Nucleosome Positioning by Excluding Genomic Energy Barriers

Alain Arneodo

Laboratoire Joliot-Curie / Laboratoire de Physique, Ecole Normale Supérieure de Lyon
Alain.Arneodo@ens-lyon.fr

Benjamin Audit
Guillaume Chevereau
Julien Moukhtar
Philippe StJean
Cédric Vaillant

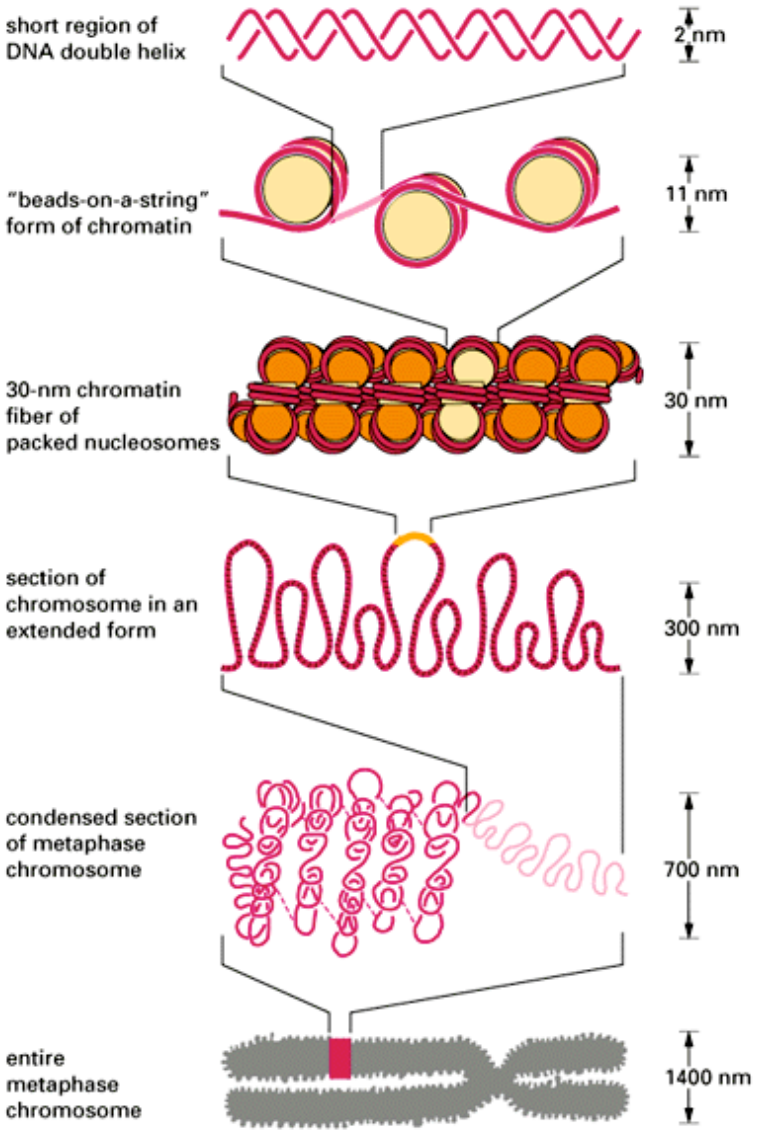
Françoise Argoul
Zofia Haftek-Terreau
Monique Marilley
Pascale Milani
Cendrine Moskalenko

ENS de Lyon, France

Yves d'Aubenton-Carafa
Claude Thermes

CGM, Gif-sur-Yvette, France

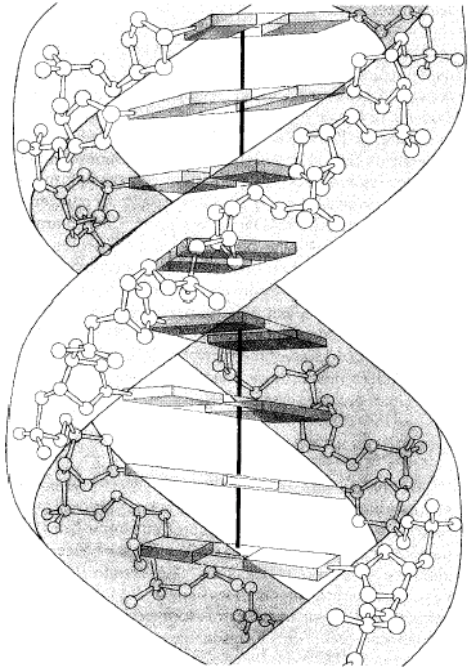
To which extent does the DNA sequence “code” for the 3D structure of chromatin?

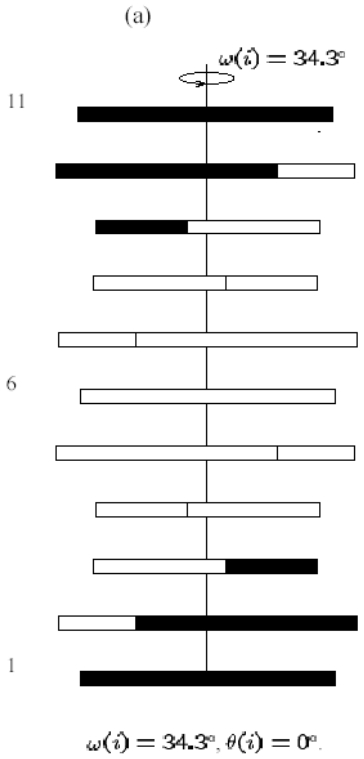


Sequencing projects result in 4 letter texts :

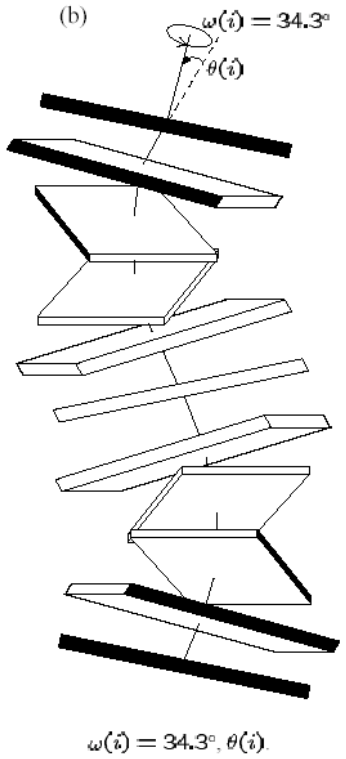
```
gtcagtttcctgaggcgggtcgggacccaggcgtgagactggagtctgcc
caggggcccagctgagccagcctcctcgtcagctgcttgggccgccagga
cgccgccgggggtgcgccgcgcttccctggatggggtgccccccactcccc
tcggagccccagggagaccccccgaactcagctcctctcaggggtgccag
ggggaccccctcaaactccactccccgcaggttcctggggagacgccccct
gctcgattcccctcagggtcccagggagacccccctaattcagctcctctc
aggggtactgggggacctctcgagctccactcccatcagggtcccaggga
gacccccccaactatgctcaggggtcccagggagatgccagcaccccaact
ccgcttccctggggcccccctcccccttacagctcaacttccctcgagagt
ctggggctggggctccgttcagttcttgagtccccttcccctcggggtgtc
ccggggccgcccacccccacactgtctgtgattccccaaggcgcgggtct
cgggccgcagcctgttccacgttctgctgctcgttctttttctggctcctt
gctttcgaaggagagaaggaggccttcgtttccagtctttttgccttttc
taatggagccctgcttttccttccgtgtcccttcaggctacttctgccag
gtttctatttttcattctttattatgacttcgcccaaaatattcttgact
tctattgagaaggattcgggggtctatttcttattcggaggcgtgtgctt
aagttccaaacagatgaggattttccagttaatccttctggggtgactta
ttgcttaatgccaccatagccagaaaatggactctcagtgtccgaaactg
cattcggctctgaagtgtctgtccttgtcacctcttgcaatgtttcgcgg
cgggaagcctgcactcgccgacgctgacgtaactgtttctgtctttcagg
tctacagcctcctgtgggtgggcgatattgacatatactttatttctata
tatgttatgaactcaatatttcttgcagcgggtctgctgataataagata
tgcctactctgcgagtctggaagccatcttaagcttaccctgtatgtgcc
ccatgcatctcttccgttacacggctcctgagttgacacctgtgtgataa
actggtaatagcaagtaaactgttttcttgtgctctgtaagctgctctag
caaattatctaggaggaggtggtcttggaaacccctgatttataagcggg
cagtcagcagtacacgtggcccagaatcgtgattggcatttgaagtgggg
gcagtagggtgggactgagcccttcacctgtggggtctgccctgctcaag
gcagtgtcagaattgaagtgaaatgttggacggtcggtgtccagagagtt
ggagaactggtttgtgtgtaaaaactnacatatttagggtcagaagtatg
...
```

Sequence induces local structural disorder along the DNA polymer





$\omega(i) = 34.3°, \theta(i) = 0°$.



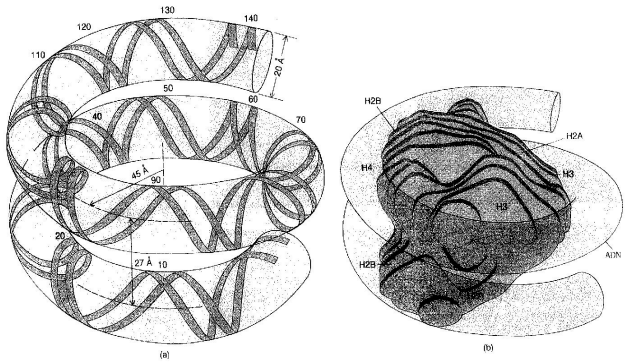
$\omega(i) = 34.3°, \theta(i)$.

frozen structure @ T = 0

Trinucleotidic structural tables based on experiments

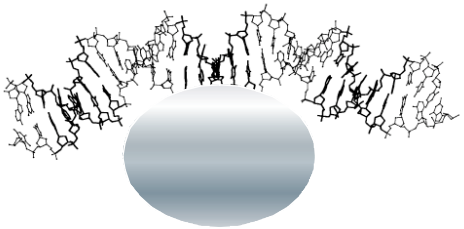
Nucleosome positioning

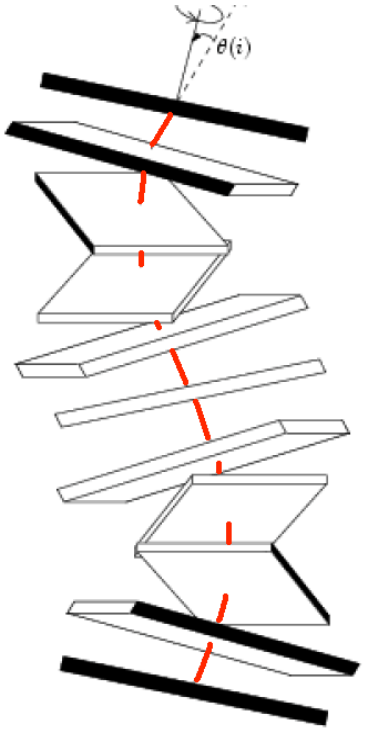
local curvature



Dnase I sensitivity

Local flexibility



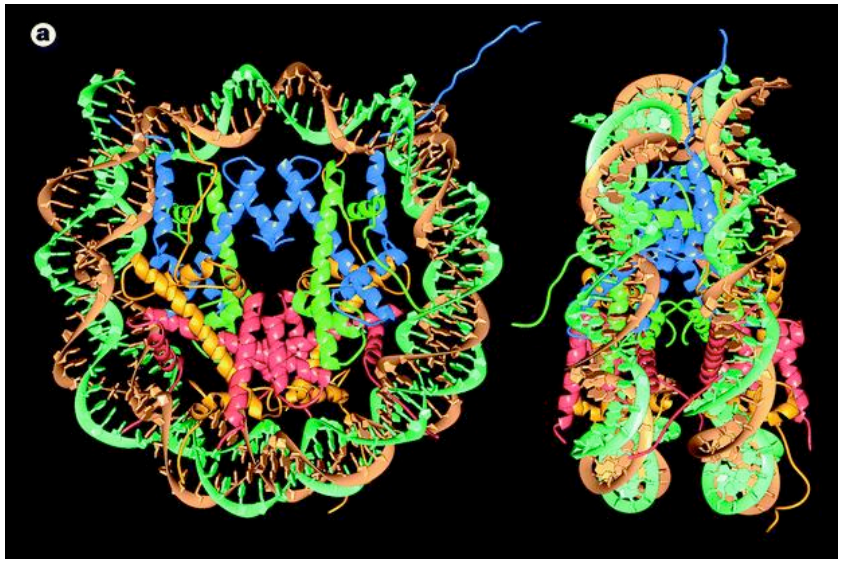


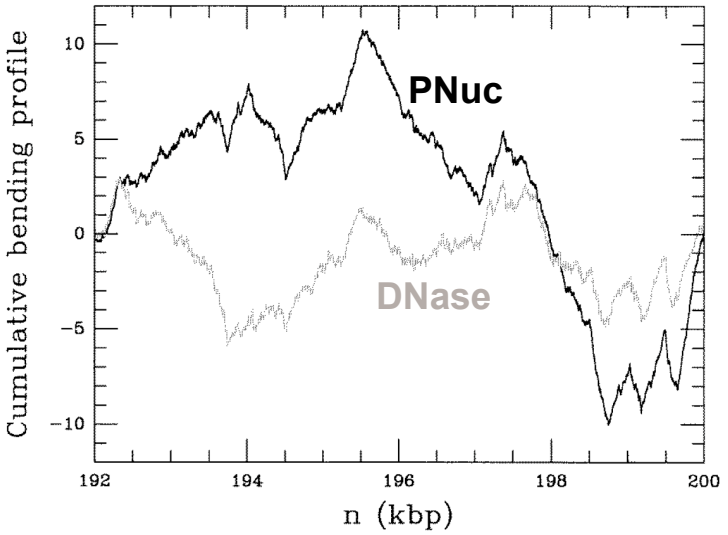
Trinucleotide	PNuc	DNase I
AAA/TTT	0.0	0.1
AAC/GTT	3.7	1.6
AAG/CTT	5.2	4.2
AAT/ATT	0.7	0.0
ACA/TGT	5.2	5.8
ACC/GGT	5.4	5.2
ACG/CGT	5.4	5.2
ACT/AGT	5.8	2.0
AGA/TCT	3.3	6.5
AGC/GCT	7.5	6.3
AGG/CCT	5.4	4.7
ATA/TAT	2.8	9.7
ATC/GAT	5.3	3.6
ATG/CAT	6.7	8.7
CAA/TTG	3.3	6.2
CAC/GTG	6.5	6.8

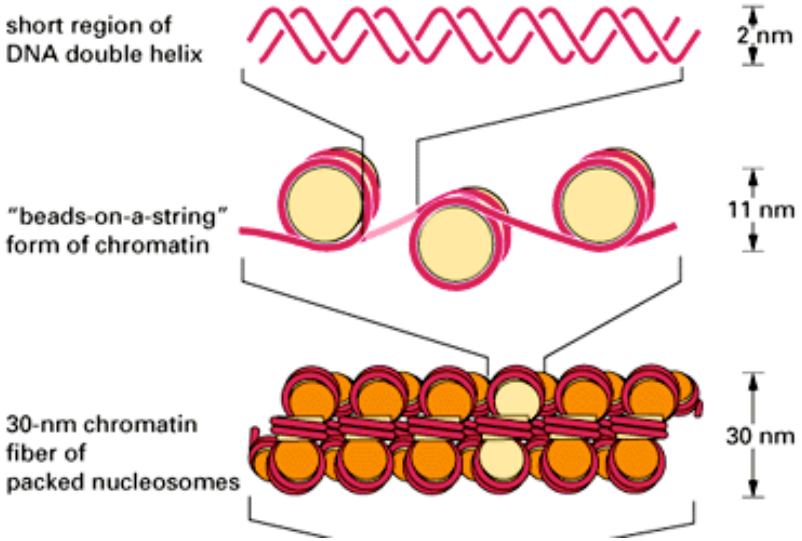
Trinucleotide	PNuc	DNase I
CAG/CTG	4.2	9.6
CCA/TGG	5.4	0.7
CCC/GGG	6.0	5.7
CCG/CGG	4.7	3.0
CGA/TCG	8.3	5.8
CGC/GCG	7.5	4.3
CTA/TAG	2.2	7.8
CTC/GAG	5.4	6.6
GAA/TTC	3.0	5.1
GAC/GTC	5.4	5.6
GCA/TGC	6.0	7.5
GCC/GGC	10.0	8.2
GGA/TCC	3.8	6.2
GTA/TAC	3.7	6.4
TAA/TTA	2.0	7.3
TCA/TGA	5.4	10.0

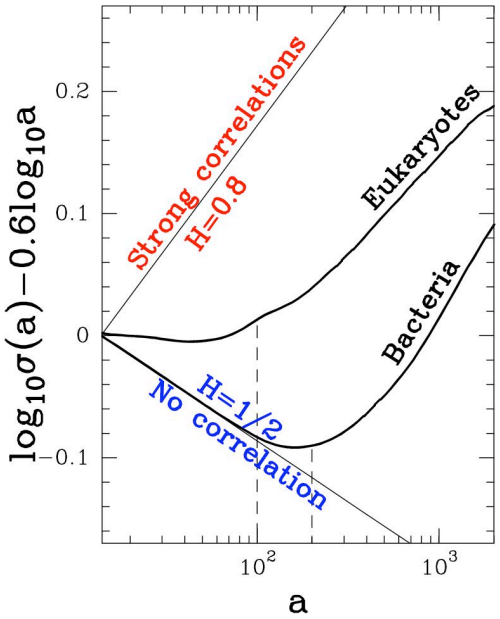
trajectory of the double helix axis

Long-range correlations in genomic DNA: A signature of the nucleosomal structure

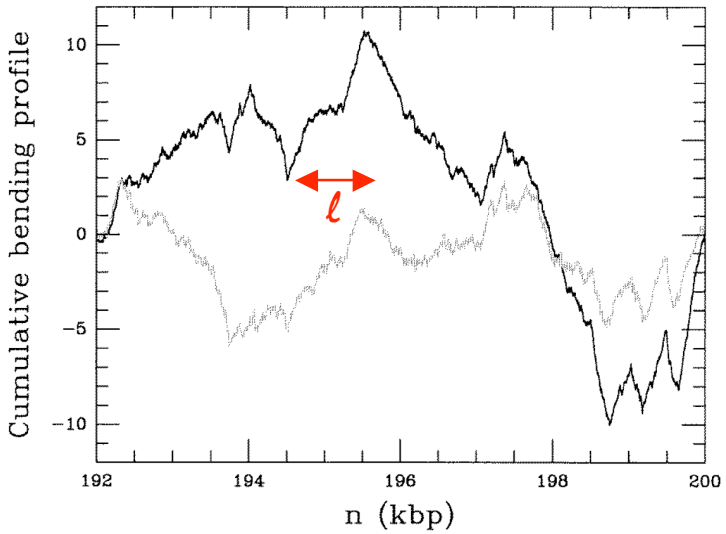








Roughness exponent



- Root-mean square of the height fluctuations

$$W(\ell) = \text{rms}\,[f(n+\ell)-f(n)] \sim \ell^{H}$$

H = roughness exponent $D_f = 2 - H$

- Power spectrum

$$S_f(k) \sim k^{-(2H+1)}$$

- Random walks
  - $0.5 < H < 1$ **LONG RANGE CORRELATIONS**
  - $H = 0.5$ **UNCORRELATED**
  - $0 < H < 0.5$ **ANTI-CORRELATIONS**

- Correlation function

$$C_f(\tau) = \langle \Delta_1 f(n)\, \Delta_1 f(n+\tau) \rangle - \langle \Delta_1 f(n) \rangle^2$$
$$\sim \tau^{2H-2}$$

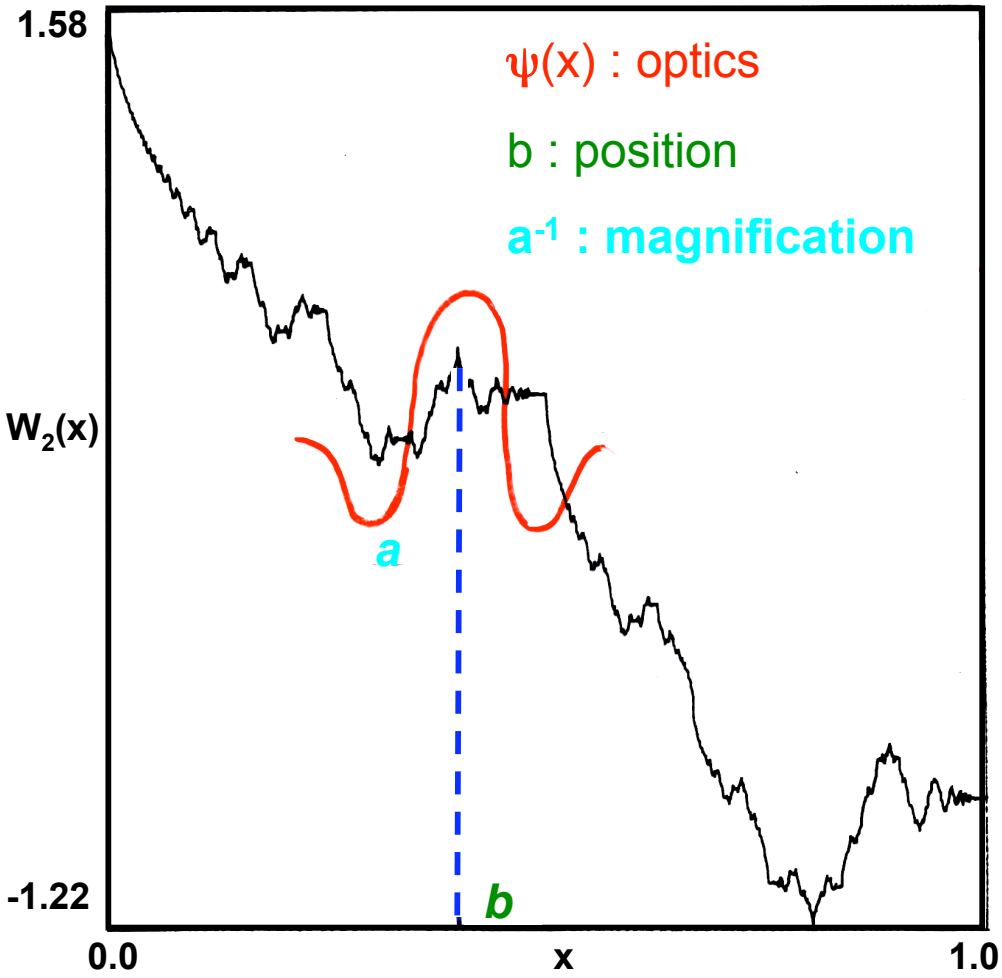
Wavelet Analysis of Fractal Signals

Definition

$$T_\psi[f(x)](b,a) = \frac{1}{a}\int_{\mathbb{R}} f(x)\ \psi(\frac{x-b}{a})\mathrm{d}x$$

a scale parameter, b space parameter, ψ analyzing wavelet

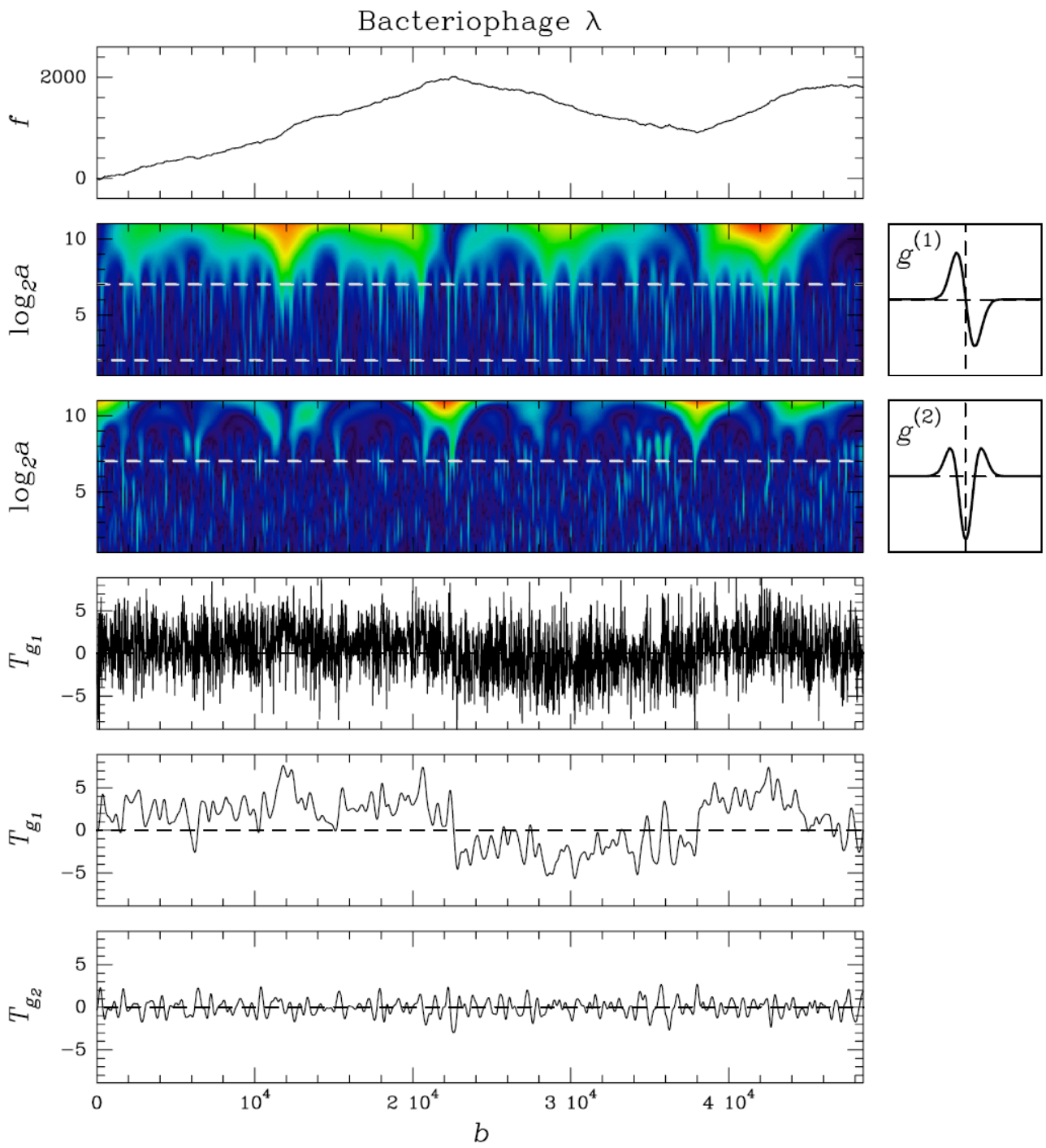
Mathematical microscope

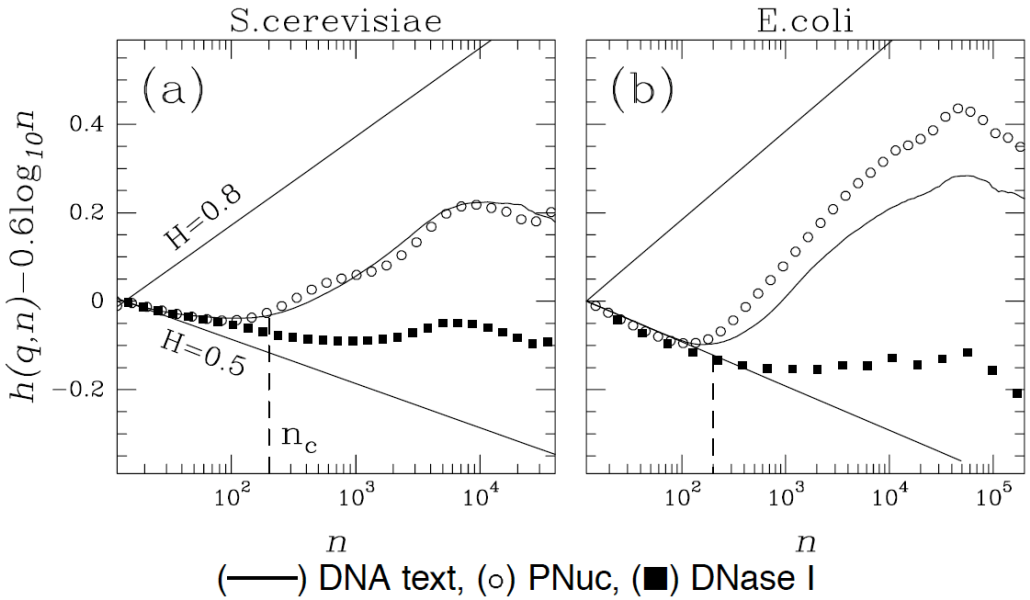


Singularity scanner

The wavelet transform allows us to **LOCATE** (**b**) the singularities of f and to **ESTIMATE** (**a**) their strength h(x) (Hölder exponent)

