Thermodynamics of Intragenic Nucleosome Ordering

G. Chevereau,[1,2] L. Palmeira,[1,2,3] C. Thermes,[4] A. Arneodo,[1,2] and C. Vaillant[1,2]

[1]*Laboratoire Joliot-Curie and Laboratoire de Physique, ENS-Lyon, CNRS, 46 Allée d'Italie, 69364 Lyon Cedex 07, France*
[2]*Université de Lyon, 69000 Lyon, France*
[3]*Laboratoire Biométrie et Biologie Evolutive, UMRS 5558, CNRS, Université Lyon 1, F-69100 Villeurbanne, France*
[4]*Centre de Génétique Moléculaire, CNRS, Allée de la Terrasse, 91198 Gif-sur-Yvette, France*

(Received 28 May 2009; revised manuscript received 11 September 2009; published 30 October 2009)

The nucleosome ordering observed *in vivo* along yeast genes is described by a thermodynamical model of nonuniform fluid of 1D hard rods confined by two excluding energy barriers at gene extremities. For interbarrier distances $\mathcal{L} \lesssim 1.5$ kbp, nucleosomes equilibrate into a crystal-like configuration with a nucleosome repeat length (NRL) $\mathcal{L}/n \sim 165$ bp, where n is the number of regularly positioned nucleosomes. We also observe ''bistable'' genes with a fuzzy chromatin resulting from a statistical mixing of two crystal states, one with an expanded chromatin (NRL $\sim \mathcal{L}/n$) and the other with a compact one (NRL $\sim \mathcal{L}/(n+1)$). By means of single nucleosome switching, bistable genes may drastically alter their expression level as suggested by their higher transcriptional plasticity. These results enlighten the role of the intragenic chromatin on gene expression regulation.

DOI: 10.1103/PhysRevLett.103.188103 PACS numbers: 87.15.Cc, 87.14.gk, 87.15.A−, 87.16.Sr

As a first step of DNA compaction inside eukaryotic nuclei, the nucleosome (~146 bp of DNA wrapped around a histone octamer) controls the accessibility of protein complexes to their target sites and as such participates to the regulation of nuclear functions [1]. In yeast, genome-wide experimental mappings of nucleosome occupancy [2–4] have revealed a patchy landscape characterized by an alternation of nucleosome depleted regions (NDR), highly organized regions with a NRL of ~165 bp and regions with no apparent organization. Understanding the mechanisms that control the chromatin pattern at these genomic loci and their coupling to the regulation of transcription and replication is a main challenge in functional genomics.

The NDR are of particular interest since they collocate with natural loci for functional and/or structural regulation by protein complexes. Promoter regions as well as other regulatory sites are generally depleted in nucleosomes [2–6]. Moreover NDR regions might also control the structure of the neighboring chromatin. Indeed, periodic ordering of nucleosomes can result from a non local effect induced by fixed boundaries (i.e., stable nucleosome exclusion zones) that can be caused by either stably bound proteins on specific sites, stable nucleosomes on strong positioning sequences or alternatively by some structurally unfavorable sequences. Such a ''statistical'' ordering naturally arises when confining non overlapping objects at the vicinity of a fixed boundary [7]. Recently, we have confirmed the relevance *in vivo* of this ordering mechanism via a thermodynamical modeling of nucleosome assembly along the yeast chromosome III that takes into account a sequence-dependent nucleosome wrapping energy and a hard-core repulsion between nucleosomes [8]. This simple equilibrium model accounts amazingly well for the *in vivo* nucleosome occupancy profiles, suggesting that the DNA sequence significantly influences the overall chromatin organization and that the observed nucleosome periodic ordering is indeed partly controlled by eviction from high sequence-directed excluding energy barriers.

