

Statistical mechanics of nucleosome positioning

Alexandre V. Morozov

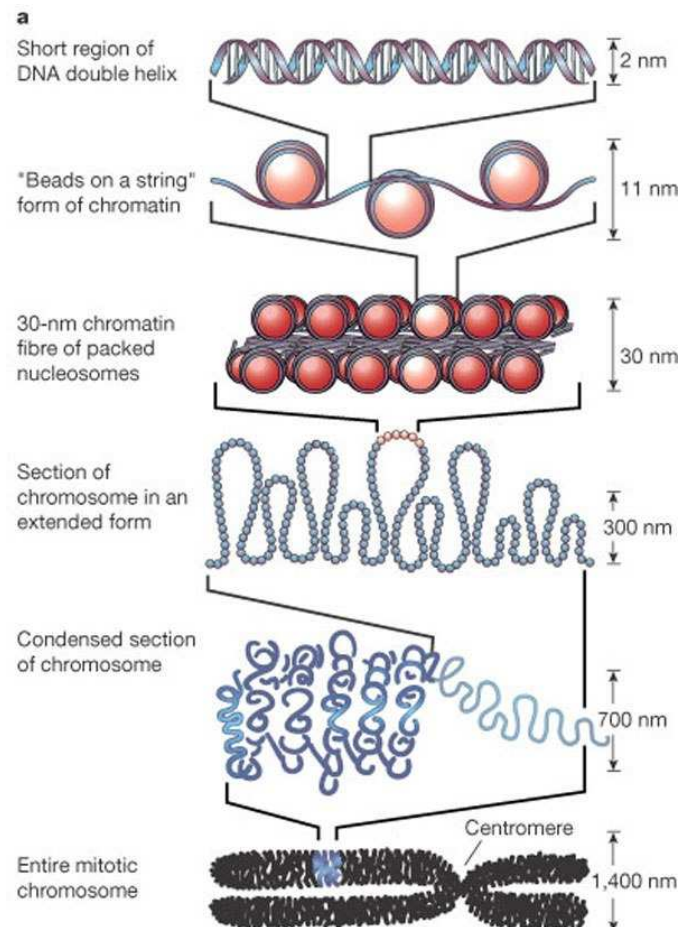
Department of Physics & Astronomy and
the BioMaPS Institute for Quantitative Biology,

