Available online at www.sciencedirect.com
SCIENCE DIRECT
Thin Solid Films 499 (2006) 207 – 212
thin solid films
www.elsevier.com/locate/tsf

An exact numerical method to calculate the base-unpairing probability for any given DNA sequence by Benham model

Ming Li [a,*], Zhong-can Ou-Yang [a,b]

[a] Institute of Theoretical Physics, The Chinese Academy of Sciences, Beijing 100080, China
[b] Center for Advanced Study, Tsinghua University, Beijing 100084, China

Available online 24 August 2005

Abstract

In this article we report an exact numerical method to calculate the partition function of Benham model which describes the stress-induced denaturation of doublestrand DNA (dsDNA). Generating function approach is employed in the algorithm. It successes to avoid the catastrophic sign cancellation which makes Benham's algorithm impractical to analyze DNA sequences longer than 10 kb. By our algorithm, the transition profile of base-unpairing probability can be efficiently computed for very long sequences on PC (32-bit, CPU frequency 1.4 GHz). The algorithm requires $O(N^3)$ operations and $O(N\log_2 N)$ memory, and by contrast Benham's algorithm needs $O(N^3\log_2 N)$ operations and $O(N_2)$ memory. Examples are given on the performance of the implementation on single CPU and multi-CPU cluster.

Keywords: dsDNA; Unwinding; Local denaturation; Generating function

1. Introduction

Genomic DNAs, either of prokaryotes or eukaryotes, are almost the longest polymers inside the cell. For example, a human genomic DNA is about 4 cm long in average whereas the whole nucleus is just several microns in diameter. If we consider the DNA as a semi-flexible polymer (the persistence length l_P is about 50 μm), its gyration radius R_G can be calculated as $R_G = \sqrt{Ll_p/6}$~18μm which is much larger than the radius of the nucleus. Since there are 46 DNAs in total, each of them must be tightly compacted in order to be accommodated inside the same nucleus. Therefore, to what a degree and how they are packaged are largely determined by their elasticity. In fact, either prokaryotic or eukaryotic DNA is bent, stretched and twisted by variety of proteins in the cell, and on the other hand its elastic response to the external force is very important in gene expression. In particular the torsional state, as the intrinsic property of double-strand DNA (dsDNA), is the leading factor to determine the elastic response since torsional stress generated on DNA backbone can significantly affect the interaction between DNA and proteins. It has been well known for a long time that circular dsDNA of prokaryotes (e.g., bacterium *E. coli*) is always unwound to a large degree by special types of proteins [1] and this unwinding stress can function in gene transcription (examples as [2,3]). For eukaryotic linear dsDNA, it is also effectively unwound by histones and the torsional stress can be held due to the peculiar organization of the DNA-contained chromatin: roughly speaking, the linear DNA is divided into a series of successive loops on which the basements are anchored to nuclear scaffold/matrix and then can serve as barriers to prevent the axial rotation of the dsDNA. Existence of the torsional stress in such cases has been testified by some experiments [4–8]. Therefore, it is intriguing to investigate how torsional stress can function in basic biological processes, e.g., gene transcription or DNA replication.

* Corresponding author. Tel.: +86 10 62561291; fax: +86 10 62562587.
E-mail address: lmmonk@itp.ac.cn (M. Li).

