Multi-Scale Modelling of the Sequence Dependent Mechanics of DNA

John H. Maddocks

Section de Mathématiques

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Kaunas, 7th 8th 9th April 2015

These talks will give an introduction to various mathematical, computational and experimental techniques that can be used to model the physics of the DNA double helix. The most basic function of DNA is to use an ordered list of the four bases A, T, C, and G within a gene to code for the production of the proteins that make organisms work. However in, for example, humans the gene coding regions are estimated to be only 2% or so of the total genome, so what does the rest of the DNA do?

There is an increasing body of evidence and scientific consensus that Nature uses special sequence-dependent mechanical properties, e.g. sequence-dependent variations in stiffnesses, natural curvature etc, on length scales of a few tens to a few hundreds of base pairs as a means to regulate the physical mechanisms of how DNA functions. Examples include TATA boxes, A-tracts, nucleosome positioning sequences, DNA loop repressors and promoters such as LAC and GAL, etc.

Methylation patterns are central to epigenetics and are implicated in cancer. And methylation is a modification of some bases that is believed to change the physical properties of DNA.

But a quantitative understanding of how sequence and modified bases effect the mechanics of DNA remains wanting.

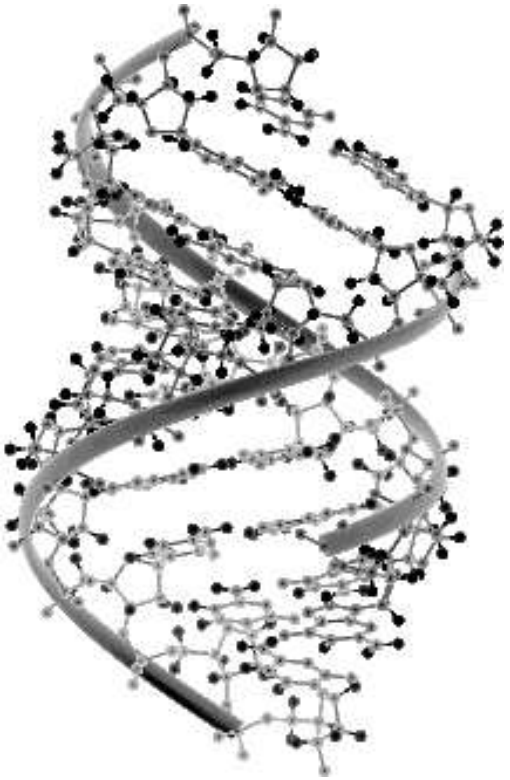
First objective is to set some basic foundations of DNA:

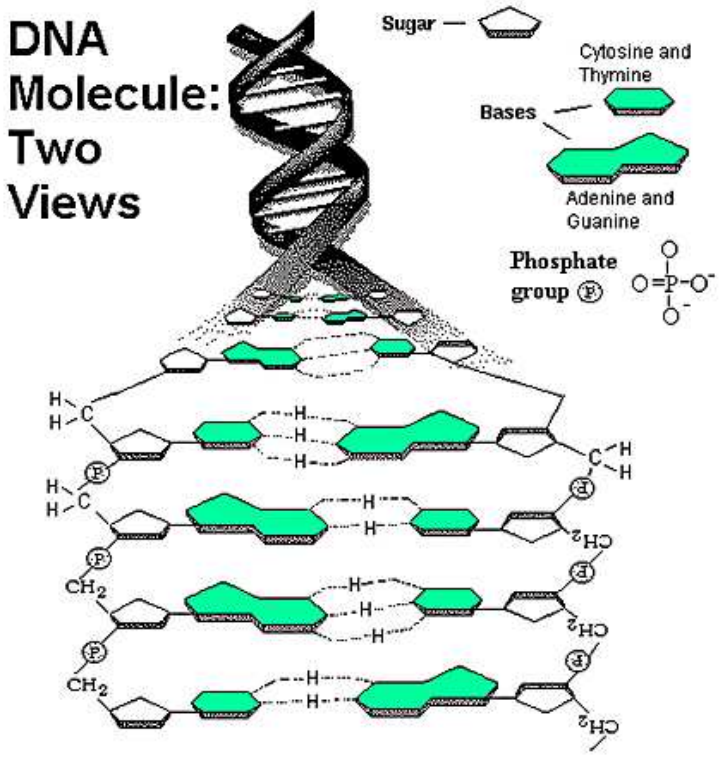
- B form of DNA and its parameters,
- purines and pyrimidines, and pairings
- elements of statistical mechanics of homo-polymers, persistence lengths, worm like chain WLC and twisted WLC or TWLC models