Rutgers University

morozov@physics.rutgers.edu

RosettaCON, August 2009

Chromatin length scales



Chromatin under the microscope

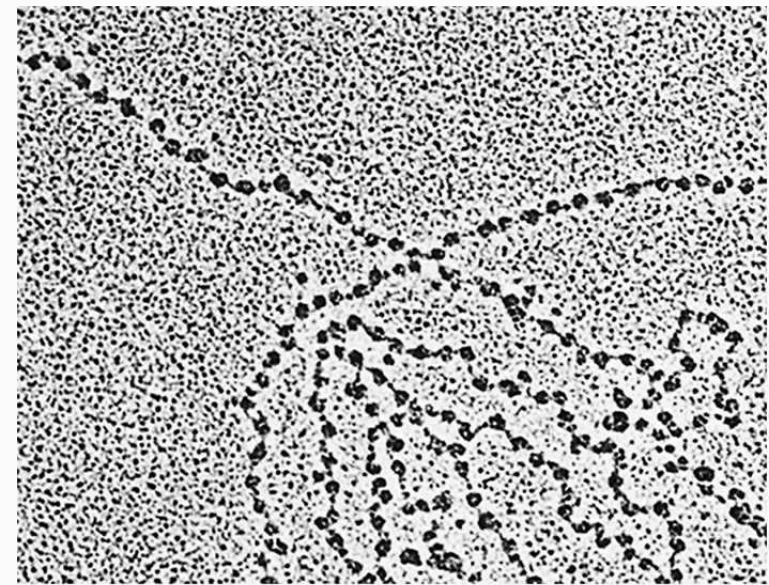
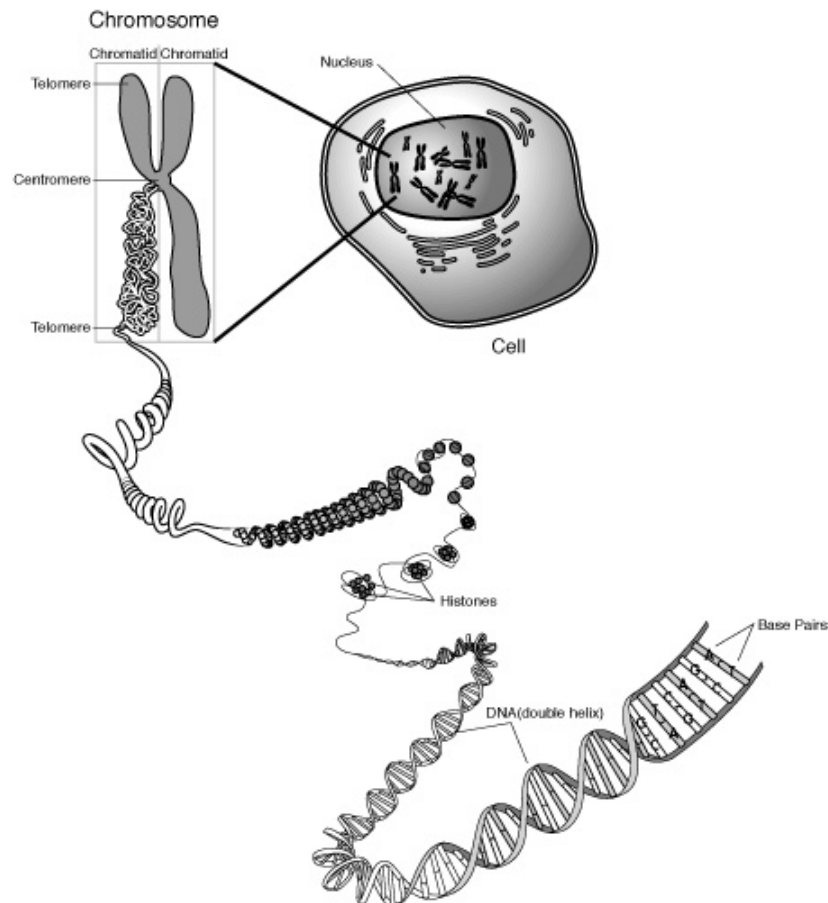
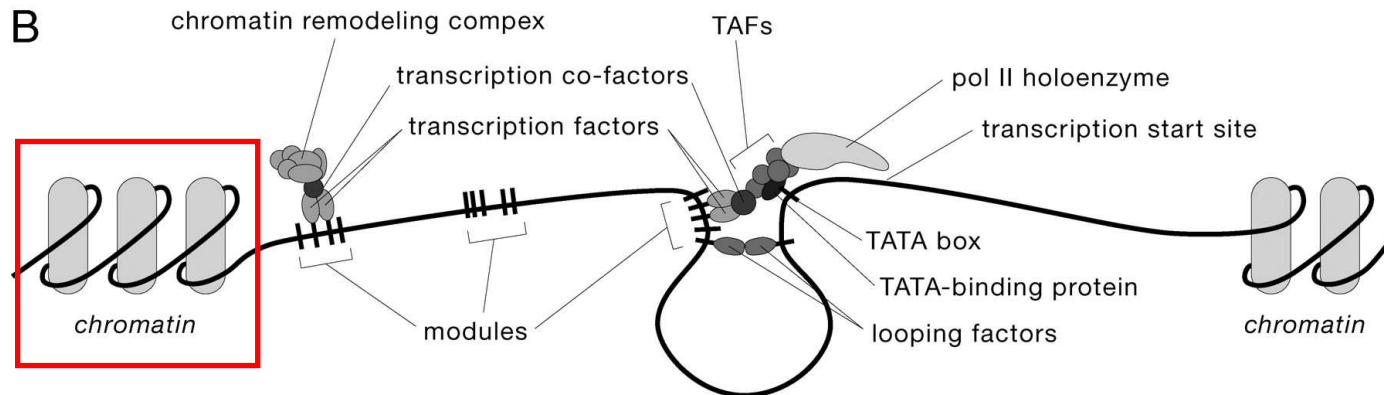
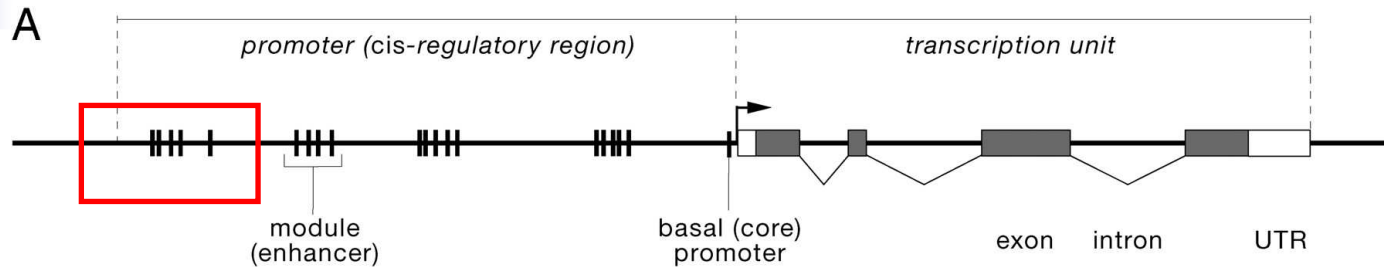


