

DNA as active polymer: long-range allosteric effect and chromatin loop structure

Ming Li^{a,*}, Zhong-can Ou-yang^b

^aGraduate University of Chinese Academy of Sciences, Beijing 100080

^bInstitute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100080

Abstract

We investigated on a possible relation between DNA elasticity and eukaryotic chromosome structure. We find, by Benham model, that (1) there exists long-range allosteric structural transition on a topologically-constrained DNA; (2) this effect seems to interplay with chromosome structure and gene regulation. We conclude that DNA, as an active polymer, may utilize its elastic properties in some cellular processes.

Key words: DNA topology; DNA elasticity; long-range allosteric; chromosome structure

1. Introduction

Compared with prokaryotic chromosome, eukaryotic chromosome is much more complicate and ordered. This observation leads to the assumption that eukaryotic chromatin structure may serve as an important regulatory mechanism in gene expression, which is not shared by prokaryote. In this article, we attempt to address this possibility from the viewpoint of DNA mechanics.

Eukaryotic DNA is highly stressed due to chromosome packaging: DNA wraps onto histone octamers in a left-handed way (so it's undertwisted) to form nucleosomes, then the nucleosome string folds into a solenoid, the solenoid is demarcated into chromatin loops by nuclear matrix [1], and finally assembles into chromosome. The base of each chromatin loop, SMARs (scaffold/matrix attachment regions), is found as duplex unwound regions (local denaturation) *in vitro* due to the presence of torsional stress on DNA[2]. Since DNA is actually un-

dertwisted *in vivo*, we assume that local denaturation could occur under the stress *in vivo* (as will be clear below), among which the SMARs are chosen as bases of chromatin loops. Once the looping structure is formed, the two boundary SMARs may conversely impose torsional constraint on the intervening DNA, keeping it in a stressed state to interfere in the basic molecular processes (transient torsional stress can affect gene transcription, e.g., see [3,4]). In this sense, DNA should be regarded as an 'active polymer': its elastic state and the chromatin loop structure are mutually affected and co-determined. We try to make some trial investigation on this assumption in the following sections.

2. DNA Topology

We first give the description on the topology, i.e. the torsional state, of circular DNA.

Linking number, Lk , is the number of double-helical turns of circular DNA when the molecule is forced into planar conformation. It's a topological invariant, i.e. an integer. $Lk_0 = N/\gamma$ is often referred as the linking number of relaxed DNA (e.g.,

* Corresponding author.

Email address: liming@gucas.ac.cn (Ming Li).

circular DNA with single-strand nicks), N is the total number of base-pairs in the DNA, and γ is the number of base-pairs per helical turn in ordinary B-form DNA (10.5 bp/turn). Superhelical density, $\sigma = (Lk - Lk_0)/Lk_0$, is to quantify the twist of DNA. It's found that σ varies within a very narrow range from -0.01 to -0.1 for mesophile (most often -0.06). Especially, since DNA segment of $\sim 167bp$ is under-twisted one helical turn per nucleosome, the superhelical density of eukaryotic DNA is around -0.06 ($\sim \gamma/167$).

Linking number can be decomposed into writhing number Wr and twisting number Tw . Wr is roughly the self-linking number of the DNA axis, and Tw is the inter-winding number of the two strands when counting along the DNA axis. In general, the following conservation law holds for any circular DNA, $Lk = Wr + Tw$ or $\Delta Lk = \Delta Wr + \Delta Tw$ ($\Delta Lk = Lk - Lk_0$), when both Wr and Tw (or ΔWr and ΔTw) vary.

For linear DNA constrained at the two terminus, one can construct quantities analogous to Lk , Wr and Tw with some assumptions and still obtain the conservation law[5]. So we will not distinguish between circular DNA and constrained linear DNA in the next sections.

3. DNA elasticity

Free energy corresponding to ΔLk has been carefully studied by electrophoresis for intact circular DNA, presented as:

$$G(\Delta Lk) = \frac{1}{2}K(\Delta Lk)^2 \quad (1)$$

$\Delta Lk = \sigma * Lk_0$, $K = qRT/N$, $q = 2200$, R is gas constant, T is the absolute temperature [6,7]. A.V.Vologodskii *et al.* pointed out that this energy form can be interpreted in terms of DNA elasticity[8].

