Nucleosomes: what, why and where?

Rob Brewster

Outline

What is a nucleosome?

- how is DNA packaged/organized in Eukaryotes?

Why do nucleosomes form?

- DNA is stiff, how do ~100 bp loops form so readily?

Where do nucleosomes form?

- What controls the spacing and structure of nucleosomes on the chromosome?

Accessibility of DNA in the nucleosomes

- How is DNA inside a nucleosomes accessed?

DNA Organization in Eukaryotes

DNA is packaged and condensed into chromosome

-Human genome is big, nucleus is small

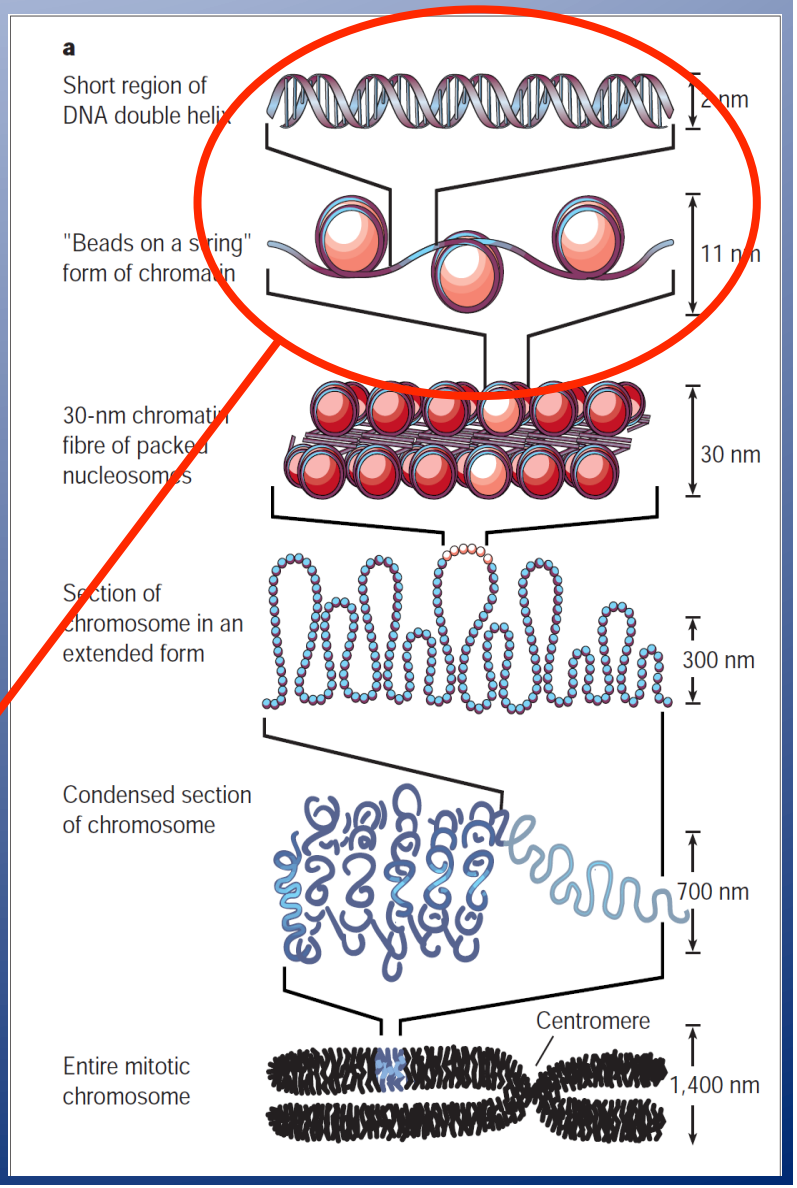
~ 2 billion basepairs ≈ 2m

nucleus radius ~ 6 µm

- Many different levels of organization
  - Compacts chromosome
  - Regulates transcription by making portions of the chromosome more/less accessible (up to 80% is inaccessible to protein binding)

(Lee W et al., Nat. Genetics 2007)

- We will focus on nucleosome formation

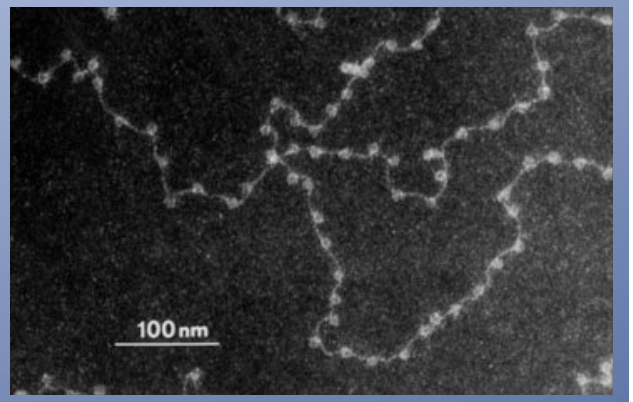


(Alberts, Essential Cell Biology 1998)

What is a nucleosome?

Electron micrograph of chromatin at low ionic strength

- Nucleosomes appear as “beads on a string

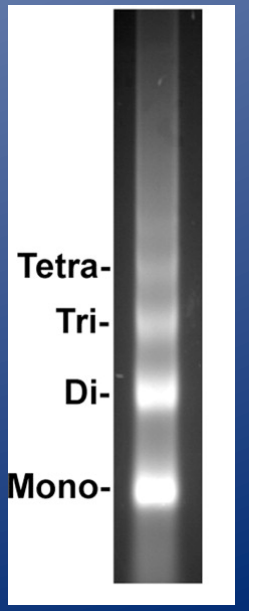


(Electron Micrograph from Olins and Olins)

Basic repeating structure can be probed (protect and seq method)

- Digestion enzyme cuts accessible regions of DNA

- DNA protected by nucleosome is not cut



(Furuyama and Biggins ,PNAS)

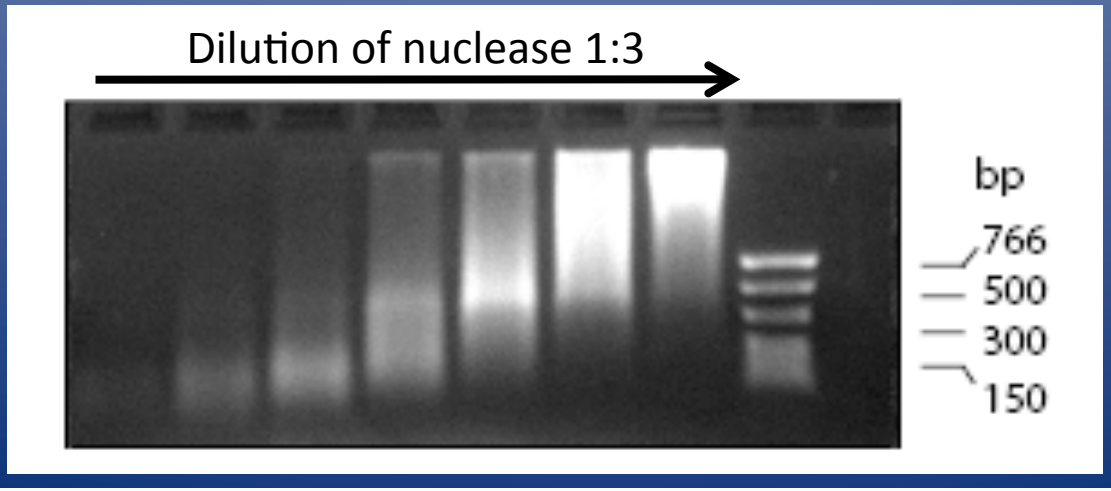
Enzymatic digests

Particular enzymes can cut double stranded DNA between basepairs

- Most have specific recognition sites

Recognition Sequence	Enzyme
AA/CGTT	AclI
A/AGCTT	HindIII
AAT/ATT	SspI SspI-HF™ SspI-HF™
/AATT	Tsp509I
A/CATGT	PciI
A/CCGGT	AgeI AgeI-HF™ AgeI-HF™

- Micrococcal nuclease cleaves everything it can (no specific seq.)



(pictures from NEB)

Structure of individual nucleosome

Repeating structure contains 5 different proteins

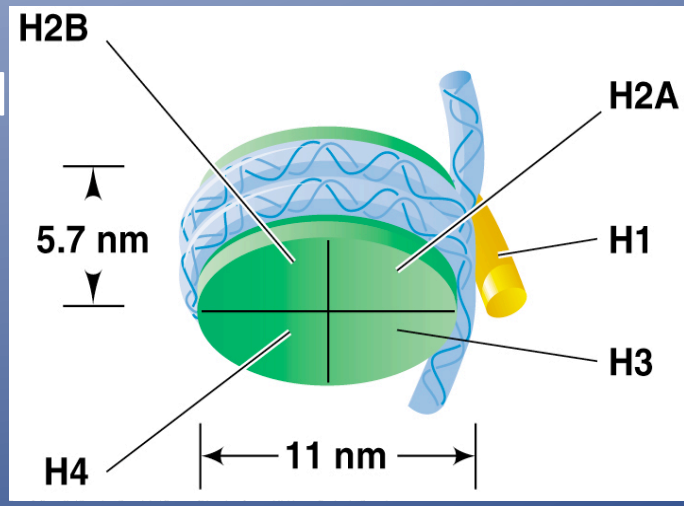
- Main body is an octomer formed from: two copies each of H2A, H2B, H3, H4

- DNA wraps 147 bp (1.75 turns) around the core

-14 non specific adhesive contacts with histone (major groove – histone)

- H1 attaches to the linker region and changes the conformation; required for chromatin formation

- H1 covers 30-50bp



(Peter J. Russell, *iGenetics*)

DNA rigidity

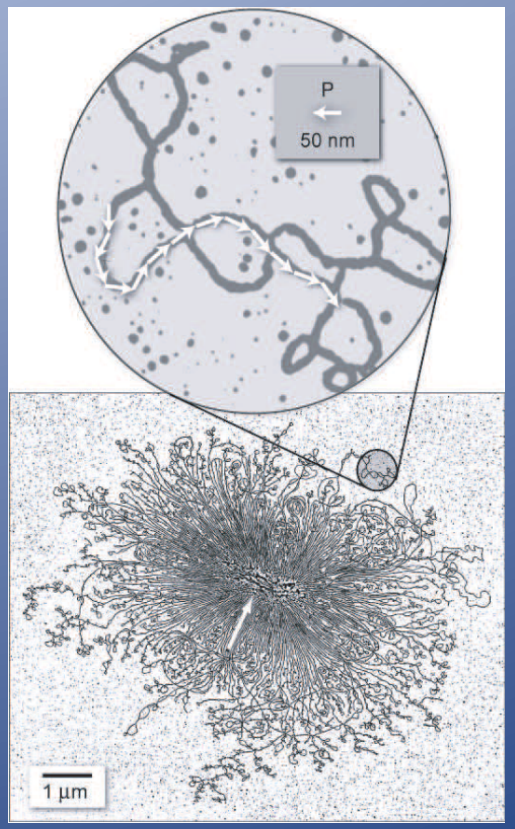
DNA is stiff…

- ~150bp persistence length, λ, for lysed bacterial genome

- loops smaller than λ should be rare

… but not that stiff

- single turn around histone core ~100 bp



Back of the powerpoint calculation

Energy paid to bend loop of 147 bp DNA in 16.5 bp circle

$$F_{\text{bend}} = \frac{1}{2}\frac{\lambda L}{R^2}KT$$

Assume λ=150 bp then: $F_{\text{bend}} \approx 40 K_{\text{B}}T$

This cost must be balanced out (and then some) by the 14 contacts so each contact must contribute ~3 KT

However, nucleosome contacts are ~ 1.5 KT (Polach and Widom, 2005; Schiessel, 2003)

Reversing this and solving for maximum stiffness for stable nucleosomes, λ < ~ 80 bp

What gives?

Bending of short DNA fragments

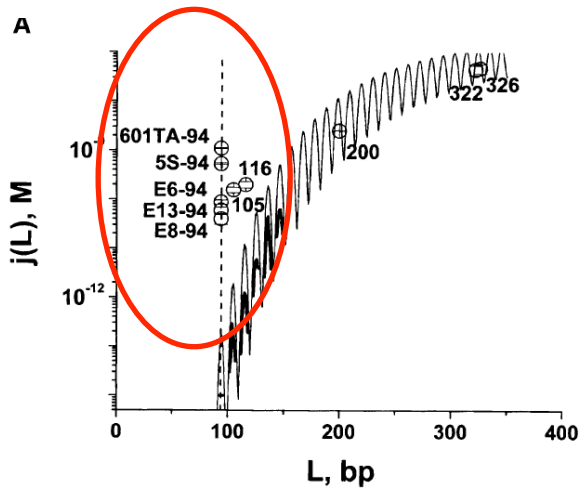
Looping probability for small fragments is larger than expected

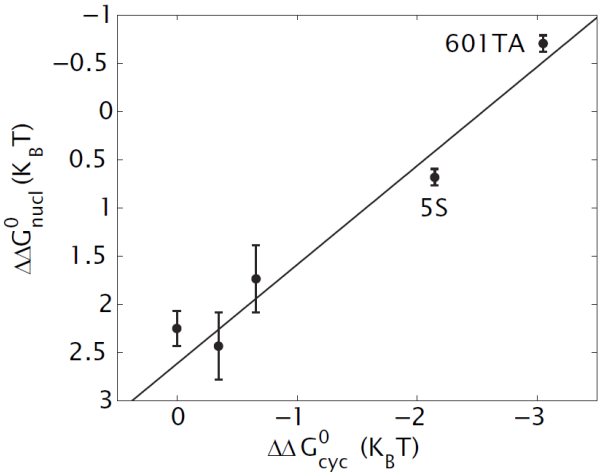
- ~100 bp fragments form loops more readily than predicted

- certain sequences are more flexible than others

- sequences which loop more readily also more readily form nucleosomes (as much as 10^3-fold difference)

Does this sequence dependence control where nucleosomes are positioned?