Figure 23-43. An electron micrograph of *D. Melanogaster* chromatin showing strings of closely spaced nucleosomes. [Courtesy of Oscar L. Miller, Jr. and Steven McKnight, University of Virginia.]

Electron micrograph of *D. Melanogaster* chromatin: arrays of regularly spaced nucleosomes, each ~80 Å across.

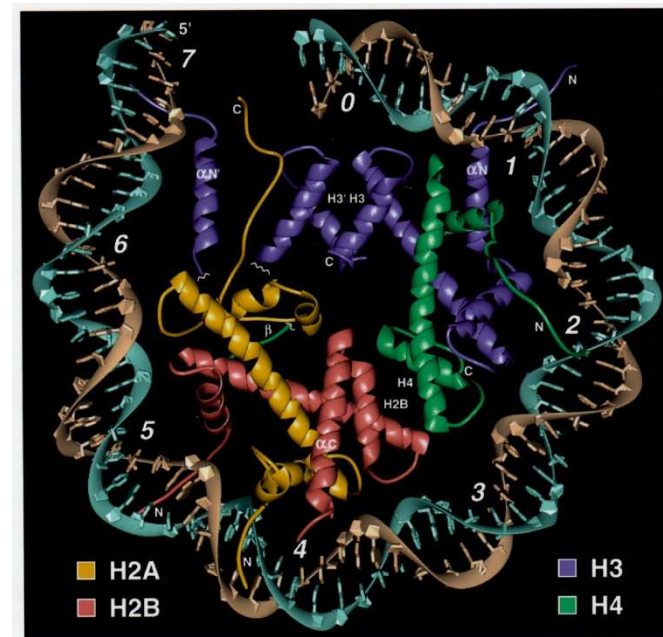
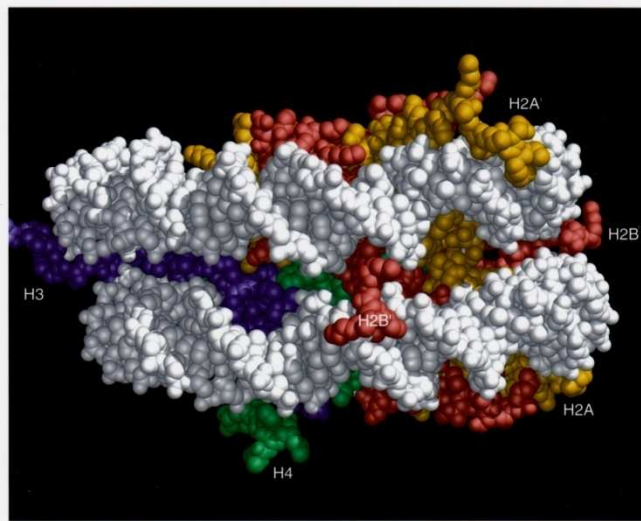
Biophysical picture of gene transcription



Wray, G. A. et al. Mol Biol Evol 2003 20:1377-1419

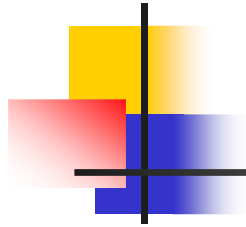
Molecular Biology
and Evolution

Nucleosome core particle



Left-handed superhelix: (1.84 turns, 147 bp, $R = 41.9 \text{ \AA}$, $P = 25.9 \text{ \AA}$)
PDB code: **1kx5**