Our purpose here is to extend this modeling to the complete yeast genome with the specific goal of studying intragenic chromatin. As the mean *in vivo* nucleosome occupancy profiles shown in Figs. 1(a) and 1(b) indicate, most yeast genes are bordered by a NDR at the transcription start (TSS) and termination (TTS) sites [4,6,9,10]. This depletion may result from the combination of some ''intrinsic'' and ''extrinsic'' sequence specific force fields: (i) the presence of poly(dA:dT) [4,10–12] might impair nucleosome formation or favor nucleosome disassembly by increasing the DNA wrapping free energy cost; (ii) particular sequences may recruit transcription factors [4] and/or other protein complexes such as chromatin regulators [13] that may compete with nucleosomes. When computing the free energy cost of nucleosome formation as explained in [8], we indeed observe an energy barrier at the NDR locations [Figs. 1(a) and 1(b)] which confirms recent theoretical investigations with some related energy models [4,6,11,14]. To compare with experiments, from this energy profile, we need to compute a nucleosome occupancy landscape. As an efficient alternative to Monte Carlo simulations [8], we solve the many-nucleosome problem with hard-core interaction by numerically implementing the Vanderlick *et al.*'s solution of the Percus equation [15]:

$$\beta\mu = \beta E(s,l) + \ln\rho(s) - \ln\left(1 - \int_s^{s+l} \rho(s')ds'\right) + \int_{s-l}^{s} \frac{\rho(s')}{1 - \int_{s'}^{s'+l} \rho(s'')ds''} ds', \quad (1)$$

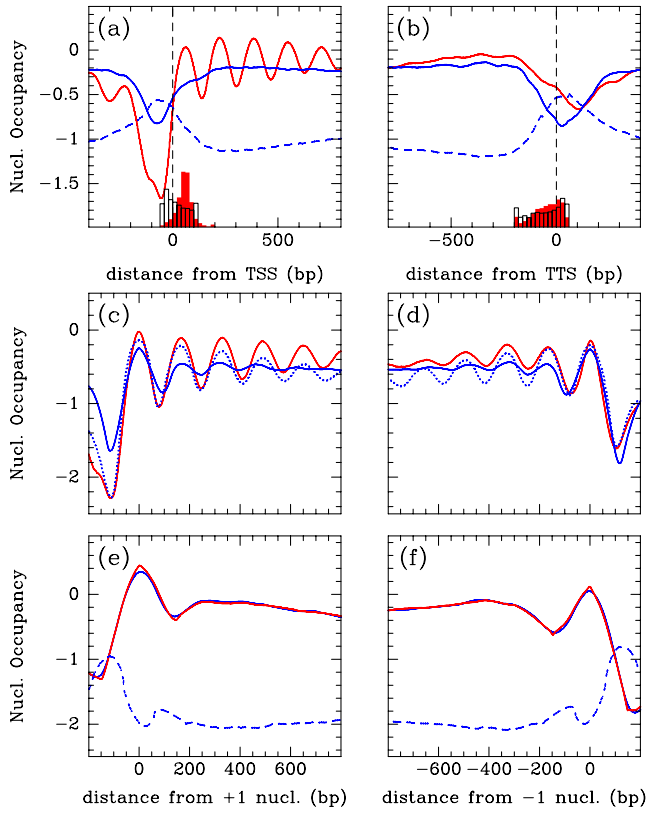

FIG. 1 (color online). Average (4554 yeast genes) *in vivo* (red, grey line) [3] and theoretical (blue, black line) nucleosome occupancy ($\log_2$) profiles, and mean theoretical energy profile (blue, dashed line) plotted around (a) the TSS and (b) the TTS. (c) and (d): same as in (a) and (b) but with individual profiles aligned on the first flanking nucleosome (+1) downstream the TSS and (−1) upstream the TTS. The dotted line corresponds to the theoretical nucleosome occupancy profile obtained when imposing an artificial excluding energy barrier at TSS and/or TTS when not predicted by the sequence (see text). (e) and (f): same as in (c) and (d) for the average *in vitro* (red, grey line) [4] and the theoretical low density nucleosome occupancy profile (blue, black line) and energy profile (blue, dashed line). The solid bars (white bars) histograms correspond to experimental (theoretical) distance between TSS and +1 nucleosome (a) and between TTS and −1 nucleosome (b). The model parameters are (a)–(d) $\delta E = \langle (E - \langle E \rangle)^2 \rangle^{1/2} = 2$ kT, $\tilde{\mu} = -1.3$ kT (blue, black line), $\delta E = 0.75$ kT, $\tilde{\mu} = 1.5$ kT (blue, black dotted line) and (e)–(f) $\delta E = 2$ kT, $\tilde{\mu} = -6$ kT. $\tilde{\mu} = \mu - \langle E \rangle$ has been adjusted so that the mean *in vivo* (a)–(d) and *in vitro* (e)–(f) nucleosome densities are correctly reproduced.