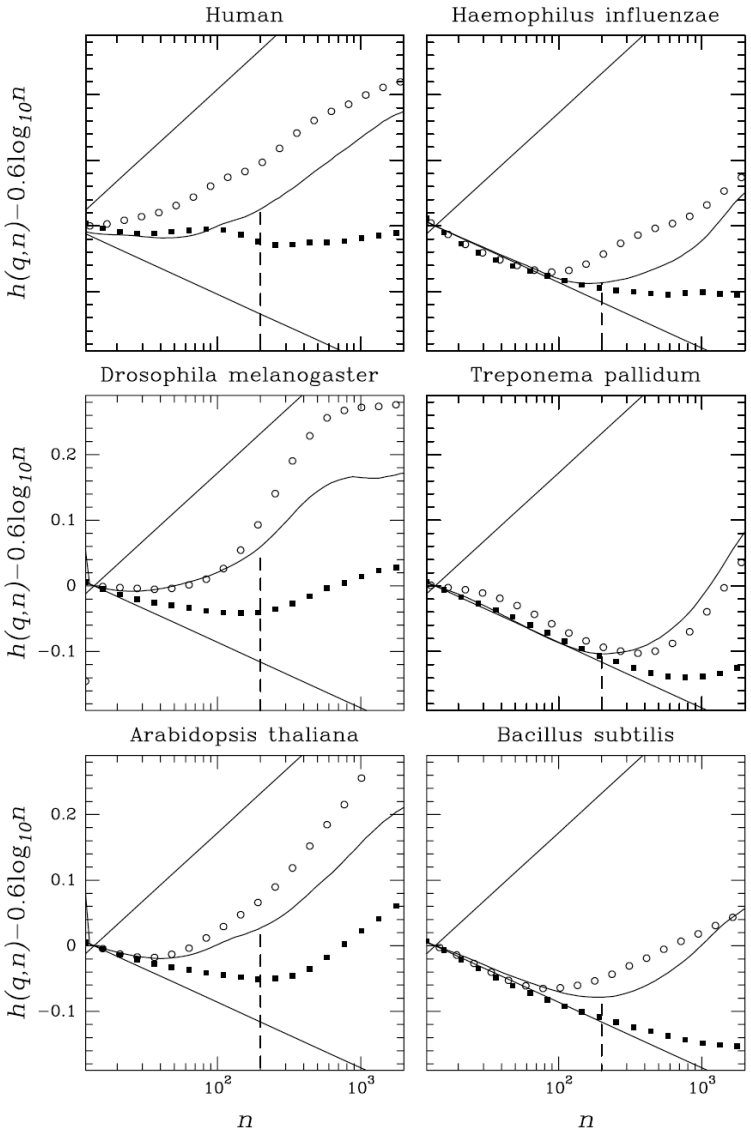
Wavelet Analysis of the sequence of bacteriophage λ





Hypothesis: LRC in the small scales regime is the signature of of the nucleosomal structure

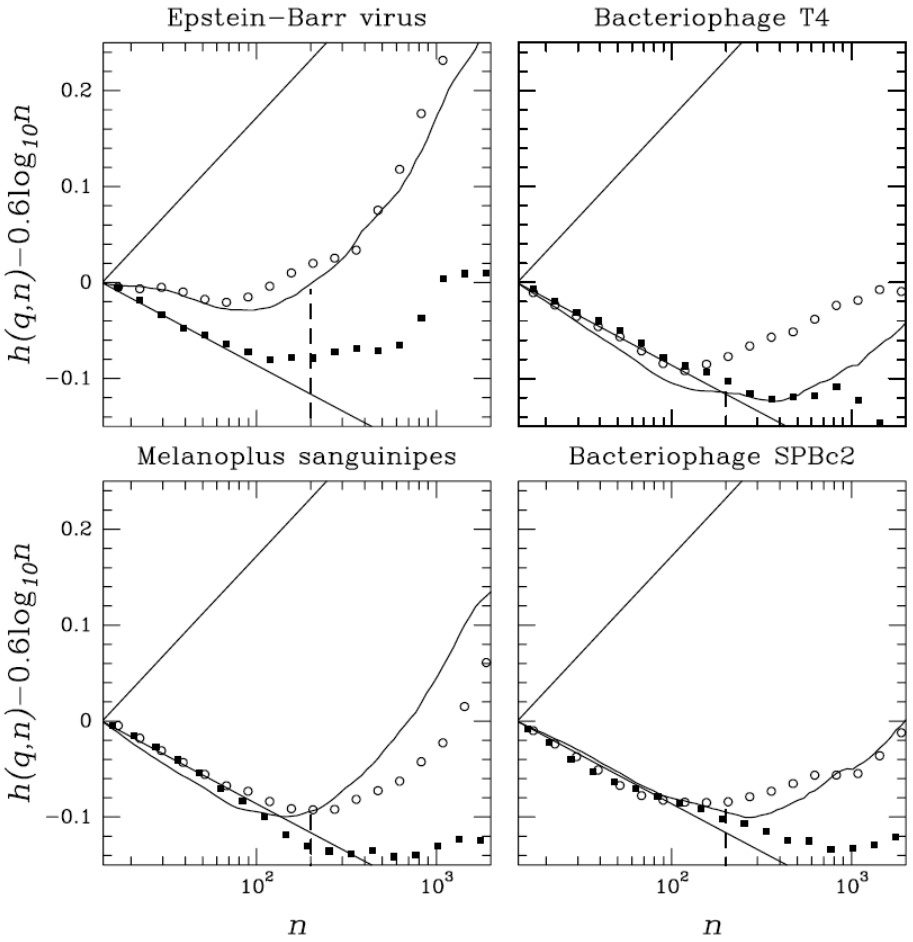
Eucaryotes Bacteria



(——) DNA text, (○) PNuc, (■) DNase I

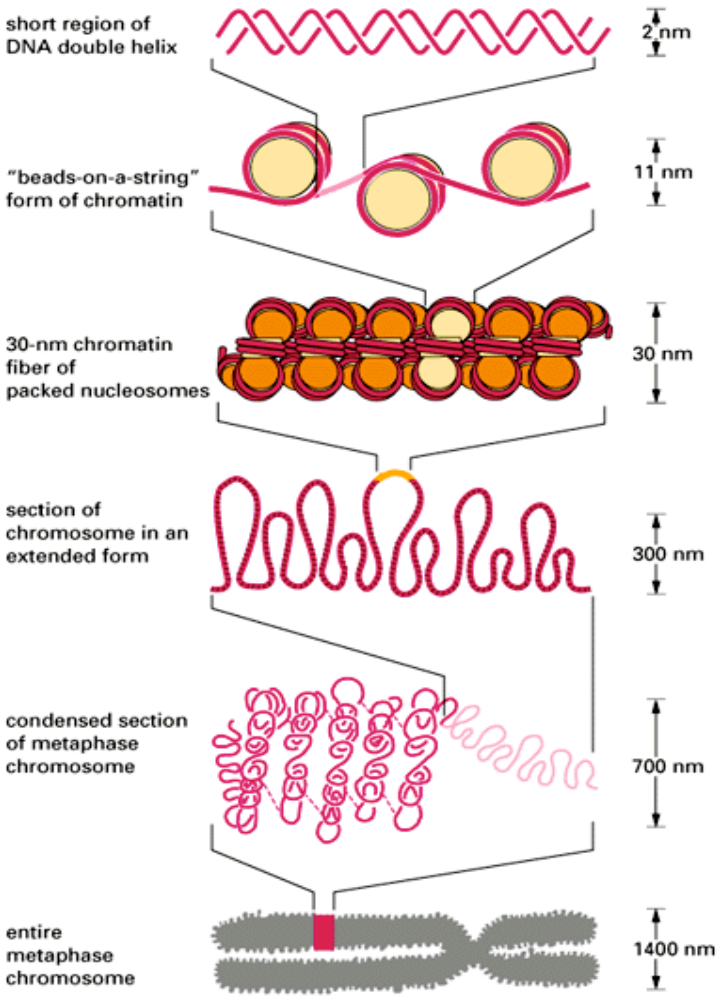
Nucleosomes No nucleosomes

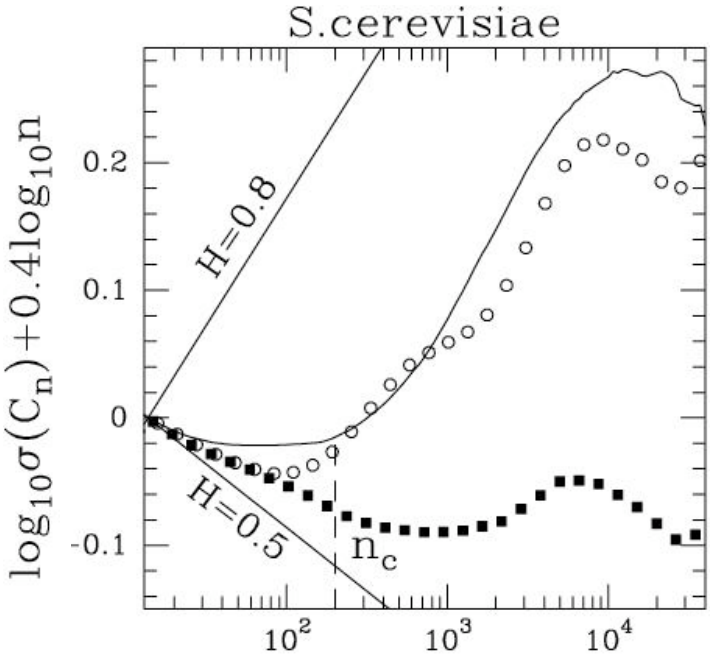
SMALL SCALES LRC ARE RELATED TO NUCLEOSOME LIKE STRUCTURES

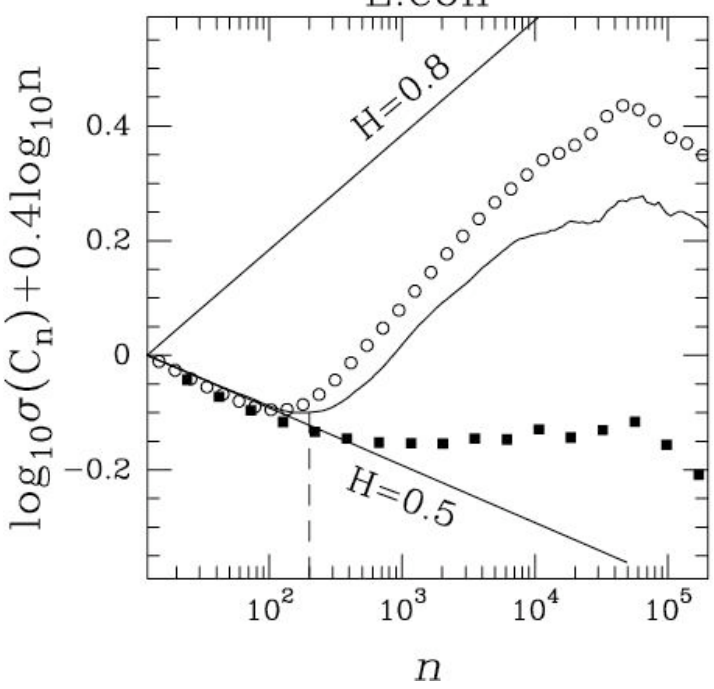


Pox virus don’t display LRC in the small scale regime

HIERARCHICAL STRUCTURE OF EUCARYOTIC DNA

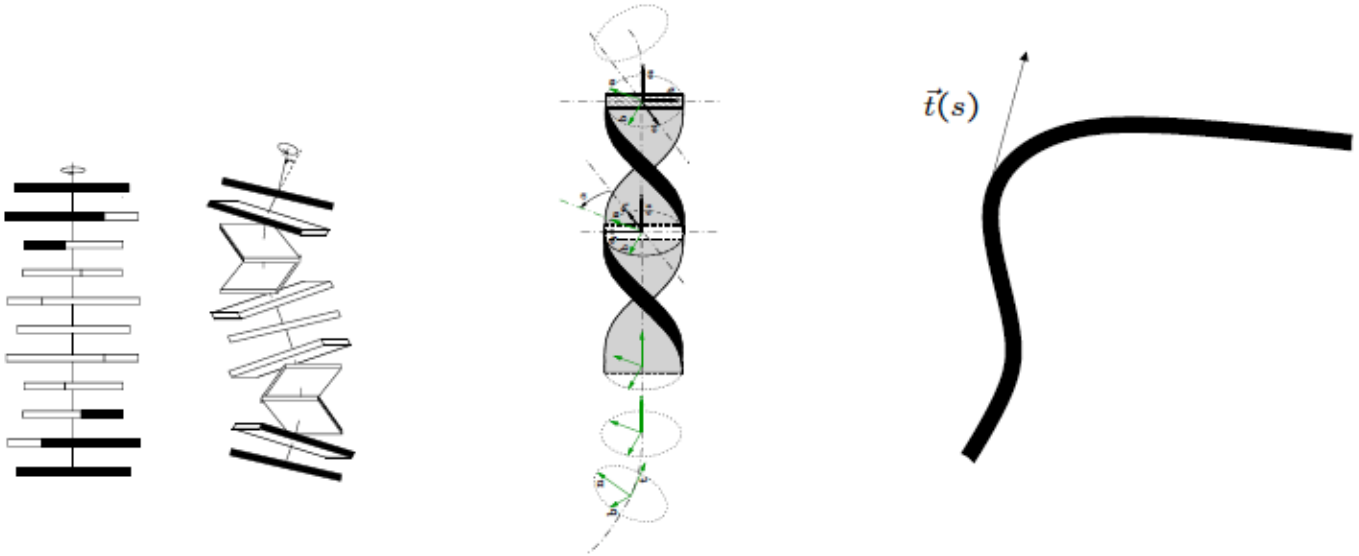






DNA in chromatin: Modelling the influence of the sequence on the structure and dynamics of chromatin

Heterogeneous elastic model for DNA



Local polymer structure locally defined by 3 angles: $(\Omega_1(s), \Omega_2(s), \Omega_3(s))$
and 3 flexibilities: $(A_1(s), A_2(s), C(s))$

$$\frac{\delta E}{k_B T} = \frac{A_1}{2}(\Omega_1 - \Omega_{o1})^2 + \frac{A_2}{2}(\Omega_2 - \Omega_{o2})^2 + \frac{C}{2}(\Omega_3 - \Omega_{o3})^2 ds$$

Equilibrium configuration: $(\Omega_{o1}(s), \Omega_{o2}(s), \Omega_{o3}(s)) \neq (0,0,0)$
$\rightarrow \Omega_{o3} \sim 34°/$ pb.
$\rightarrow \Omega_{oi}$ vs séquence: structural tables (crystallography, AFM, molecular dynamics)

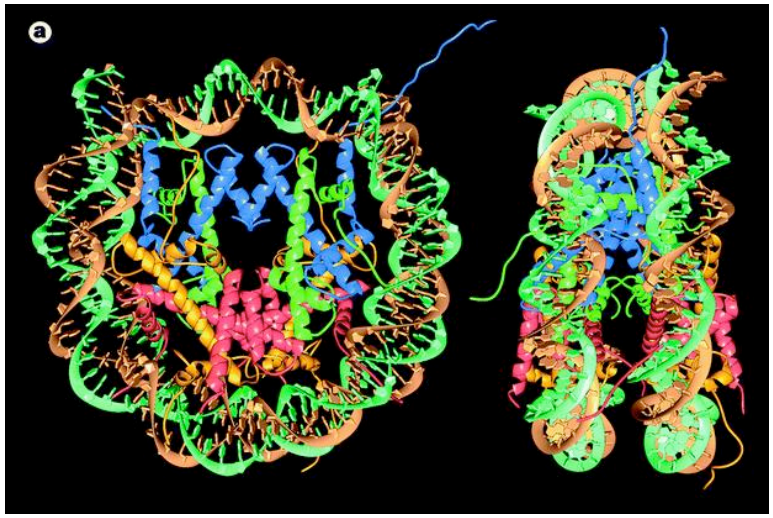
DNA at large scale: Ideal rod like chain model

$\beta\delta E = \frac{A}{2}(\frac{dt}{ds})^2 + \frac{C}{2}(\Omega_3)^2$

persistence length: $l_p = \ln < t(s).t(s+l) > /l = A \sim 50$ nm, $C \sim 75 - 110$ nm

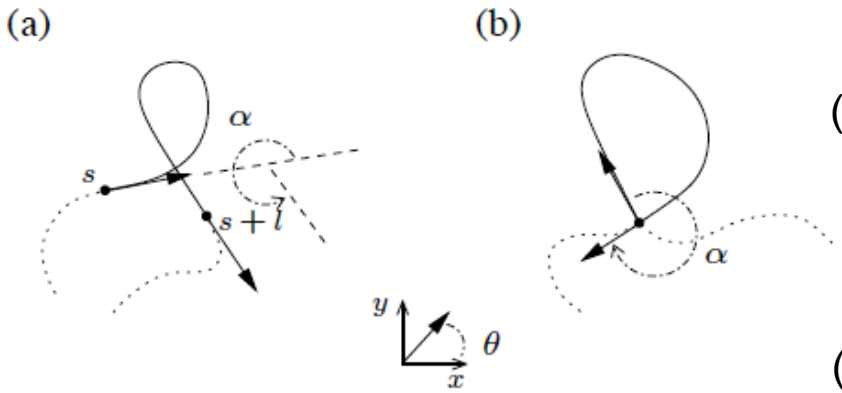
2D DNA loops

- **Nucleosome ~ 2 coplanar loops of DNA**



- **Inhomogeneous 2D elastic model**

$$E(s,l) = \frac{A}{2}\int_s^{s+l}(\dot{\theta}(u) - \dot{\theta}_o(u))^2 du$$



(a) Winding constraint

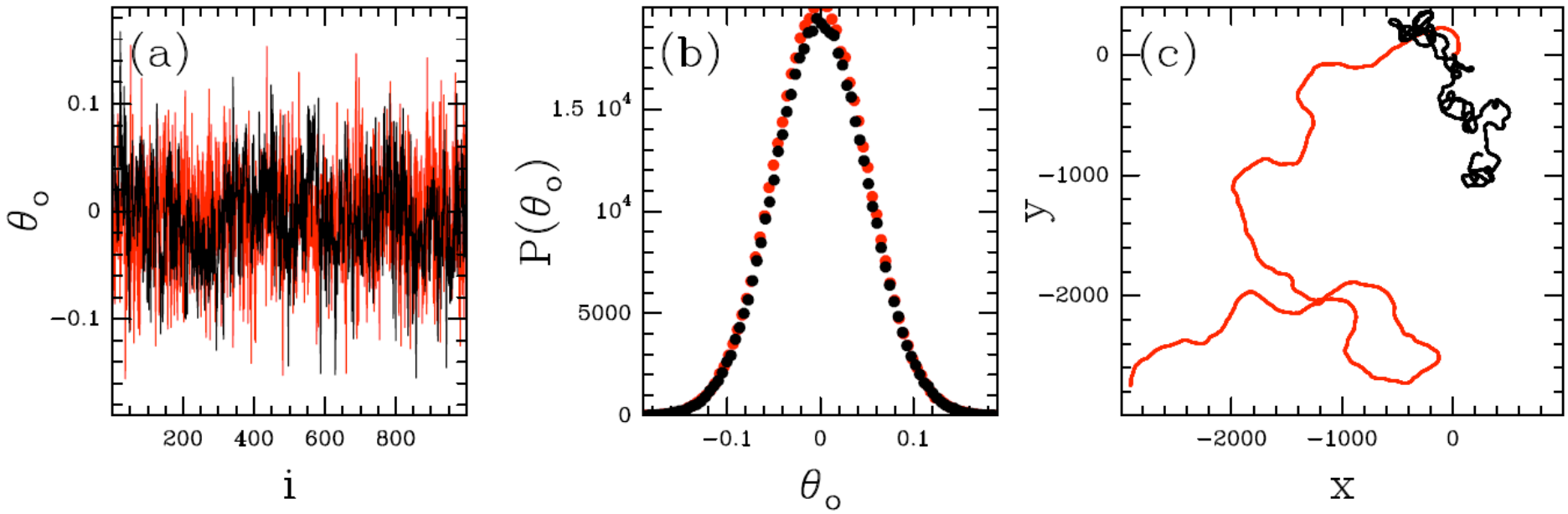
$$\Delta\theta(s,l) = \theta(s+l) - \theta(s) = \alpha$$

(b) Cyclisation constraint

$$\Delta\theta(s,l) = \alpha \;\&\; \int_s^{s+l}\cos\theta = \int_s^{s+l}\sin\theta = 0$$

Intrinsic curvature

$H = 0.5$ $H > 0.5$



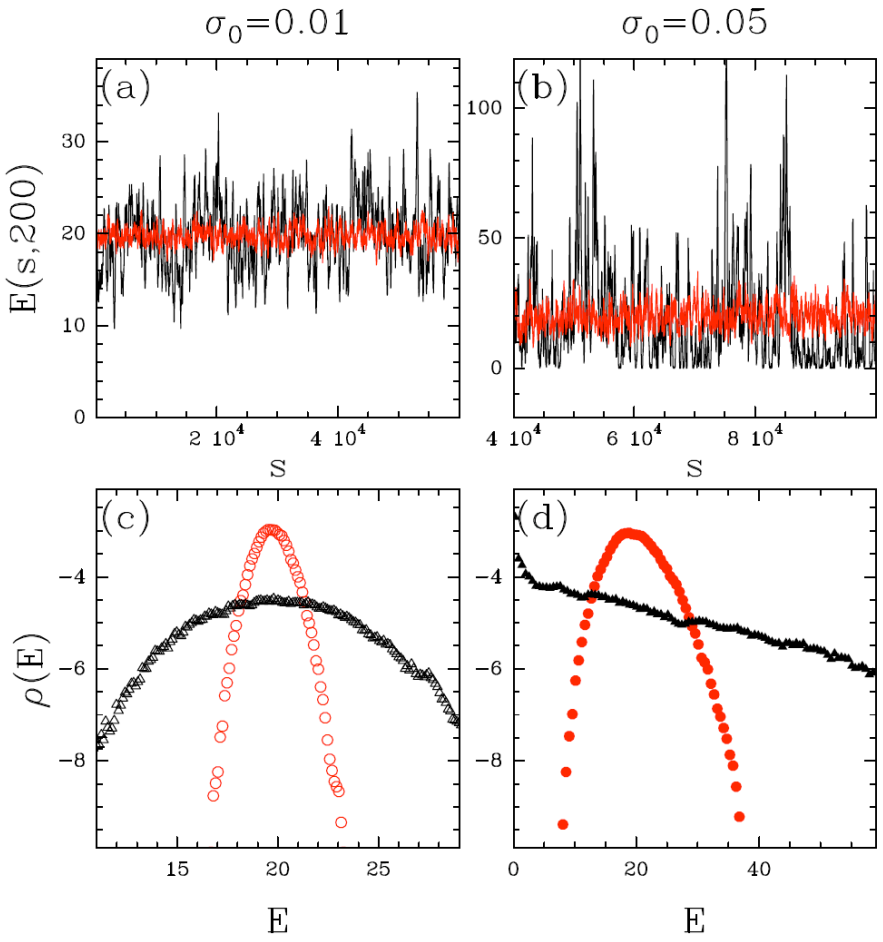
Intrincic curvature $\dot{\theta}_o(s)$: Gaussian noise $\mathcal{N}(0, \sigma_o)$

Intrinsic winding $\Delta\Theta_o(s,l) = \int_s^{s+l} \theta_o(u)du$: $|\Delta\Theta_o(s,l)| \sim \sigma_o l^H$

long-range correlations ⇒ polymer intrinsically more compact

Loop energy profiles

$H = 0.5$ $H > 0.5$



- Energy

$$E(s,l) = \frac{\tilde{A}}{2l}[\alpha^2 - 2\alpha\Delta\Theta_o(s,l) + \Delta\Theta_o^2(s,l)]$$

- Free energy

$$\beta f(s,l) = \beta E(s,l) - \Delta S(s,l)$$

$\Delta S(s,l) = -7/2 \ln l + b$ Independently of position *s*

- Weak disorder: Gaussian statistics

$$\beta\overline{f}(l) = \frac{A}{2}\frac{\alpha^2}{l} + \frac{A}{2}\sigma_o^2 l^{2H-1} + c \ln l - b$$

$$\Lambda(l) = A^2\alpha^2\sigma_o^2 l^{2H-2}$$

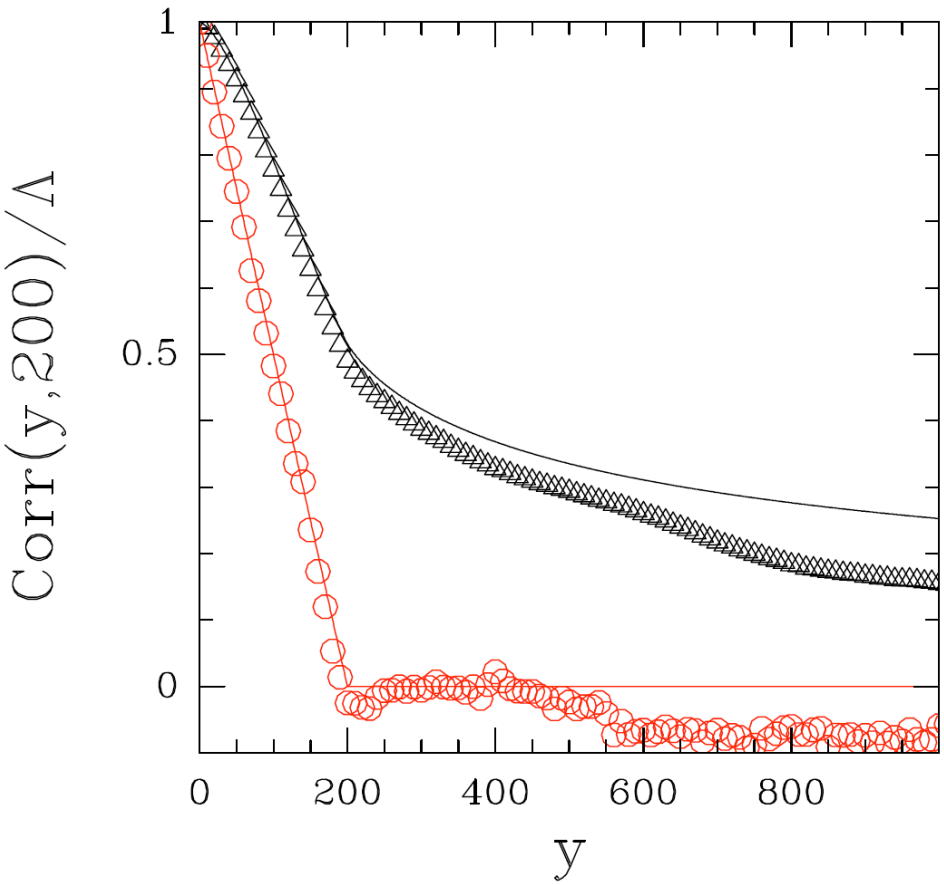
- Strong disorder, H ↗ :
  exponential distribution of energetic barriers

long-range correlations ⇒ larger fluctuations of energy profiles

Correlations in loop energy profiles

$H = 0.5$

$H > 0.5$



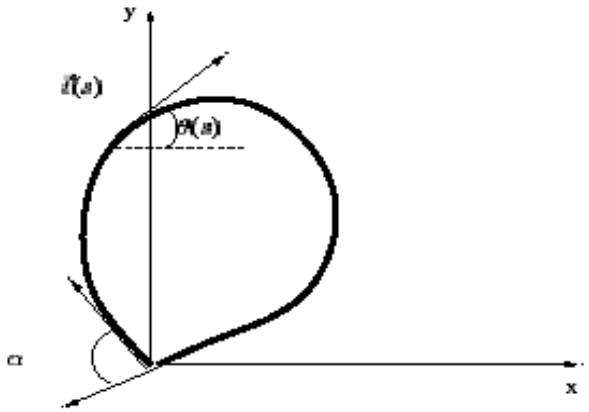
$$Corr(y/l) \underset{y\to\infty}{\sim} 2H(2H-1)\Lambda(l)(y/l)^{2H-2}$$