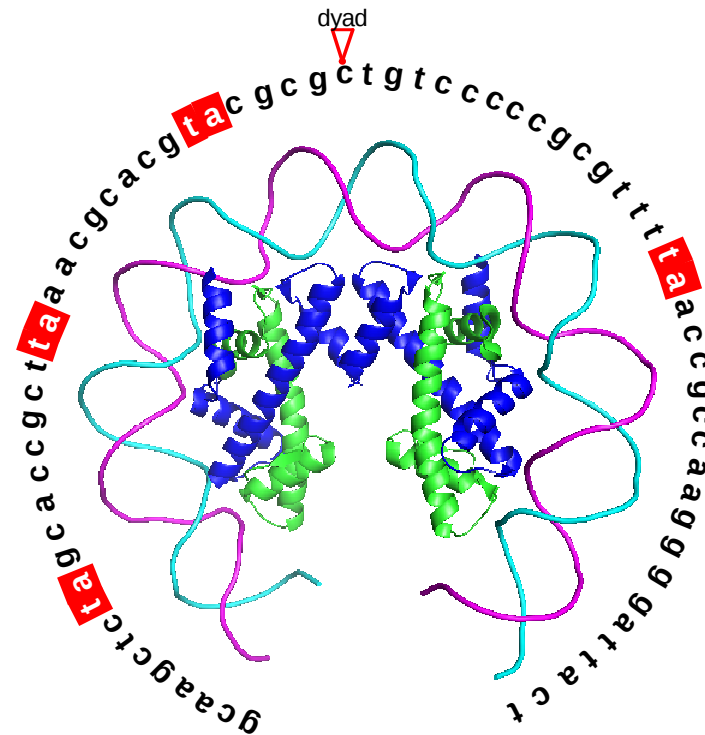
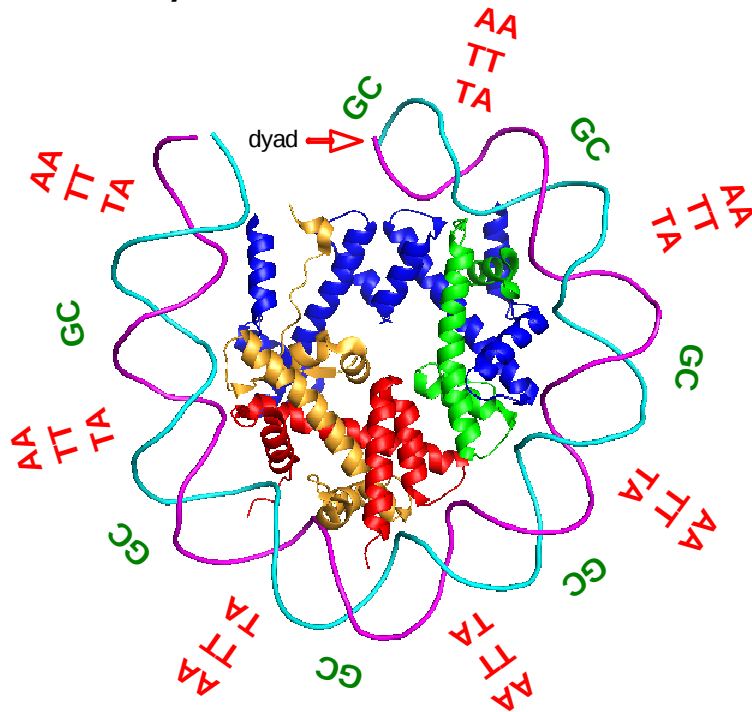
T.J.Richmond: K.Luger et al. *Nature* 1997 (2.8 $\text{\AA}$); T.J.Richmond & C.A.Davey *Nature* 2003 (1.9 $\text{\AA}$)



What factors determine
nucleosome positioning on
genomic DNA?

