Physica A 418 (2015) 126–153

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

# Biophysics of protein–DNA interactions and chromosome organization

John F. Marko *

*Department of Physics & Astronomy, Northwestern University, Evanston, IL 60208, USA*
*Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA*

## HIGHLIGHTS

- Review of mechanical properties of DNA and DNA–protein complexes.
- Pedagogical presentation of statistical-mechanics of DNA response to tension and twist.
- Discussion of DNA topology and topology control in cells.

## ARTICLE INFO

*Article history:*
Received 25 March 2014
Received in revised form 9 July 2014
Available online 18 July 2014

*Keywords:*
DNA
DNA–protein interactions
Chromosome structure
Chromosome dynamics

## ABSTRACT

The function of DNA in cells depends on its interactions with protein molecules, which recognize and act on base sequence patterns along the double helix. These notes aim to introduce basic polymer physics of DNA molecules, biophysics of protein–DNA interactions and their study in single-DNA experiments, and some aspects of large-scale chromosome structure. Mechanisms for control of chromosome topology will also be discussed.



## 1. Introduction

DNA molecules in cells are found in double helix form, consisting of two long polymer chains wrapped around one another, with complementary chemical structures. The double helix encodes genetic information through the sequence of chemical groups—the "bases" adenine, thymine, guanine and cytosine (A, T, G and C). Corresponding bases on the two chains in a double helix bind one another according to the complementary base-pairing rules A=T and G ≡ C. These rules follow from the chemical structures of the bases, which permit two hydrogen bonds to form between A and T (indicated by =), versus three that form between G and C (indicated by ≡). Each base pair has a chemical weight of about 600 Daltons (Da).

The presence of the two complementary copies along the two polynucleotide chains in the double helix provides redundant storage of genetic information and also facilitates DNA replication, via the use of each chain as a template for assembly of a new complementary polynucleotide chain.

### 1.1. Basic physical properties of the DNA double helix

The structure of DNA gives rise to a number of interesting physical properties.

*Stiffness:* The DNA double helix is a moderately stiff semiflexible polymer, with a persistence length of about 50 nm (containing 150 base pairs or bp; there are approximately 0.34 nm per base pair along the double helix). The thickness of the double helix is about 2 nm, so a persistence length of double helix DNA is long and thin.

* Correspondence to: Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA.
*E-mail address:* John-MARKO@Northwestern.edu.