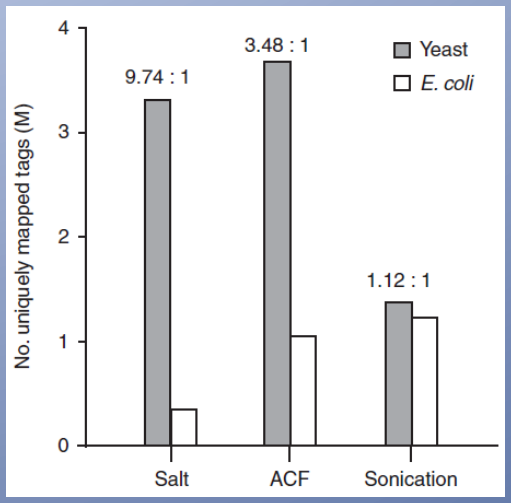
(Cloutier and Widom, Molecular Cell 2004)

Where are Nucleosomes?

The Eukaryotic genome!

- Formation on Yeast genome is more probable than for e. coli genome

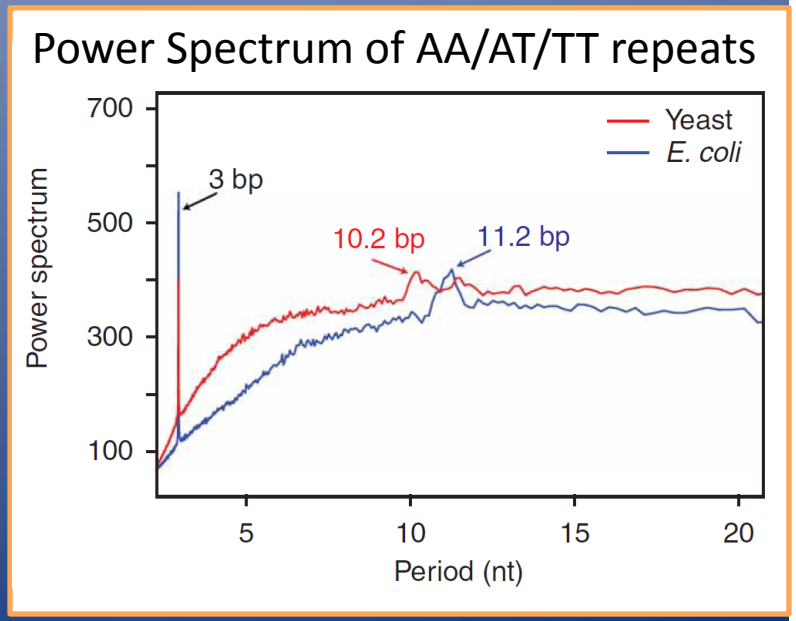
- Implies Euk. Genome has sequence preferential for nucleosomes



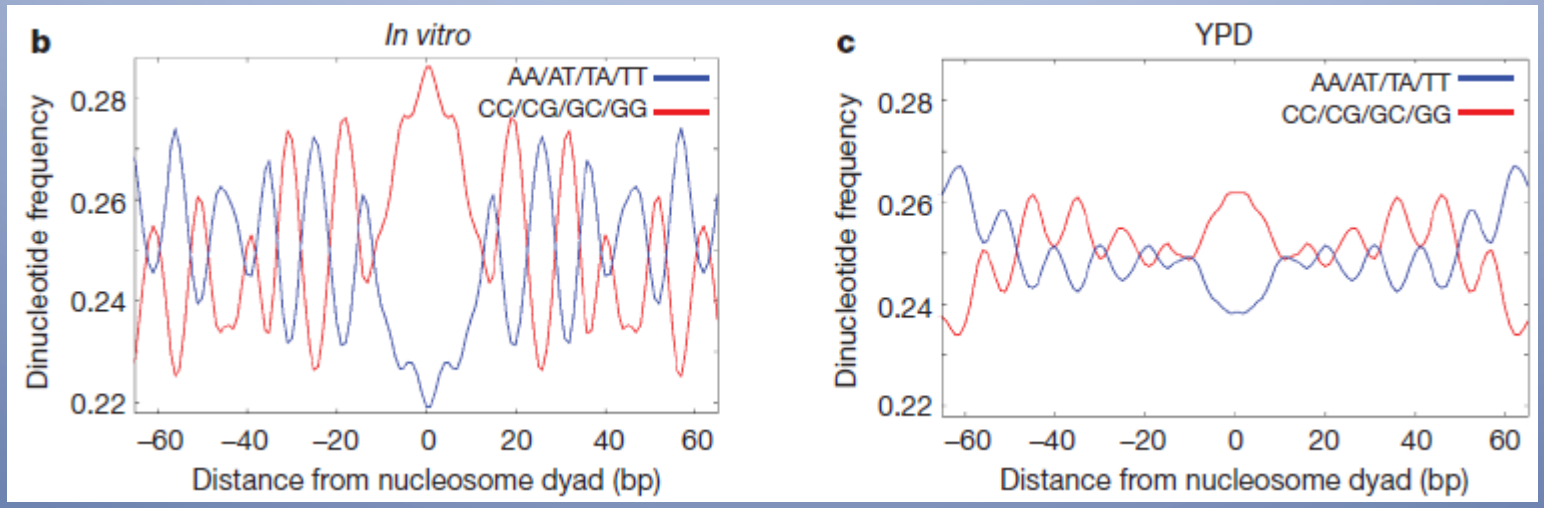
(Zhang et al., Nature struct. & mol. bio., 2009)

Eukaryotic genome shows specific patterns

- AA/AT/TT dinucleotide frequency ~10 bp (one DNA twist)



Nucleosome code



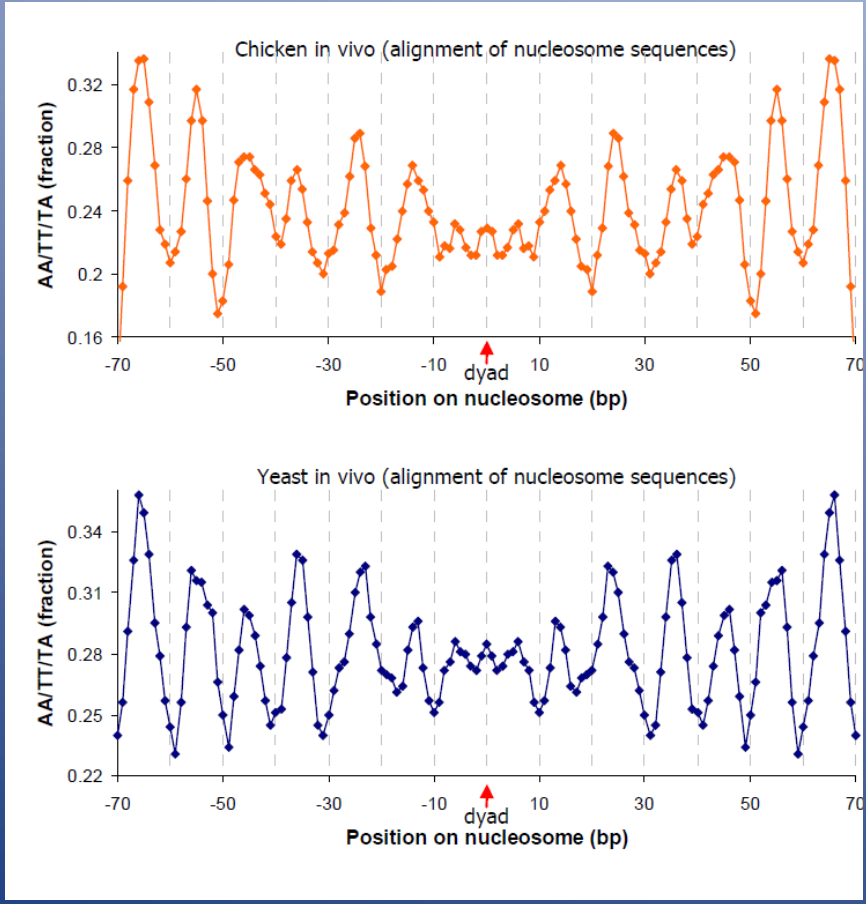
(Kaplan et al, nature 2009)

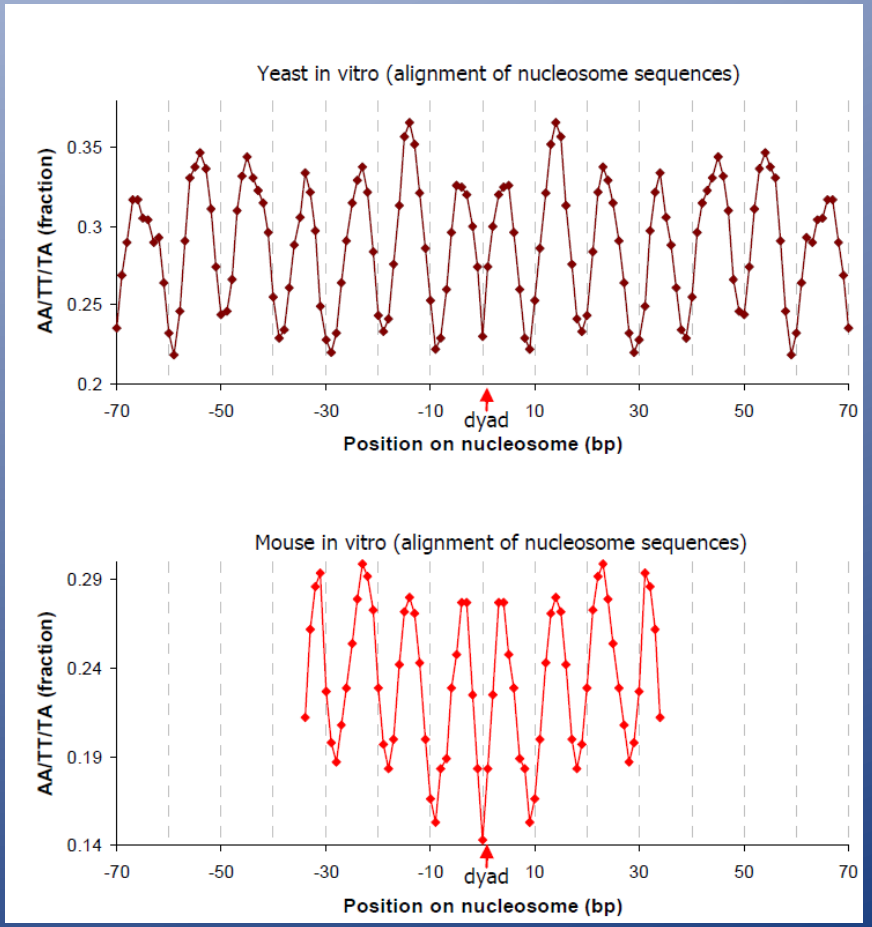
10bp Periodicity is evident both in vivo and in vitro

- GC is 5bp out of phase with AT dinucleotides

- due to bending differences in GC and AT?