where μ is the chemical potential, β the Boltzmann temperature, $E(s, l)$ the sequence-dependent nucleosome free energy [8], ρ the nucleosomal density, s the sequence abcyssa, and l the hard-core interaction length. Solving this equation gives the density of nucleosomes in a 1D potential [16]. When comparing our theoretical nucleosome occupancy profiles to *in vivo* data, the correlation obtained over the 16 *S. cerevisiae* chromosomes is quite significant: 0.5 with Kaplan's data [4] and 0.3 with Lee's data [3]. However, as shown in Fig. 1, if our model reproduces well the NDR at TTS, it fails to faightfully describe the average chromatin pattern at TSS where a pronounced NDR flanked (especially towards the ORF gene) by stretches (5–6) of well positioned nucleosomes with a periodic (∼165 bp) ordering is observed *in vivo*. Our modeling accounts, both at TTS and TSS, for a depletion zone with no ordering on the sides very similar to what is experimentally observed at gene end [Fig. 1(b)], but significantly smaller and wider than the one observed at TSS [Fig. 1(a)]. We interpret this discrepancy as a consequence of a less precise phasing of the theoretical NDR with respect to the TSS leading to the averaging out of the periodic ordering and a widening of the depletion. This is confirmed in Fig. 1(c) and 1(d) where when now averaging the nucleosome occupancy profiles around the +1 and −1 nucleosomes, we recover theoretically a periodic ordering bordering the NDR at both TSS and TTS.

As shown in Fig. 2, in addition to a small phase shift of the order of few helical pitches between predicted and observed NDRs, there are a number of gene's NDRs that our model does not predict; namely, 50% (20%) of TSS (TTS) NDRs are not accounted for by our model. This strongly suggests that our sequence-directed nucleosome energy field does not capture all the sequence specificity of nucleosome assembly *in vivo*. Importantly the comparison of our modeling to recent *in vitro* data [4] after adjusting the chemical potential $\tilde{\mu}$ (∼30% density as compared to 75% *in vivo*), yields nucleosome occupancy profiles that remarkably reproduce the experimental data (Fig. 2): the

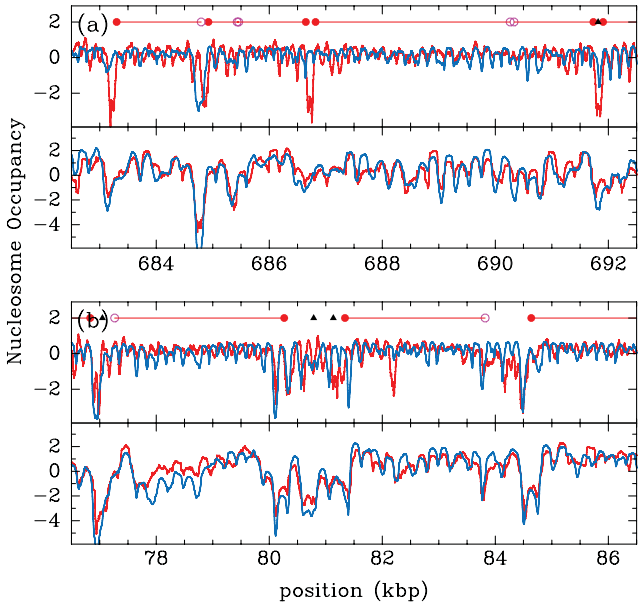

FIG. 2 (color). (a) Nucleosome occupancy ($\log_2$) profile along the yeast chromosome II genome: *in vivo* [3] (top red curve), *in vitro* [4] (bottom red curve), theoretical profiles for $\delta E = 2$ kT, $\tilde{\mu} = -2$ kT (top blue curve) and $\tilde{\mu} = -6.7$ kT (bottom blue curve). (b) Same as in (a) for chromosome XII. In (a) and (b) are indicated the positions of TSS (red dots), TTS (red circles) and of TF binding sites (black triangles).

direct correlation is as high as 0.75 genome-wide (a result which is as good as the correlation value 0.77 obtained by recent discrete models [4,11]) and the experimental nucleosome patterns at both TSS and TTS are predicted without any phase shift [Figs. 1(e) and 1(f)]. A careful examination of the *in vivo*, *in vitro* and theoretical nucleosome occupancy profiles in Fig. 2, shows that the main discrepancy between the *in vivo* and predicted NDRs at TSSs arises from sequence-dependent action of external factors such as transcription factors (a NDR is observed at the majority of transcription factor binding sites out of our model predictions) and remodelers (some phase shift is experimentally observed with respect to the sequence-induced nucleosome occupancy profile) [13].