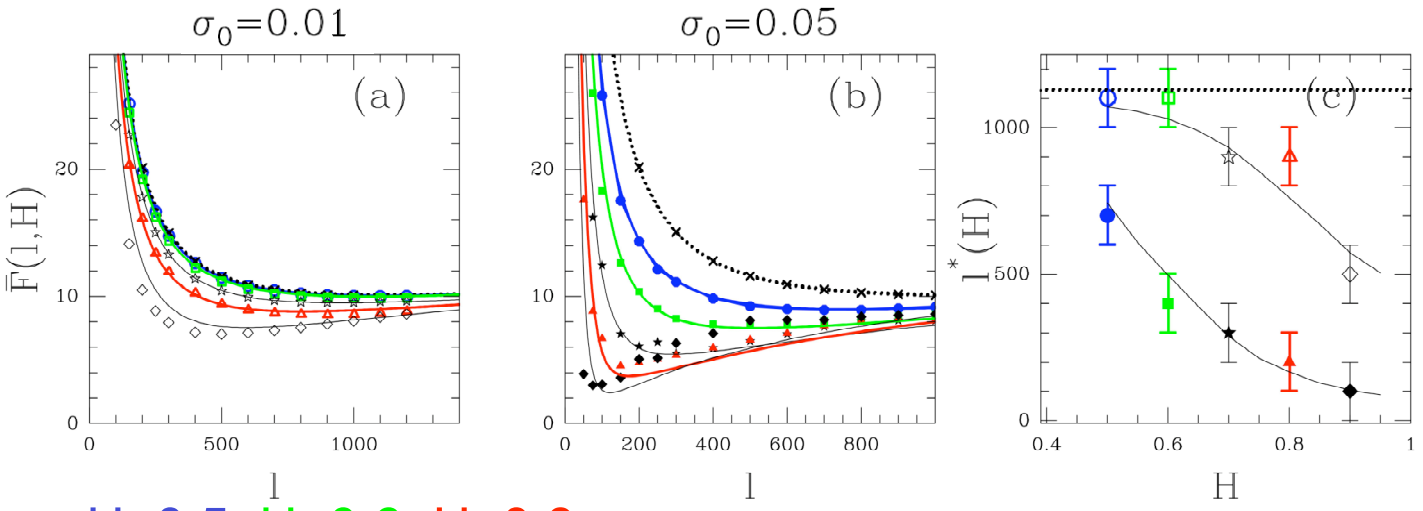
long-range correlations ⇒ persistent fluctuations of energy profiles

Formation of DNA loops

Free-energy for the formation of a loop of length *l* in a semi-flexible chain of length *L* >> *l*
(**2D elastic model**)

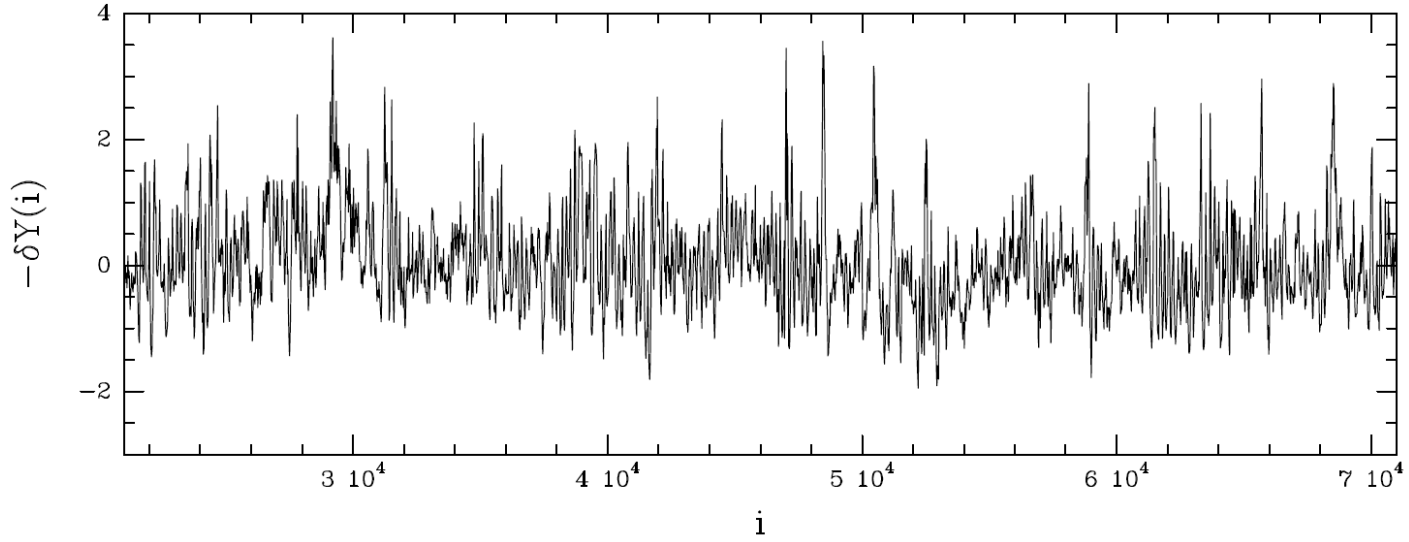
Equilibrium: $\beta F(l) = -\ln\left(\int_0^L ds \exp -[\beta f(s,l)]\right)$



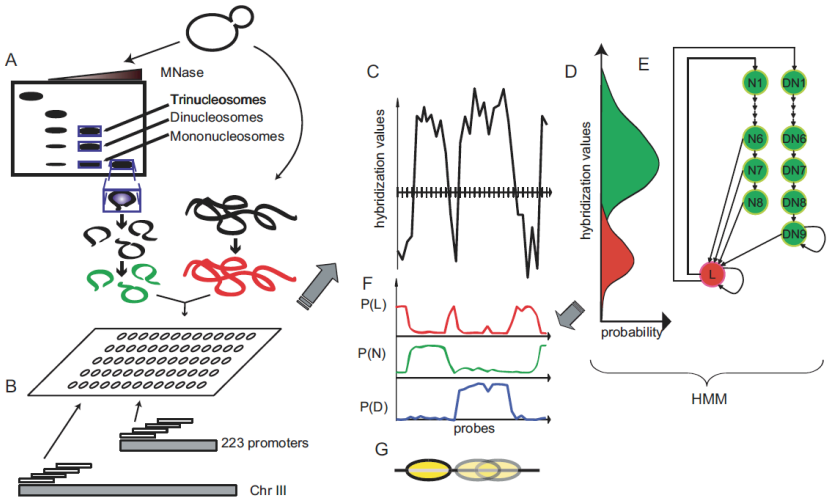


long-range correlations ⇒ facilitates the formation of small loops

Genome-Scale Identification of Nucleosome Positions in *S. cerevisiae* (Yuan et al., Science 309)

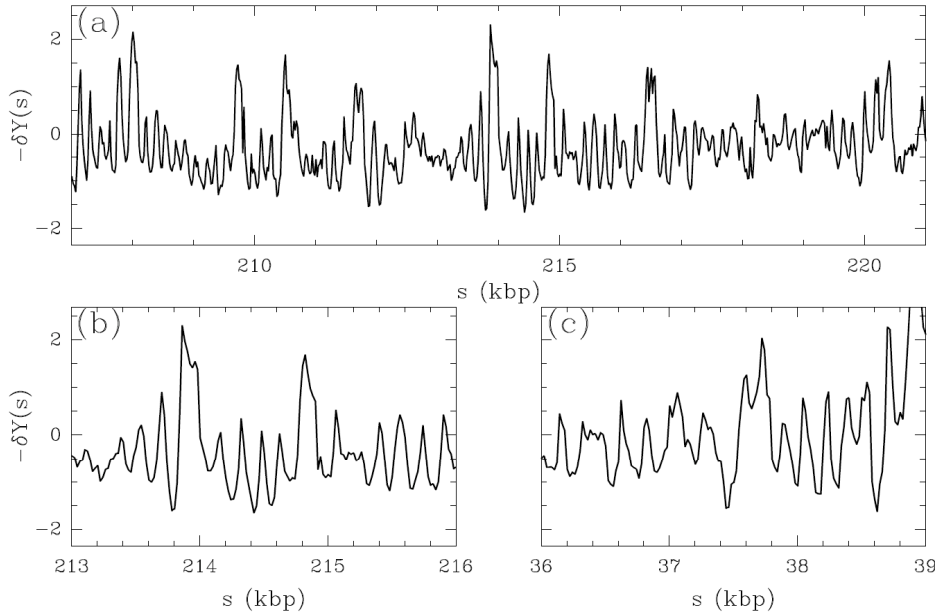


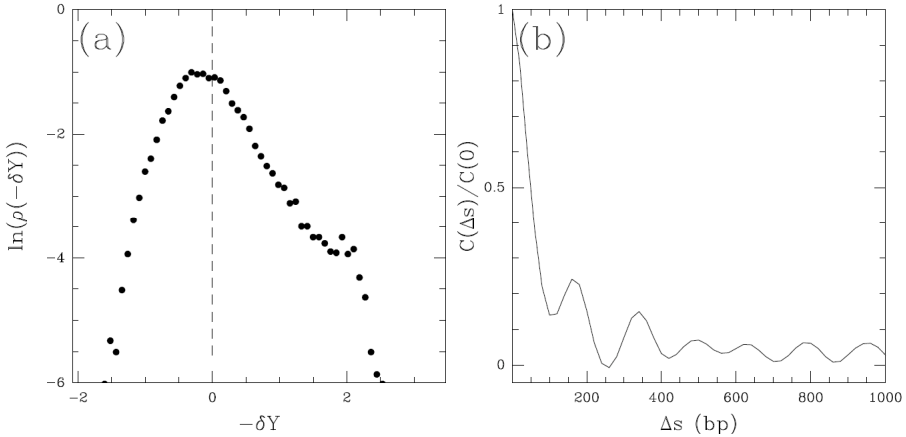
Rando Fig S1

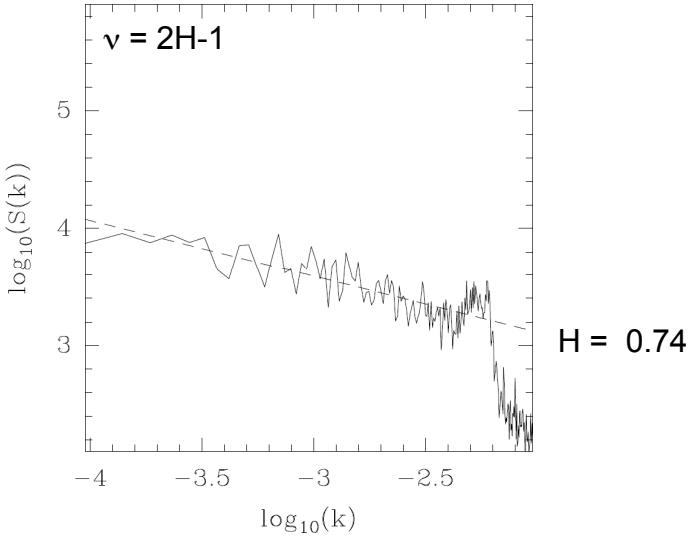


1. Hybridisation of nucleosomal DNA on chip
2. Results: $Y(i) \sim \log_2 [P(i \in nuc)]$
3. $P(i \in nuc) = e^{-\beta E_{eff}(i)}/Z, -Y(i) \propto E_{eff}(i)$

Statistical characterization : I Experiment







H = 0.74

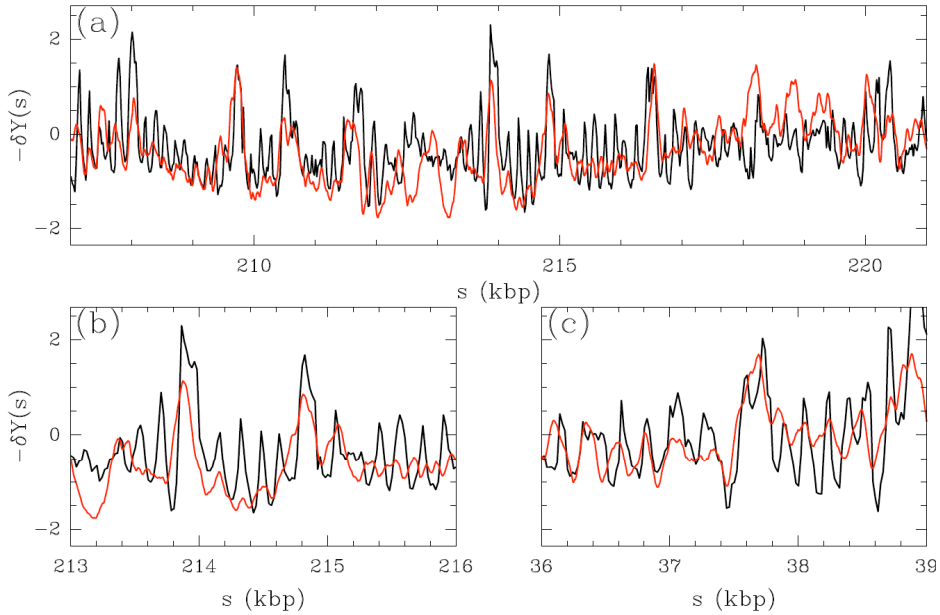
Large-scale disorder of nucleosome positioning

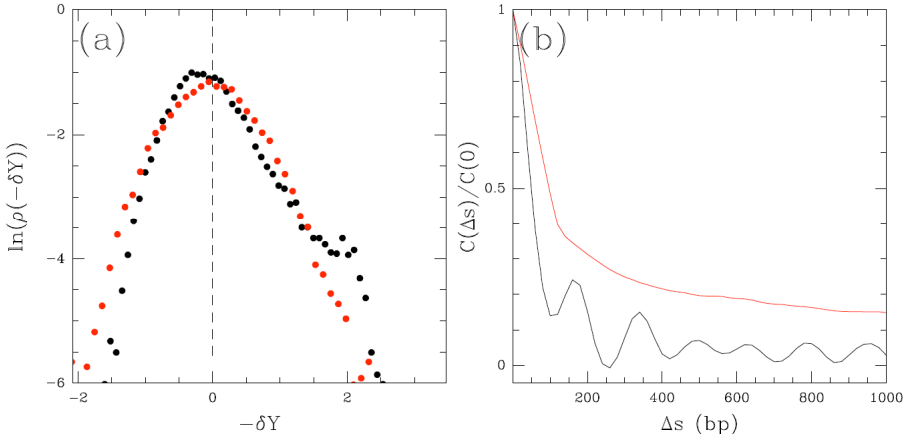
Long-range correlations between 200 and 1000bp with H~0.8

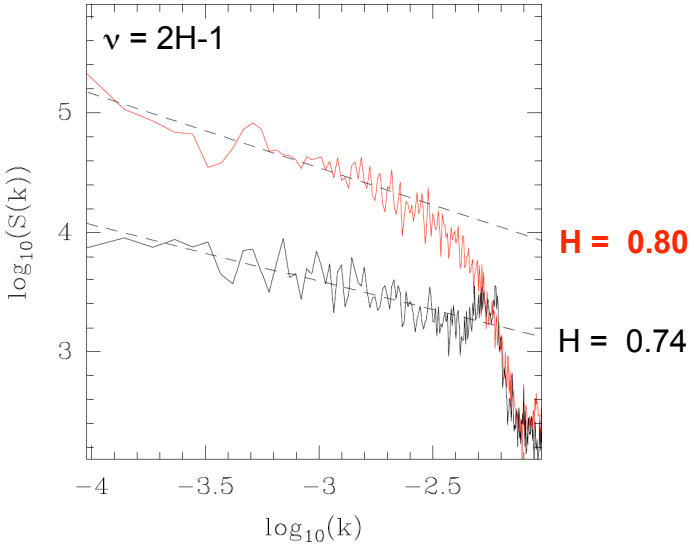
Small-scale periodic component at 160 bp

Weak positioning: fat tail (exponential) statistics

Statistical characterization : II One nucleosome model



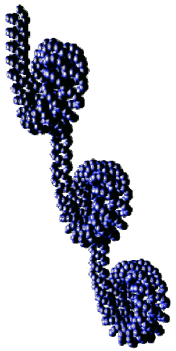




H = 0.80

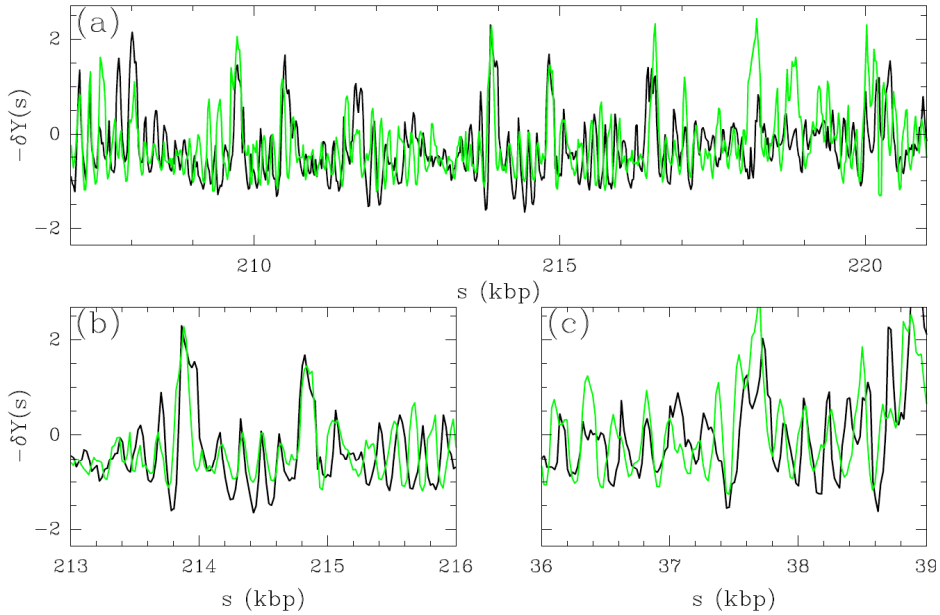
H = 0.74

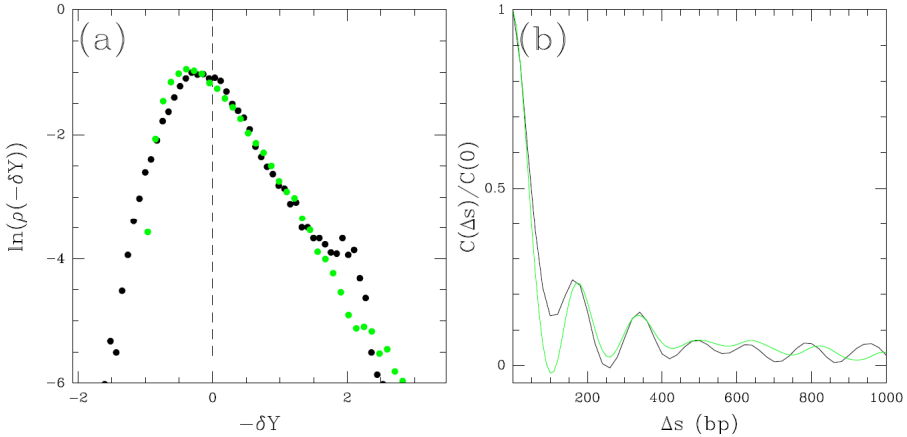
One nucleosome energy landscape based on a simple sequence dependent helical model

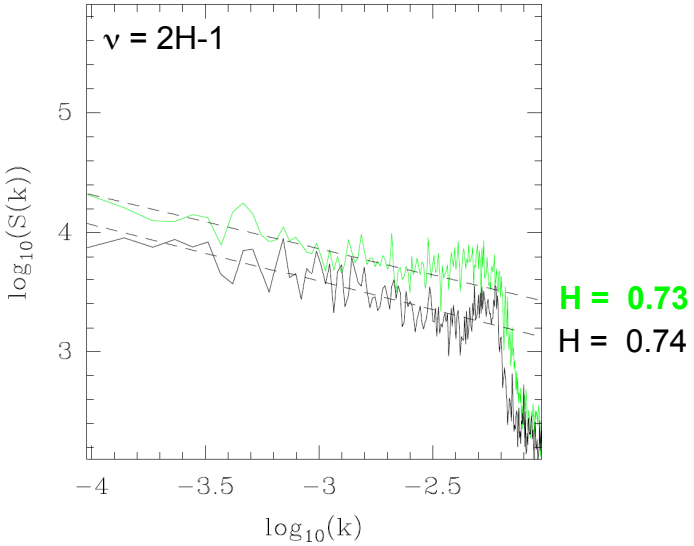


Radius	R = 4.19 nm
Pitch	P = 2.49 nm
Total length	ℓ = 125bp

Statistical characterization: III Many nucleosomes simulations







H = **0.73**

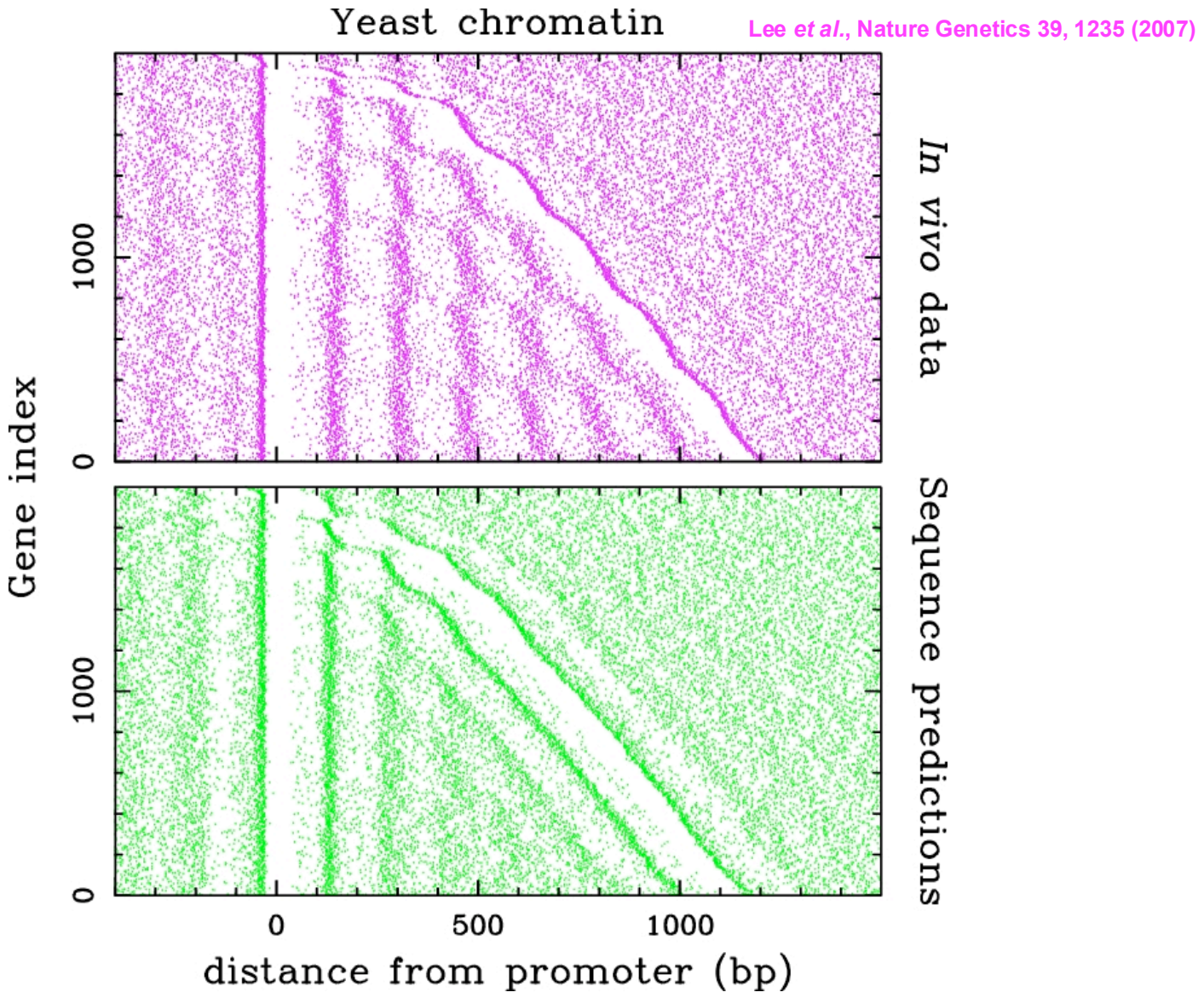
H = 0.74

Many nucleosomes Grand Canonical Monte-Carlo simulations of the nucleosomal array

Chemical potential: 1nuc / 250bp

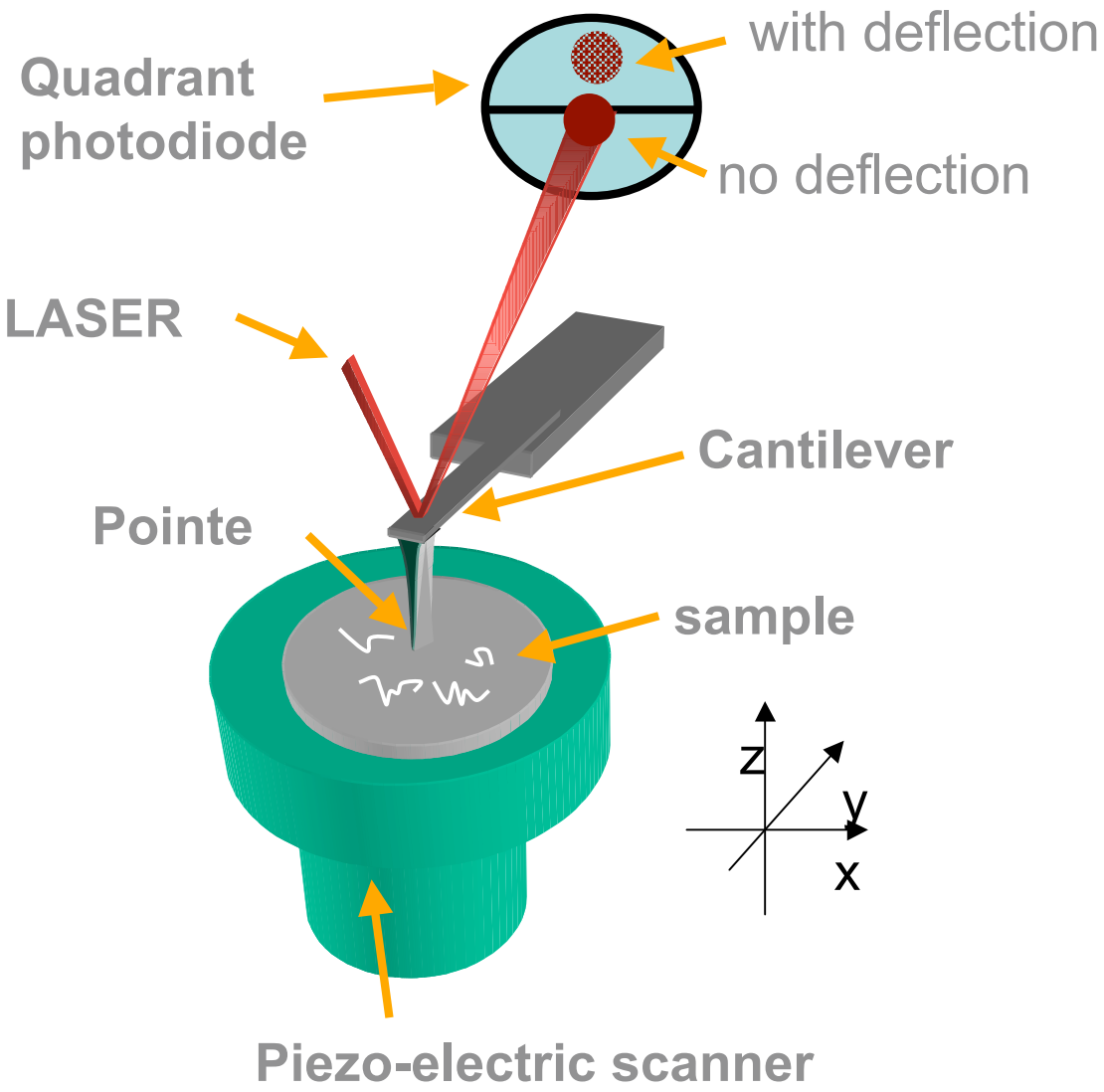
Experimental evidence that long-range correlations influence nucleosomal organisation