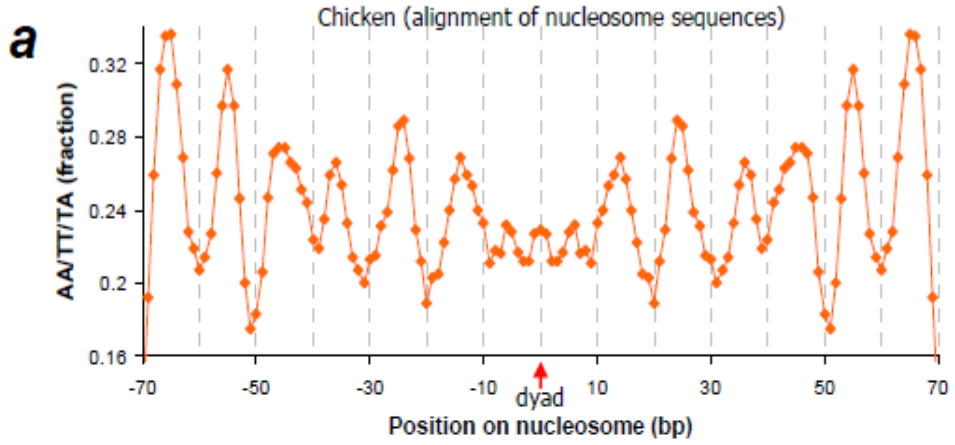
More examples

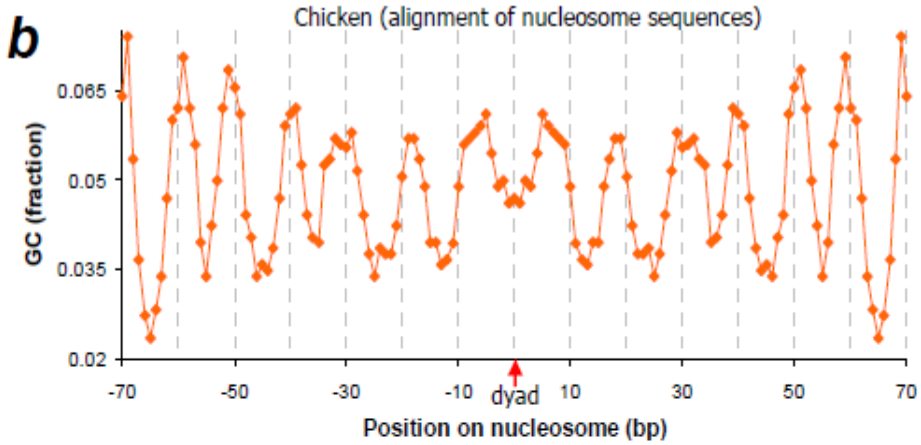


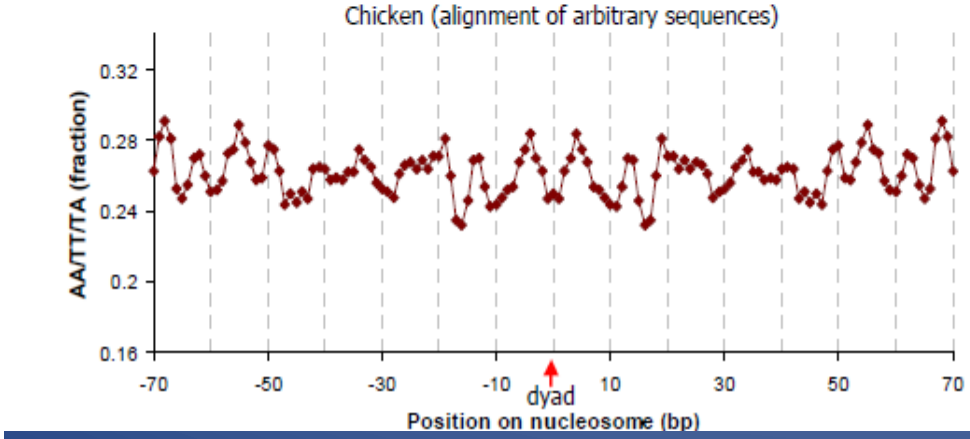


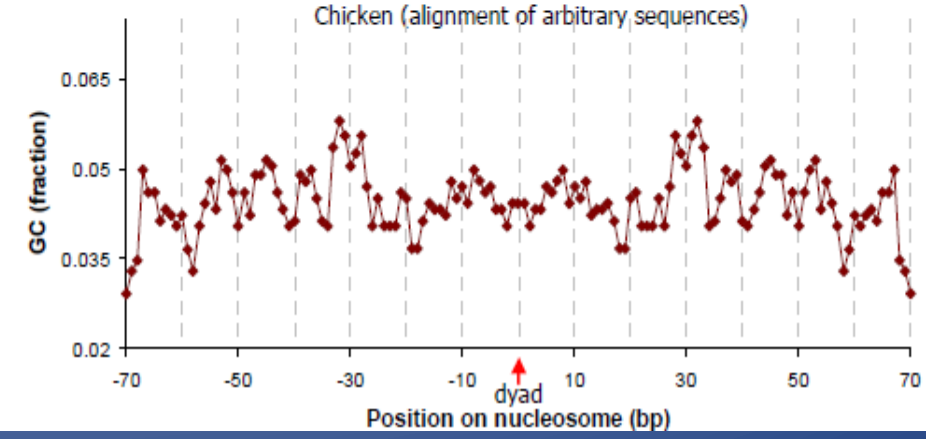
(Segal et al., Nature , 2006)

Nucleosome vs random sequence









Periodicity is missing from arbitrary sequences

(Segal et al., Nature , 2006)

Probabilistic model of nucleosome occupancy

Method: From in vivo occupancy data, calculate the probability of any dinucleotide pair at a given nuc. position.

AA = 1	CA = 1	GA = 0	TA = 5
AC = 0	CC = 0	GC = 0	TC = 4
AG = 1	CG = 0	GG = 0	TG = 0
AT = 2	CT = 4	GT = 2	TT = 10

Represent as conditional counts ↓

Nucleotide in position i	Nucleotide in position i-1			
	A	C	G	T
A	1	1	0	5
C	0	0	0	4
G	1	0	0	0
T	2	4	2	10

Derive conditional probabilities ↓

Nucleotide in position i	Nucleotide in position i-1			
	A	C	G	T
A	0.25	0.2	0	0.263
C	0	0	0	0.211
G	0.25	0	0	0
T	0.5	0.8	1	0.526

Probability that 147bp sequence S is wrapped:

$$P(\boldsymbol{S}) = P_1(S_1)\prod_{i=2}^{147} P_i(S_i \mid S_{i-1})$$

The statistical weight of a longer sequence with multiple nucleosomes is just the product of the probability for every basepair to be in that state,

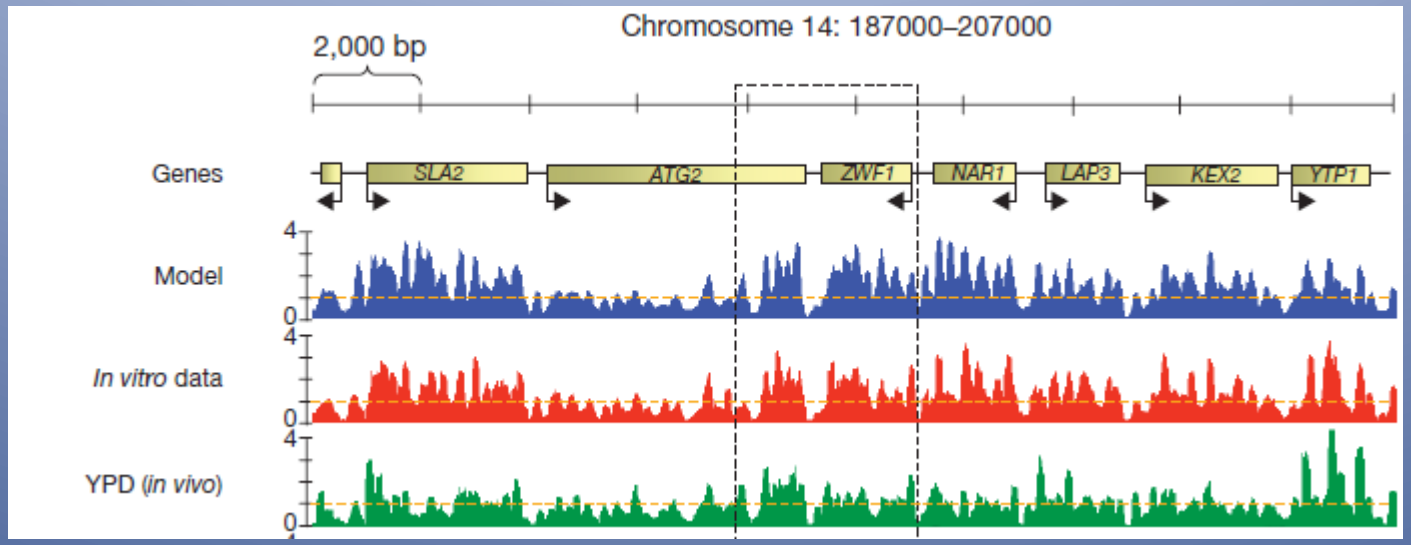
$$W_c[\mathrm{S}] = P_B(S_1, S_{c[1]-1}) \cdot \left(\prod_{i=1}^{k} \tau P(S_{c[i]}, S_{c[i]+146})\right) \cdot \left(\prod_{i=1}^{k-1} P_B(S_{c[i]+147}, S_{c[i+1]-1})\right) \cdot P_B(S_{c[k]+147}, S_N)$$

The probability must now be computed computationally due to the enormous number of possible configurations (C)

$$P(W_c[\mathrm{S}]) = \frac{W_c[\mathrm{S}]}{\sum_{c' \in \mathrm{C}} W_{c'}[\mathrm{S}]}$$

(Segal et al., Nature , 2006)

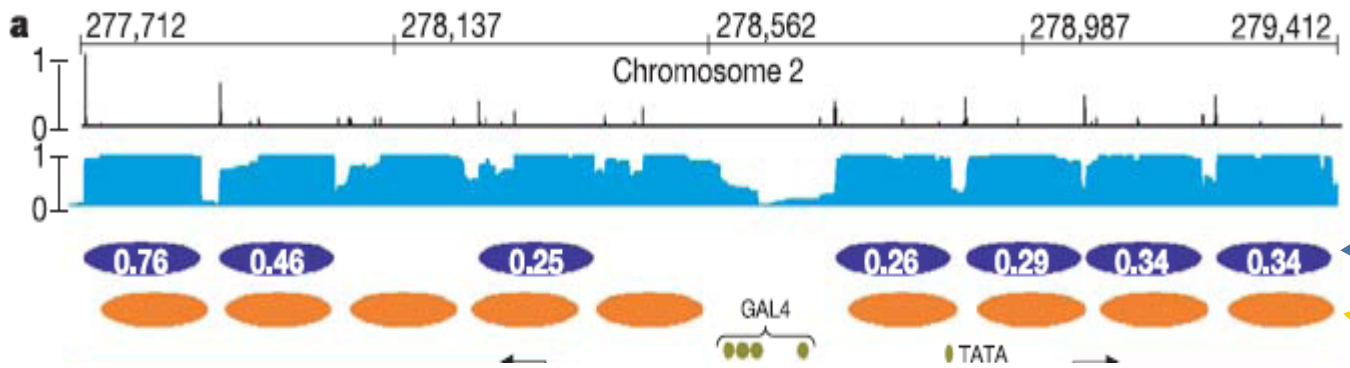
Probabilistic landscape for occupancy



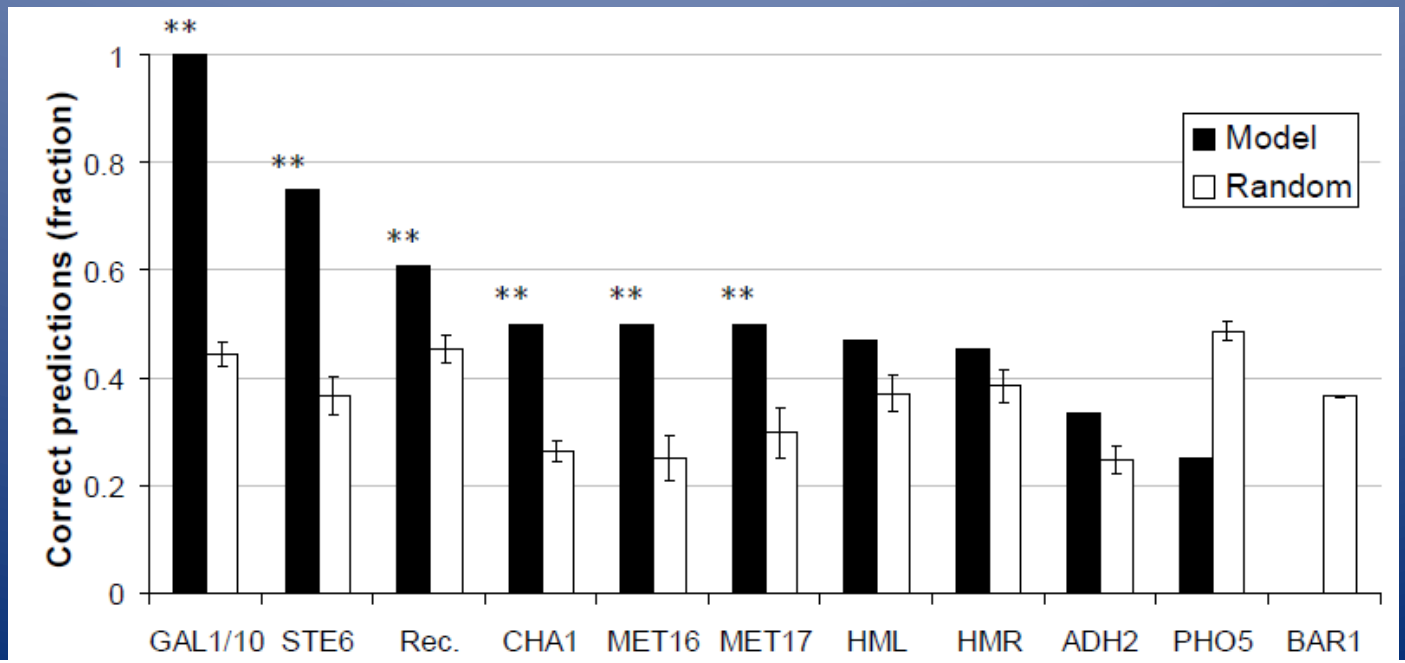
At a particular nucleosomal coverage can predict where stable nucleosomes will be located