Whatever the origin of the effective nucleosome energy barriers that impair nucleosome formation at the observed NDRs, they result in a confining and an ordering of flanking nucleosomes and as such, condition the chromatin organization inside the genes. When ordering yeast genes by the distance L that separates the first $(+1)$ and last (-1) nucleosomes, we obtain a striking organization of the nucleosome distribution [Fig. 3(a)]. Small genes ($L \leq$ 1.5 kbp) present a clear periodic packing in between the two bordering NDRs with a well-defined number n of regularly spaced nucleosomes [Fig. 3(b)]. As the inter-distance L increases, these ''crystallized'' genes cluster into L-domains with genes having the same number of nucleosomes, from $n = 2$ to about 9 nucleosomes. For rather large gene sizes ($L \geq 1.5$ kbp), the nucleosome positioning appears periodic essentially at the two boundaries and fuzzy in the middle where the confinement induced by both boundaries is probably too weak to constrain the positioning of the central nucleosomes [Fig. 3(a)]. In our modeling, whenever a NDR observed *in vivo* at either the TSS or TTS is not accounted for by a genomic energy barrier, we locally impose the presence in $E(s, l)$ of an artificial excluding barrier of trapezoidal shape to mimic the effect of external factors. For $\tilde{\mu} = 1.5$ kT, we get mean nucleosome occupancy profiles at TSS and TTS [Figs. 1(c) and 1(d)] and a corresponding 2D map [Fig. 3(b)] in remarkable agreement with the *in vivo* experimental data. This periodic nucleosome ordering is not observed in the *in vitro* mean nucleosome occupancy profiles in Figs. 1(e) and 1(f), indicating that the main positioning signal is not specified by the intragenic high affinity positioning sequences but rather by the long-range confinement induced by the bordering excluding barriers.

When simplifying even more the nucleosome formation energy landscape by assuming that it is flat (no sequence effect) inside the genes, the problem of stacking hard-rods of 146 bp in a box with infinite wall boundaries is then analytically tractable [Figs. 4(a)–4(c)]. From the theoretical density probability for a gene to be in a n-nucleosome crystal configuration according to its size $\mathcal{L}$ ($\simeq L + 188$ bp) [Fig. 4(d)], we compute the gene nucleosome occupancy profile as the weighted sum of each $n = 1, 2, \ldots$ occupancy profiles [Figs. 4(a)–4(c)]. ''Crystal'' genes are those that are characterized by a single dominating crystal n configuration. This model predicts that in between the crystal n [Fig. 4(a)] and $n+1$ [Fig. 4(c)] domains, there is a coexistence domain where these two (or more) crystalline configurations contribute statistically to a seemingly fuzzy occupancy profile [Fig. 4(b)]. From the spectral analysis of the experimental nucleosome occupancy profiles, we assign genes to one of the three categories: crystal-like, bi-stable, other. Genes presenting a positioning profile with a single and dominant periodic contribution, i.e., a single well-defined NRL, are considered as crystal genes. As shown in Fig. 4(e), in agreement with experimental observations, our simple model predicts the existence of $\mathcal{L}$ domains of ''crystallization'' characterized by a NRL: $140 < l_{\min} < \text{NRL} \sim \mathcal{L}/n < l_{\max} < 200$ bp. When considering as ''bi-stable'' genes those whose experimental nucleosome occupancy power spectrum presents at least two dominating peaks of comparable magnitude, we observe, in full agreement with the theoretical predictions, that they concentrate on $\mathcal{L}$ domains alternating with the successive crystal domains [Fig. 4(e)]. Furthemore they present a fuzzy mean occupancy profile [Fig. 3(a)] very

