleosomal and naked DNA suggest caution in the interpretation of nucleosome positioning signals in regions with anomalous MNase degradation profile.

The high correlation of MNase degradation profiles of nucleosomal and naked DNA and with unusual stiffness properties indicates that (without dismissing the importance of the cellular machinery for control of chromatin structure) intrinsic physical properties of naked DNA determine major nucleosome location signals in yeast, especially those at TSS and TTS. This hypothesis is indirectly supported by very recent studies [64], where nucleosome positioning signals are clearly identified after genome-wide nucleosome reconstitution *in vitro*.

Essential regions for gene regulation like TSSs and TTSs are characterized by unusual physical properties that disfavor positioning of nucleosomes and therefore expose DNA to interaction with regulatory proteins. This property of regulatory regions is quite general across the genome. The genes with well-defined CLRs at regulatory regions did not differ from those with more diffuse signals in terms of Gene Ontology analysis [65], promoter architecture, transcription rate or their dependence on regulatory proteins. Accordingly, we can infer that unusual physical properties are perhaps a general property of gene regulatory regions that can confer a basal mechanism of gene regulation. Furthermore, we speculate that additional specific signals were evolutionarily conferred to enable proteins to directly read DNA sequences in those genes that might require a finer regulatory mechanism.

All conclusions drawn here have been derived from the analysis of yeast genome and thus concerns exist whether they can be validated for higher eukaryotes with a different sequence composition at regulatory regions. Therefore, we compared the sequence-dependent physical properties of the *Drosophila melongaster* genome with the high-resolution genomic nucleosome map available [66]. The comparative analysis is shown in Additional File 1: Figure S7, which revealed that coverage and stiffness profiles at TSS are conserved between such distant organisms like yeast and fruit fly [67]. Extension of conclusions to vertebrates is more complex, due to the higher impacts of epigenetic factors. Nevertheless unusual physical properties are also remarkable in human promoters [31]. All these findings prompt us to believe in the general conclusion that nucleosome-depleted and enriched regions are signalled by unusual physical properties, which define the core of an evolutionarily conserved mechanism of gene regulation.

Methods

DNA sample preparation

Both nucleosomal and genomic (histone-free) DNA were isolated from *Saccharomyces cerevisiae* BY4741 strain, (an outline of the experimental procedure is presented in Figure 1, adapted from a previously described procedure) [50]. For nucleosomal DNA preparation, exponentially growing yeast cells were first cross-linked with formaldehyde, spheroplasted with zymolase and finally subjected to a MNase partial digestion to generate core nucleosomes containing DNA fragments of around 147 bp (see Additional File 1: Additional Methods). Agarose gel electrophoresis confirmed that more than 90% of the isolated DNA corresponded to mono-nucleosomal fragments (Additional File 1: Figure S8). Naked DNA was prepared from overnight grown culture by spheroplasting the cells with zymolase and subsequently incubated with SDS and RNase for an efficient protein and RNA depletion. DNA samples were analyzed by fluorometry and UV spectrophotometry to ensure that proteins and RNA were completely removed from the DNA (Additional File 1: Figure S8). The purified DNA was then sheared following two different approaches (MNase digestion and sonication) that yielded fragments of approximately 150 bp in both cases (additional details in Additional File 1: Additional Methods). To guarantee that results were not dependent on MNase concentration, experiments were repeated using two MNase concentrations (0.04 and 0.12 U) (data not shown, but available upon request). The original, the corrected degradation maps and MNase cutting preferences did not show any differences between the two MNase concentrations. Accordingly in this study only the data obtained with high MNase concentration are reported. These degradation conditions ensure that in nucleosomal DNA experiments only the linker DNA is digested, most of the degraded sample corresponds to mononucleosomes, and integrity of DNA bound to histones is preserved.

DNA sequencing

Cleaved DNA samples were sequenced on the Illumina/Solexa Genome Analyzer (GA) IIx to generate reads of 54 bp length. Data were processed with standard GA base calling pipeline to convert initial raw images into sequences. All sequencing experiments were done in duplicates. Pooled data highly converged, as the reproducibility of individual experiments was very large in all cases (Additional File 1: Additional Methods). Reads are available in Short Read Archive of NCBI under Accession Number SRA030453.