(Kaplan et al, nature 2009)

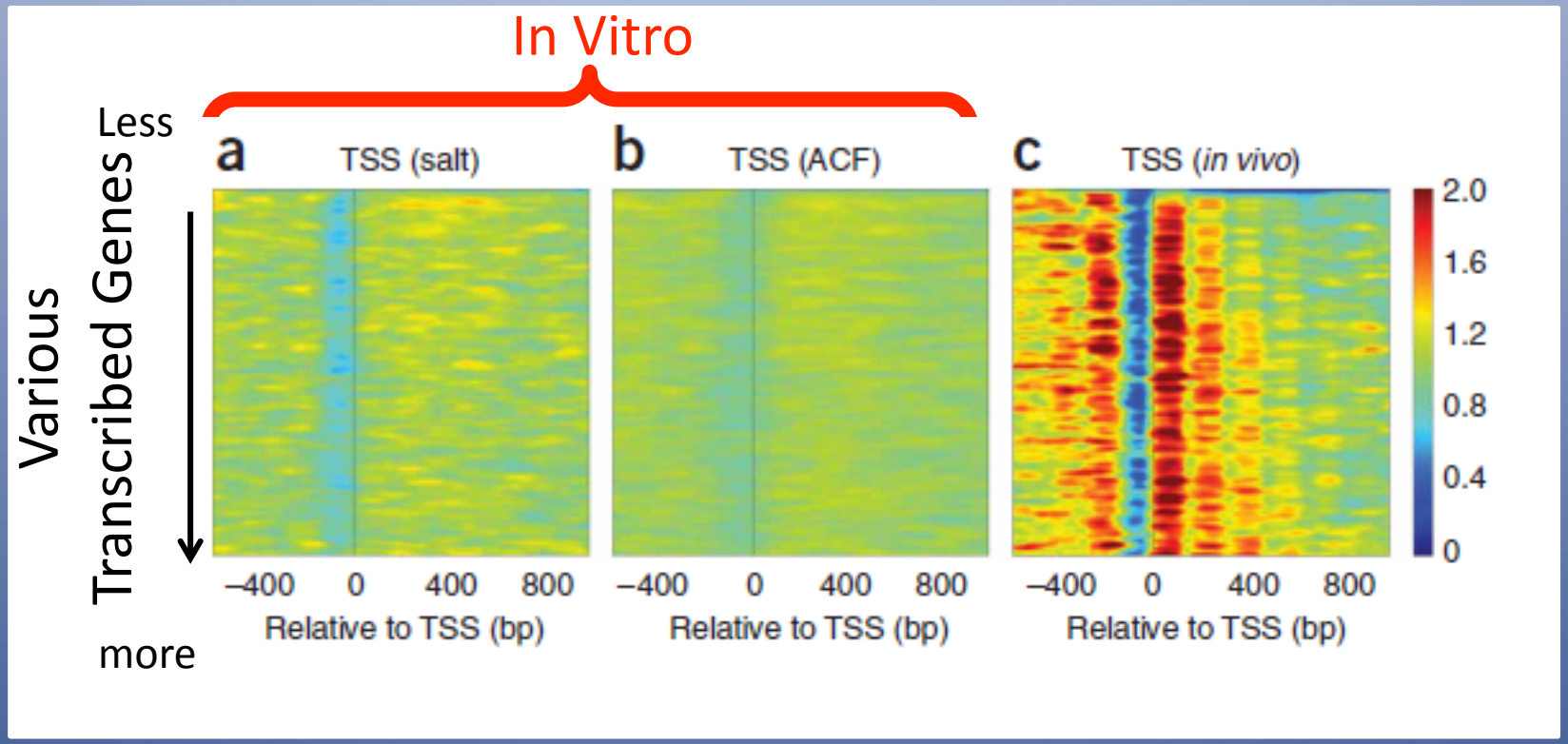
Predictions from thermodynamic model



Success rate within 32bp of predicted position



Statistical Positioning – A (semi)competing view



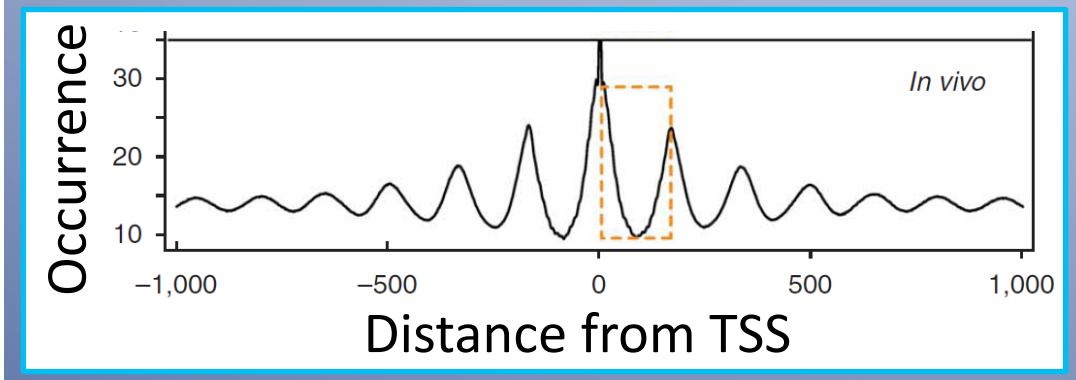
(Zhang et al., Nature struct. & mol. bio., 2009)

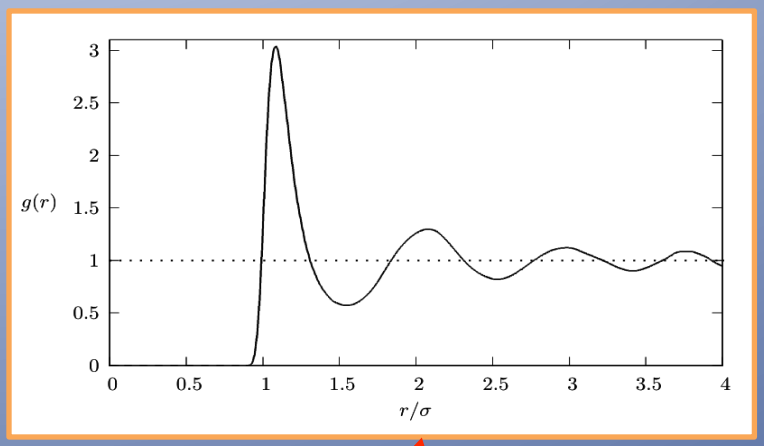
Nucleosomes pattern emerges from steric exclusion on a line

- Promoter region is always “nucleosome free”
- DNA sequence appears to play only a small role

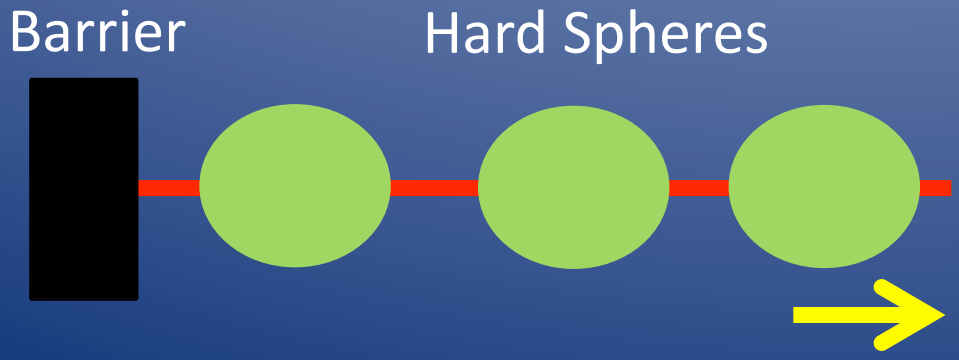
Random 1D hard sphere gas?

(Zhang et al., Nature struct. & mol. bio., 2009)





Phasing in vivo resembles RDF of (for instance) LJ Gas



Can a simple model of a 1D hard sphere gas predict nucleosome positioning pattern?

Conclusions on positioning

DNA sequence matters for nucleosomes positions…

- preferentially bind to particularly repeating sequences

- avoid long tracks of A/T

… however, this effect seems to be minimized in vivo

- Simple 1D gas model fits in vivo nucleosome positions well (data not shown)

How does one resolve the apparent conflict in these views?

Accessing nucleosomal DNA

How is protected DNA accessed?

- DNA replication, repair and transcription all require access to occluded DNA

- Specialized motors called “remodeling factors” disassemble and perturb nucleosomes to allow access

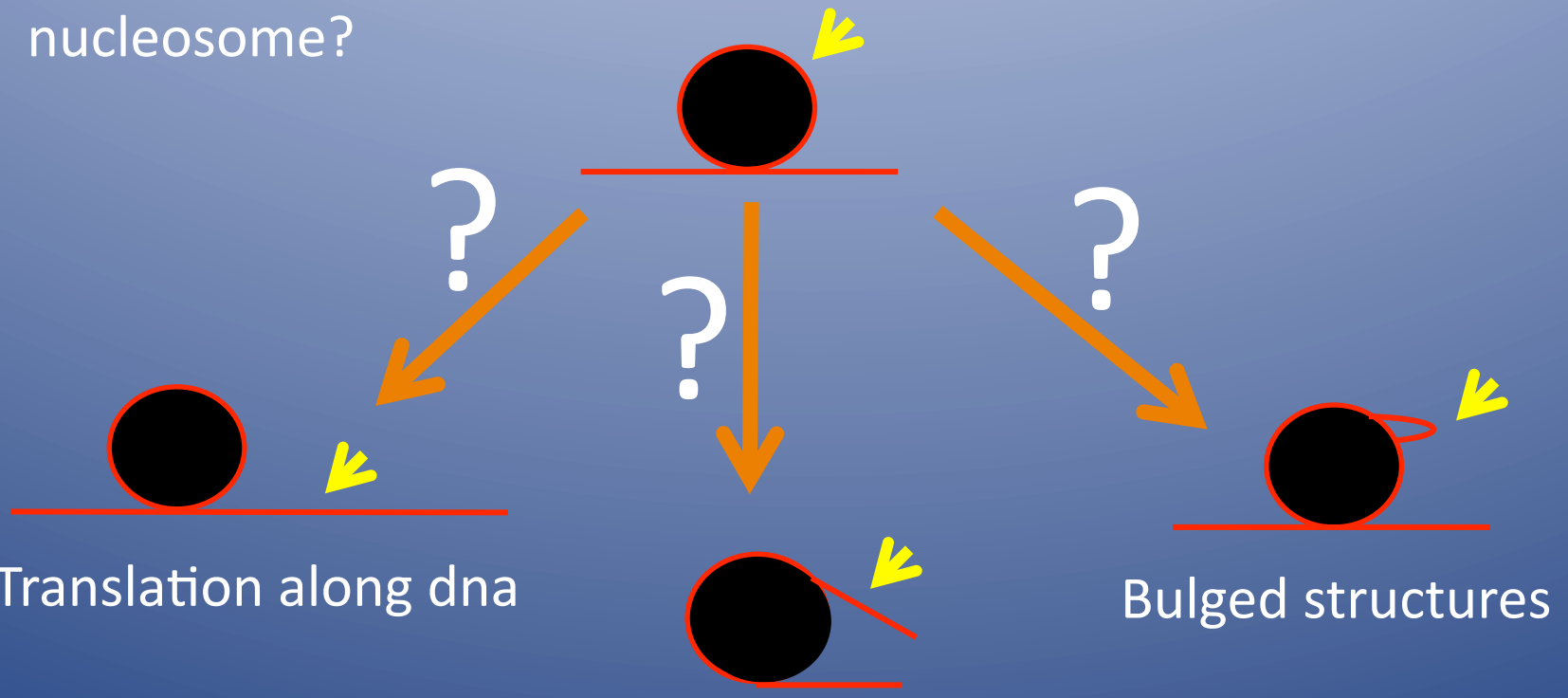
-However, even without these motors, wrapped DNA has some accessibility