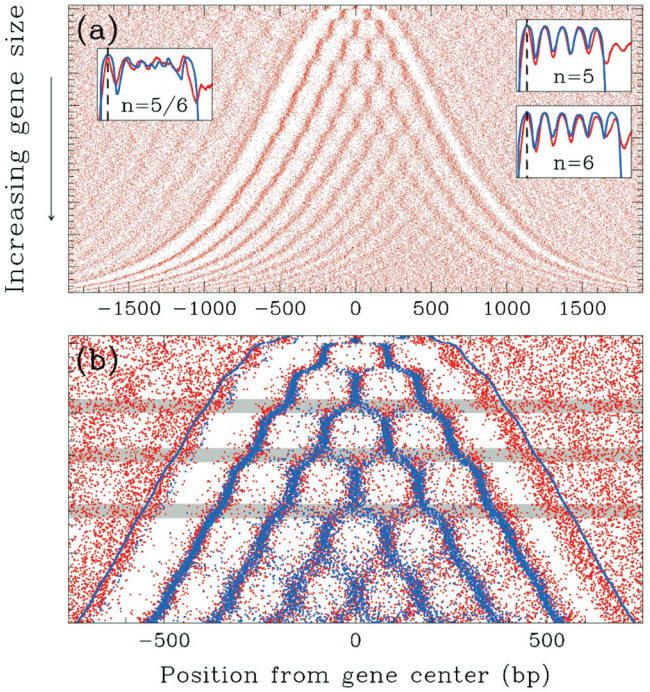

FIG. 3 (color). (a) 2D map of local minima (red) of the experimental *in vivo* nucleosome occupancy profile at yeast genes [3]; genes are ordered vertically by the distance L between the $+1$ and -1 nucleosomes. Insets: mean experimental (red) and one individual theoretical (blue) nucleosome occupancy profiles for crystal genes harboring 5 nucleosomes (right, top), 6 nucleosomes (right, bottom) and the bi-stable genes with 5/6 nucleosomes. (b) Zoom on the first 2000 genes in (a); on the top of the experimental data (red) are superimposed the predictions of our physical modeling (blue); horizontal grey-shaded bands correspond to some bi-stable L domains. Same model parameters as in Figs. 1(a)–1(d).

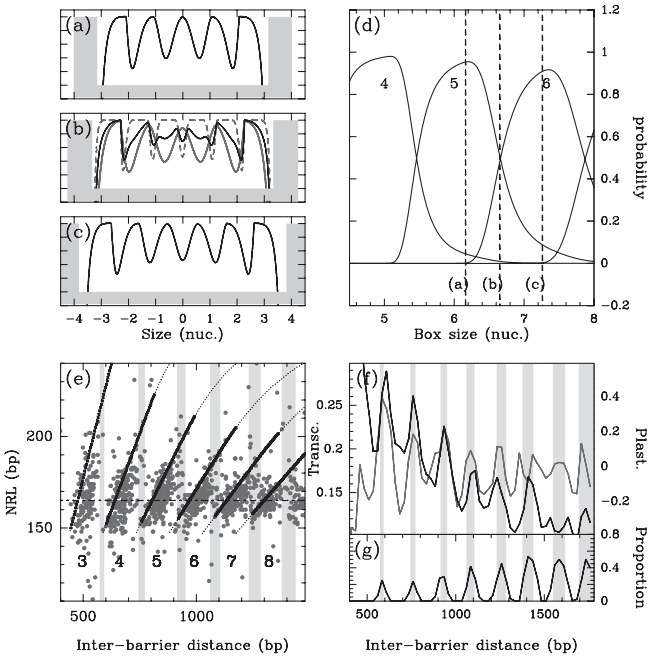

FIG. 4. Theoretical probability of nucleosome occupancy at each point of a box bordered by two infinite walls mimicking excluding barriers at gene extremities: (a) box large enough to shelter $n = 5$ nucleosomes (black); (b) larger box where the two $n = 5$ and 6 configurations (grey) are possible; the weighted average of these crystal-like profiles yields a *fuzzy*-looking profile (black); (c) larger box where 6 nucleosomes can be inserted (black). (d) Probability of crystal configurations with a fixed number n of nucleosomes *vs.* the box size. Vertical dashed lines correspond to the interbarrier distances used in (a), (b) and (c) respectively. (e) NRL dependency on the box size: thin black dotted lines correspond to a fixed number n of nucleosomes and the thick black dotted lines to the NRL at a fixed nucleosome density (∼85%); grey dots correspond to individual crystal gene values. Vertical grey-shaded bands correspond to the experimental bi-stable $\mathcal{L}$-domains. (f) Average transcription rate [20] (black) and transcriptional plasticity (square $\log_2$ expression ratio) [17] (grey) from various microarray experiments [21]. (g) Proportion of bi-stable genes.

similar to the one theoretically predicted as the statistical mixture of neighboring crystal patterns [Fig. 4(b)].