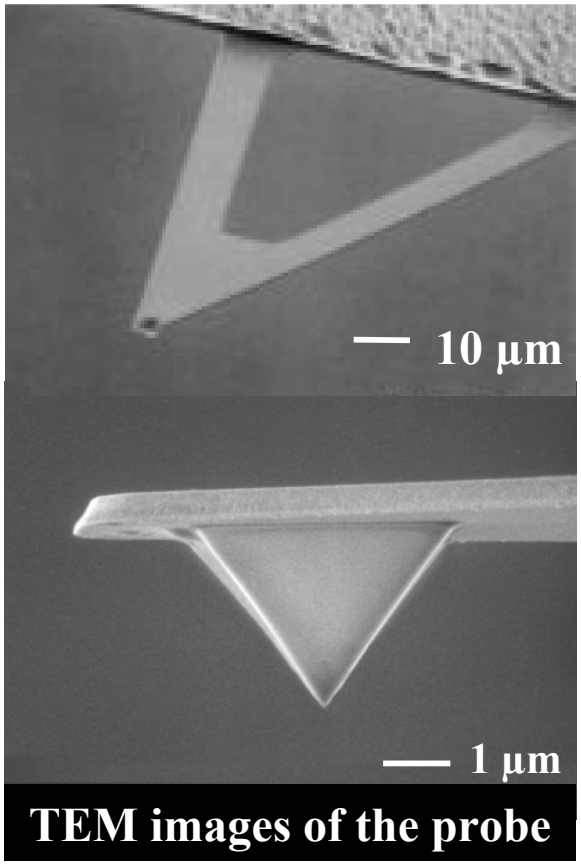
Nucleosome occupancy for the 2000 shortest *S. Cerevisiae* genes



Probing long-range correlations in eukaryotic DNA with Atomic Force Microscopy

Atomic Force Microscopy





TEM images of the probe

AFM image of 2200 pb virus DNA fragments

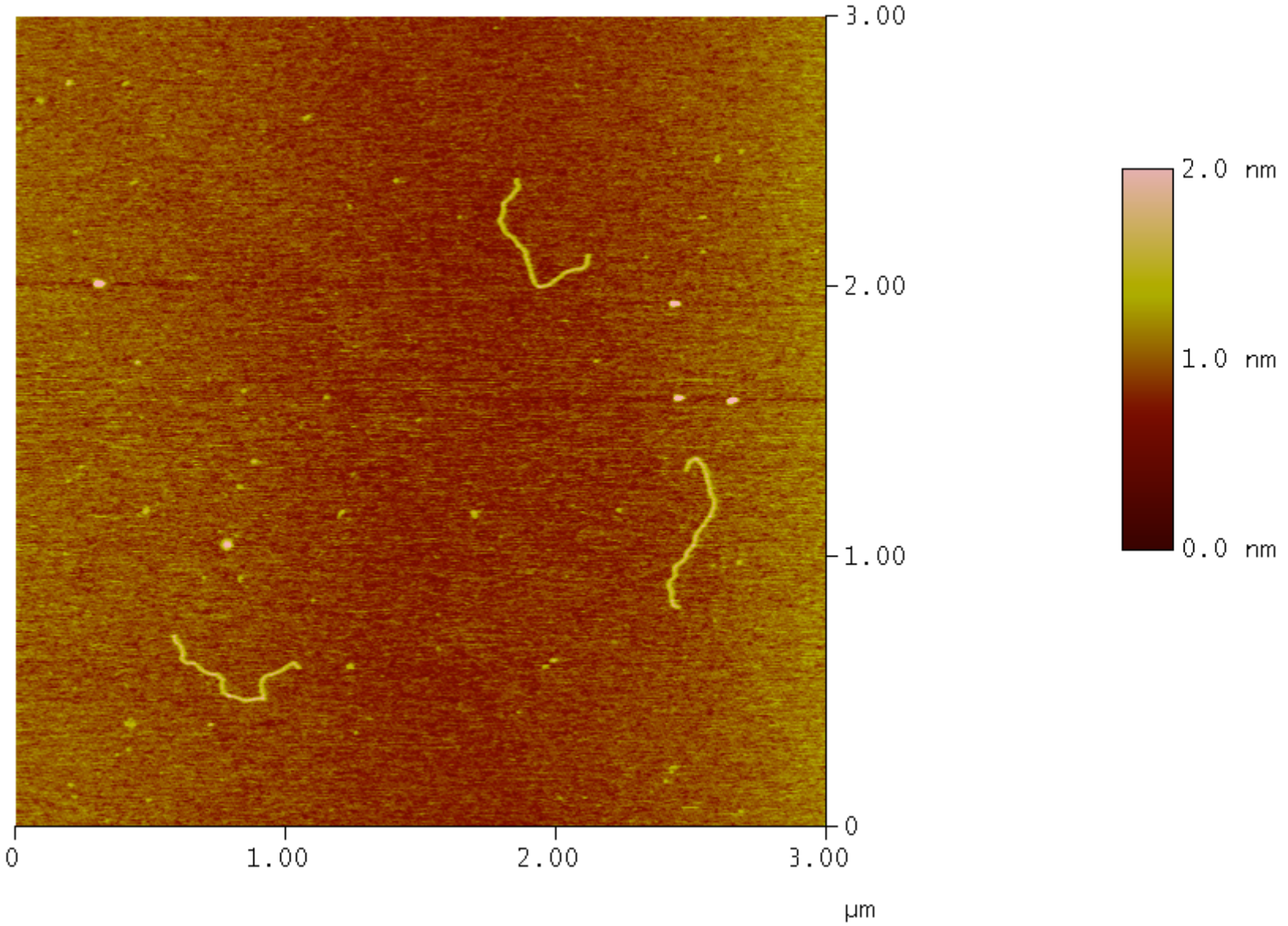
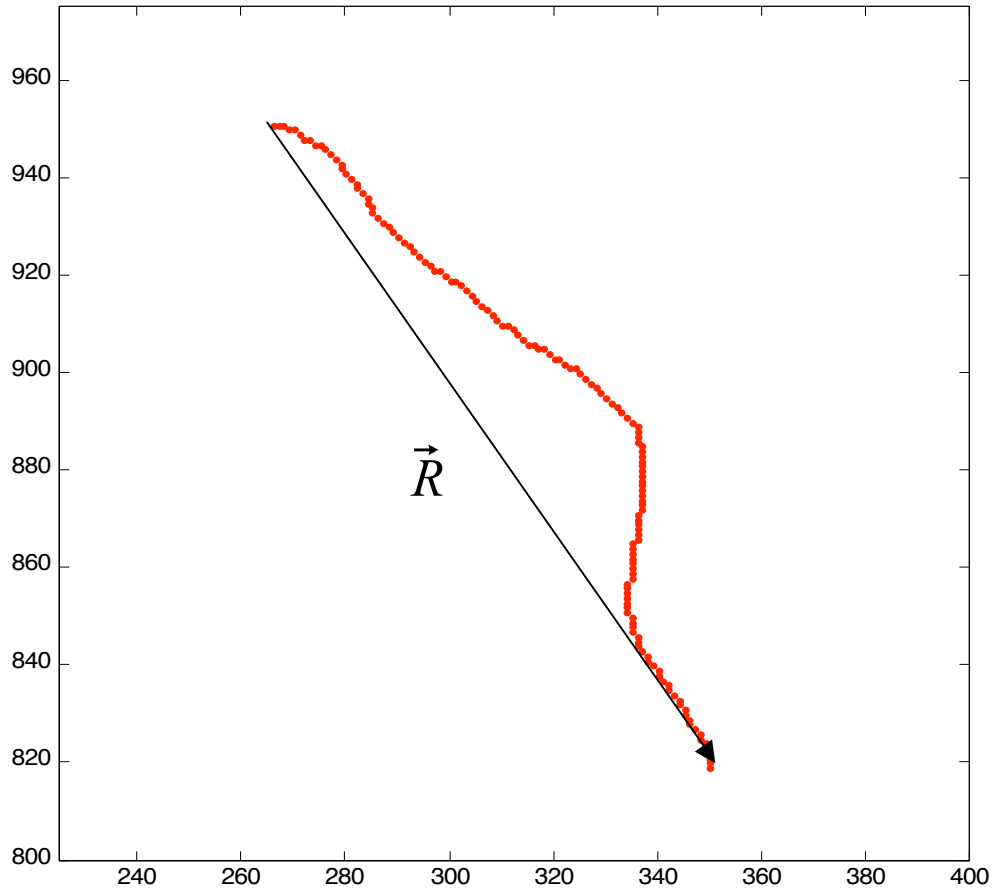


Image analysis

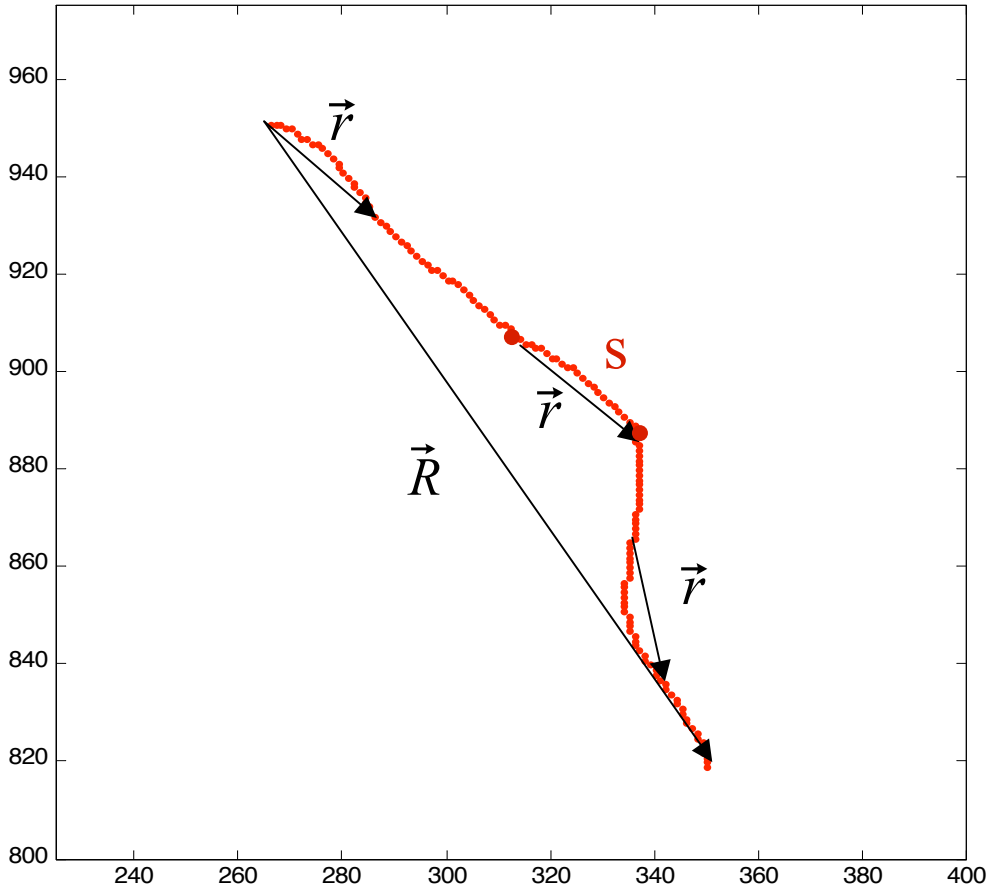
local persistence length calculation



$$l_p = \lim_{L\to\infty} \frac{\langle \vec{R}^2 \rangle}{2L}$$

Image analysis

local persistence length calculation



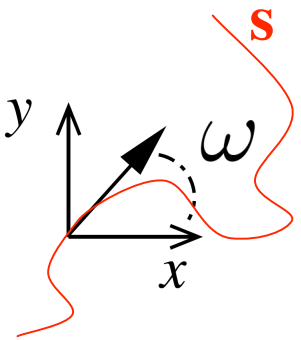
$$l_p = \lim_{L\to\infty} \frac{\langle \vec{R}^2 \rangle}{2L}$$

$$l_p(s) = \frac{\langle \vec{r}^{\,2} \rangle}{2s}$$

Introducing sequence effects in polymer elasticity modeling

• **semi-flexible polymer, WLC model**

$$\frac{E_{el}}{k_B T} = \frac{A}{2}\int_s^{s+l}\left(\frac{\partial \omega(u)}{\partial u}\right)^2 du$$



persistence length : $L_p = \lim_{L\to\infty}\frac{\langle \vec{R}^2 \rangle}{2L} = A$ *Elastic modulus*

• **taking into account sequence induced structural disorder in the WLC model**

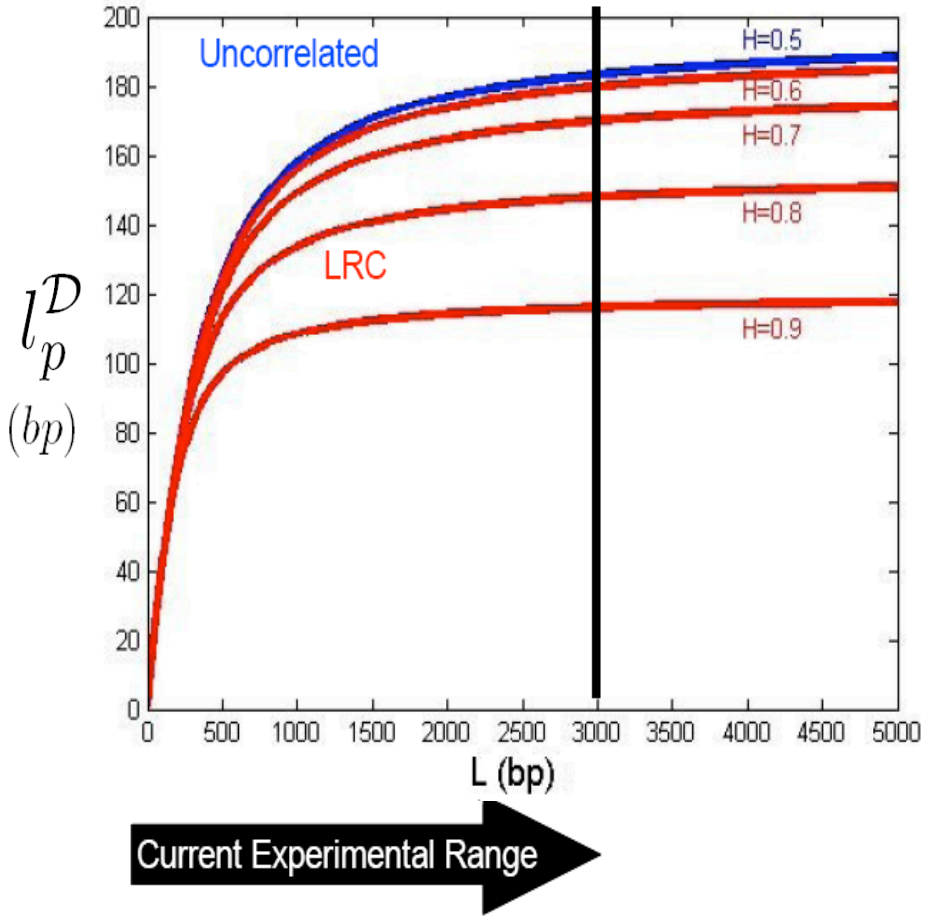
$$\frac{E_{el}}{k_B T} = \frac{A}{2}\int_s^{s+l}\left(\frac{\partial}{\partial u}(\omega(u) - \omega_0(u))\right)^2 du$$

$\rightarrow$ $l_p = A_{eff}$

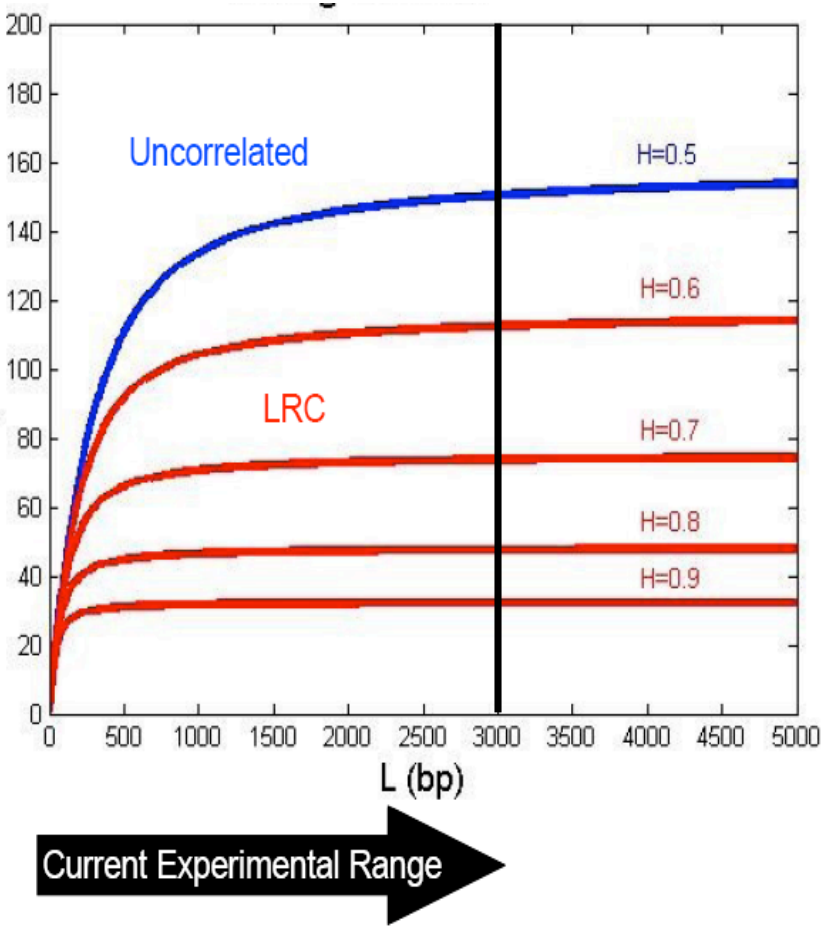
Influence of LRC on the measurement of persistence length

$$A_d = 198\text{bp}$$

Weak disorder $\sigma_l = 0.01$

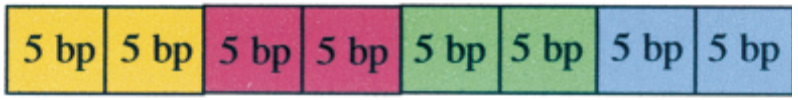


Strong disorder $\sigma_l = 0.05$

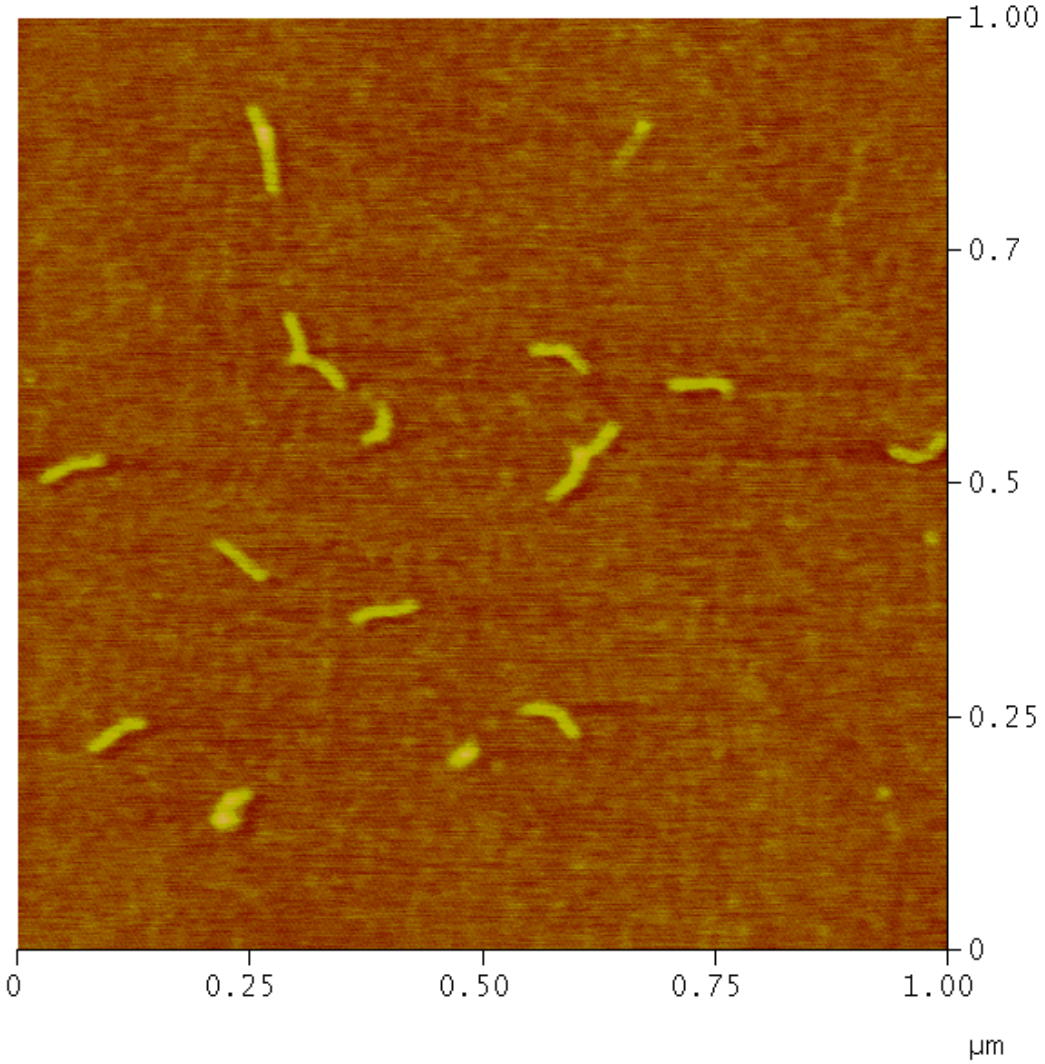


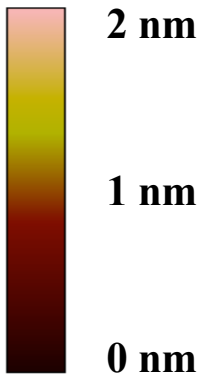
Intrinsically straight DNA fragment (200 bp)

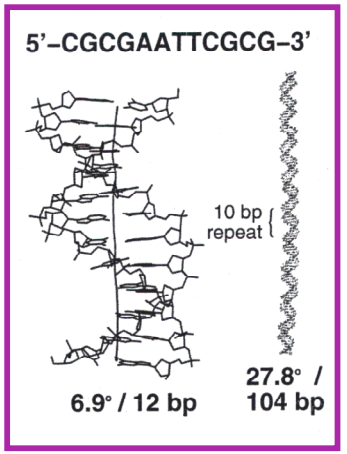
Construction of a 'straight' artificial DNA :



M. Vologodskaia and A. Vologodskii, J. Mol. Biol. (2002)





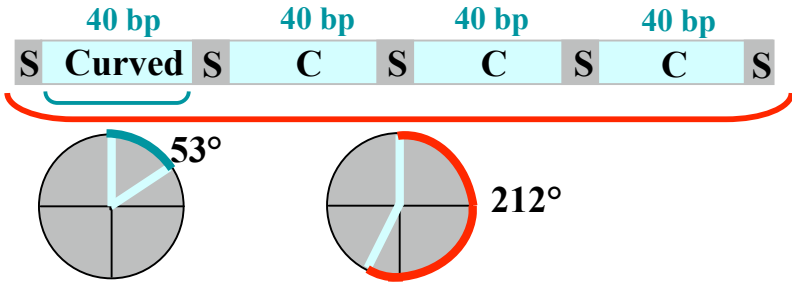


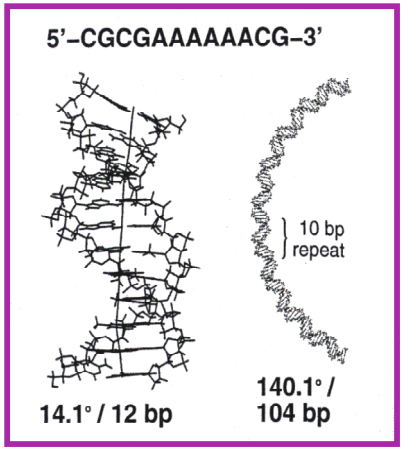
No intrinsically disorder > 10 pb

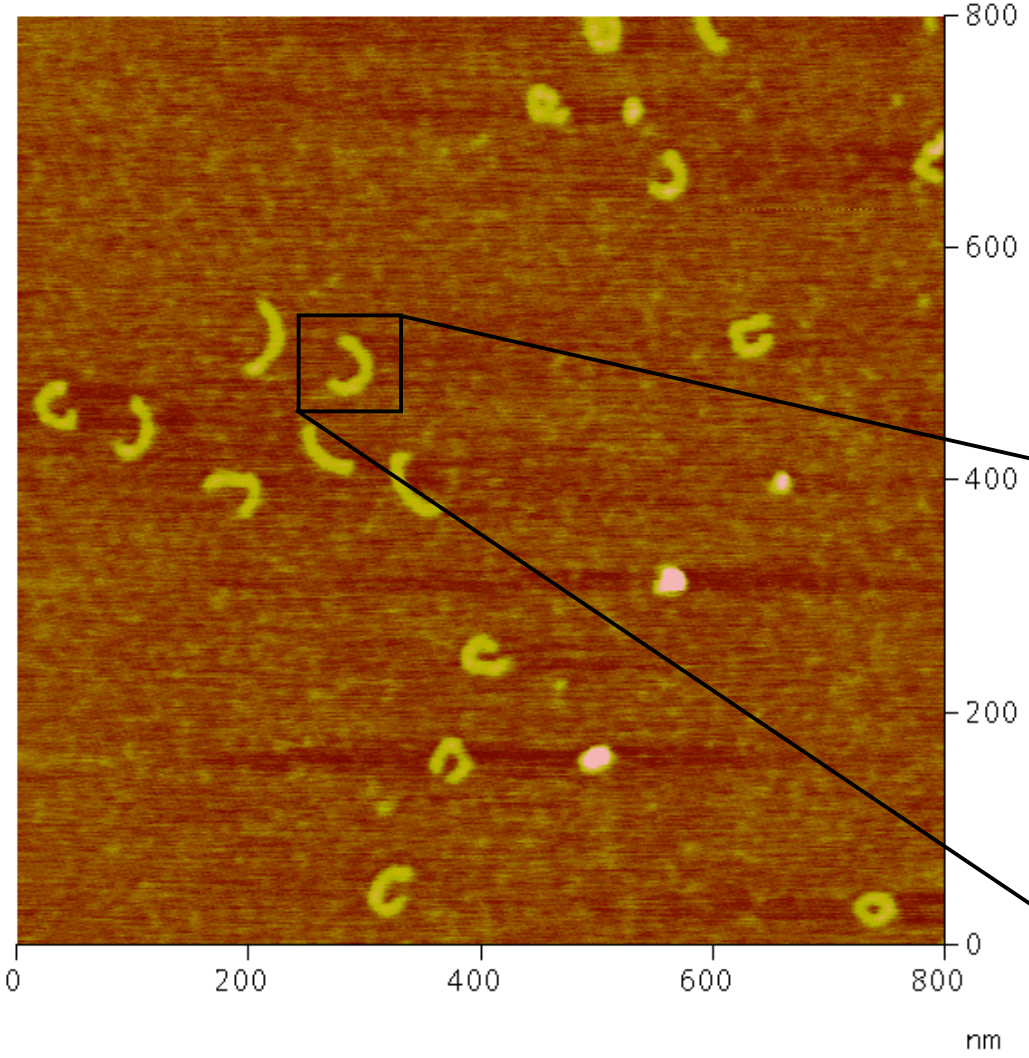
$$A_{eff} = A_{dyn}$$

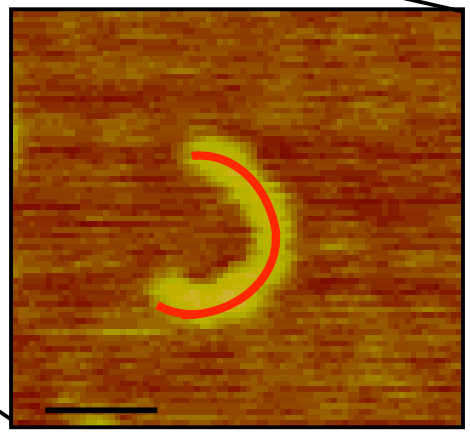
Highly curved DNA fragment (200 bp)