Imaging DNA as an isotropic rod, its elastic energy comes from (1)bending of the axis (persistence length noted as l_b); (2)twisting of the rod(persistence length l_c). When linear DNAs transform into covalently-closed circular DNAs, there will be a distribution of Lk (or ΔLk) rather than a single Lk value among the products, since DNA undergoes thermal fluctuation. On one hand, the distribution is determined by the elastic energy $G(\Delta Lk)$. On the other hand, it is also determined by the fluctuation of ΔWr and

ΔTw . Noting that $\Delta Lk = \Delta Wr + \Delta Tw$, ΔWr and ΔTw are independent before the closure of DNA, thus $\langle \Delta Lk \rangle = \langle \Delta Wr \rangle + \langle \Delta Tw \rangle = 0$ and $\langle \Delta Lk^2 \rangle = \langle \Delta Wr^2 \rangle + \langle \Delta Tw^2 \rangle$ always holds. Assuming that ΔLk obeys Gaussian distribution, i.e. $P(\Delta Lk) \propto \exp(-\Delta Lk^2/2\langle \Delta Lk^2 \rangle)$, one can easily get $q = N/\langle \Delta Lk^2 \rangle$ (compared with Eq.1). For a long chain DNA, it's well documented that $\langle \Delta Tw^2 \rangle = L/(2\pi)^2 l_c$ and $\langle \Delta Wr^2 \rangle = \mu L/(2\pi)^2 l_b$, here $L = Nh$ ($h = 0.34nm$ is the pitch length of DNA helix). Thus we have

$$q = (2\pi)^2 [h(\frac{\mu}{l_b} + \frac{1}{l_c})]^{-1} \quad (2)$$

Taking the mostly cited values, $l_b = 50nm$, $l_c = 80nm$, and $\mu \simeq 2$ (simulation result, see [9]), one get $q = 2209$, which is almost the same as that given in Eq.1.

It's worth noting that the above discussion also applies to constrained linear DNA, except that μ may differ. Computer simulation shows that μ decreases(q increases), while DNA is stretched. Thus one can guess that q and σ are possibly the effective (mechanic) variables *in vivo* for cells to control chromosome structure. In next sections, we take q as a changeable parameter, though only an illustrative value is shown.

4. Benham model

Stress-induced local denaturation can be described by Benham model. A state variable s is designated to each base-pair, $s = 0$: intact (B-DNA region); $s = 1$: unpaired (denatured region). The total energy of the molecule consists of four energy terms: (1)topological elastic energy, (2)base-unpairing energy, (3)boundary energy(energy cost of boundary formation between intact B-DNA and the adjacent denatured region), and (4)inter-strand twisting energy of denatured regions. The first one has been given in Section(3). Here we just give a sketch on the other ones. Details can be found in [11,12].

At physiological condition, base unpairing energy is taken as $u_{AT} = 0.36 \text{ kcal} \cdot \text{mol}^{-1}$, and $u_{GC} = 1.41 \text{ kcal} \cdot \text{mol}^{-1}$. The total base-unpairing energy is $G = \sum_{j=1}^N u_j \cdot s_j$, s_j denoting the state of j -th base-pair. Boundary energy B is taken as $B=10.8 \text{ kcal} \cdot \text{mol}^{-1}$. Suppose there are r unbound regions in DNA, it's easy to show that $r = \sum_{j=1}^N s_j(1 - s_{j+1})$,

thus the corresponding energy is $G = B \cdot r = \sum_{j=1}^N B s_j (1 - s_{j+1})$. The interwinding energy of the two strands in denatured regions is assumed as harmonic oscillator potential: $G = \frac{1}{2} C n \tau^2$, where $n = \sum_{j=1}^N s_j$ is the total number of unpaired base pairs, C is the torsional stiffness coefficient (about $1.4 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{rad}^{-2}$ under Benham's experimental condition), τ is the twist per base pair of the denatured regions.