- Digestion assays probe equilibrium accessibility

- Fret probes dynamic accessibility

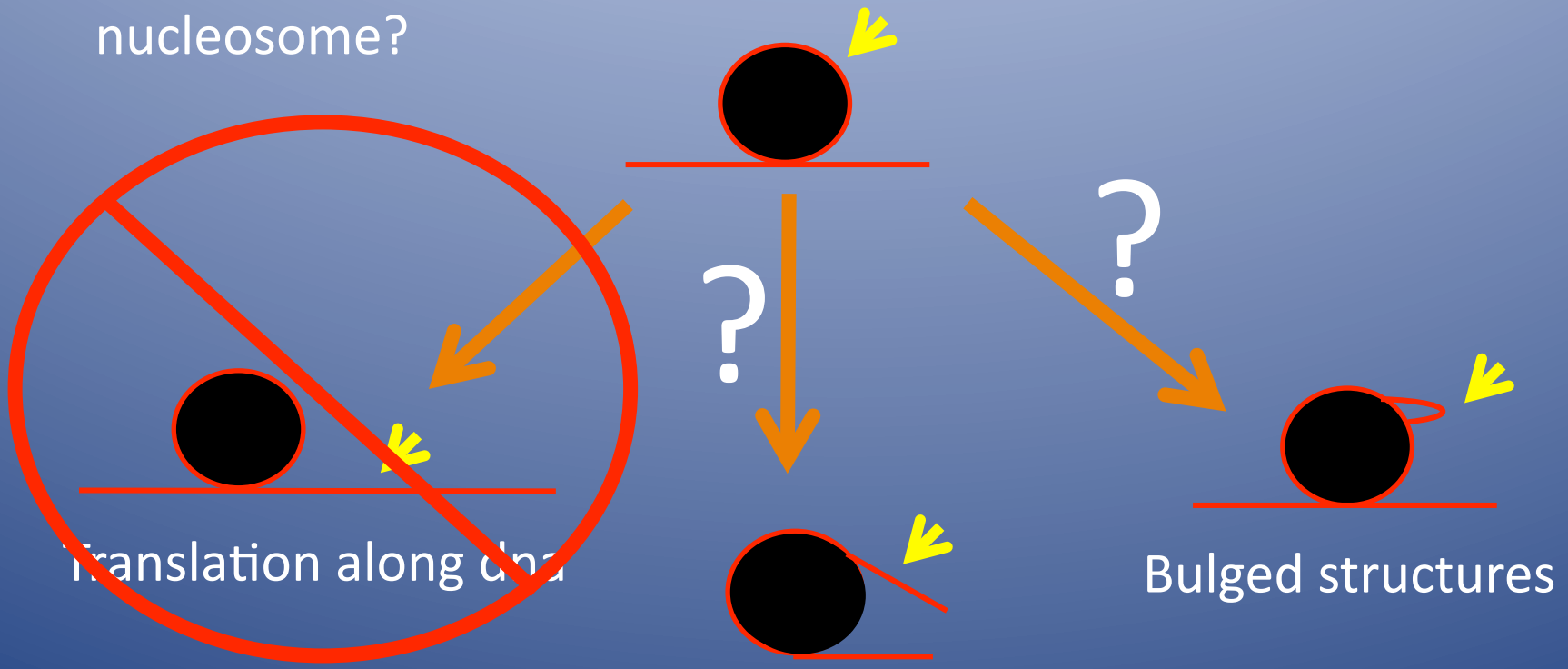
In vitro accessibility mechanism

What is the mechanism to access a site buried deep in the nucleosome?



In vitro accessibility mechanism

What is the mechanism to access a site buried deep in the nucleosome?



In vitro accessibility does not show length dependence beyond ~100 bp

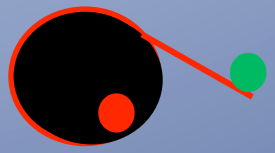
(Anderson et al, 2002)

Fret on nucleosomal complex

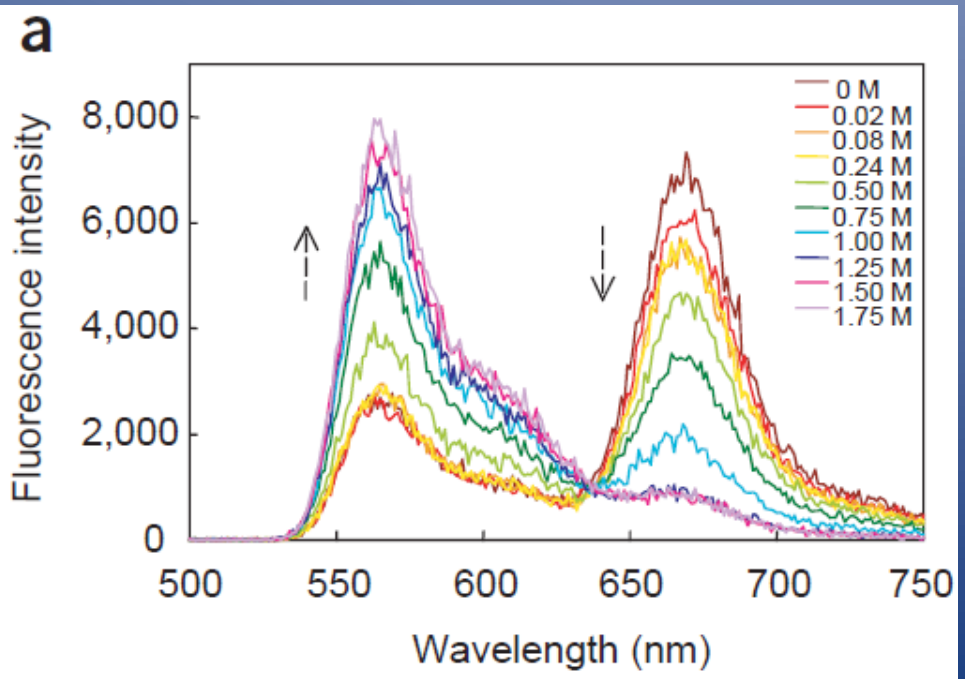
One fret dye is put on the nucleosome, one is put on the end of the DNA strand

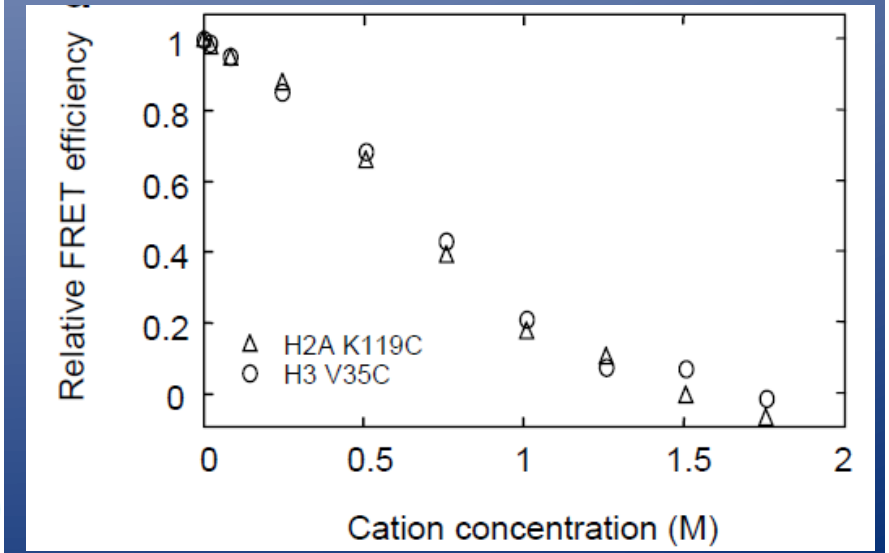
Dyes are far – signal ~680nm is low

Excite at low wavelengths:



Dyes are close – signal ~680nm is high

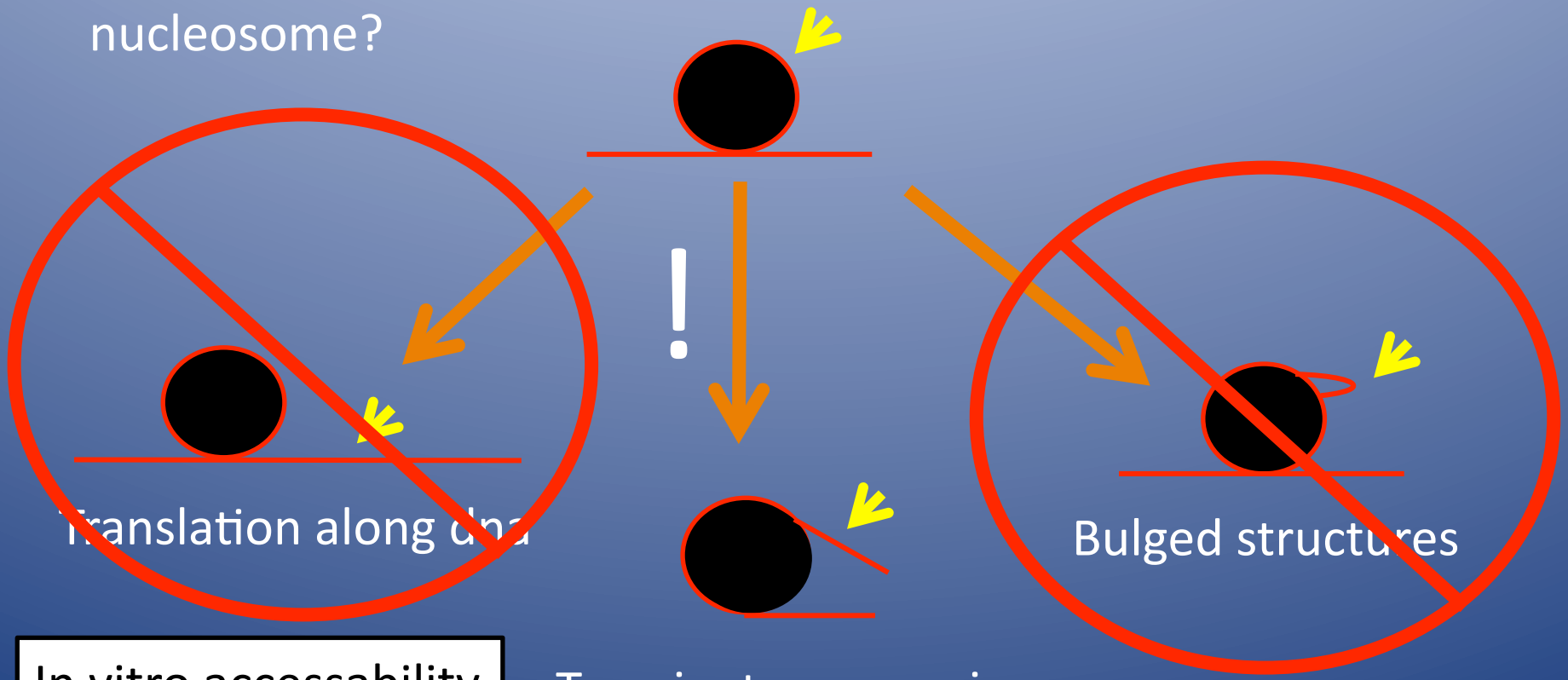




(Li and Widom, 2004)

In vitro accessibility mechanism

What is the mechanism to access a site buried deep in the nucleosome?