Construction of an artificial DNA with very high curvature:



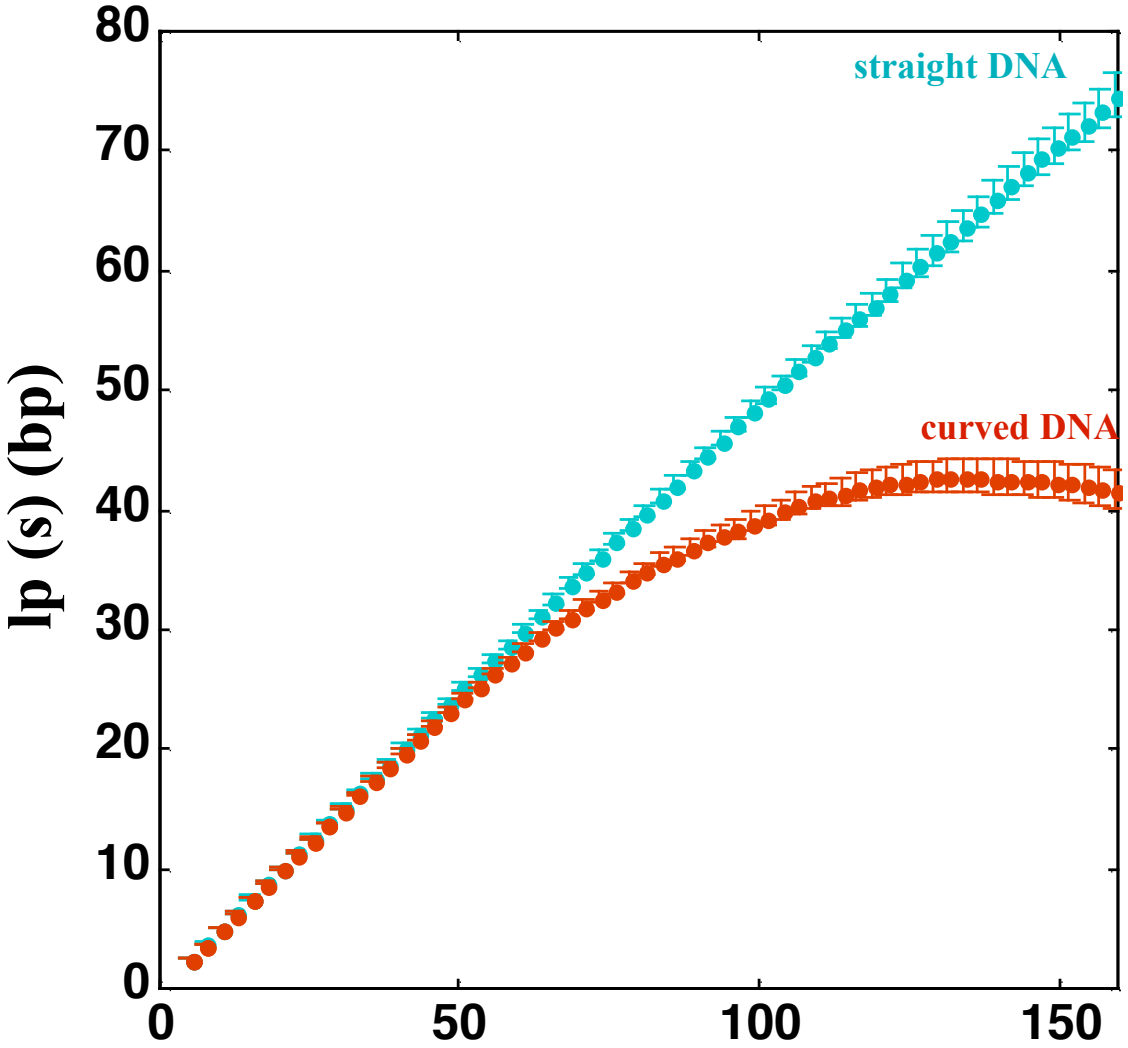






50 nm

Image analysis of 200bp straight and highly curved DNA fragments



we are able to distinguish the two fragments by measuring the persistence length

BUT

200 bp is too short to accurately estimate A

need for longer fragments

DNA selection and preparation

• **DNA with no structural disorder ($\sigma \sim 0$) :**

1 intrinsically straight DNA in the absence of thermal fluctuations (length = 800 bp).

• **DNA with uncorrelated structural disorder (H = 0.5) :**

1 DNA fragment extracted from Hepatitis C RNA virus genome (length = 2200 bp).

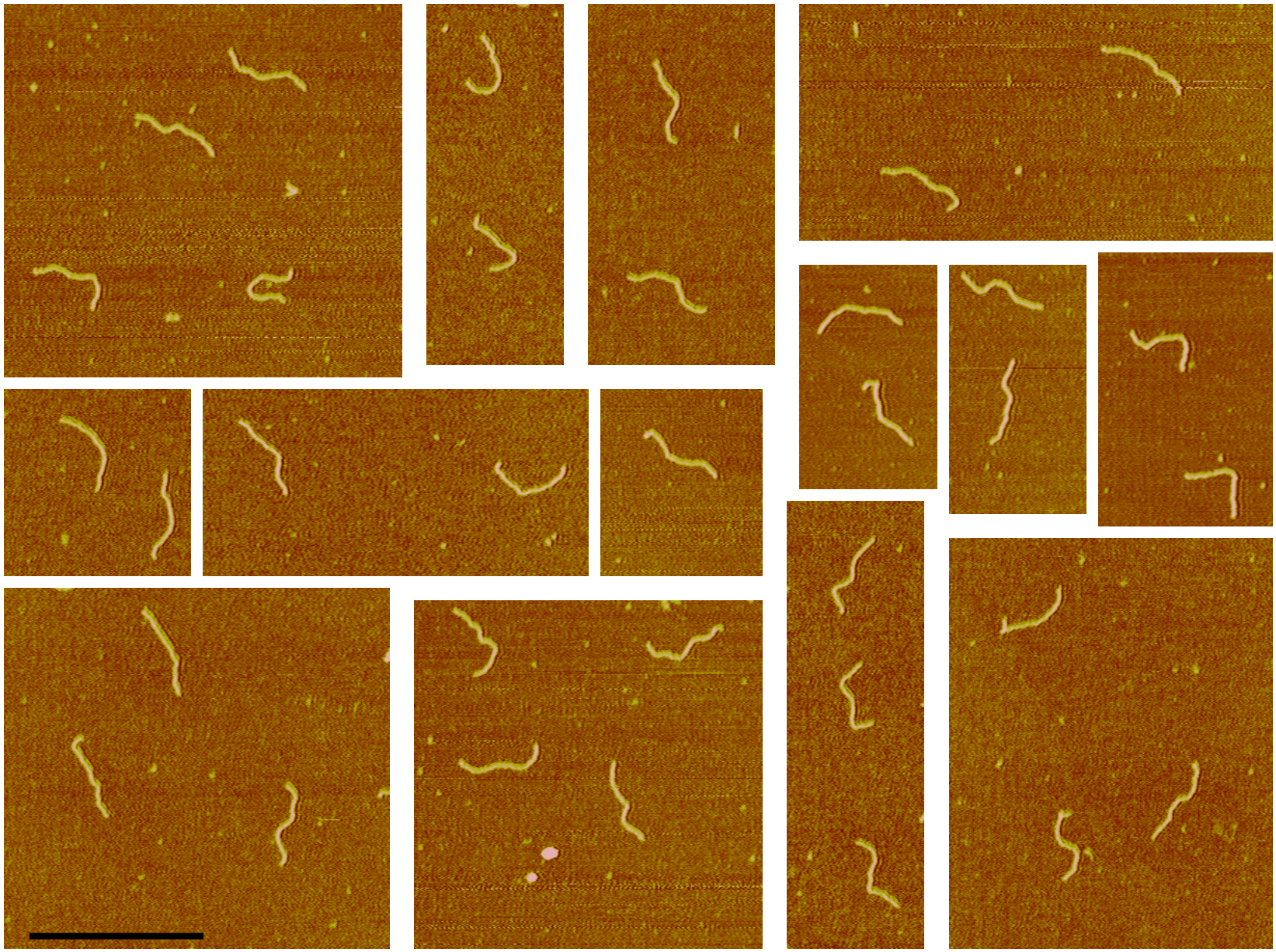
• **DNA having long range correlated structural disorder (0.5 < H < 1) :**

2 DNA fragments extracted from the human genome :

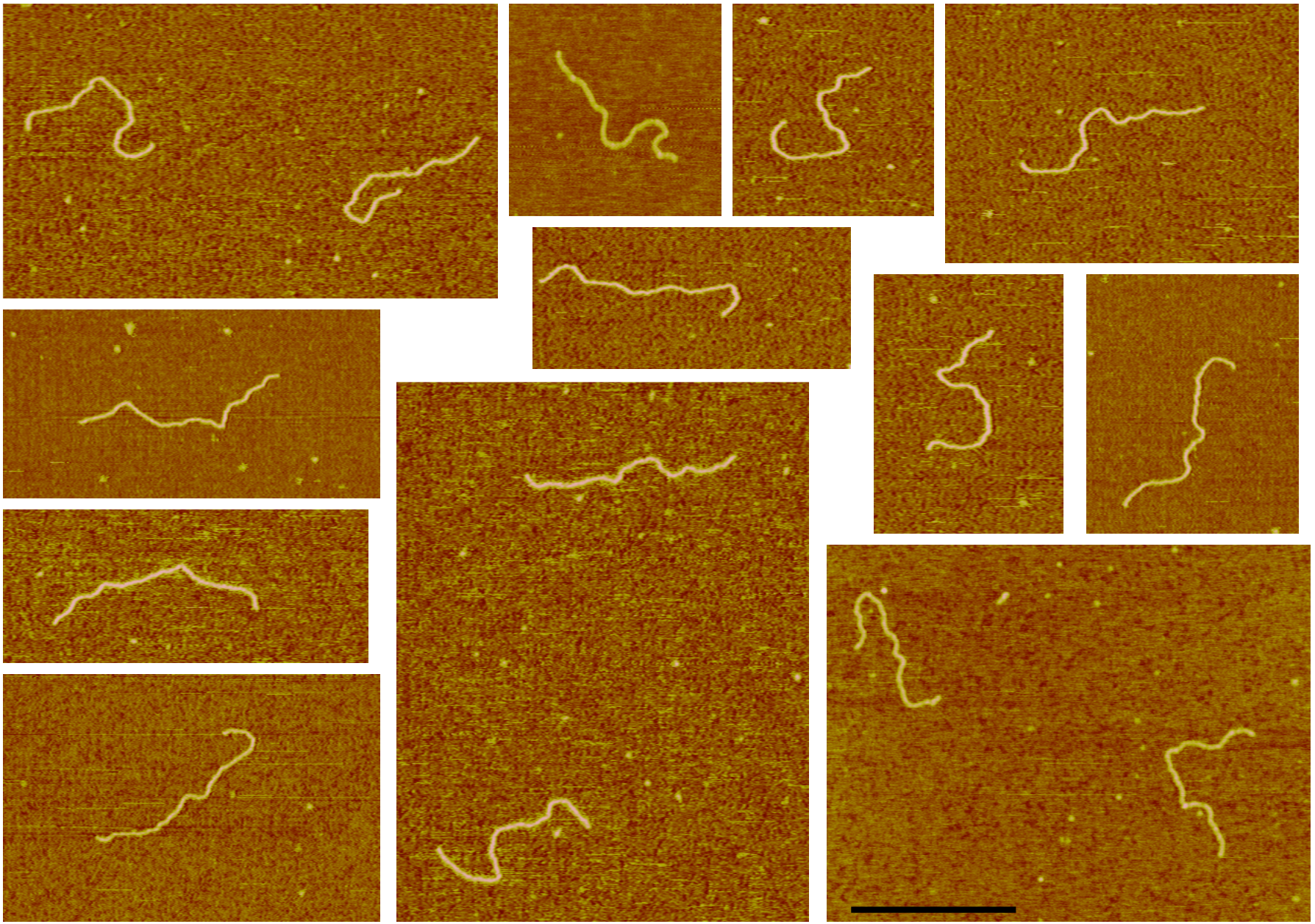
chromosome 21 and chromosome 8

(length = 2200 bp).

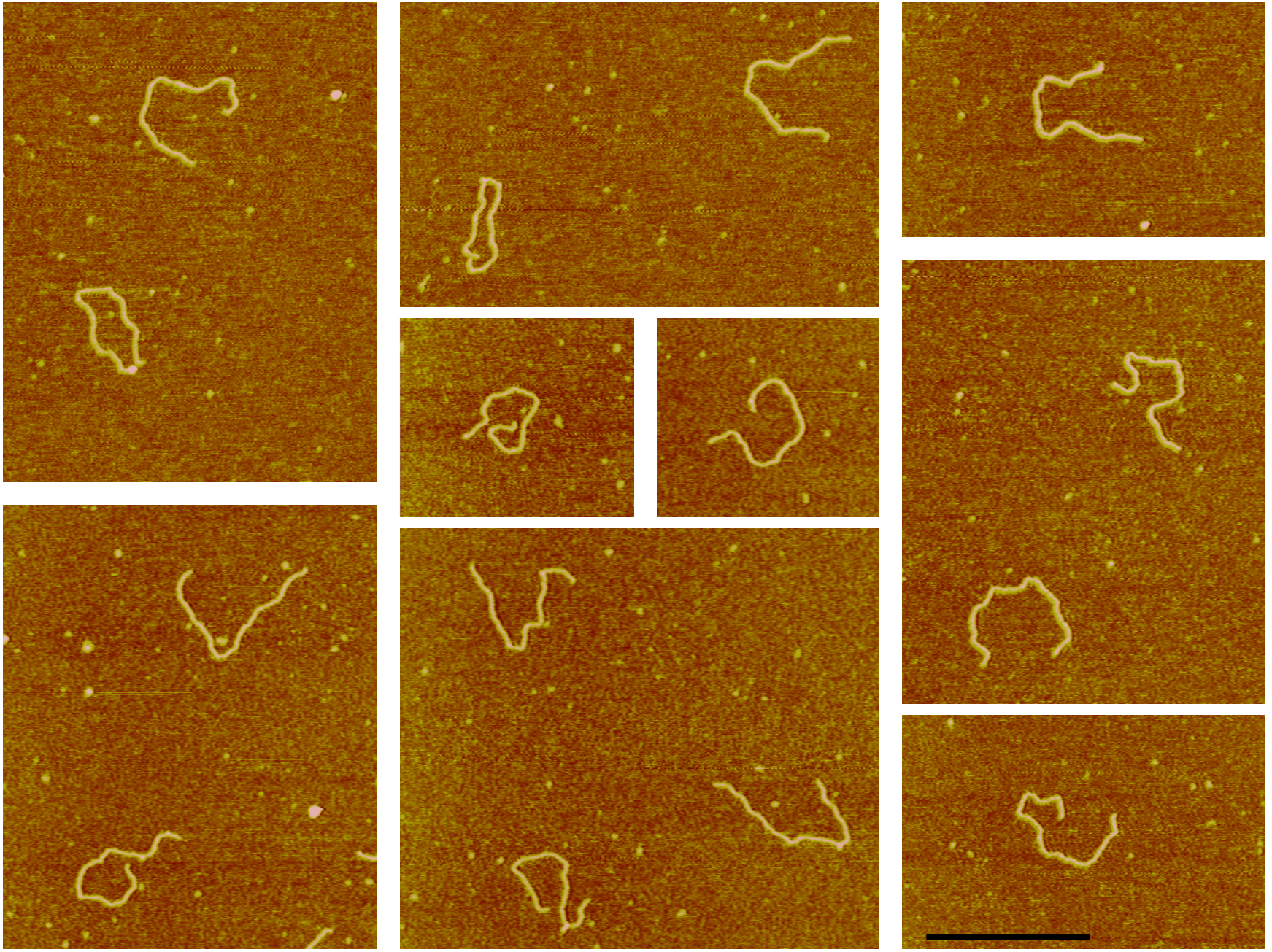
AFM images of 800 bp straight DNA (no structural disorder, $\sigma \sim 0$) in 2D



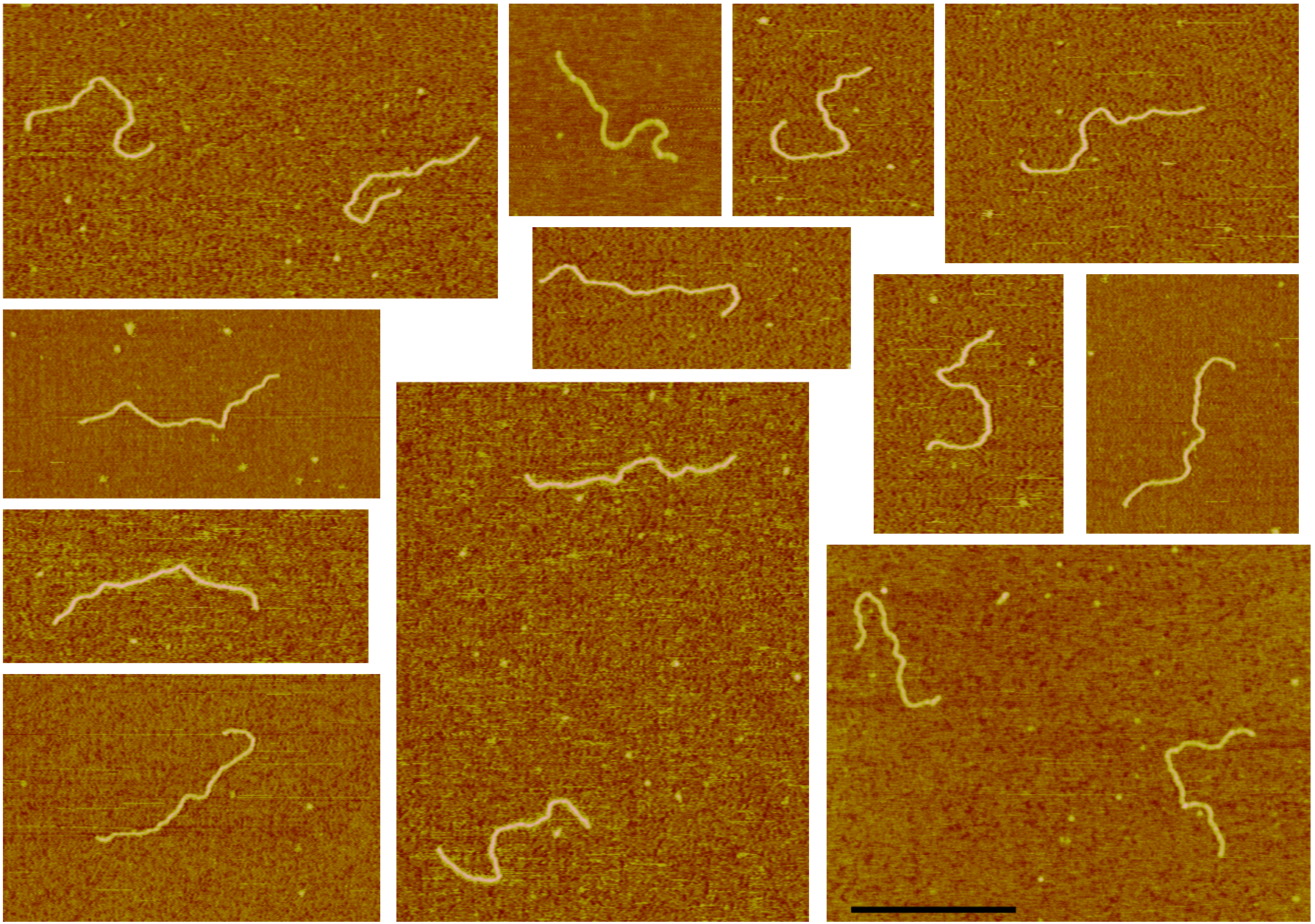
AFM images of 2200 bp virus DNA (uncorrelated structural disorder H =0.5) in 2D



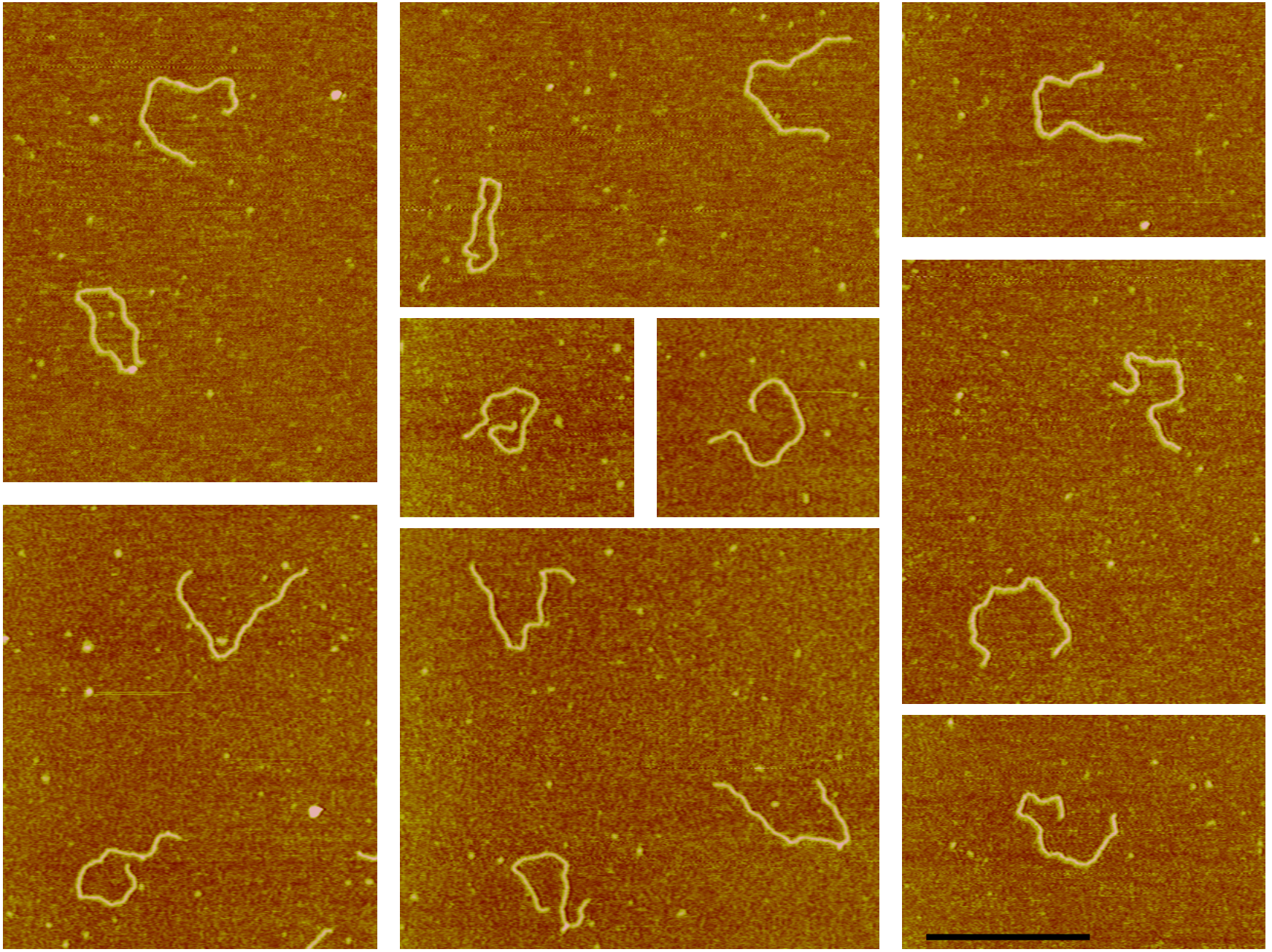
AFM images of 2200 bp human DNA from chromosome 21 (LRC, H~0.8) in 2D