Mapping reads to genome

GA reads were aligned to the *Saccharomyces cerevisiae* reference genome using the Bowtie software [68],

allowing up to three mismatches per read. Short reads with multiple alignments were mapped to all possible places, thus avoiding the generation of artificial depleted regions. Largely over-represented reads were eliminated to reduce PCR amplification artifacts. Coverage values were calculated for each position on the genome, normalized and converted to reads per million (r.p.m.) (Additional File 1: Additional Methods).

Nucleosome calling and MNase bias correction

Nucleosomes were defined as regions flanking ± 74 bases the peaks detected in the coverage maps. Peak detection was performed using a recently published algorithm *nucleR* [69] (Additional File 1: **Additional Methods**). Correction of nucleosomal digestion profiles was done by using the degradation profiles obtained for naked DNA as background (Additional File 1: Additional Methods).

Identification of cut sites and low coverage regions

MNase cut sites were extracted from mapped reads, taking two bases upstream and the two bases downstream of every read end. Low coverage regions (LRs) account for regions where MNase degradation has been especially extensive. Low coverage regions (LRs) were detected in both nucleosomal and naked DNA as genomic segments with non-zero coverage below certain thresholds (Additional File 1: Additional Methods).

Derivation of physical descriptors

Parameters describing the equilibrium geometry and deformability of naked DNA were derived from long atomistic MD simulations of a reduced number of short oligonucleotides (displaying all unique dinucleotide or tetranucleotide steps) in solvent water by using a newly developed force-field [70]. Base pair and base step structures of DNA can be described as a set of three translations (shift, slide and rise) and three rotations (tilt, roll and twist), while the deformability along those directions can be described by their associated stiffness constants (K_i), considering the equilibrium conformation as the origin of energies following the approach suggested by Lankas and others [34,32,42,41]. In brief, the covariance matrix defining the deformability of helical parameters of a given DNA segment (for example a dinucleotide step) is computed from the ensemble of molecular dynamics simulations and inverted to determine 6×6 stiffness matrix for each fragment (for example each of the ten unique dinucleotide steps, or the ten dinucleotide steps adapted to all tetramer environments). Pure stiffness constant associated to individual helical deformations (k_{tilt} , k_{roll} , k_{shift} , k_{tilt} , k_{rise} and k_{slide}) are taken from the diagonal of the matrix. K_{total} is obtained as the product of the six pure stiffness constants and gives a rough global estimate of the flexibility of each base pair step (Additional File 1: Additional Methods)

Calculation of nucleosome deformation energy

The energetic cost of wrapping a 147 bp DNA fragment was determined by using an harmonic approach: $E = 0.5 X^T \Theta X$; where Θ is the stiffness matrix derived from MD simulations; X (or X^T) is the deformation vector (or its transposed), given by translating a relaxed DNA fiber into the coiled nucleosome core DNA conformation as described for averaging and smoothing of X-ray structures (Additional File 1: Additional Methods). Note that no training is performed and therefore deformation energies are fully *ab initio* descriptors. The scripts used to perform deformation energy calculations are available upon request to the authors.

Additional material

Additional file 1: Additional Methods, Additional Figures and Additional Tables. PDF document with detailed methods and additional results.

List of abbreviations

MNase: micrococcal nuclease; DNase I: DNA nuclease I; MD: molecular dynamics; LR: low coverage region; CLR: common low coverage region; TSS: transcription start site; TTS: transcription termination site; RNase: RNA nuclease.

Acknowledgements

We thank D. Rossell and E. Planet for their support in pre-processing the high-throughput sequencing data and F. Azorín for assistance in the experimental assays. This work was supported by the Spanish Ministry of Science and Innovation (BIO2009-10964 and Consolider E-Science), Instituto de Salud Carlos III (INB-Genoma España and COMBIOMED RETICS) and Fundación Marcelino Botín. AP is an EMBO fellowship holder (ALTF 1107). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author details

¹Institute for Research in Biomedicine and Barcelona Supercomputing Center Joint Research Program on Computational Biology. Baldiri i Reixac 10. Barcelona 08028. Spain. ²Lauffer center for physical and quantitative Biology, Stony Brook University, Stony Brook, NY 11794, USA. ³Department of Biochemistry and Molecular Biology. University of Barcelona. Avinguda Diagonal. Barcelona 08028. Spain. ⁴Instituto Nacional de Bioinformática. Parc Científic de Barcelona. Baldiri i Reixac 10. Barcelona 08028. Spain.