Basics of DNA 1

- Four bases, two Purines A and G, and two Pyrimidines T and C
- Purines are bigger (two covalently closed rings), and Pyrimidines are smaller (one covalently closed ring)
- A pairs with T with two hydrogen bonds, C pairs with G with three hydrogen bonds

Individual atoms can be grouped into relatively rigid units

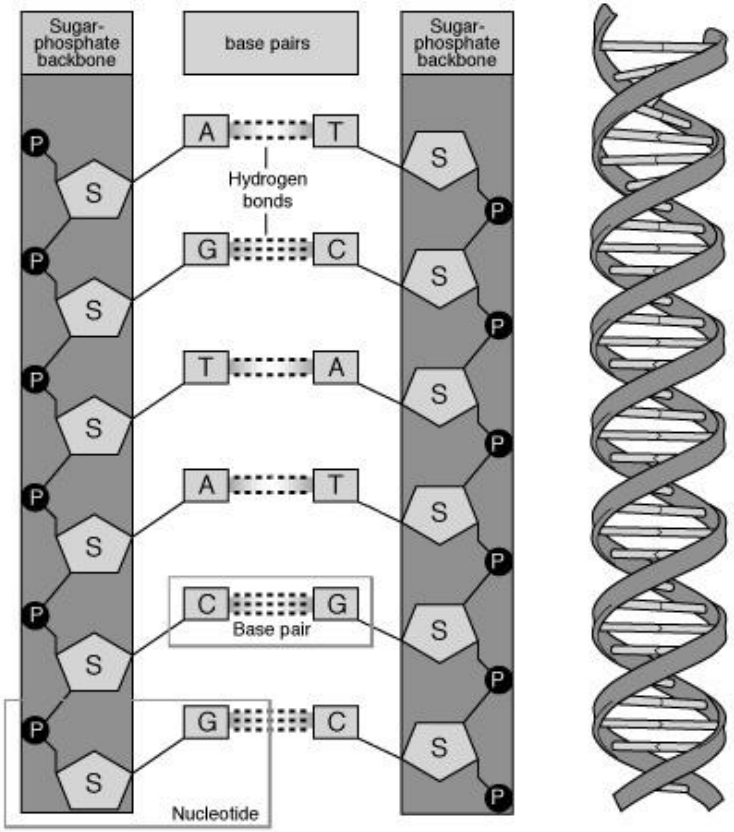




Basics of DNA 2

- Geometry of AT pair almost identical with geometry of CG pair. Each pair forms a flat rectangle, with links to the backbones at the two adjacent corners of a long side, and there is a rotational axis of symmetry through the short diameter.

- Leads to the idealized, but Nobel prize winning, Crick-Watson B-form double-helix (the secondary structure) with arbitrary base sequence along one backbone (the primary structure).



Approximate sequence-independent parameters of the B-form double helix: Diameter 2 nm, rise per base-pair 0.34 nm, one full turn every 10.5 base pairs.