AFM images of 2200 bp virus DNA (uncorrelated structural disorder H =0.5) in 2D

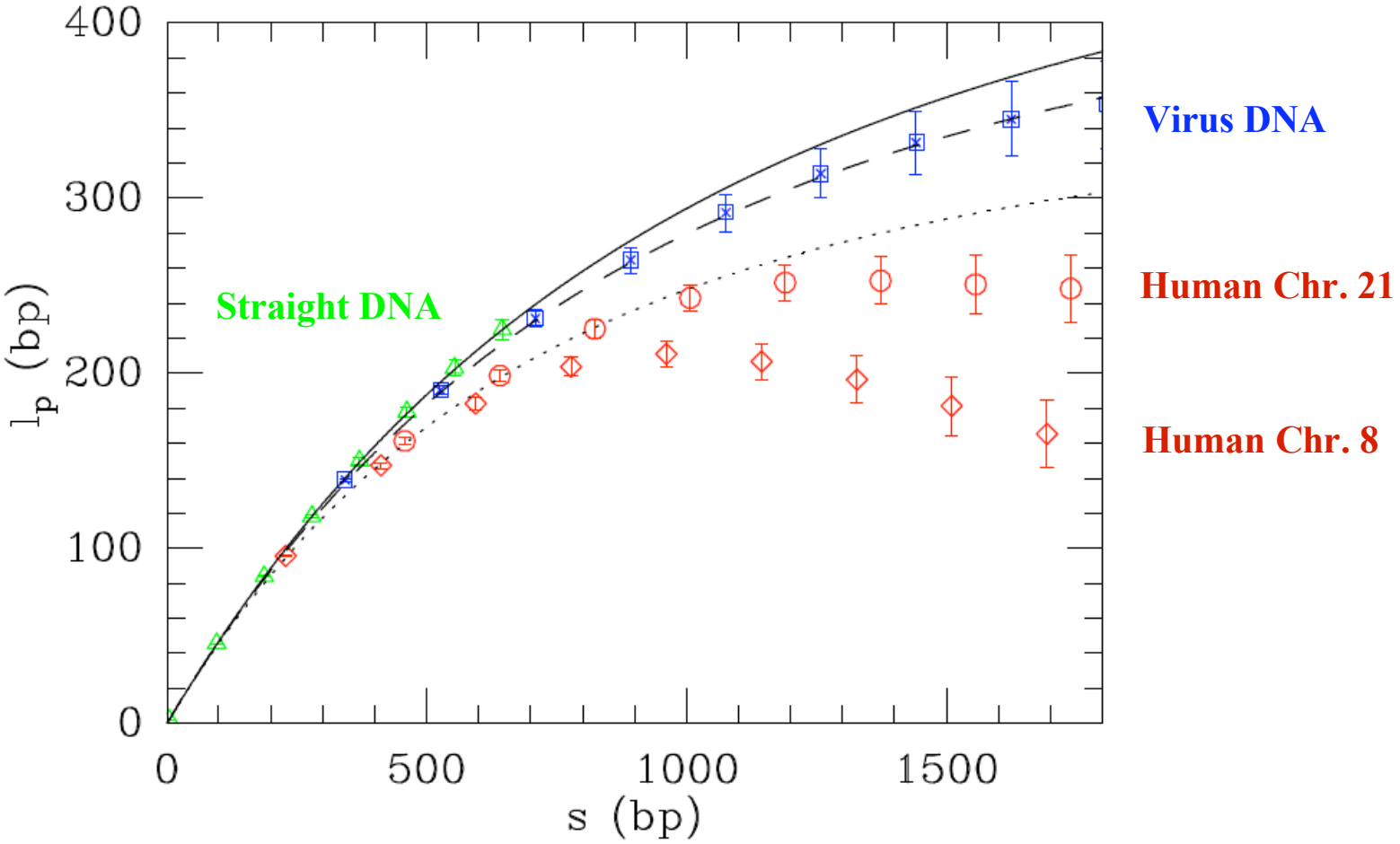


AFM images of 2200 bp human DNA from chromosome 21 (LRC, H~0.8) in 2D

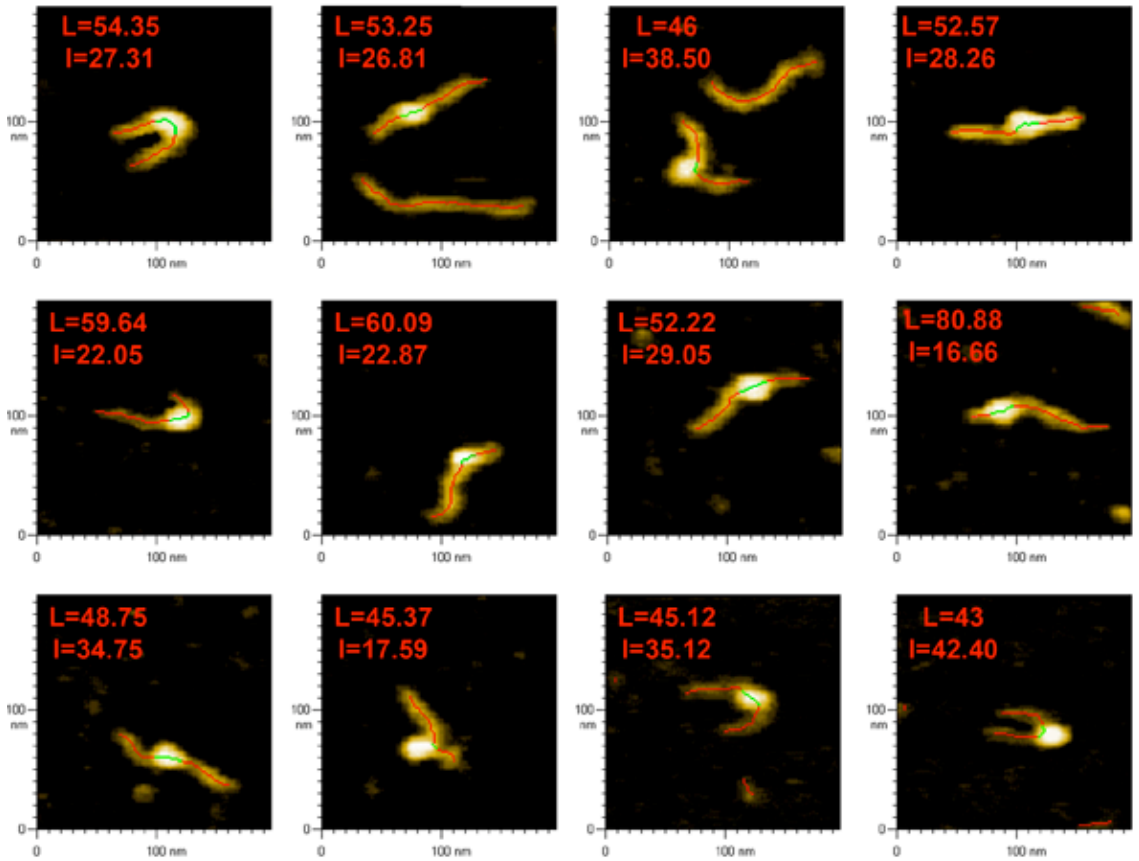


Influence of LRC on persistence length measurement

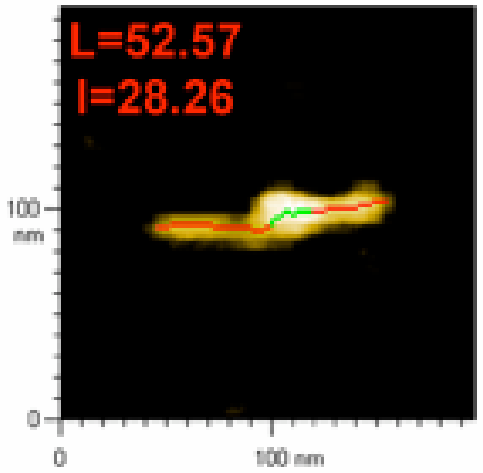
Persistence length calculation for AFM trajectories



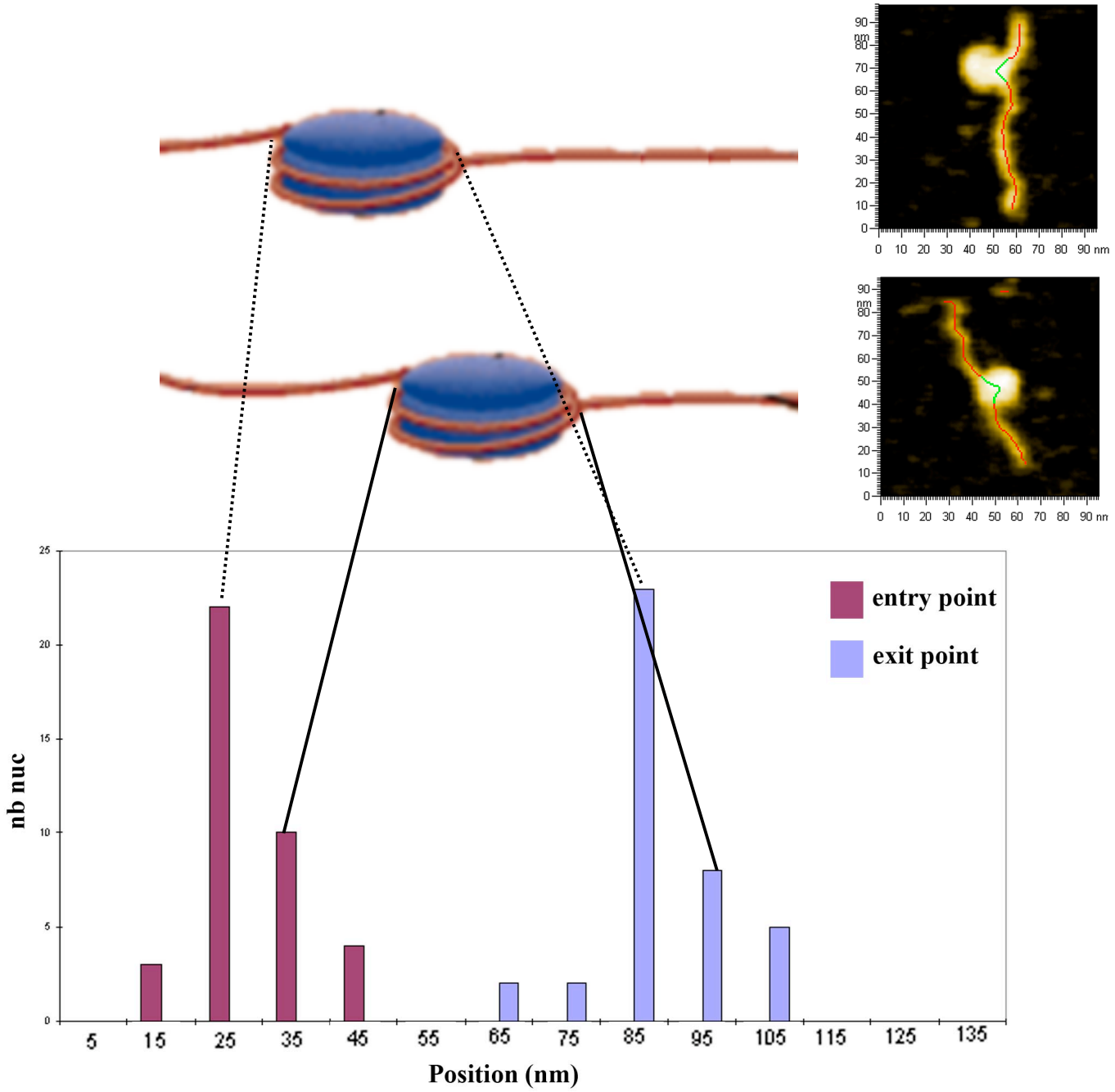
Experimental evidence of sequence dependent nucleosome positioning



AFM imaging in liquid



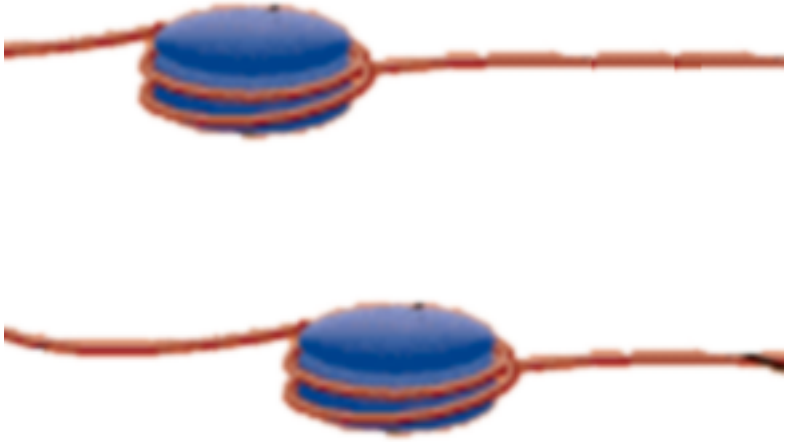
Experimental evidence of sequence dependent nucleosome positioning

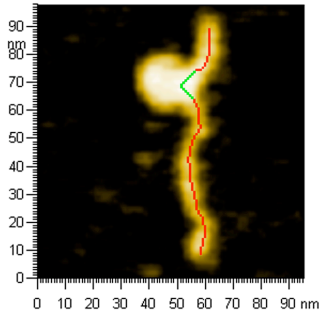


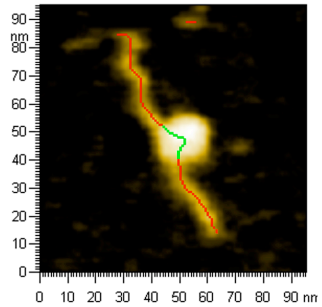
Experimental evidence of sequence dependent nucleosome positioning

1

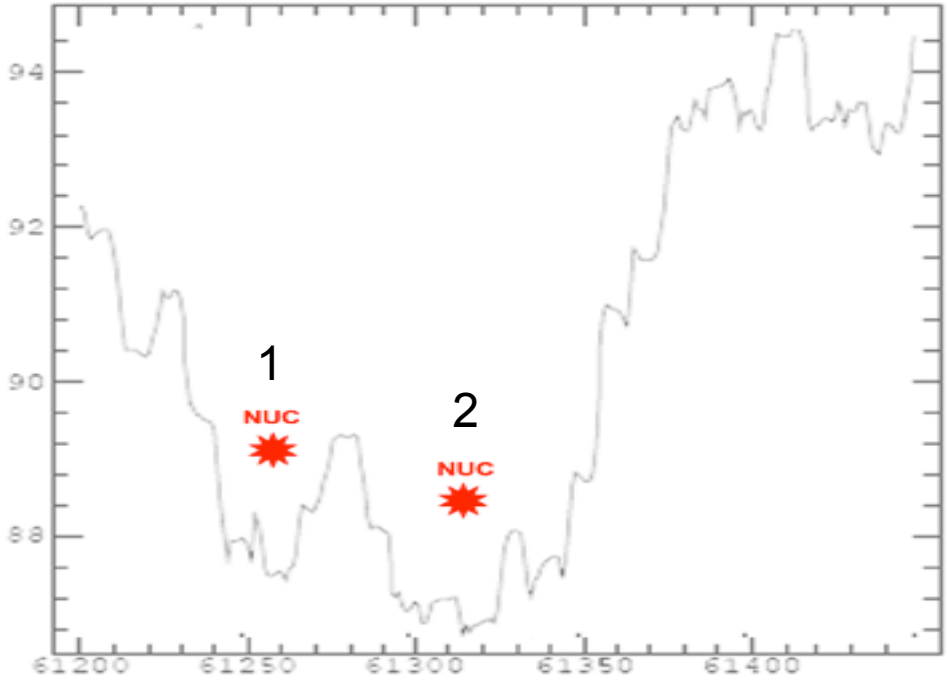
2



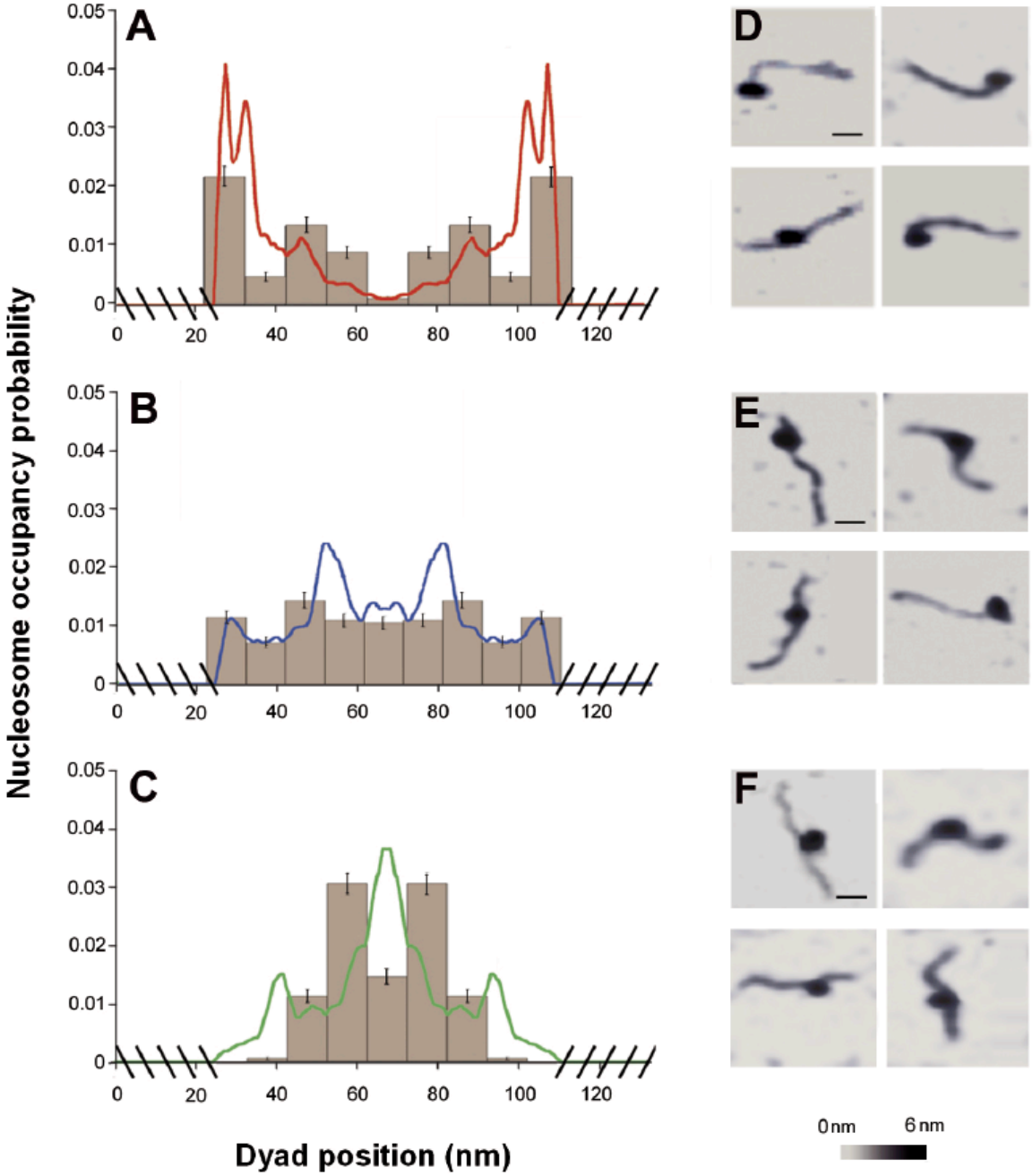




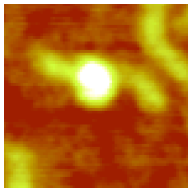
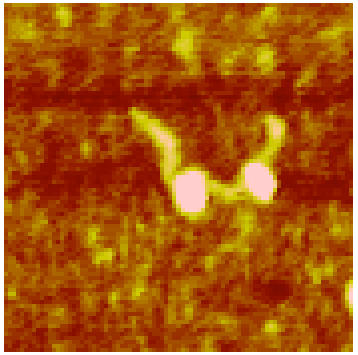
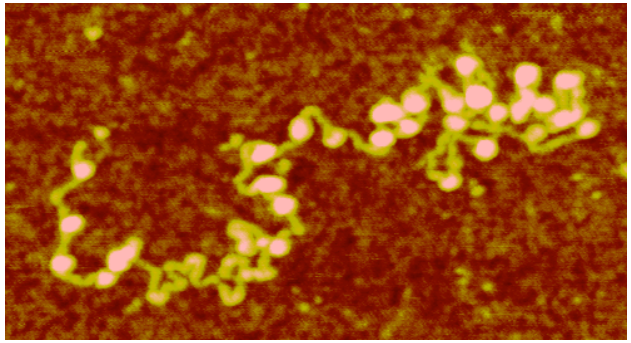
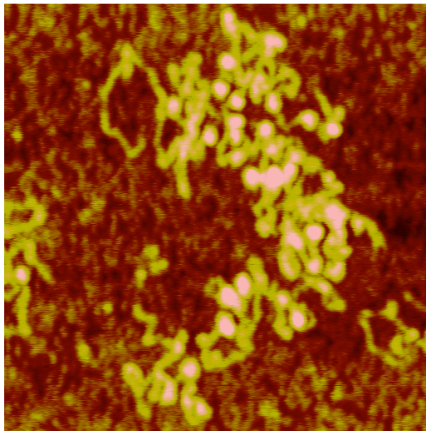
Predicted nucleosome energy landscape



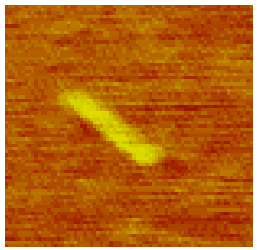
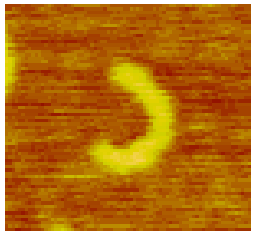
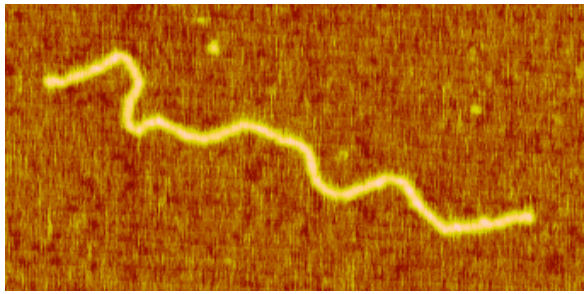
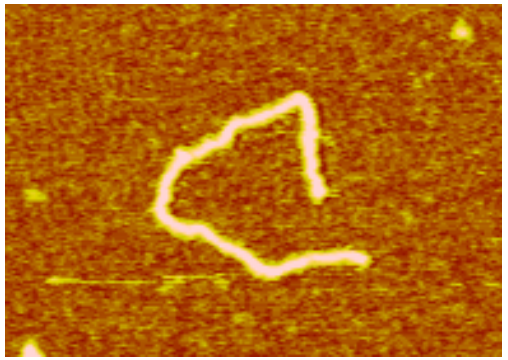
Experimental evidence of sequence dependent nucleosome positioning



Structure and dynamics of nucleosomes : histone variants, and chromatin remodeling

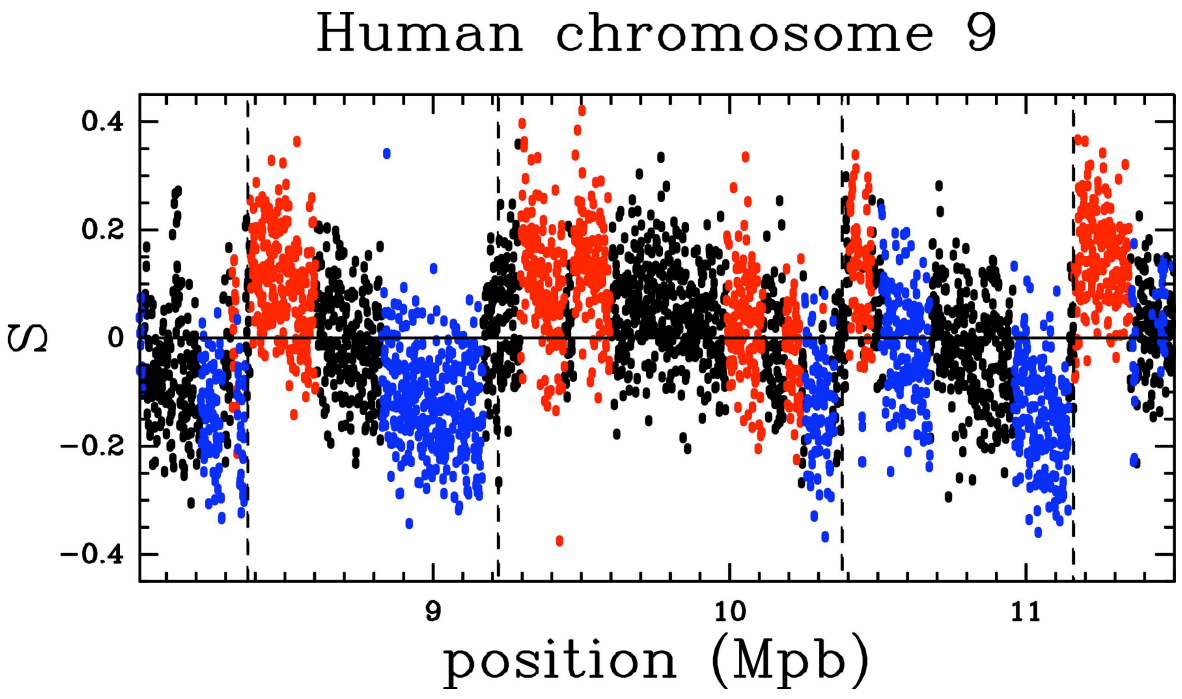
DNA sequence effects on the structural and mechanical properties of the double helix

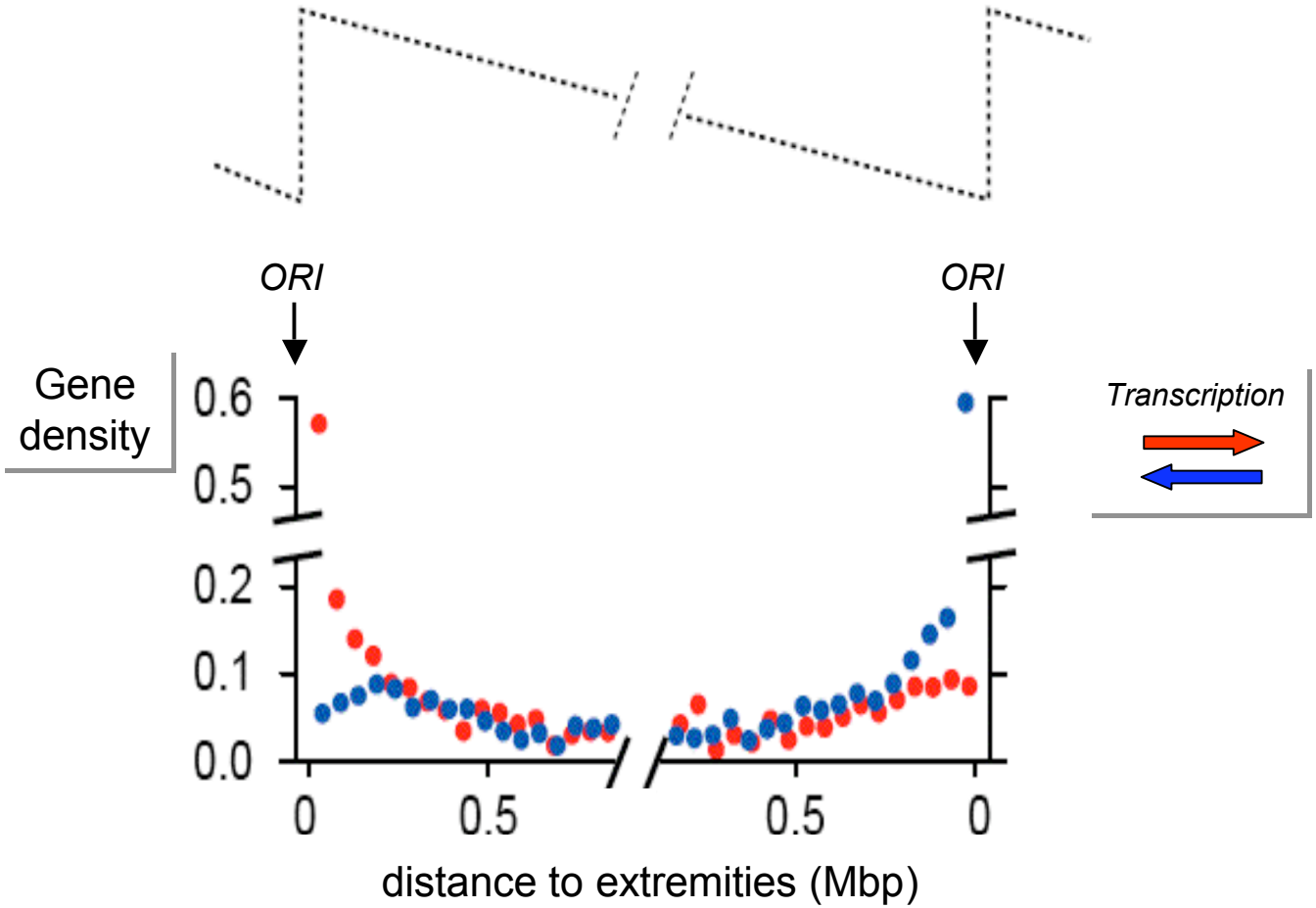
The sequence is highly predictive of open chromatin around putative replication origins