Authors' contributions

The authors have made the following declarations about their contributions: MO had the idea and make the general planning of the study. Conceived and designed the experiments: OD MS MO. Performed the experiments: OD. Analyzed the data: OF FB AP MO. All authors contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Competing financial interests

The author(s) declare that they have no competing interests

Received: 21 June 2011 Accepted: 7 October 2011

Published: 7 October 2011

References

- Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, Bickmore WA: Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* 2004, **118**:555-566.

2. Holde Kvan, Zlatanova J: Unusual DNA structures, chromatin and transcription. *BioEssays: news and reviews in molecular, cellular and developmental biology* 1994, **16**:59-68.
3. Jiang C, Pugh BF: Nucleosome positioning and gene regulation: advances through genomics. *Nature reviews Genetics* 2009, **10**:161-72.
4. Lodhi N, Ranjan A, Singh M, Srivastava R, Singh SP, Chaturvedi CP, Ansari SA, Sawant SV, Tuli R: Interactions between upstream and core promoter sequences determine gene expression and nucleosome positioning in tobacco PR-1a promoter. *Biochimica et biophysica acta* 2008, **1779**:634-44.
5. Choi JK, Kim Y-J: Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nature genetics* 2009, **41**:498-503.
6. Schwartz S, Ast G: Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. *The EMBO Journal* 2010, **29**:1629-1636.
7. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE: High-resolution mapping and characterization of open chromatin across the genome. *Cell* 2008, **132**:311-22.
8. Ling G, Sugathan A, Mazor T, Fraenkel E, Waxman DJ: Unbiased, Genome-wide in vivo Mapping of Transcriptional Regulatory Elements Reveals Sex Differences in Chromatin Structure Associated with Sex-specific Liver Gene Expression. *Molecular and cellular biology* 2010, MCB.00601-10.
9. Micheli E, Martufi M, Cacchione S, Santis PDe, Savino M: Self-organization of G-quadruplex structures in the hTERT core promoter stabilized by polyaminic side chain perylene derivatives. *Biophysical chemistry* 2010.
10. Ioshikhes IP, Albert I, Zanton SJ, Pugh BF: Nucleosome positions predicted through comparative genomics. *Nature genetics* 2006, **38**:1210-5.
11. Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, Segal E: The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 2009, **458**:362-6.
12. Chung H-R, Vingron M: Sequence-dependent nucleosome positioning. *Journal of molecular biology* 2009, **386**:1411-22.
13. Cui F, Zhurkin VB: Structure-based analysis of DNA sequence patterns guiding nucleosome positioning in vitro. *Journal of biomolecular structure & dynamics* 2010, **27**:821-41.
14. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ: Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 1997, **389**:251-60.
15. Richmond TJ, Davey CA: The structure of DNA in the nucleosome core. *Nature* 2003, **423**:145-50.
16. Hall MA, Shundrovsky A, Bai L, Fulbright RM, Lis JT, Wang MD: High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nature structural & molecular biology* 2009, **16**:124-9.
17. Makde RD, England JR, Yennawar HP, Tan S: Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* 2010, **467**:562-566.
18. Zhang Y, Moqtaderi Z, Rattner BP, Euskirchen G, Snyder M, Kadonaga JT, Liu XS, Struhl K: Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nature structural & molecular biology* 2009, **16**:847-52.
19. Lowary PT, Widom J: New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of molecular biology* 1998, **276**:19-42.
20. Tolstorukov MY, Colasanti AV, McCandlish DM, Olson WK, Zhurkin VB: A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. *Journal of molecular biology* 2007, **371**:725-38.
21. De Santis P, Morosetti S, Scipioni A: Prediction of nucleosome positioning in genomes: limits and perspectives of physical and bioinformatic approaches. *Journal of biomolecular structure & dynamics* 2010, **27**:747-64.
22. Trifonov EN, Sussman JL: The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proceedings of the National Academy of Sciences of the United States of America* 1980, **77**:3816-20.
23. Xu F, Olson WK: DNA architecture, deformability, and nucleosome positioning. *Journal of biomolecular structure & dynamics* 2010, **27**:725-39.
24. Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B: The role of DNA shape in protein-DNA recognition. *Nature* 2009, **461**:1248-53.
25. Travers A, Hiriart E, Churcher M, Caserta M, Di Mauro E: The DNA sequence-dependence of nucleosome positioning in vivo and in vitro. *Journal of biomolecular structure & dynamics* 2010, **27**:713-24.