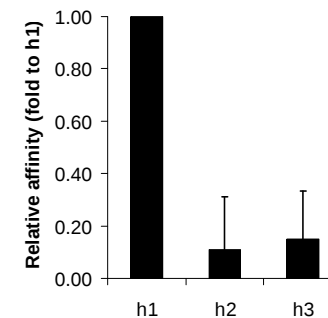
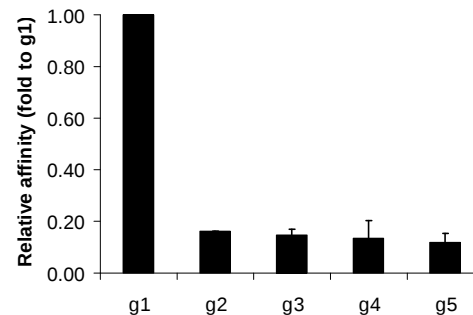
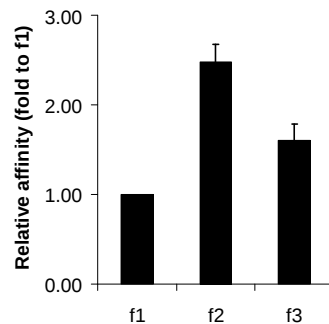
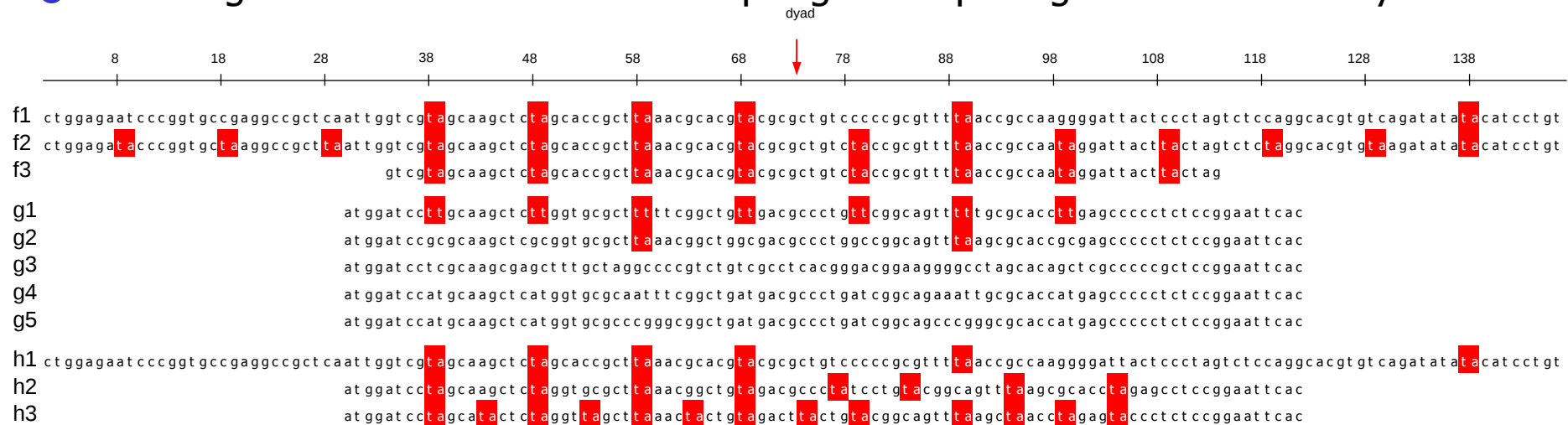
DNA elastic energy explains histone-DNA interactions

- Nucleosome affinity depends on the presence and spacing of key dinucleotide motifs (e.g. TA,CA)
- Nucleosome affinity can be explained by sequence-dependent DNA flexibility