We have demonstrated that a thermodynamical model of nucleosome assembly at equilibrium accounts very well for the evolution of the chromatin pattern as a function of the gene size as measured by the interbarrier distance. Thus far, the role of chromatin structure on gene expression regulation has been mostly investigated at the level of transcription initiation and different regulation strategies associated with different kinds of promoter structural design have been revealed [3,17–19]. Analysis of genomic and epigenetic data suggests that the intragenic chromatin architecture also plays significant role in gene expression regulation at the level of the elongation process. Indeed, as compared to crystal-like genes that rather present a constitutive expression level, bi-stable genes show a higher transcriptional plasticity, i.e., a higher expression level variability under environmental condition changes [Fig. 4(f)]. As the transcription rate (Pol. II density) tends to increase when the nucleosome linker size decreases [Fig. 4(f)], bi-stable genes may drastically alter their expression level in response to various stimuli by switching from a weakly expressed diluted (n nucleosomes) chromatin state to a higher expressed more compact ($n + 1$ nucleosomes) structure (and vice versa) via the gain (or removal) of a single nucleosome. To summarize, our study provides a very simple interpretation of the intragenic nucleosomal organization observed *in vivo* in yeast, as resulting from an equilibrium ordering induced by inhibitory energy barriers located at gene extremities. This suggests that a thermal-like equilibrium (likely resulting from the action of chromatin remodeling factors) is attained along most genes with a characteristic relaxation time much shorter than the typical time separating the successive chromatin alterations induced by the elongating transcription machinery.

This work was supported by Conseil Regional Rhônes-Alpes (Emergence 2005) and the Agence Nationale de la Recherche (ANR-06-PCVI-0026 and ANR NT05-3-41825).

[1] R. D. Kornberg and Y. Lorch, Cell **98**, 285 (1999).
[2] G.-C. Yuan *et al.*, Science **309**, 626 (2005).
[3] W. Lee *et al.*, Nat. Genet. **39**, 1235 (2007).
[4] N. Kaplan *et al.*, Nature (London) **458**, 362 (2009).
[5] E. Segal *et al.*, Nature (London) **442**, 772 (2006).
[6] V. Miele *et al.*, Nucleic Acids Res. **36**, 3746 (2008).
[7] R. D. Kornberg and L. Stryer, Nucleic Acids Res. **16**, 6677 (1988).
[8] C. Vaillant, B. Audit, and A. Arneodo, Phys. Rev. Lett. **99**, 218103 (2007).
[9] S. Shivaswamy *et al.*, PLoS Biol. **6**, e65 (2008).
[10] T. N. Mavrich *et al.*, Genome Res. **18**, 1073 (2008).
[11] Y. Field *et al.*, PLoS Comp. Biol. **4**, e1000216 (2008).
[12] E. Segal and J. Widom, Curr. Opin. Struct. Biol. **19**, 65 (2009).
[13] P. D. Hartley and H. D. Madhani, Cell **137**, 445 (2009).
[14] G.-C. Yuan and J. S. Liu, PLoS Comp. Biol. **4**, e13 (2008).
[15] T. K. Vanderlick, L. E. Scriven, and H. T. Davis, Phys. Rev. A **34**, 5130 (1986).
[16] Note that similar methods based on discrete thermodynamical nucleosome positioning models have been already employed in [4,5] and very recently by A. V. Morozov *et al.*, Nucleic Acids Res. **37**, 4707 (2009).
[17] I. Tirosh and N. Barkai, Genome Res. **18**, 1084 (2008).
[18] F. H. Lam, D. J. Steger, and E. K. O'Shea, Nature (London) **453**, 246 (2008).
[19] H. Boeger, J. Griesenbeck, and R. D. Kornberg, Cell **133**, 716 (2008).
[20] E. J. Steinmetz *et al.*, Mol. Cell **24**, 735 (2006).
[21] J. Ihmels *et al.*, Nat. Genet. **31**, 370 (2002).