0040-6090/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.tsf.2005.07.007

Great progress has been achieved on this issue. It has been widely accepted that the unwinding stress can alter the local structures of DNA (secondary structure), thus alter the interaction between DNA and transcriptional factors, and finally affect the transcription process [3,9,10]. The stress can induce several types of non-B-form structures, including Z-DNA, cruciform structure, local denaturation (break of the hydrogen bond between the paired bases), etc. Local denaturation may be the most common one whereas other types are subject to more strict sequence specificity. Actually, there are variety of proteins interacting with single-strand DNA in vivo but not with dsDNA, which gives a hint that denatured regions (i.e. bubbles), may play peculiar and important roles in biological processes. Kowalski et.al. succeeded to demonstrate that bubbles could be induced on pBR322 plasmid topoisomers (the same circular dsDNAs at different torsional state) upon increasing of unwinding stress. In addition, their results indicated that the observed bubbles coincide with functional elements such as promoters or terminators [9]. It strongly supports the idea that stress-induced secondary structures bear important functions. Based on this experiment, Benham proposed a phenomenological model to describe the denaturing behavior of circular dsDNA in 1990 [11]. Subsequent studies on both experiment and computer simulation offer further supports to the model. Since the model is essentially in terms of statistical physics, the key problem to predict the melting behavior of real DNA sequences (especially very long DNA) is how the partition function can be calculated fast and precisely. Fye and Benham presented the first numerically exact method for this problem in 1999 [12], in which the number of arithmetic operations is claimed as $O(N^2)$ and the memory as $O(N)$, N is the length of the DNA sequence. It encounters, however, fast loss of numeric precision during the computation process because the use of complex numbers in the calculation inevitably leads to a severe sign cancellation. To solve this problem, the authors appealed to high precision arithmetic in which the needed (minimal) number of decimal digits D is roughly proportional to N (Fig. 1 in Ref. [12]). Since the basic multiplication of two real numbers in high precision arithmetic is implemented by Fast Fourier Transformation (FFT), the time consumption per multiplication is $O(D\log_2 D)$ or $O(N\log_2 N)$. Therefore the actual number of basic operations (i.e. addition or multiplication of real numbers of machine precision) in Fye's algorithm is not $O(N^2)$ but $O(N^3\log_2 N)$, and the actual memory cost is not $O(N)$ but $O(N^2)$. It becomes impractical to implement the algorithm on a 32-bit PC for relatively long sequences (>10 kb). To circumvent the problem, we propose another exact numerical method avoiding catastrophic sign cancellation. It turns out highly efficient to deal with very long sequences (e.g. 100 kb) on a 32-bit PC. In the following, we will first introduce Benham model, and then the implementation of our algorithm to compute the partition function and related quantities.

2. Model and algorithm

A brief review will be given first on the basic ideas of Benham model. Readers should refer to Refs. [13,14] for details, e.g., the experimental estimation on those parameters.

2.1. Basic concepts

Several parameters are used to describe the torsional state of dsDNA. The linking number Lk is roughly the number of helical turns of dsDNA. $Lk_0 = N/\gamma$ is the linking number designated to relaxed DNA (uniform B-DNA, the standard model proposed by Waston and Crick), N is the total number of base-pairs in the DNA, and γ is the number of base-pairs per helical turn in B-DNA (often 10.5 bp/turn). Superhelical density σ is introduced as $\sigma = (Lk - Lk_0)/Lk_0$ to measure how much the dsDNA is twisted. $\Delta Lk = Lk - Lk_0$ is the linking difference. The observed values of σ is almost limited in a narrow range from -0.01 to -0.1 for mesophile (most common -0.06). Especially, the superhelical density of eukaryotic DNA is about $-0.06(\sim 1/167)$ since DNA segment of 167 bp is unwound one helical turn by histone proteins in a single nucleosome. It is the linking difference that introduces unwinding stress to alter the secondary structures, as displayed schematically in Fig. 1: unwinding a native dsDNA segment and ligating it into a circle, then two possible configuration will be adopted by the circular DNA to release the stress (i.e., the negative linking difference): writhing of the axis (termed as negative supercoiling) or local denaturation. The equilibrium

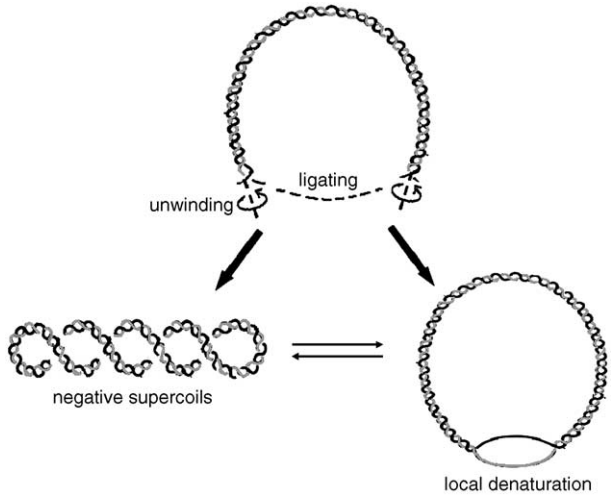

Fig. 1. If a linear dsDNA segment is unwound and ligated into a circle, the unwinding stress can induce two possible configuration: axis supercoiling or local denaturation. Since dsDNA is right-handed double-helical, the imposed negative linking difference σ drives the axis into a form of right-handed superhelix (negative supercoils). When the value of σ is large enough, the dominant configuration shifts from axis supercoiling to local melting. The latter is the focus of our interest since it often occurs preferentially at special sites on a real DNA molecule.