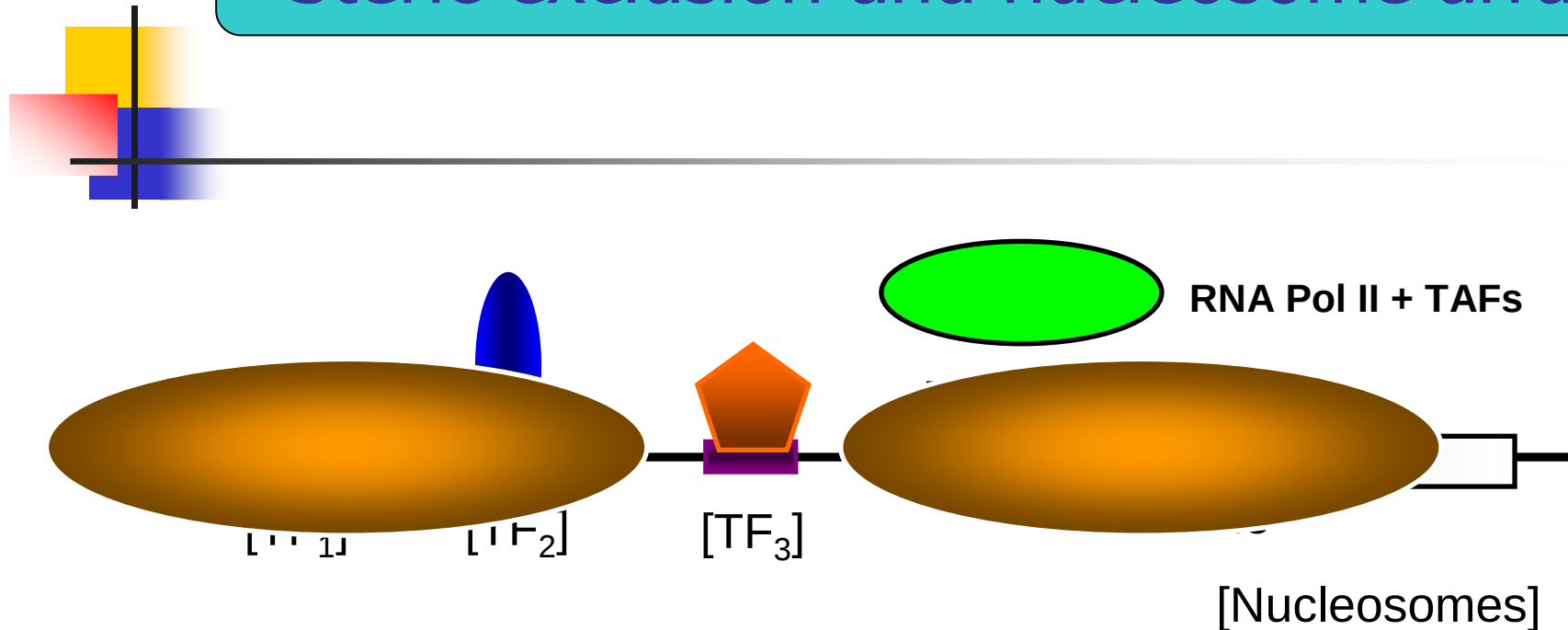
Nucleosome formation is sequence-dependent

- Adding key dinucleotide motifs increases nucleosome affinity
- Deleting dinucleotide motifs or disrupting their spacing decreases affinity