http://dx.doi.org/10.1016/j.physa.2014.07.045
0378-4371/

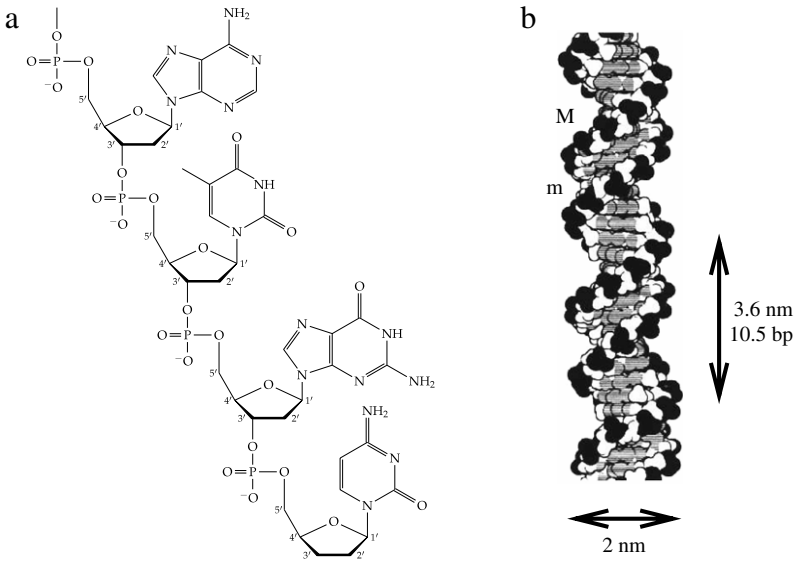


**Fig. 1.** DNA double helix structure. (a) Chemical structure of one DNA chain, showing the deoxyribose sugars (note numbered carbons) and charged phosphates along the backbone, and the attached bases (A, T, G and C following the 5′ to 3′ direction from top to bottom). (b) Space-filling diagram of the double helix. Two complementary-sequence strands as in (a) noncovalently bind together via base-pairing and stacking interactions, and coil around one another to form a regular helix. The two strands can be seen to have directed chemical structures, and are oppositely directed. Note the different sizes of the major (M) and minor (m) grooves, and the negatively charged phosphates along the backbones (dark groups). The helix repeat is 3.6 nm, and the DNA cross-sectional diameter is 2 nm.
*Source:* Image reproduced from Ref. [1].

*Length:* Double-helix DNAs *in vivo* are generally very long polymers: the chromosome of the λ bacteriophage (a virus that infects *E. coli* bacteria) is 48 502 bp or about 16 microns in length; the *E. coli* bacterial chromosome is $4.6 \times 10^6$ bp (4.6 Mb) or about 1.5 mm long; small *E. coli* "plasmid" DNA molecules used in genetic engineering are typically 2–10 kb (0.7–3 microns) in length; and the larger chromosomal DNAs in human cell nuclei are roughly 200 Mb or a few cm in length.

*Electrical charge:* The environment in the cell is essentially aqueous solution, in which DNA molecules are ionized, so as to carry essentially one electron charge per base ($2e^-$/bp $\approx 6e^-$/nm, each negative charge coming from an ionized phosphate on the DNA backbone, see Fig. 1(a)). The high electric charge density along the double helix makes it a strong polyelectrolyte, and gives it strong electrostatic interactions with other electrically charged molecules. Notably, in cells, the univalent salt concentration is 100–200 mM, making the Debye length shorter than 1 nm ($\lambda_D \approx 0.3\,\text{nm}/\sqrt{M}$ where $M$ is the concentration of 1:1 salt in mol/l = M): thus electrostatic interactions with DNA, while strong, are essentially short-ranged. Electrostatic repulsions give rise to an effective hard-core diameter of dsDNA of ≈3.5 nm under physiological salt conditions [2].

*Helical structure:* The DNA double helix is really *two* polymers wrapped around one another, with one right-handed turn every ≈10.5 bp, or about 0.6 radian/bp (Fig. 1(b)). This, combined with the moderate strength of the base-pairing interactions holding the two strands together (about $2.5k_BT$ per base pair when averaged over base-pair sequence) gives rise to the possibility of stress-driven structural defects ("bubbles" of locally base-unpaired single-strands) or transitions (stress-driven strand-separation). In addition, the two-strand structure implies the possibility of trapping a fixed linking number of the two strands when a DNA is closed into a loop. Constraint of strand linking number – a topological property of DNA – gives rise to a rich array of phenomena.

### *1.2. Proteins and DNA*

DNA molecules by themselves are already quite interesting objects for biophysical study. However, the functions of DNA *in vivo* cannot be realized without the action of a huge number of *protein* molecules. Proteins are the workhorse molecules of the cell, and are themselves polymers of amino acids, folded into specific shapes by the action of relatively complex amino-acid–amino-acid interactions. Most proteins are in the range of 100–1000 amino acids in length (since amino acids are ≈100 Da on average, this corresponds to masses from $10^4$ to $10^5$ Da), and since each amino acid is about a cubic nanometer in volume, folded proteins are from a few to a few tens of nm in size.

DNAs in cells are covered with proteins, some of which interact rather specifically with short (<20 bp) specific base-pair sequences, and some of which are less discriminating, interacting with DNA of essentially any sequence. Proteins that bind DNA tend to have positively charged patches on them to allow them to stick to the double helix (many DNA-binding proteins have a net positive charge in solution). Many proteins that bind DNA have hydrophobic amino acids which insert between bases, or hydrogen-bonding groups which link to corresponding hydrogen-bonding groups on the bases.