When denaturation is very weak, i.e., $n \ll N$, the total free energy of a partially denatured DNA is assumed as

$$G = \frac{1}{2} K (\Delta L k + \frac{n}{\gamma} - \frac{n\tau}{2\pi})^2 + \frac{1}{2} n C \tau^2 + \sum_{j=1}^N [(B + u_j) s_j - B s_j s_{j+1}] \quad (3)$$

5. Calculating the transition profile

We are interested in the mean value of s_j , i.e., the unpairing probability p_j of the j -th base pair,

$$p_j = \sum_{\{s\}} s_j e^{-\beta G(s)} / Z \quad (4)$$

s : state of the sequence. $\{s\}$: the state space (2^N states). Z is the partition function. Assuming a mechanical equilibrium between the elastic energy (1) and (4) (see Section (4)), one can eliminate the variable τ , and obtain

$$Z = \sum_{\{s\}} e^{-\beta G(s)} = \sum_{n=0}^N \Omega(n) H(n) \quad (5)$$

$$\Omega(n) = \exp \left[\frac{-2\pi^2 \beta C K}{4\pi^2 C + K n} \left(\Delta L k + \frac{n}{\gamma} \right)^2 \right] \quad (6)$$

$$H(n) = \sum_{\{s\}} \exp \left\{ -\beta \sum_{j=1}^N [(B + u_j) s_j - B s_j s_{j+1}] \right\} \cdot \delta \left(\sum_j s_j - n \right) \quad (7)$$

Similarly, we also have

$$p_j = Z_j / Z$$

$$Z_j = \sum_{\{s\}} s_j e^{-\beta G(s)} = \sum_{n=0}^N \Omega(n) H_j(n) \quad (8)$$

$$H_j(n) = H(n) \delta(s_j - 1)$$

$\Omega(n)$ can be calculated directly (numerically), the difficulty is how to calculate $H(n)$ and $H_j(n)$. To impose the constraint $\delta(\sum_j s_j - n)$, we define

$$M_j = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + x \begin{pmatrix} 0 & 0 \\ \exp[-\beta(B + u_j)] & \exp[-\beta u_j] \end{pmatrix}$$

Once the numeric calculation $\text{Tr}[\prod_{j=1}^N M_j]$ is finished, one can obtain a polynomial of x , $F(x) = \sum_{n=0}^N H(n) x^n$. $H(n)$ is just the one defined by Eq.7. Z_j can also be computed in the same way: one need only substitute the first matrix of M_j by a null matrix (each element is zero). In this way, we can calculate p_j precisely for very long DNA sequences. For details of the algorithm and its performance, readers should refer to Ref.[13].

6. Results and Discussion

Here we give one example to show how one can relate the transition profile to chromosome structures. DNA sequences are taken from databank NCBI. SMARs data is taken from databank S/MARt DB [14]. Sequences in all the profiles are coordinated from 5'- to 3'-terminal (left to right). Parameter values are illustrative. Choosing other values of σ or q doesn't change the conclusion.

The example is the histone gene cluster of *Drosophila melanogaster*, in which the five genes are arranged as (arrows indicating the transcriptional orientation)

$$5' - \overleftarrow{\text{H3}} - \overleftarrow{\text{H4}} - \overleftarrow{\text{H2a}} - \overleftarrow{\text{H2b}} - \overleftarrow{\text{H1}} - 3'$$

The five-gene cluster repeats for a hundred times on the chromosome, so we construct a sequence of two adjacent repeats from the original sequence of a unit (NCBI: X14215) for the first round of calculation (other constructions, e.g. sequence including more units, give the same result). The profile is shown in Fig.1(a). In this profile, the two identical peaks indicate two denatured regions which are in accordance with the experimentally identified SMAR(SM0000037). This result offers a theoretical evidence to Bode's observation that some SMARs are prone to denaturation under unwinding stress [2]. It also raises a further question, why does SMAR locate at these specific sites? It has been suggested that SMARs can serve as topological barrier to maintain the torsional state of the loop, so we should expect new melting events to occur elsewhere