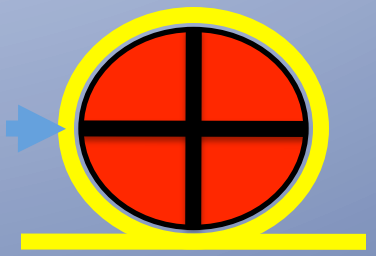


In vitro accessability does not show length dependence beyond ~100 bp

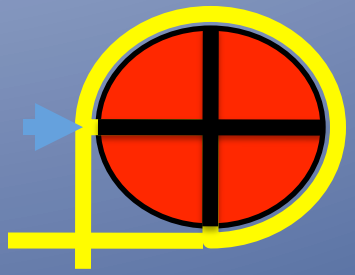
FRET experiments by Li and Widom

Probing accessibility

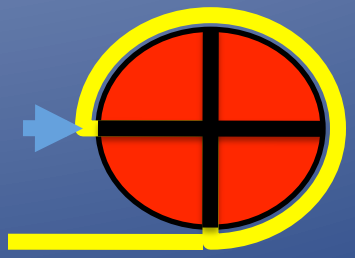
Specific restriction enzyme binding site



DNA is cleaved when site becomes accessible from unwrapping



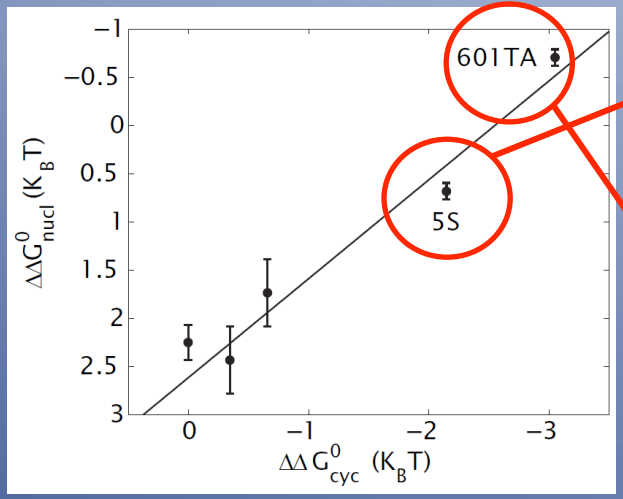
Measure probability for being cut as a function of burial depth

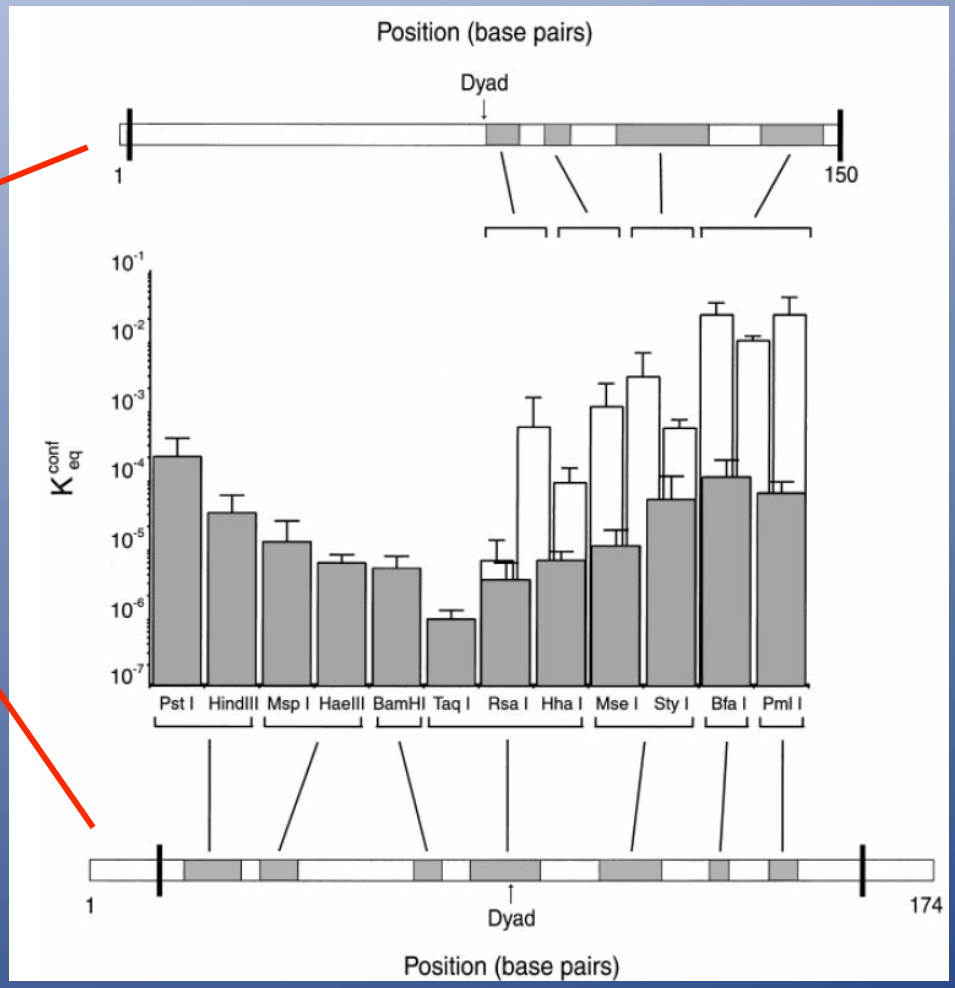


Accessibility is reduced (compared to naked dna)
sites buried shallow 10-100x, deep ~10^5x

(Polach and Widom, 2000)

Measuring unwrapping



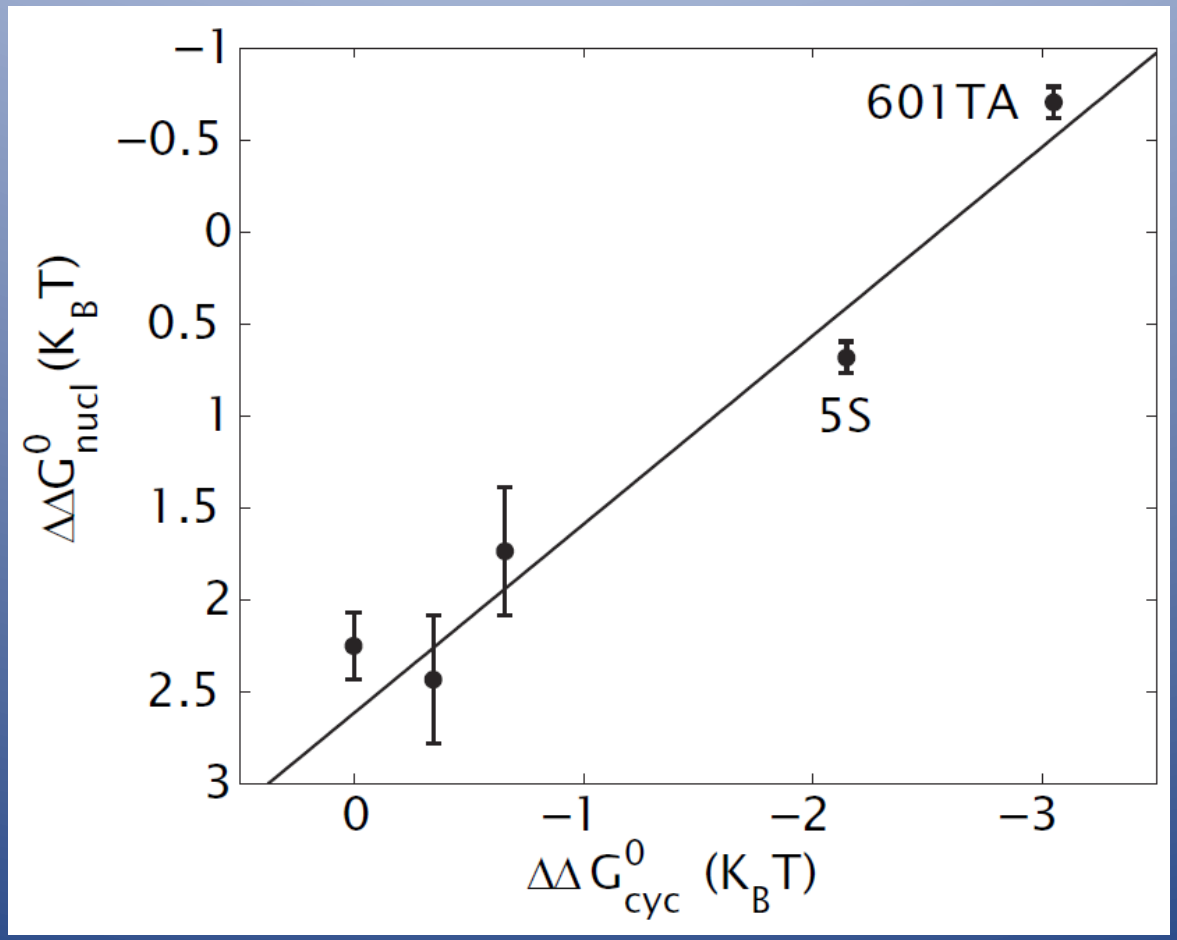


The more flexible sequence is less accessible!

$$K_{\text{conf}} = e^{-\Delta G^0_{\text{conf}}/KT}$$

Where ΔG is the change in free energy associated with opening the binding site

Theory Interlude

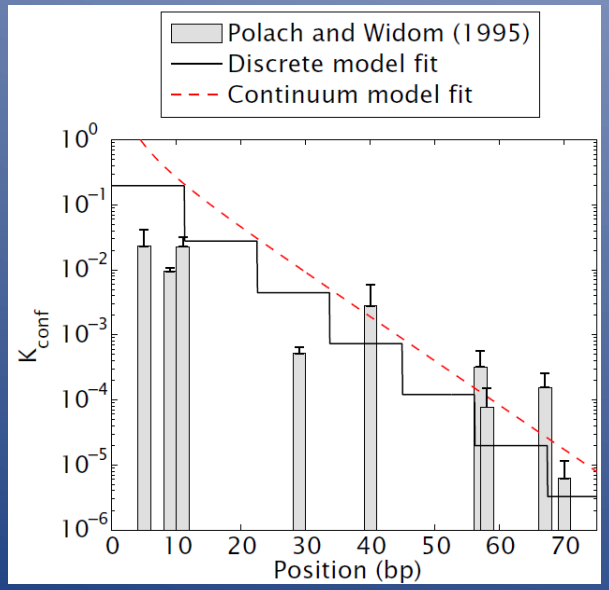


Nucleosome preference is entirely bending?!

Fitting bending energy

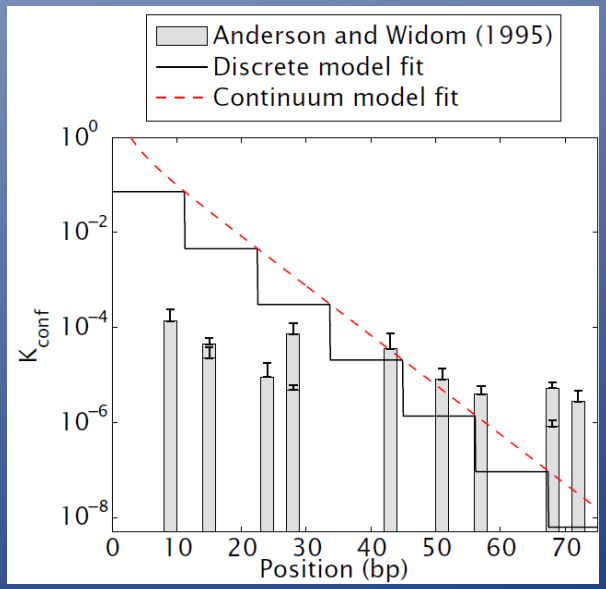
$$K_{\mathrm{conf}} = \frac{1}{\exp\left[-\gamma\alpha a/kT\right] - 1}$$

More rigid sequence (5S)



γ =- 0.16 kT/bp

More flexible sequence (601TA)



γ = -0.24 kT/bp

Per contact energies:

$$F = F_{\text{bend}} + F_{\text{adh}} = \gamma a$$

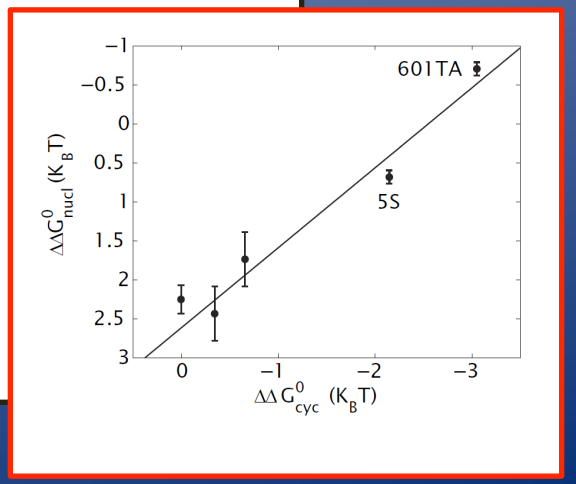
$$\Delta F_{\text{bend}} = a(\gamma_{601TA} - \gamma_{5S})$$

$$F_{\text{bend}} = \frac{1}{2}\frac{\lambda L}{R^2} KT$$

$$\lambda_{601\text{TA}} - \lambda_{5\text{S}} = \frac{R^2 a}{L}\left(\gamma_{601\text{TA}} - \gamma_{5\text{S}}\right) = -13.5\text{bp}$$