Genome organisation around putative replication origins

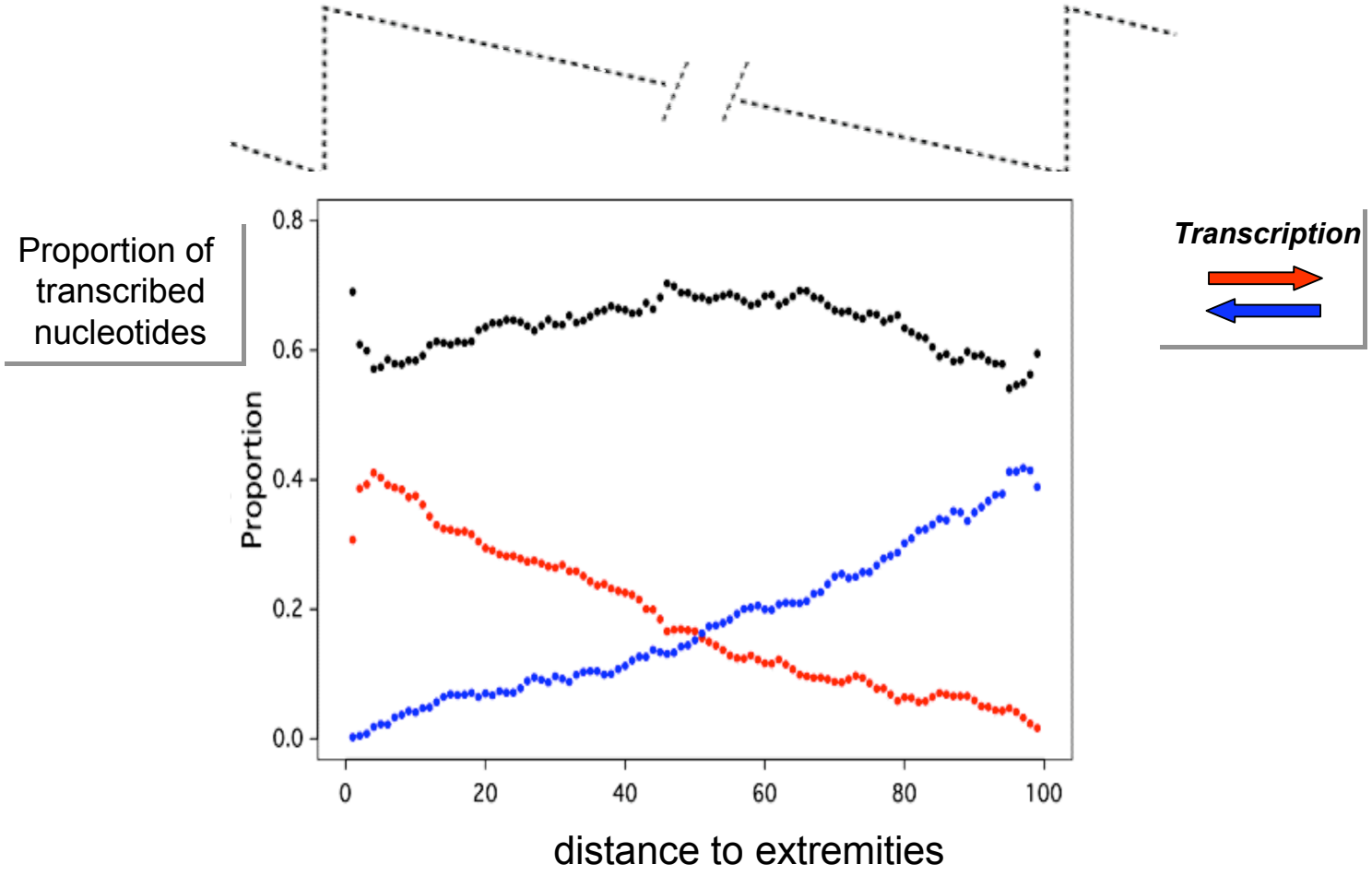
Strand compositional asymmetry: $S = \frac{n(T) - n(A)}{n(T) + n(A)} + \frac{n(G) - n(C)}{n(G) + n(C)}$



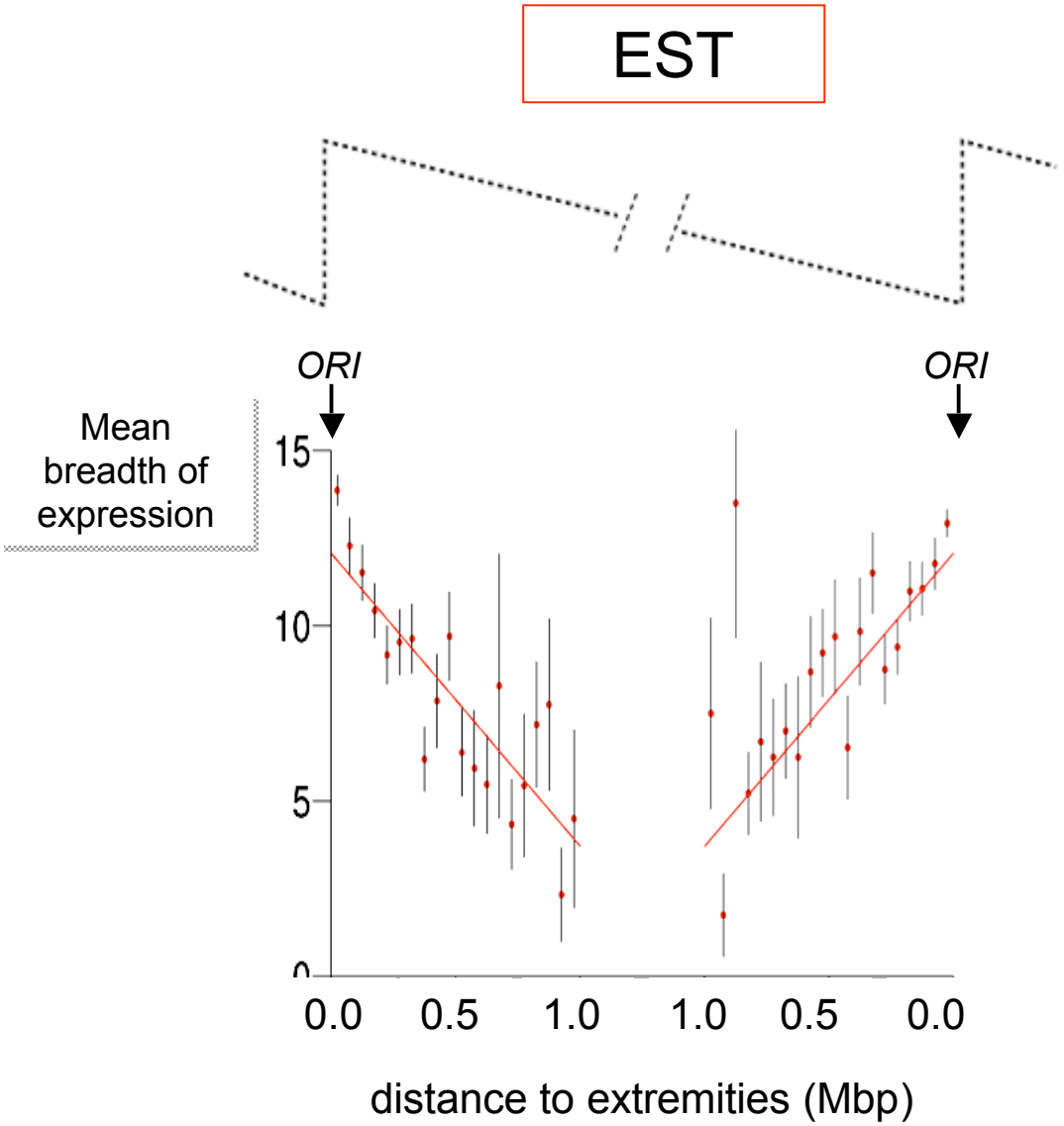
GENE DENSITY IN N-DOMAINS



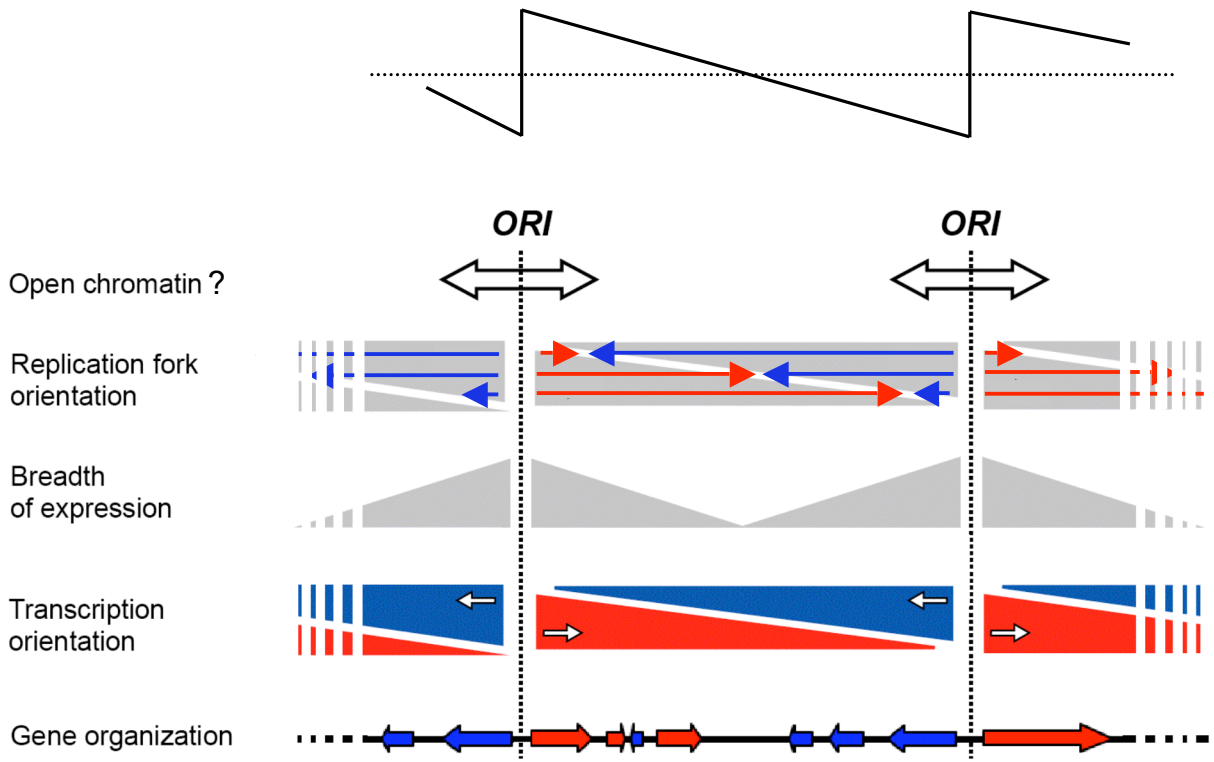
ORGANIZATION OF TRANSCRIPTION IN N-DOMAINS



ORGANIZATION OF TRANSCRIPTION IN N-DOMAINS



UNIFIED VISION OF GENE ORGANIZATION IN THE N-DOMAINS



MINIMISATION OF COLLISIONS BETWEEN DNA ET RNA POLYMERASES

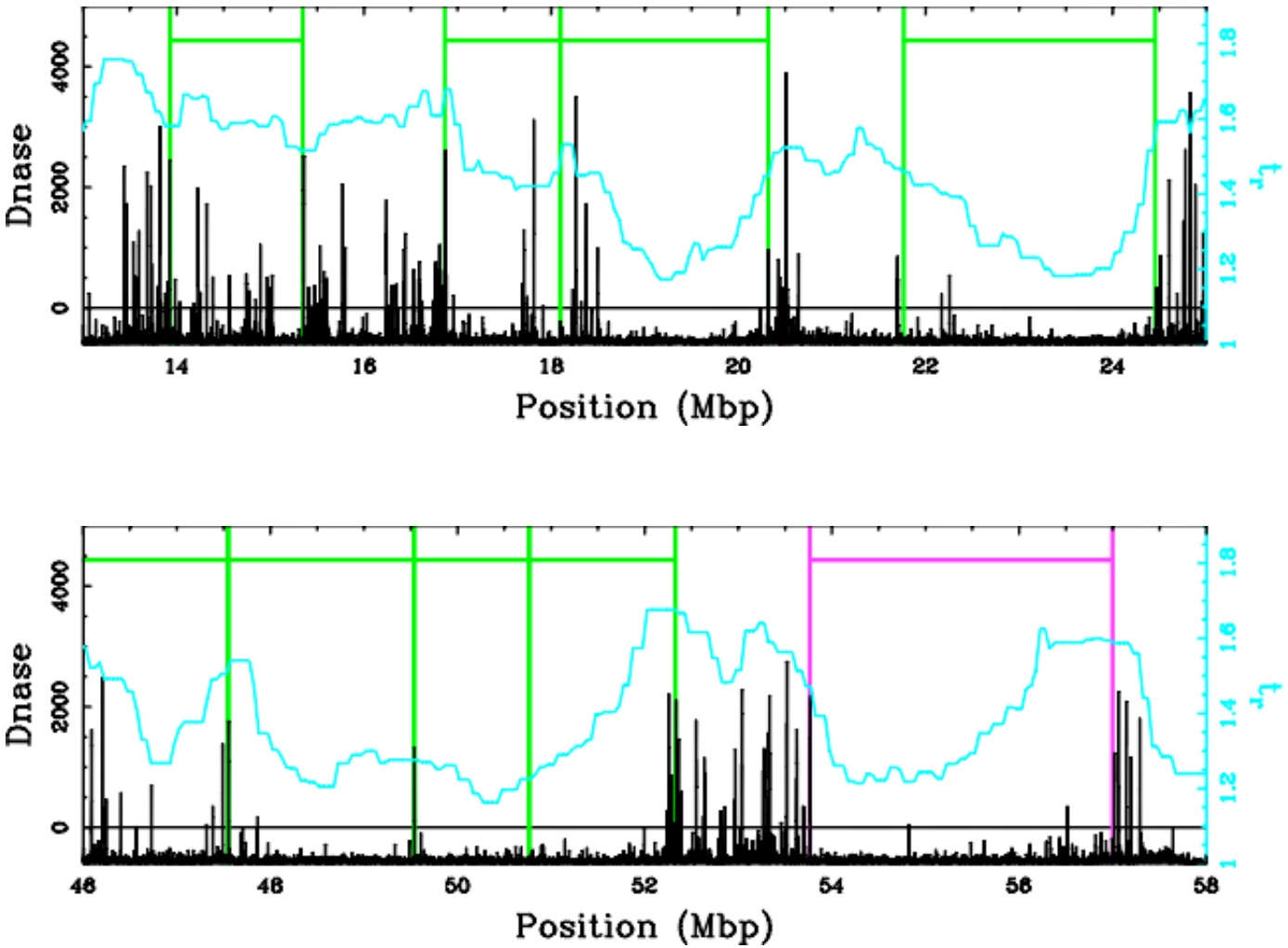
⇩

COORIENTATION OF REPLICATION AND TRANSCRIPTION

High-resolution mapping of Open Chromatin

Boyle et al., Cell 132, 311 (2008)

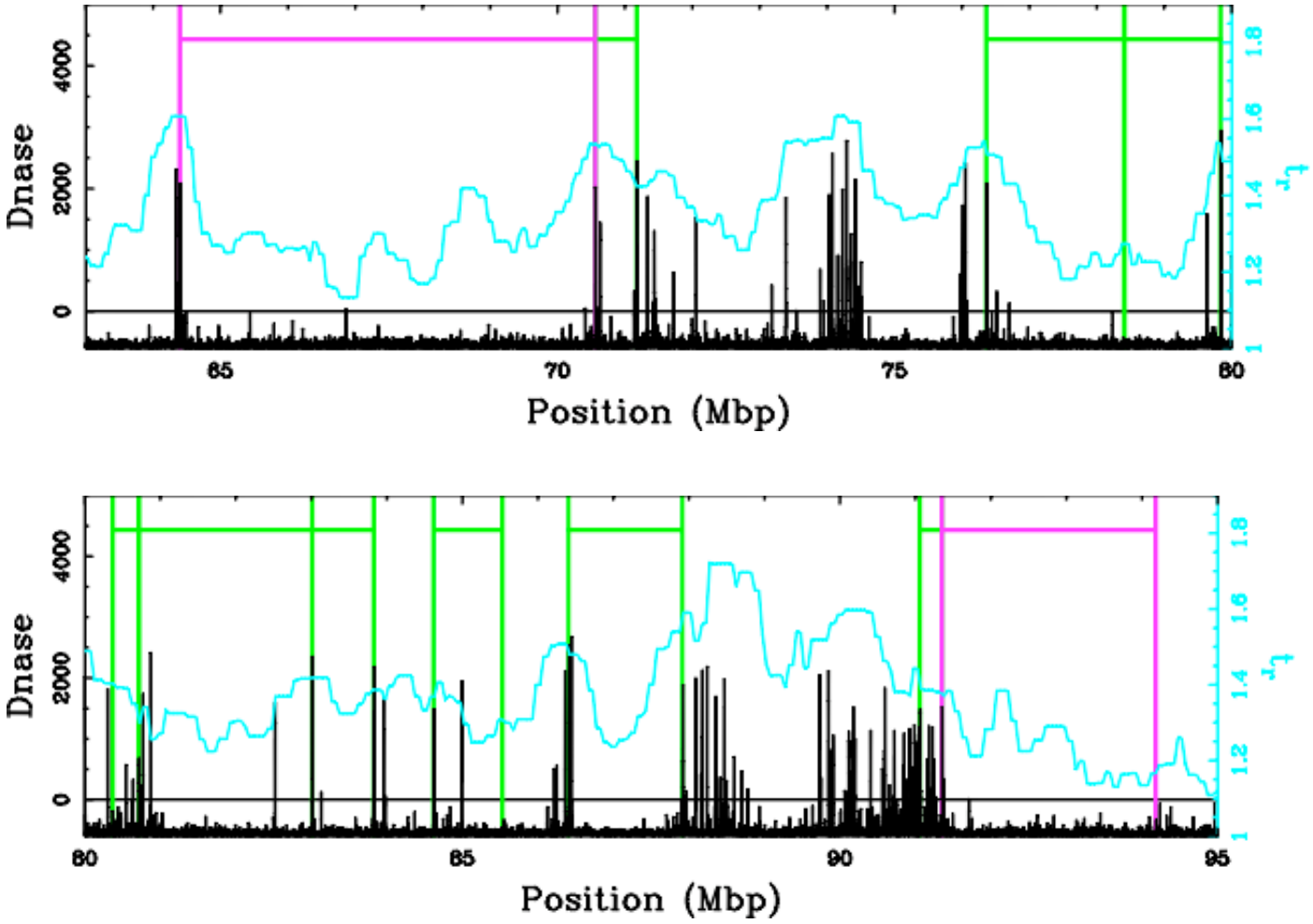
Human chromosome 6



High-resolution mapping of Open Chromatin

Boyle et al., Cell 132, 311 (2008)

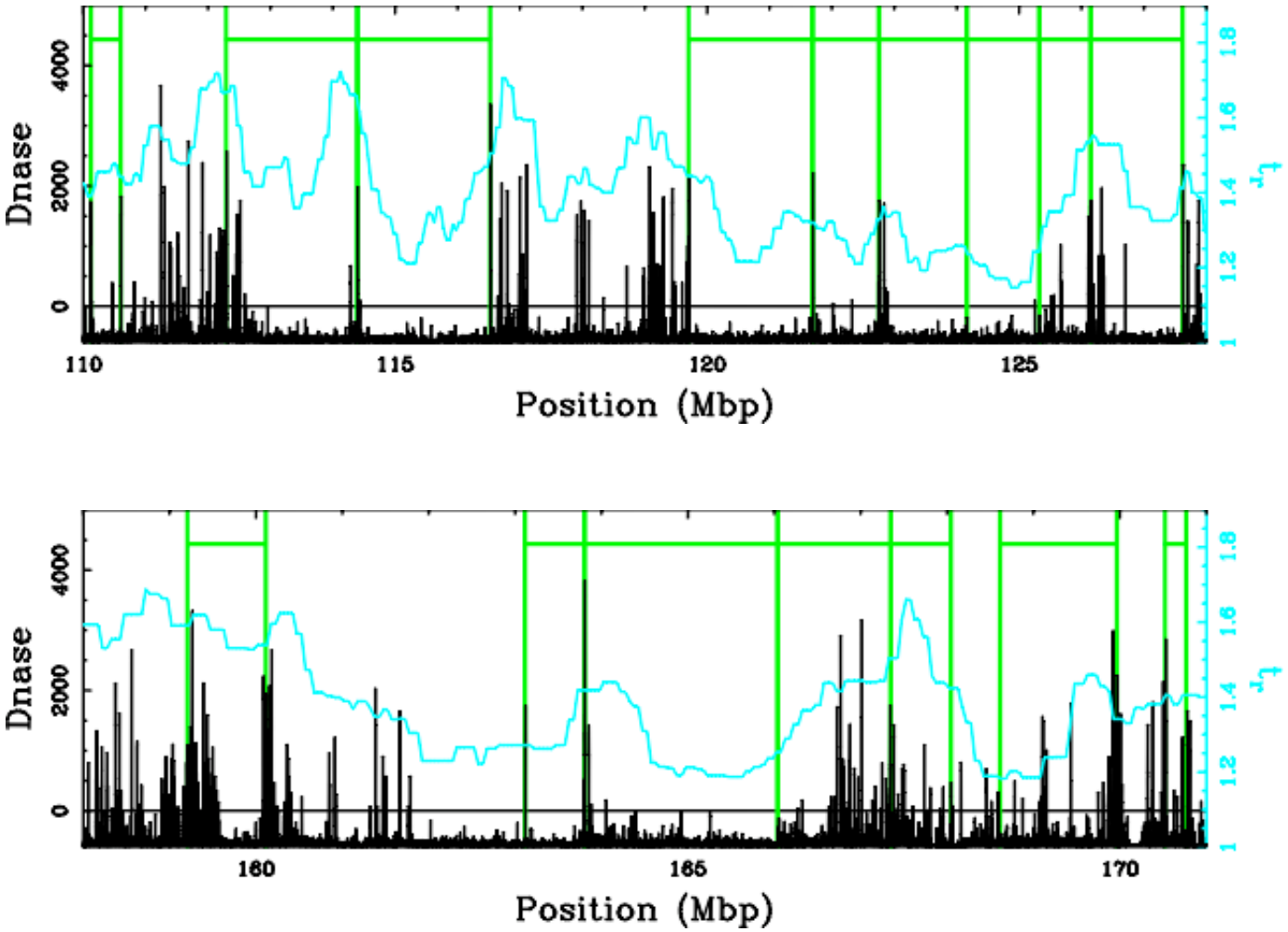
Human chromosome 6



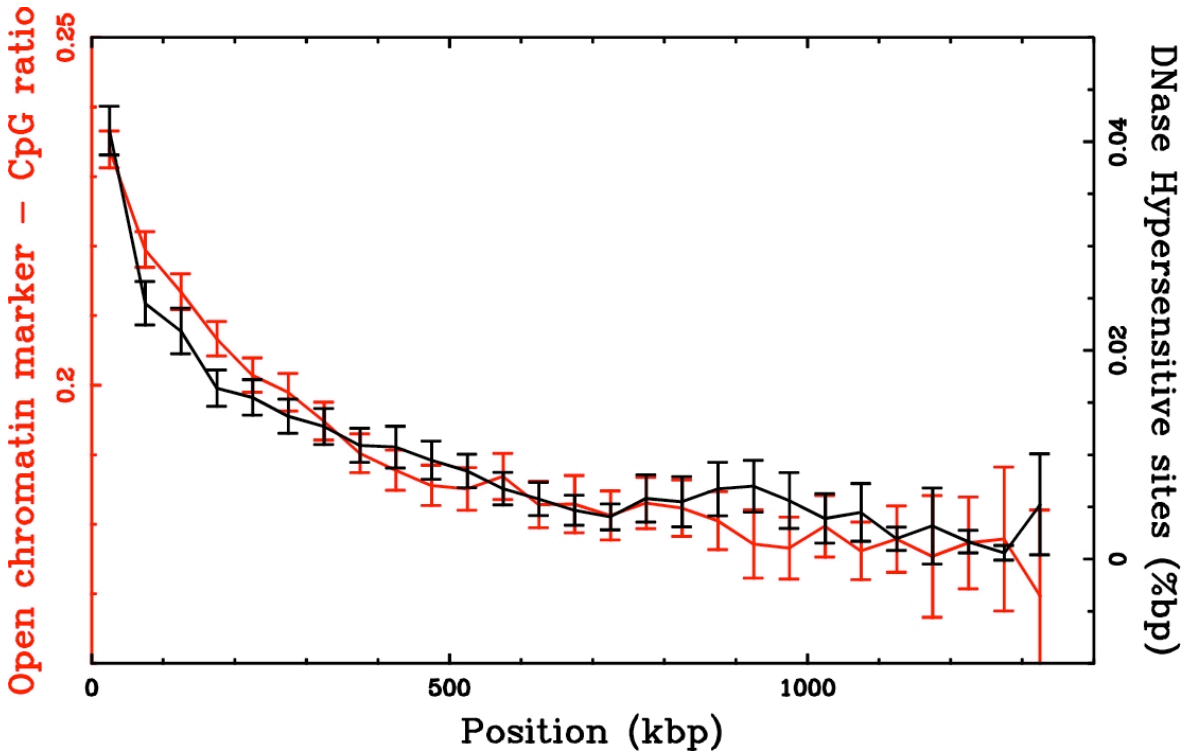
High-resolution mapping of Open Chromatin

Boyle et al., Cell 132, 311 (2008)