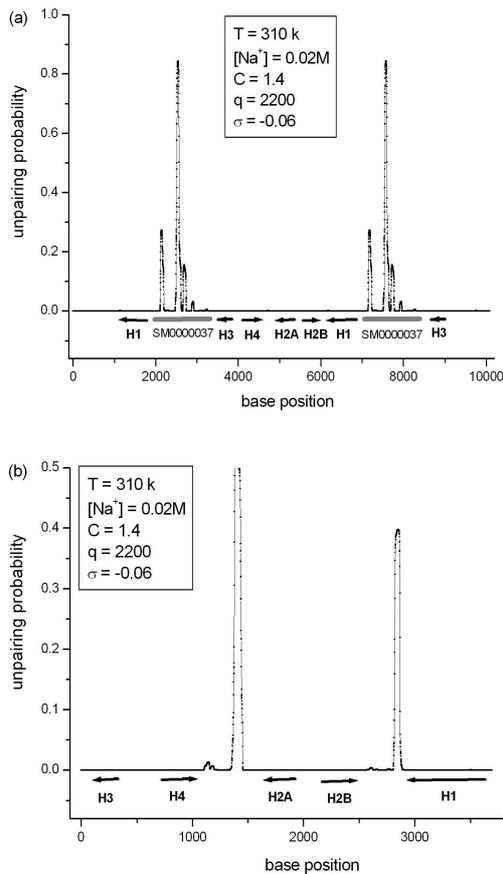


Fig. 1. (a) Profile for two adjacent units of histone gene cluster of *D.melanogaster* (X14215). The orientation of the five genes, as well as their relative positioning to SMAR, are indicated. (b) Profile for one gene cluster. Parameters used in the calculation are given in both (a) and (b).

due to the retained unwinding stress. Therefore we take out the intervening region, i.e., the single unit without the flanking SMARs, for a second round of calculation. (It's equivalent to impose the torsional constraint.) The profile is shown in Fig.1(b). Two new peaks emerge. Compared with Fig.1(a), one can see there is competition among distinct sites to release the torsional stress. This kind of 'long-range allosteric transition' may have biological meaning, as shown below.

Noting H4 and H2A (or H2B and H1) are convergently transcribed, it's conceivable that these two melting event may be involved in transcriptional termination, and the co-occurrence of these two events may be important to the coordinate transcription of the whole gene cluster. Indeed, noting that RNA polymerase can generate positive supercoils forward and negative supercoils backward when it tracks

on DNA along the transcriptional orientation[15], the two melted regions locating downstream to each gene may act as supercoiling 'absorber'. If it's true, it will give strong support to our assumption that strong SMARs could serve as topological barriers to maintain the stressed state of the intervening DNA. Further support of this conjecture is given by the same analysis of *Drosophila hydei*'s homologous histone gene locus (data not shown. For this example and others, readers can refer to Ref.[16]).

7. Summary

Our preliminary analysis indicates that the elastic state of DNA and the large-scale structure of chromosome may be co-determined, and can both affect gene transcription. We hope the strategy combined mechanics analysis and homolog analysis can provide a new way to investigate the structure and function of chromosomes.

References

- [1] J.R. Paulson, U.K. Laemmli, Cell 12 (1977) 817.
- [2] J. Bode, Y. Kohwi, L. Dickinson, et al., Science 255 (1992) 195.
- [3] M. Dunaway, E.A. Ostrander, Nature 361 (1993) 746.
- [4] J. Dorman, Mol. Gen. Genet. 256 (1997) 93.
- [5] Gert H.M. van der Heijden, M.A. Peletier, R. Planqué, arxiv.org/abs/math-ph/0310057
- [6] D.E. Pulleyblank, M. Shure, D. Tang, et al., Proc. Natl. Acad. Sci. 72 (1975) 4286.
- [7] R.E. Depew, J.C. Wang, Proc. Natl. Acad. Sci. 72 (1975) 4275.
- [8] A.V. Vologodskii, V.V. Anshelevich, A.V. Lukashin, et al., Nature 280 (1979) 294.
- [9] K.V. Klenin, A.V. Vologodskii, V.V. Anshelevich, et al., J. Biomol. Struct. Dyn. 6 (1989) 707.
- [10] V. Rossetto, A.C. Maggs, J. Chem. Phys. 118 (2003) 9864.
- [11] W.R. Bauer, C.J. Benham, J. Mol. Biol. 234 (1993) 1184.
- [12] W.R. Bauer, H. Ohtsubo, E. Ohtsubo, et al., J. Mol. Biol. 253 (1995) 438.
- [13] M. Li, Z.C. Ou-Yang, Thin. Solid. Films. 499 (2006) 207.
- [14] I. Liebich, J. Bode, M. Frisch, et al., Nuc. Acids. Res. 30 (2002) 372.
- [15] L.F. Liu, J.C. Wang, Proc. Natl. Acad. Sci. 84 (1987) 7024.
- [16] M. Li, AAPPS Bulletin. 16 (2006) 34.