MOST IMPORTANT POINT

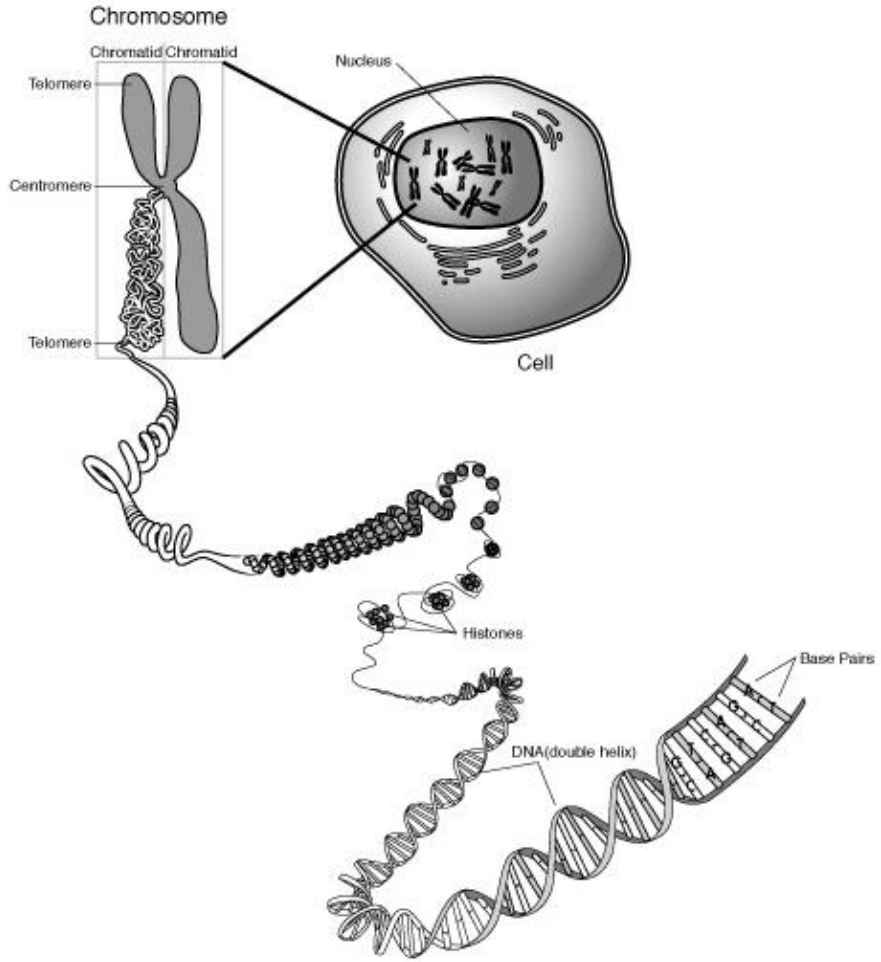
Because of the differences between purines and pyrimidines, and because of the different numbers of hydrogen bonds, the base pair sequence modulates both the intrinsic shape, and local stiffness properties of the double helix—this is the tertiary structure.

Quantifying these modulations of the physical properties of DNA via modelling and experimental verification is the basic objective of Maddocks (and several other) groups' work on DNA modelling.

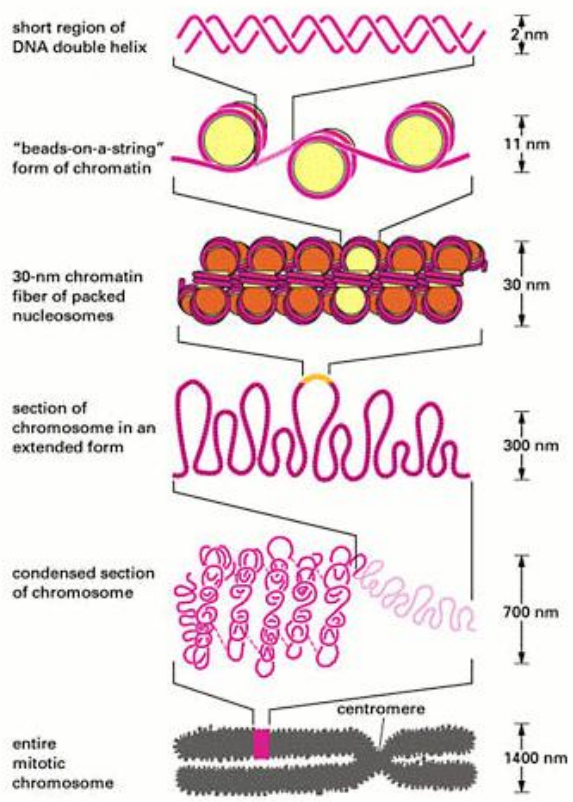
Basics of DNA 3

- Regarded as an isotropic, homopolymer DNA has a persistence length of around 170 base pairs (bps) or so.
- at much longer scales entropic effects dominate, largely sequence-independent
- $0.34 \times 10^{-9} \times 3 \times 10^{9} = 1$ (actually we are diploid so 2m of DNA in each nucleus.)
- a single chromosome, i.e. a covalently bonded molecule, can be 10 centimetres long, 3cm more typical.

Basics of Chromosome 1



Basics of Chromosome 2



Good general principle:

The answer should depend on the question...

Thus to identify the appropriate level of coarse-graining need to understand the scale of the experimental data that is to be modelled.

Today I'll only discuss naked, Crick-Watson, B-form double helical DNA at various length scales.