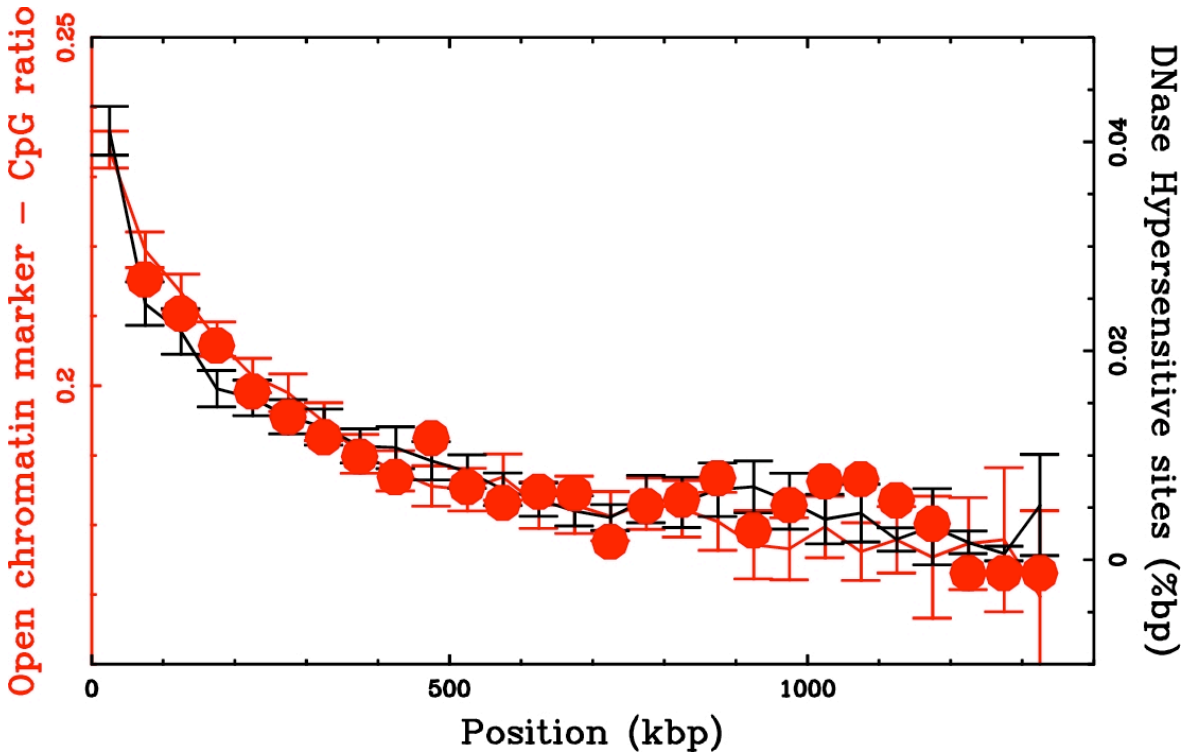
Human chromosome 6



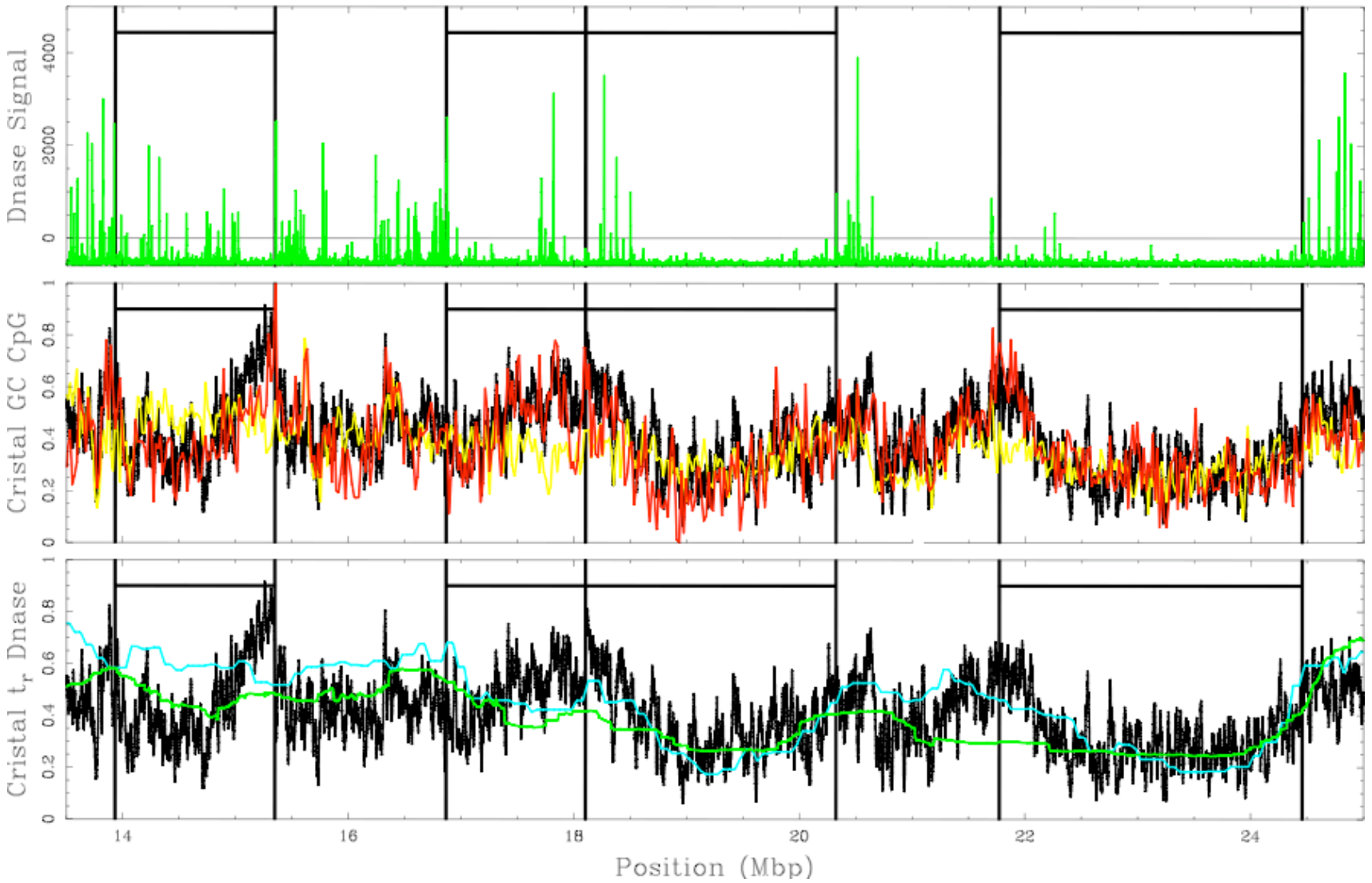
The sequence is highly predictive of open chromatin around putative replication origins



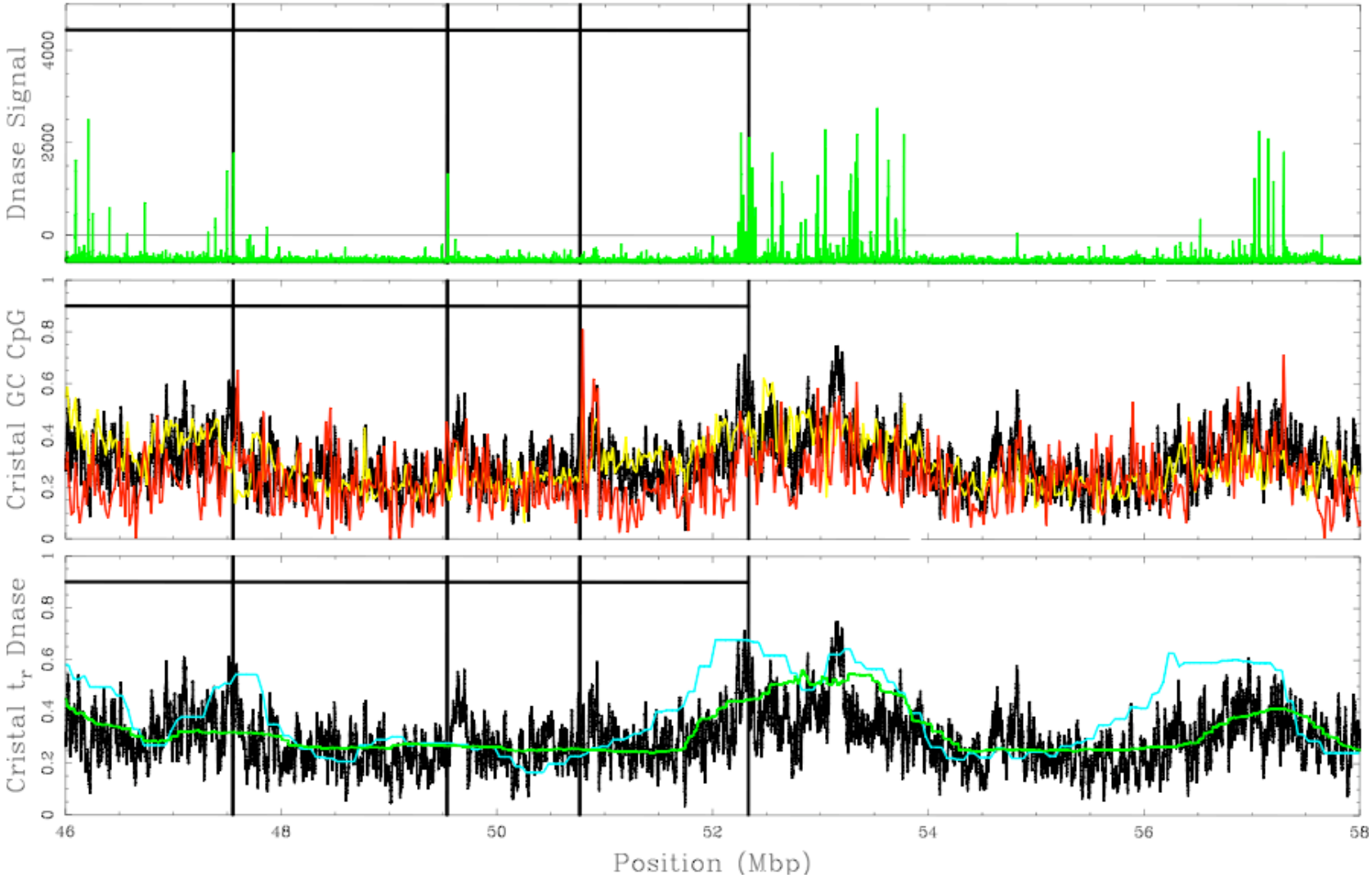
The sequence is highly predictive of open chromatin around putative replication origins



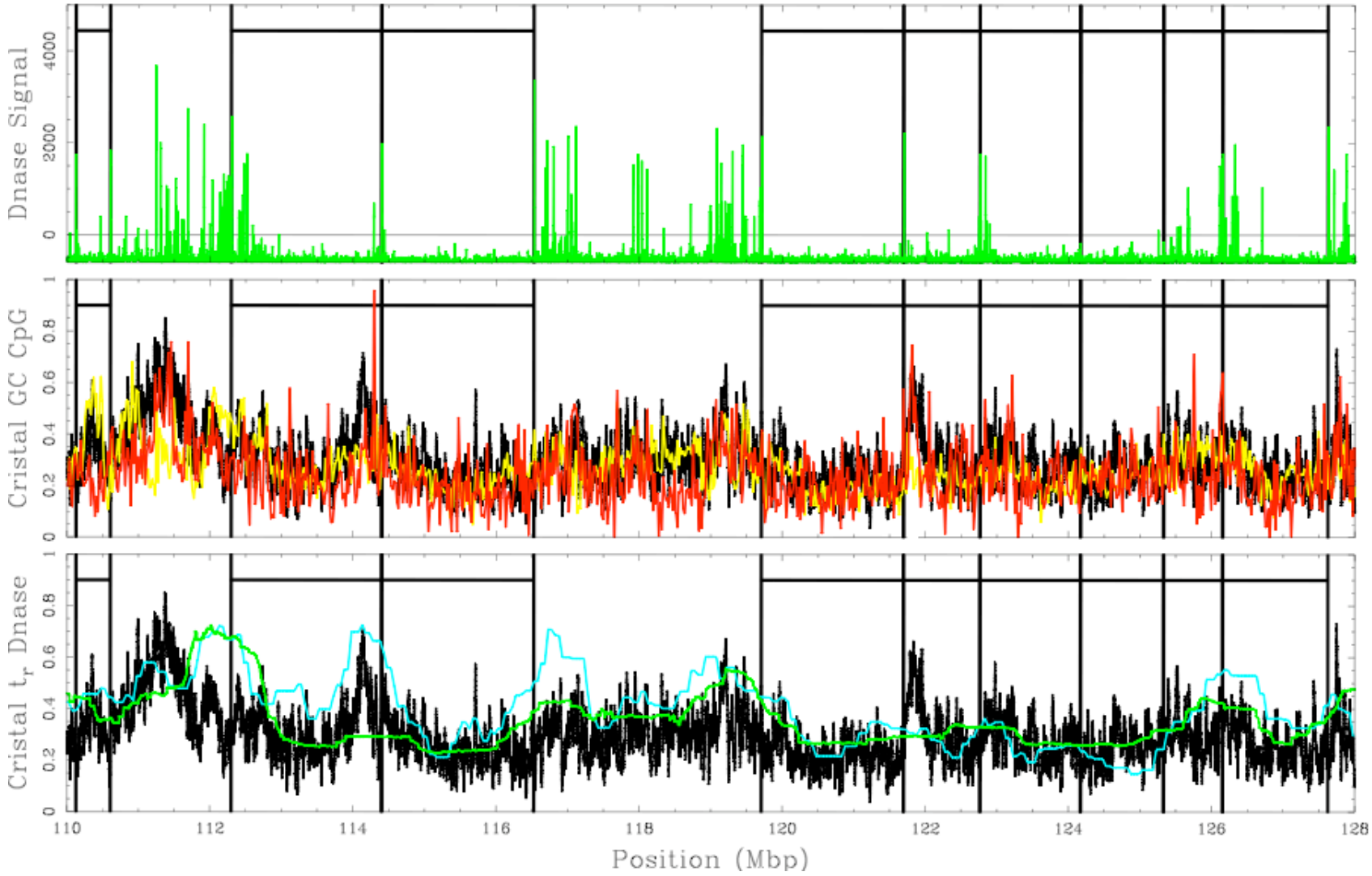
The sequence is highly predictive of open chromatin around putative replication origins



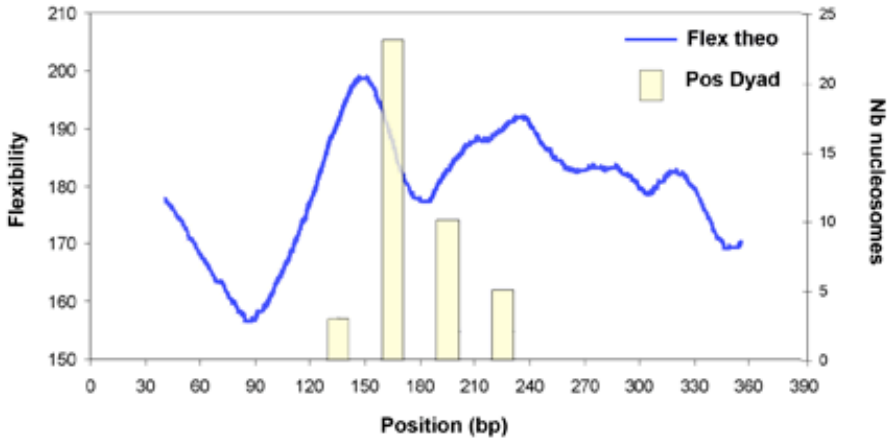
The sequence is highly predictive of open chromatin around putative replication origins

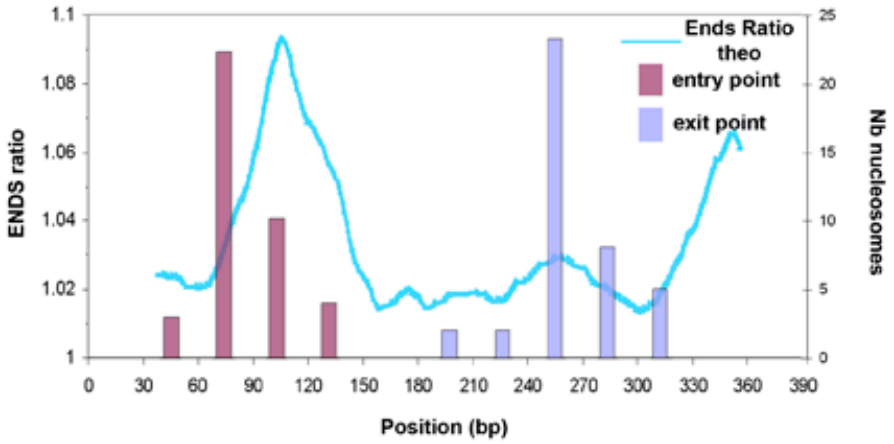


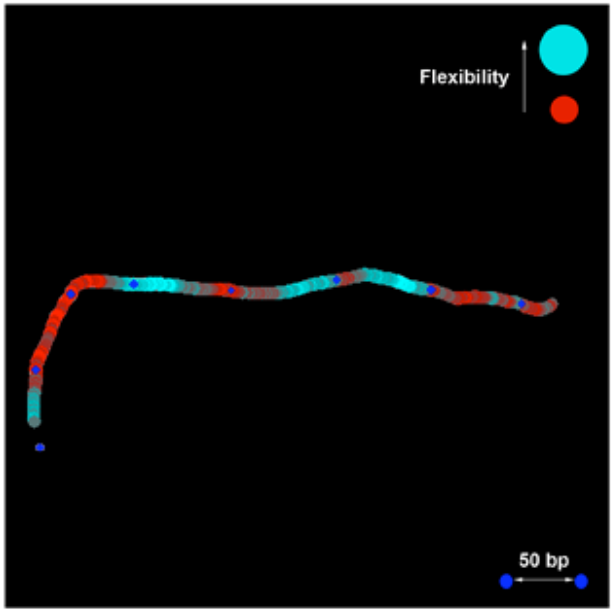
The sequence is highly predictive of open chromatin around putative replication origins

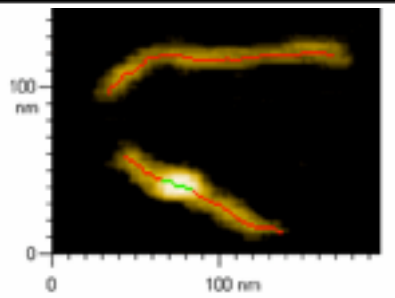


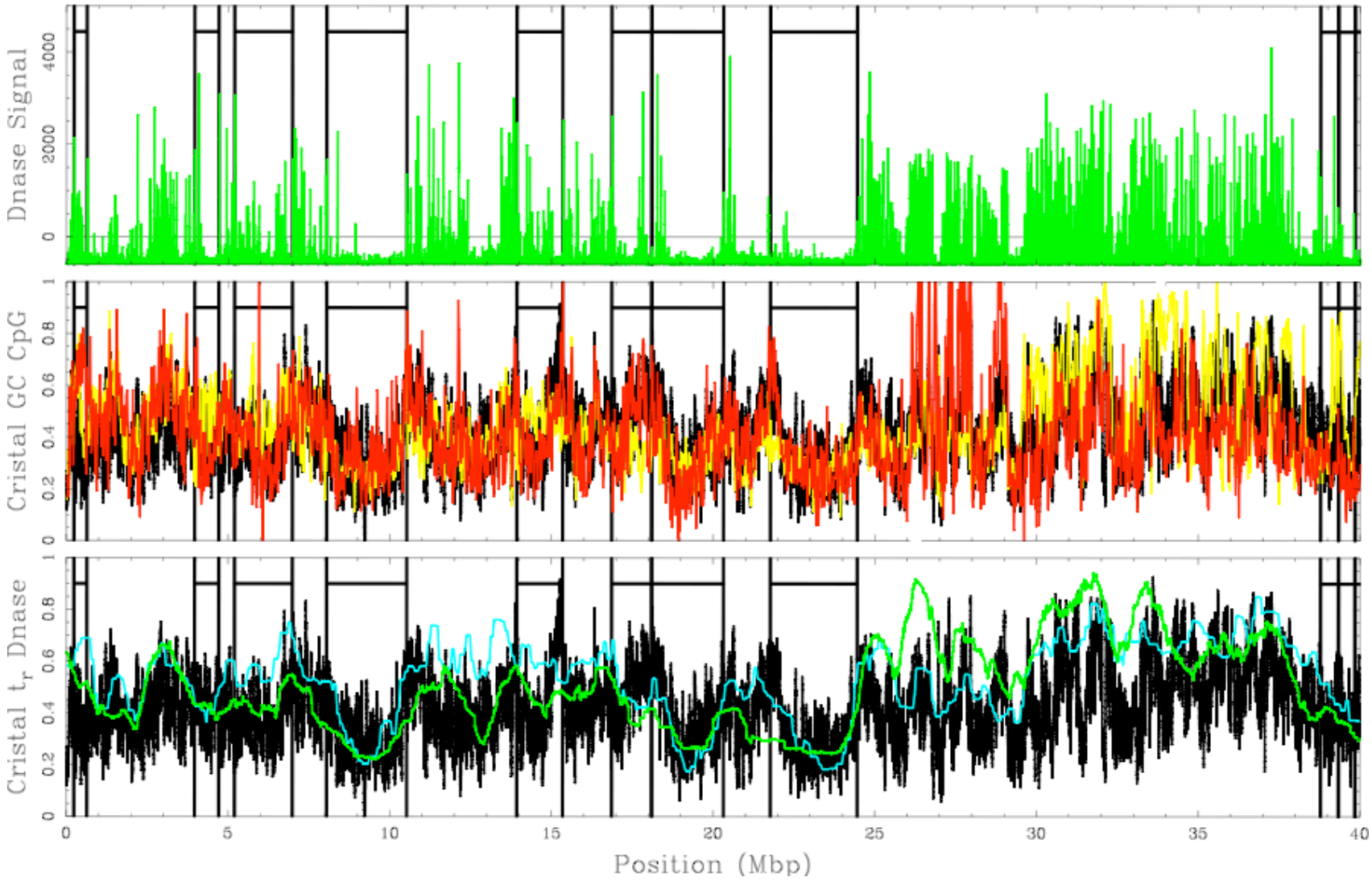
DNA mechanical properties induced nucleosome positioning

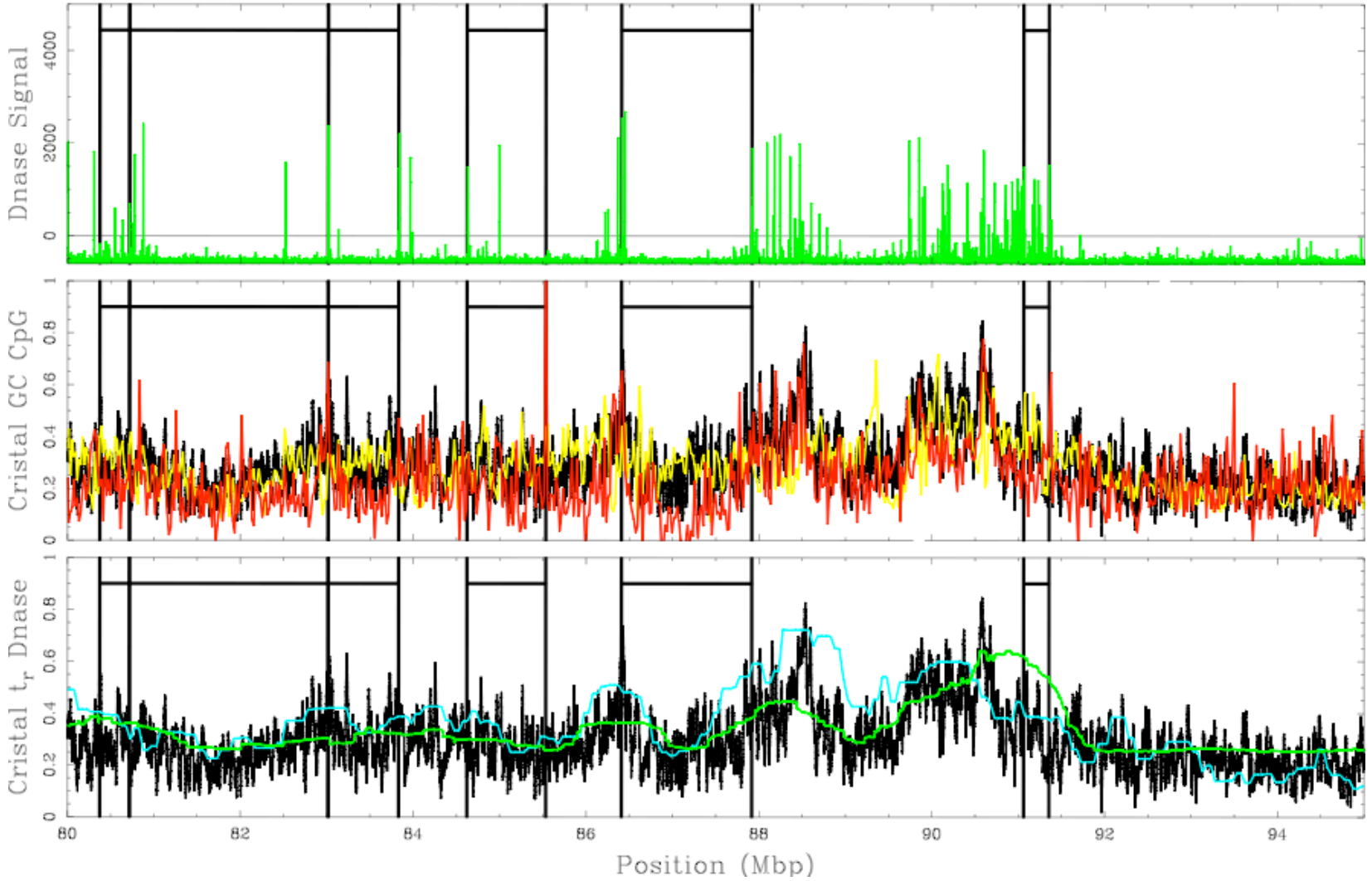


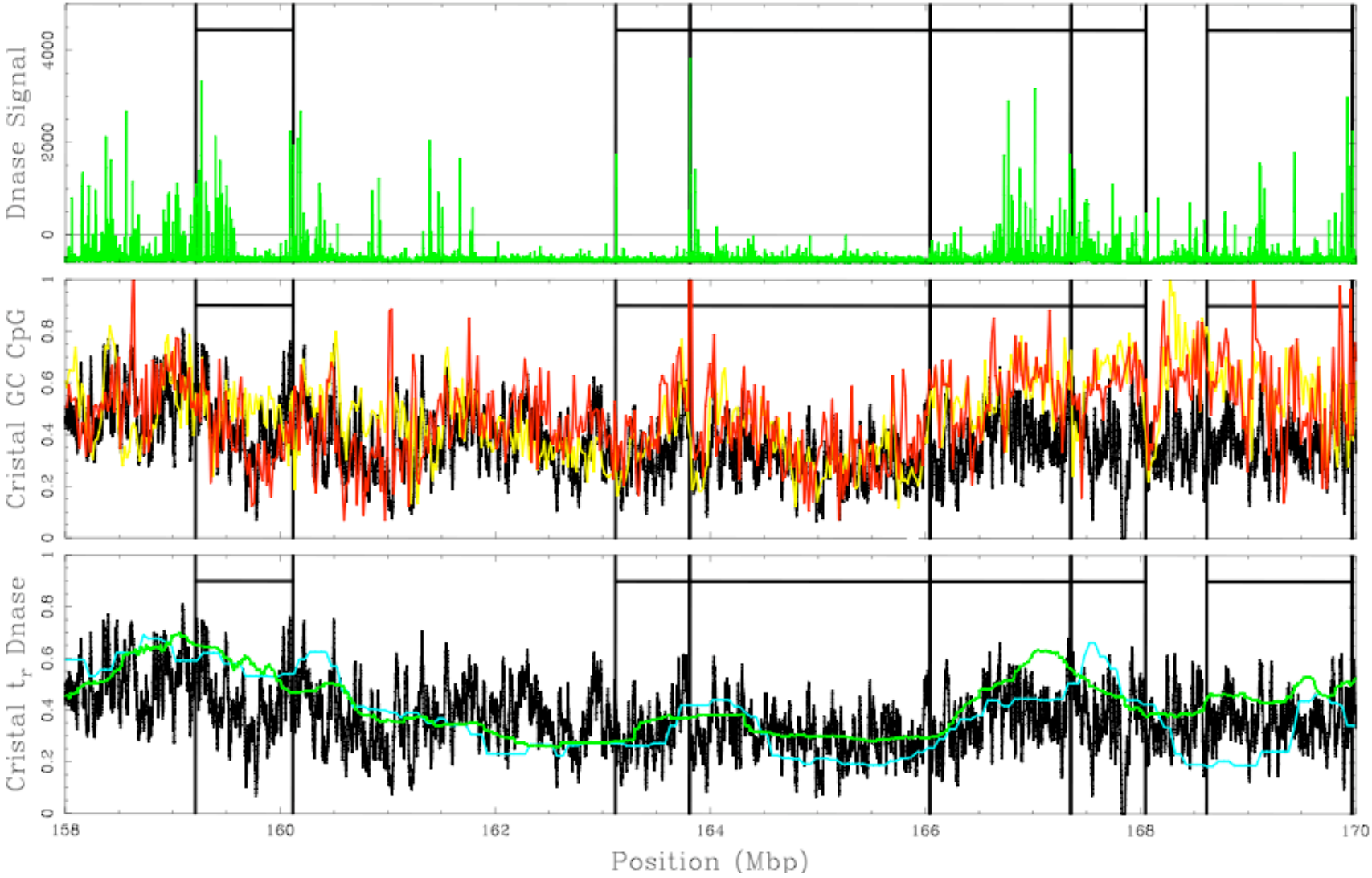







Dnase Signal
0
2000
4000
Cristal GC CpG
0
0.2
0.4
0.6
0.8
1
Cristal t_r Dnase
0
0.2
0.4
0.6
0.8
1
0
5
10
15
20
25
30
35
40
Position (Mbp)


Dnase Signal
0
2000
4000
Cristal GC CpG
0
0.2
0.4
0.6
0.8
1
Cristal t_r Dnase
0
0.2
0.4
0.6
0.8
1
80
82
84
86
88
90
92
94
Position (Mbp)


Dnase Signal
0
2000
4000
Cristal GC CpG
0
0.2
0.4
0.6
0.8
1
Cristal t_r Dnase
0
0.2
0.4
0.6
0.8
1
158
160
162
164
166
168
170
Position (Mbp)