However, for Cloutier and Widom data

$$\lambda_{601\text{TA}} - \lambda_{5\text{S}} \approx -2.6\text{bp}$$

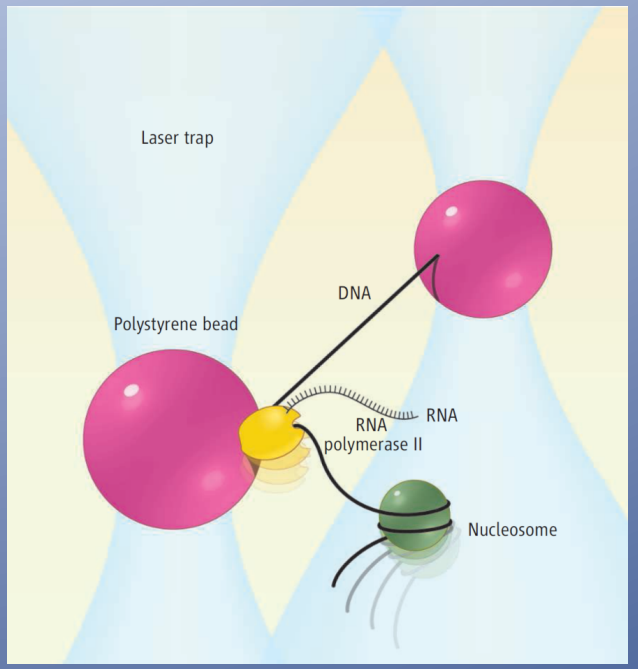


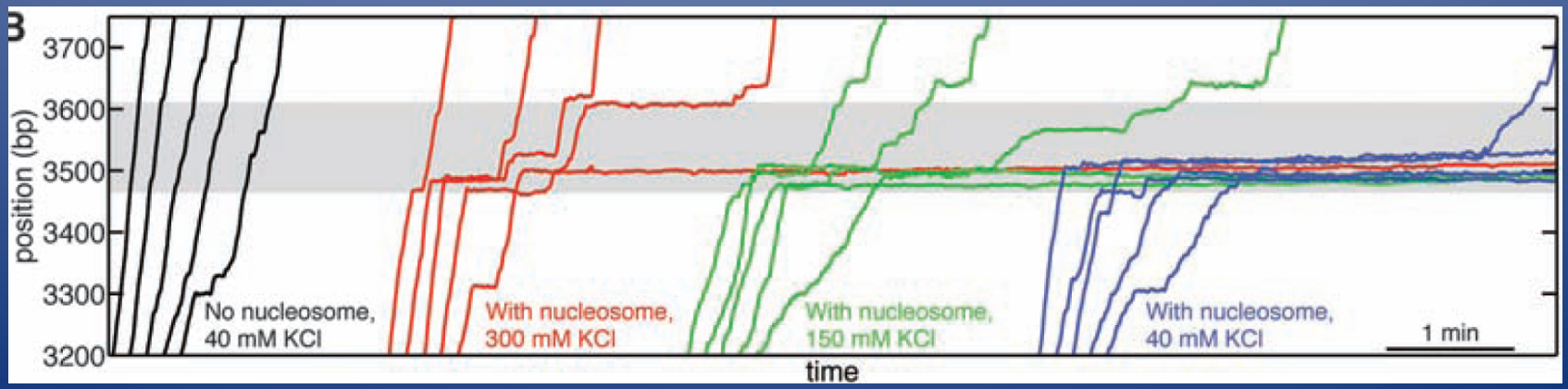
RNAP transcription through the nucleosome

Bead attached to end of DNA and RNAP

-RNAP movement through the nucleosome can be measured

- Distance between beads is a readout for RNAP depth in nucleosome





(Hodges et al, 2009)

Model for RNAP transcription through nucleosome

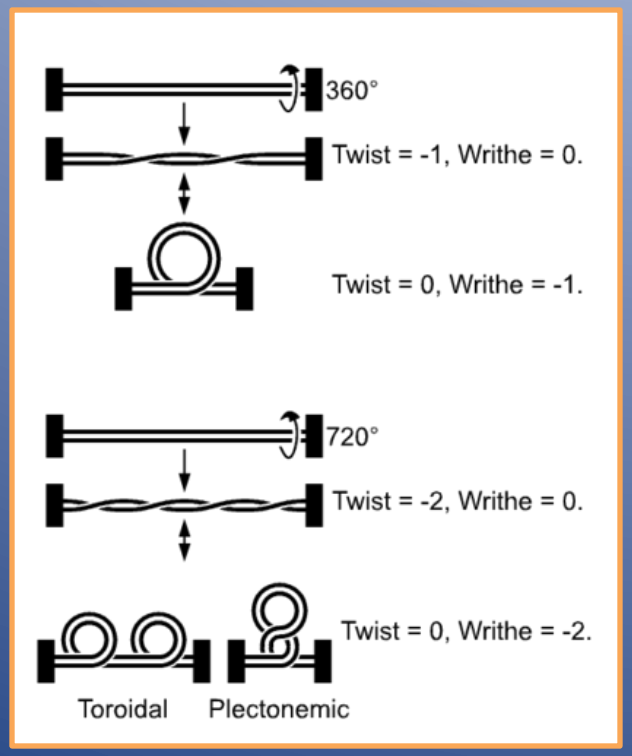
Conclusions

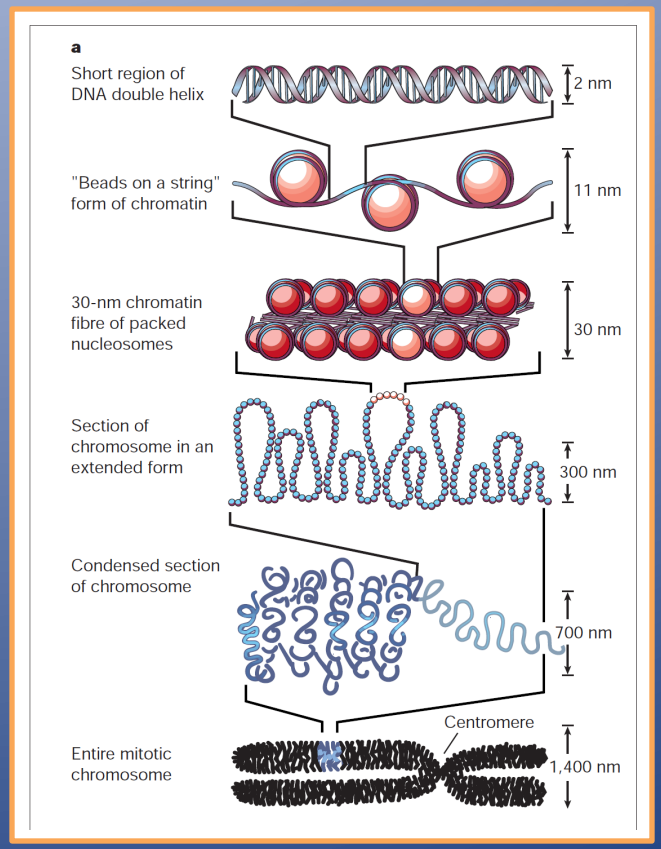
The End

DNA organization

E. Coli genome ~ 4 million bp ≈ 1.2 mm
E. Coli dimensions ~ 2 µm^3

Human genome ~ 6 billion bp ≈ 2000 mm
Nucleus dimensions ~ 750 µm^3





(Matthews, *DNA Structure Prerequisite Information*. 1997)

Sequence preference of nucleosomes

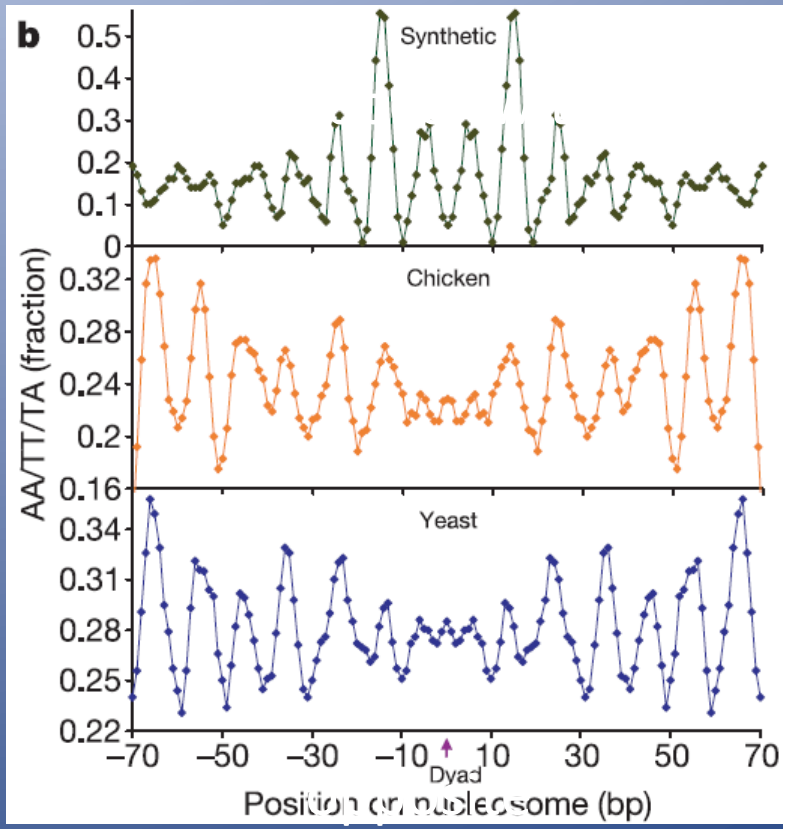
Nucleosome code

- AA/TT/TA dinucleotide repeats periodicity

-GC repeats 5bp out of phase

Due to difference in bending of dinucleotides?

- Helicity of DNA means bending at major and minor groove is



(Segal et al., Nature , 2006)

- Nucleosomes may position AA/TA/TT at minor/major, GC at major/minor groove due to difference in bending

What is a Nucleosome (how do we know)?