26. Arya G, Maitra A, Grigoryev SA: **A structural perspective on the where, how, why, and what of nucleosome positioning.** *Journal of biomolecular structure & dynamics* 2010, **27**:803-20.
27. Trifonov EN: **Nucleosome positioning by sequence, state of the art and apparent finale.** *Journal of biomolecular structure & dynamics* 2010, **27**:741-6.
28. Abeel T, Saeys Y, Bonnet E, Rouzé P, Van de Peer Y: **Generic eukaryotic core promoter prediction using structural features of DNA.** *Genome research* 2008, **18**:310-23.
29. Goñi JR, Fenollosa C, Pérez A, Torrents D, Orozco M: **DNAlive: a tool for the physical analysis of DNA at the genomic scale.** *Bioinformatics (Oxford, England)* 2008, **24**:1731-2.
30. Miele V, Vaillant C, D Aubenton-Carafa Y, Thermes C, Grange T: **DNA physical properties determine nucleosome occupancy from yeast to fly.** *Nucleic acids research* 2008, **36**:3746-56.
31. Goñi JR, Pérez A, Torrents D, Orozco M: **Determining promoter location based on DNA structure first-principles calculations.** *Genome biology* 2007, **8**:R263.
32. Lankas F, Sponer J, Langowski J, Cheatham TE: **DNA basepair step deformability inferred from molecular dynamics simulations.** *Biophysical journal* 2003, **85**:2872-83.
33. Morozov AV, Fortney K, Gaykalova DA, Studitsky VM, Widom J, Siggia ED: **Using DNA mechanics to predict in vitro nucleosome positions and formation energies.** *Nucleic acids research* 2009, **37**:4707-22.
34. Olson WK: **DNA sequence-dependent deformability deduced from protein-DNA crystal complexes.** *Proceedings of the National Academy of Sciences* 1998, **95**:11163-11168.
35. Araúzo-Bravo MJ, Fujii S, Kono H, Ahmad S, Sarai A: **Sequence-dependent conformational energy of DNA derived from molecular dynamics simulations: toward understanding the indirect readout mechanism in protein-DNA recognition.** *Journal of the American Chemical Society* 2005, **127**:16074-89.
36. Yamasaki S, Terada T, Shimizu K, Kono H, Sarai A: **A generalized conformational energy function of DNA derived from molecular dynamics simulations.** *Nucleic acids research* 2009, **37**:e135.
37. Fujii S, Kono H, Takenaka S, Go N, Sarai A: **Sequence-dependent DNA deformability studied using molecular dynamics simulations.** *Nucleic acids research* 2007, **35**:6063-74.
38. Flick JT, Eissenberg JC, Elgin SC: **Micrococcal nuclease as a DNA structural probe: its recognition sequences, their genomic distribution and correlation with DNA structure determinants.** *Journal of molecular biology* 1986, **190**:619-33.
39. Hörz W, Altenburger W: **Sequence specific cleavage of DNA by micrococcal nuclease.** *Nucleic acids research* 1981, **9**:2643-58.
40. Alexander M, Heppel LA, Hurwitz J: **The purification and properties of micrococcal nuclease.** *The Journal of biological chemistry* 1961, **236**:3014-9.
41. Pérez A, Lankas F, Luque FJ, Orozco M: **Towards a molecular dynamics consensus view of B-DNA flexibility.** *Nucleic acids research* 2008, **36**:2379-94.
42. Lavery R, Zakrzewska K, Beveridge D, Bishop TC, Case DA, Cheatham T, Dixit S, Jayaram B, Lankas F, Laughton C, Maddocks JH, Michon A, Osman R, Orozco M, Perez A, Singh T, Spackova N, Sponer J: **A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA.** *Nucleic acids research* 2010, **38**:299-313.
43. Oszolak F, Song JS, Liu XS, Fisher DE: **High-throughput mapping of the chromatin structure of human promoters.** *Nature biotechnology* 2007, **25**:244-8.
44. Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J, Segal E: **Distinct modes of regulation by chromatin encoded through nucleosome positioning signals.** *PLoS computational biology* 2008, **4**:e1000216.
45. Bai L, Morozov AV: **Gene regulation by nucleosome positioning.** *Trends in genetics TIG* 2010, **26**:476-83.
46. Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I, Pugh BF: **A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome.** *Genome research* 2008, **18**:1073-83.
47. Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, Wang J-PZ, Widom J: **A genomic code for nucleosome positioning.** *Nature* 2006, **442**:772-8.
48. Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C: **A high-resolution atlas of nucleosome occupancy in yeast.** *Nature genetics* 2007, **39**:1235-44.
49. Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF: **Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome.** *Nature* 2007, **446**:572-6.
50. Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ: **Genome-scale identification of nucleosome positions in *S. cerevisiae*.** *Science (New York, N.Y.)* 2005, **309**:626-30.
51. Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ: **Genome-scale identification of nucleosome positions in *S. cerevisiae*.** *Science (New York, N.Y.)* 2005, **309**:626-3010, 1126/science.1112178.
52. Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C: **A high-resolution atlas of nucleosome occupancy in yeast.** *Nature genetics* 2007, **39**, 1235-4410.1038/ng2117.
53. Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J, Segal E: **Distinct modes of regulation by chromatin encoded through nucleosome positioning signals.** *PLoS computational biology* 2008, **4**:e100021610, 1371/journal.pcbi.1000216.
54. Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF: **Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome.** *Nature* 2007, **446**:572-610.1038/nature05632.
55. Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, Segal E: **The DNA-encoded nucleosome organization of a eukaryotic genome.** *Nature* 2009, **458**:362-610, 1038/nature07667.
56. Chung H-R, Dunkel I, Heise F, Linke C, Krobitsch S, Ehrenhofer-Murray AE, Sperling SR, Vingron M: **The effect of micrococcal nuclease digestion on nucleosome positioning data.** *PLoS one* 2010, **5**:e1575410, 1371/journal.pone.0015754.
57. Rando OJ, Ahmad K: **Rules and regulation in the primary structure of chromatin.** *Current opinion in cell biology* 2007, **19**:250-6.
58. Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS: **Origins of specificity in protein-DNA recognition.** *Annual review of biochemistry* 2010, **79**:233-69.
59. Peckham HE, Thurman RE, Fu Y, Stamatoyannopoulos JA, Noble WS, Struhl K, Weng Z: **Nucleosome positioning signals in genomic DNA.** *Genome research* 2007, **17**:1170-7.
60. Weiner A, Hughes A, Yassour M, Rando OJ, Friedman N: **High-resolution nucleosome mapping reveals transcription-dependent promoter packaging.** *Genome research* 2010, **20**:90-100.
61. Locke G, Tolkunov D, Moqtaderi Z, Struhl K, Morozov AV: **High-throughput sequencing reveals a simple model of nucleosome energetics.** *Proceedings of the National Academy of Sciences* 2010, **107**:3838107.
62. Segal E, Widom J: **From DNA sequence to transcriptional behaviour: a quantitative approach.** *Nature reviews Genetics* 2009, **10**:443-56.
63. Fan X, Moqtaderi Z, Jin Y, Zhang Y, Liu XS, Struhl K: **Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3'-end formation.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**:17945-50.
64. Zhang Z, Wippo CJ, Wal M, Ward E, Korber P, Pugh BF: **A Packing Mechanism for Nucleosome Organization Reconstituted Across a Eukaryotic Genome.** *Science* 2011, **332**:977-98010, 1126/science.1200508.
65. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nature genetics* 2000, **25**:25-9.
66. Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC, Gilmour DS, Albert I, Pugh BF: **Nucleosome organization in the *Drosophila* genome.** *Nature* 2008, **453**:358-62.
67. Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC, Gilmour DS, Albert I, Pugh BF: **Nucleosome organization in the *Drosophila* genome.** *Nature* 2008, **453**:358-6210, 1038/nature06929.
68. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome biology* 2009, **10**:R25.
69. Flores O, Orozco M: **nucleR: a package for non-parametric nucleosome positioning.** *Bioinformatics (Oxford, England)* 2011, **27**:2149-215010, 1093/bioinformatics/btr345.

70. Pérez A, Marchán I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M: **Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers.** *Biophysical journal* 2007, **92**:3817-29.

doi:10.1186/1471-2164-12-489

Cite this article as: Deniz *et al.*: Physical properties of naked DNA influence nucleosome positioning and correlate with transcription start and termination sites in yeast. *BMC Genomics* 2011 **12**:489.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