Digestion with

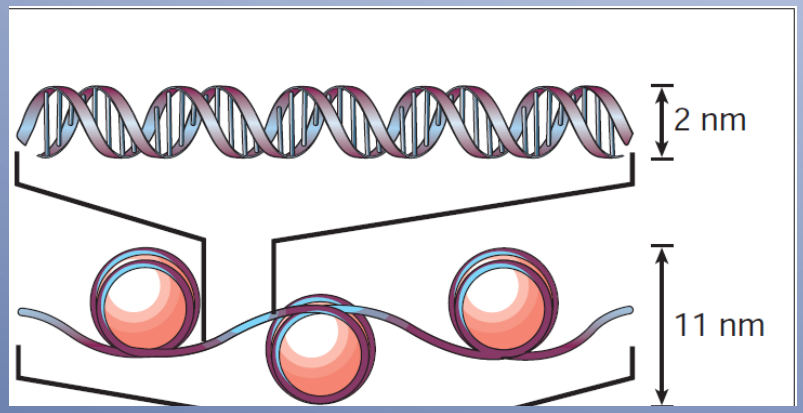
-Human genome is big, nucleus is small

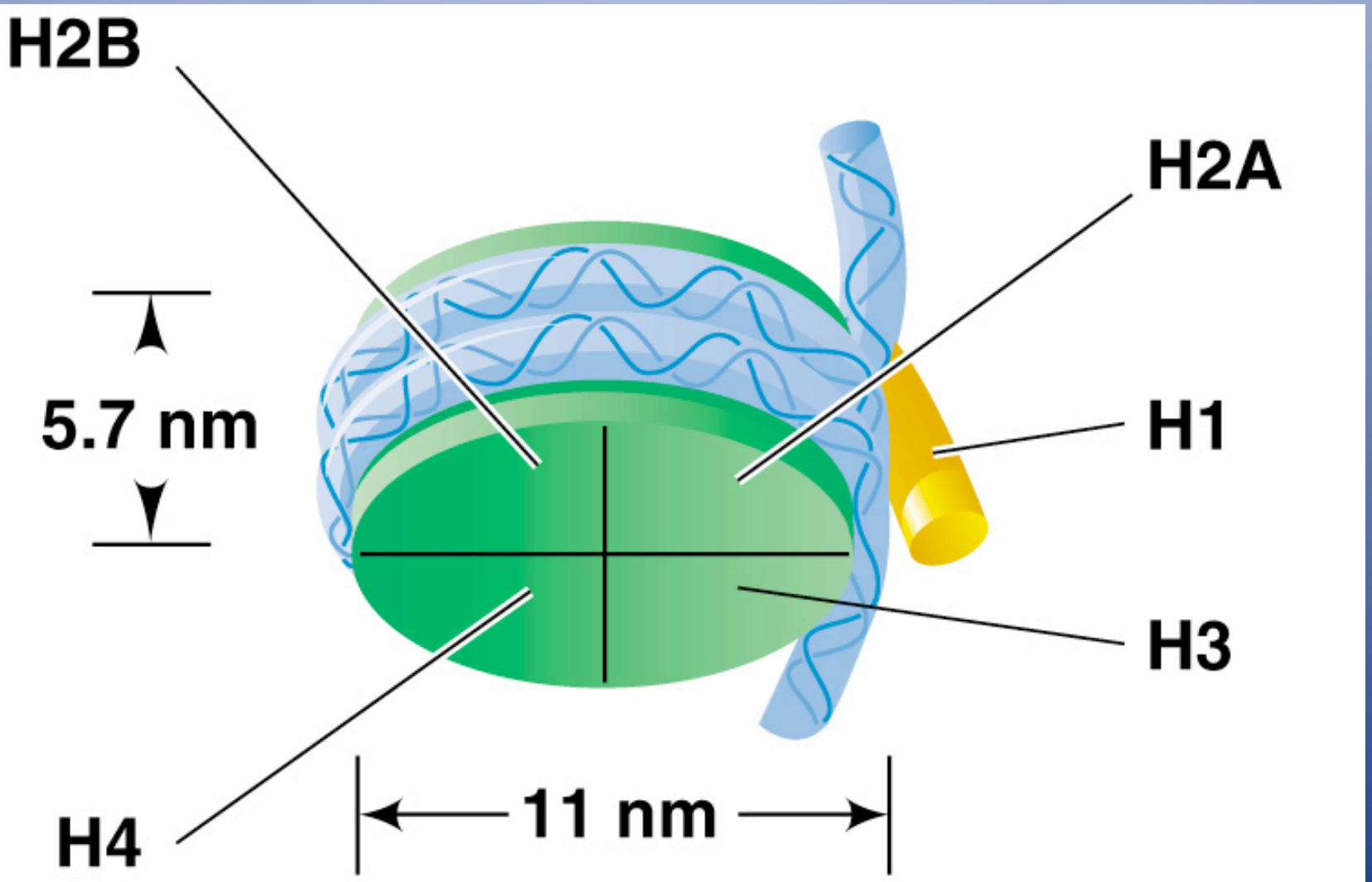
~ 2 billion basepairs ≈ 2m

nucleus radius ~ 6 µm

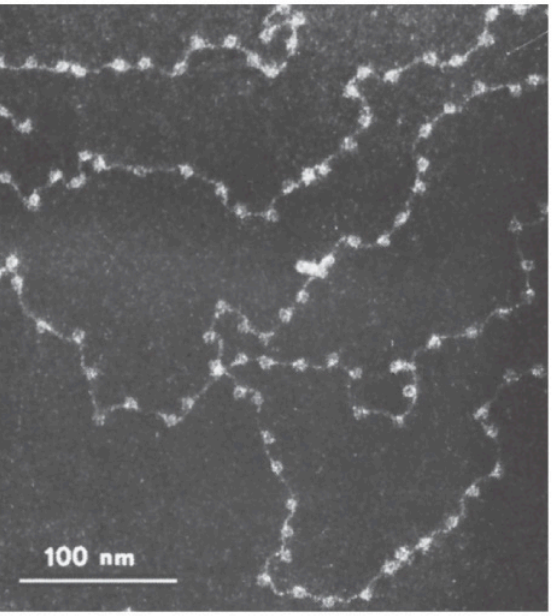
- Many different levels of organization

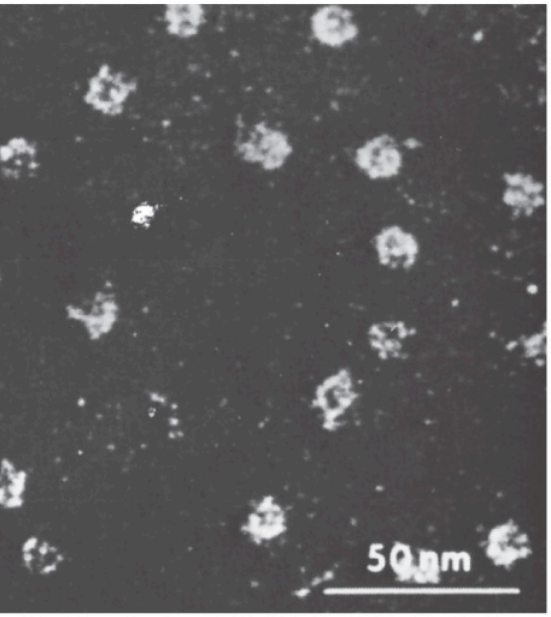
- We are interesting in Nucleosome formation


2 nm
11 nm


H2B
H2A
H1
H3
H4
5.7 nm
11 nm

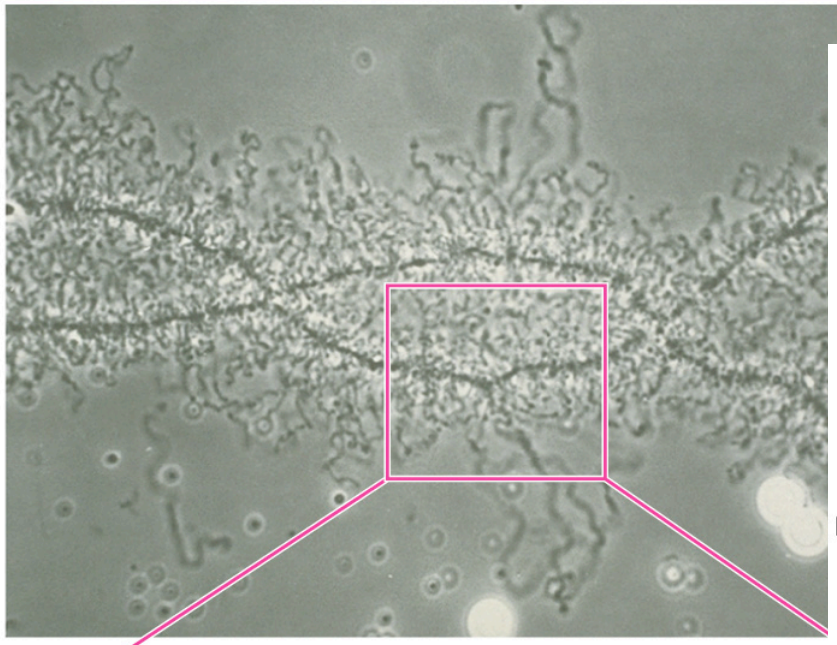
DNA o



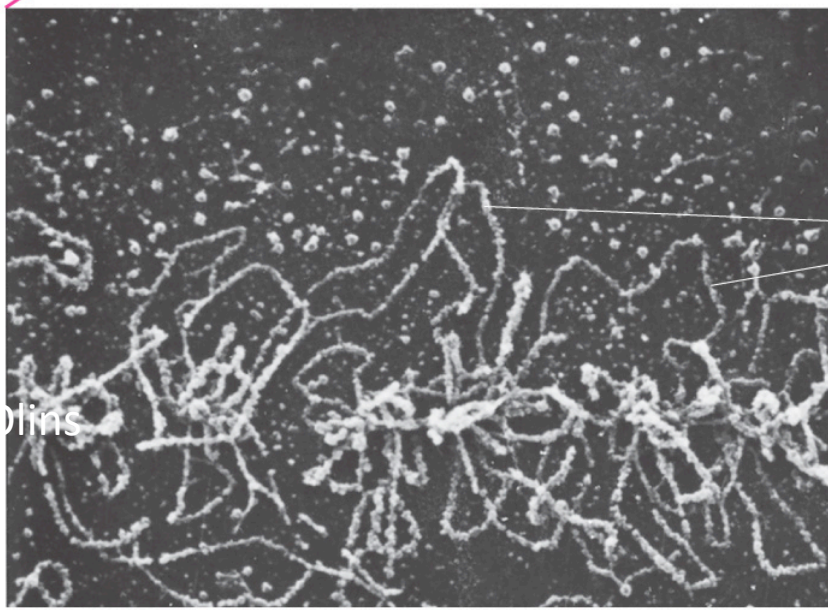


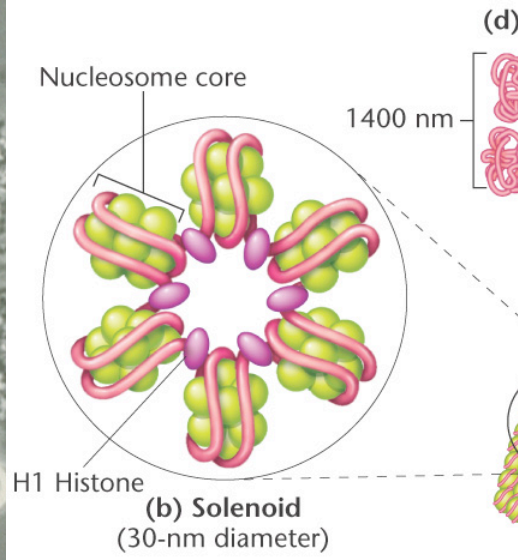
Olins and Olins

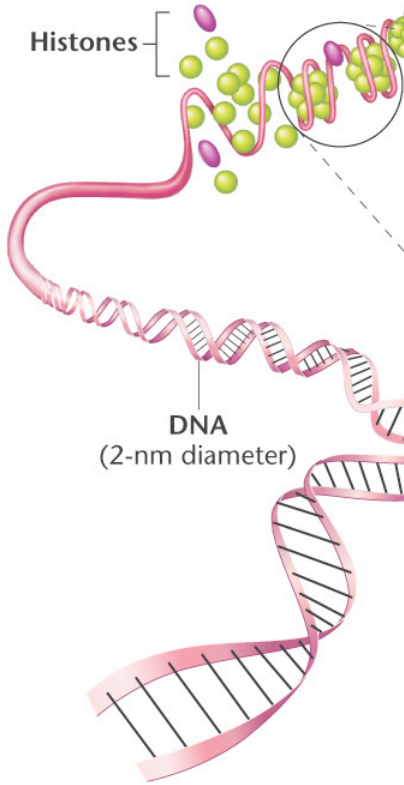
(a)



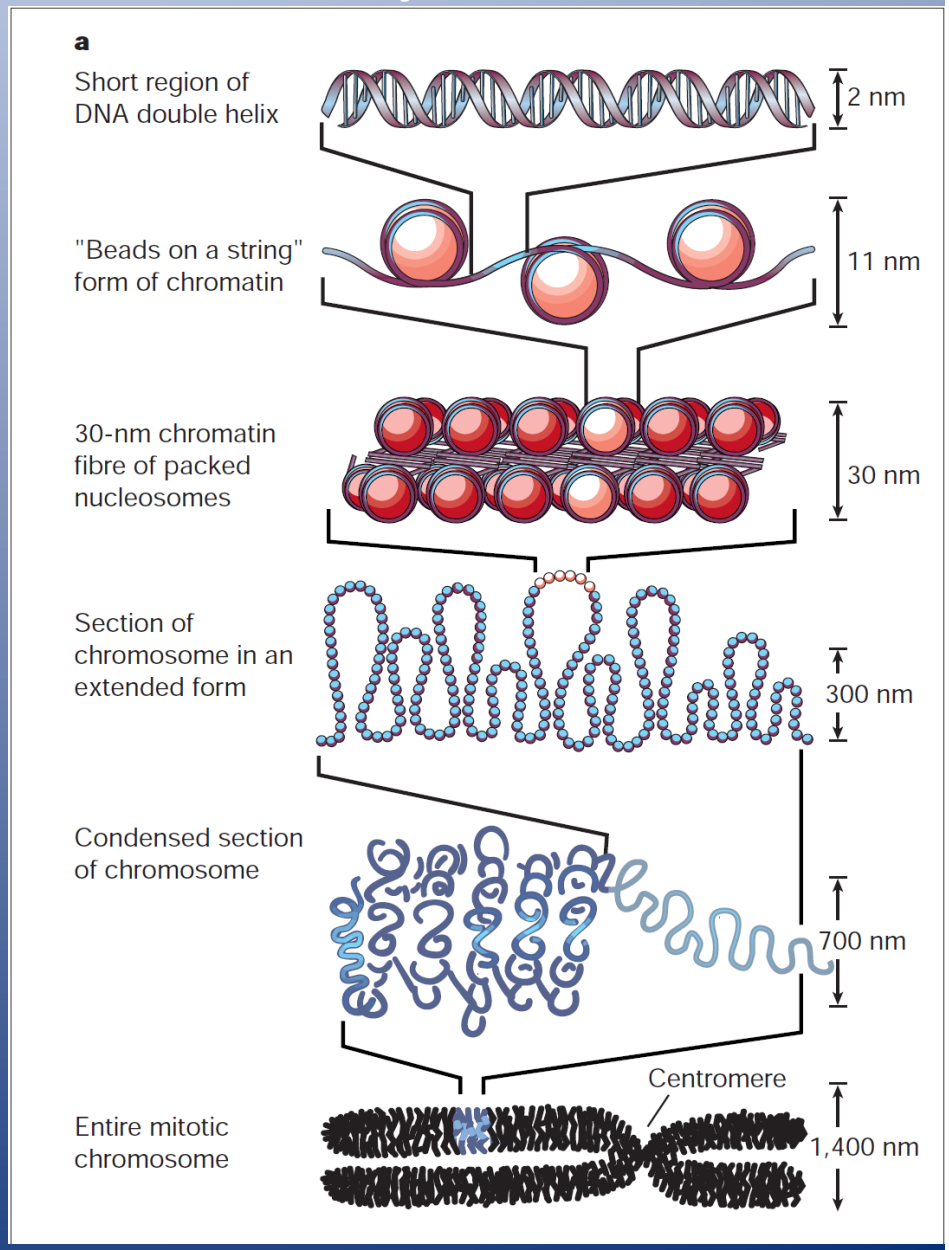
(b)







DNA organization in Eukaryotes



Topics:

-General introduction to Nucleosomes

-compaction computation?

-Protect and Seq: Widom http://www.wisdom.weizmann.ac.il/~eran/NucleosomeModel.pdf

weissman http://www.nature.com/nmeth/journal/v6/n4/full/nmeth0409-244b.html

-Lacra stuff, pulling exp

-Widom Accessibility