Segal et al., Nature 2006

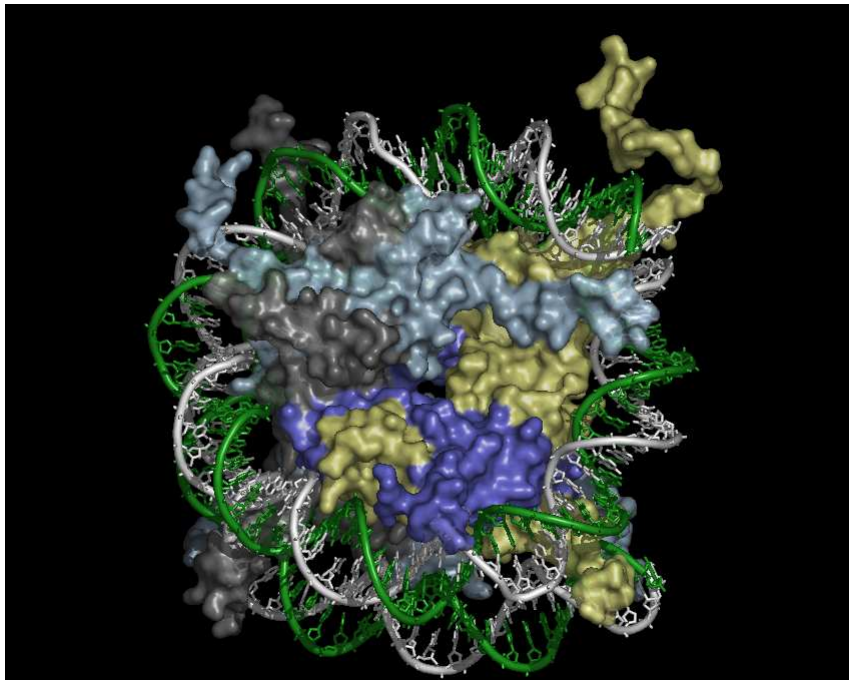
Steric exclusion and nucleosome arrays



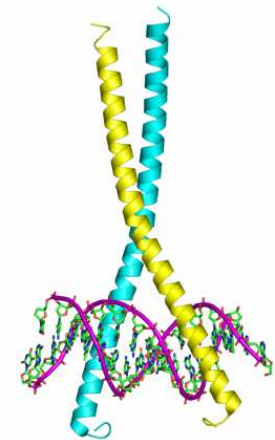
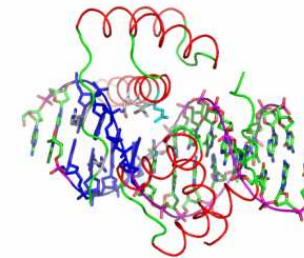
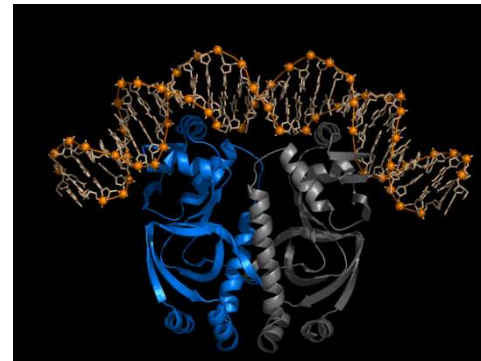
- 📖 Nucleosomes interfere with DNA function
- 📖 Nucleosomes form an approx. 1D liquid of particles of size $L=147$ bp (~80% of yeast genome is nucleosome-covered) – regular arrays are created simply by steric exclusion

Nucleosomes affect transcription factor binding to DNA

Thermodynamic competition for binding sites:



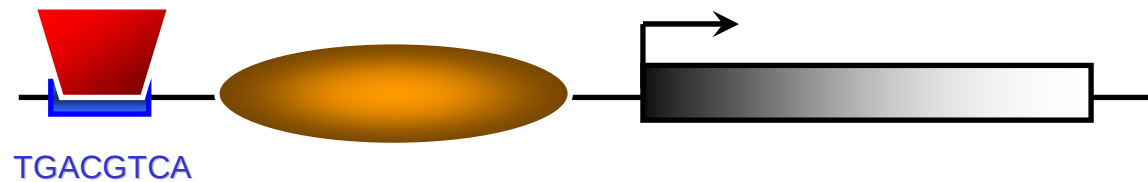
Nucleosome
e



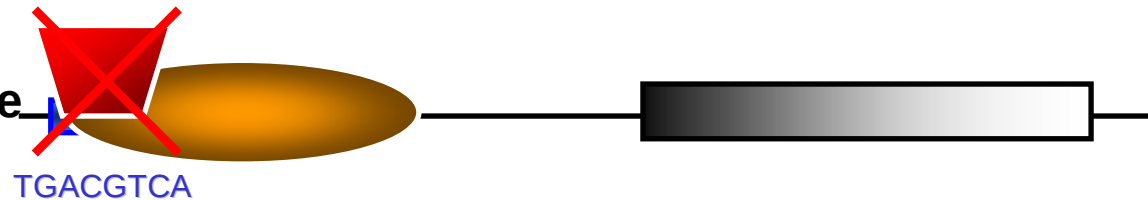
Transcription
factors

Nucleosome occupancy is affected by other DNA-binding factors

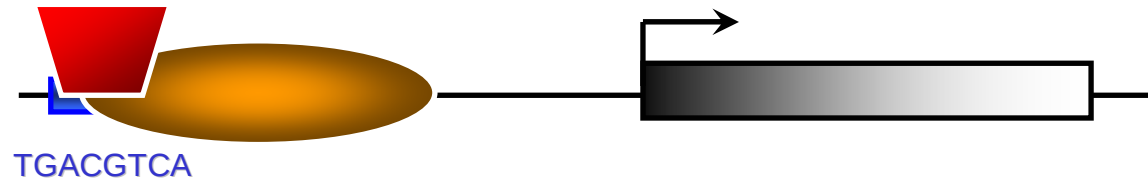
Nucleosome-free site



Nucleosome-occluded site



Nucleosome is displaced through TF binding



Chromatin-remodeling enzymes and post-translational modifications

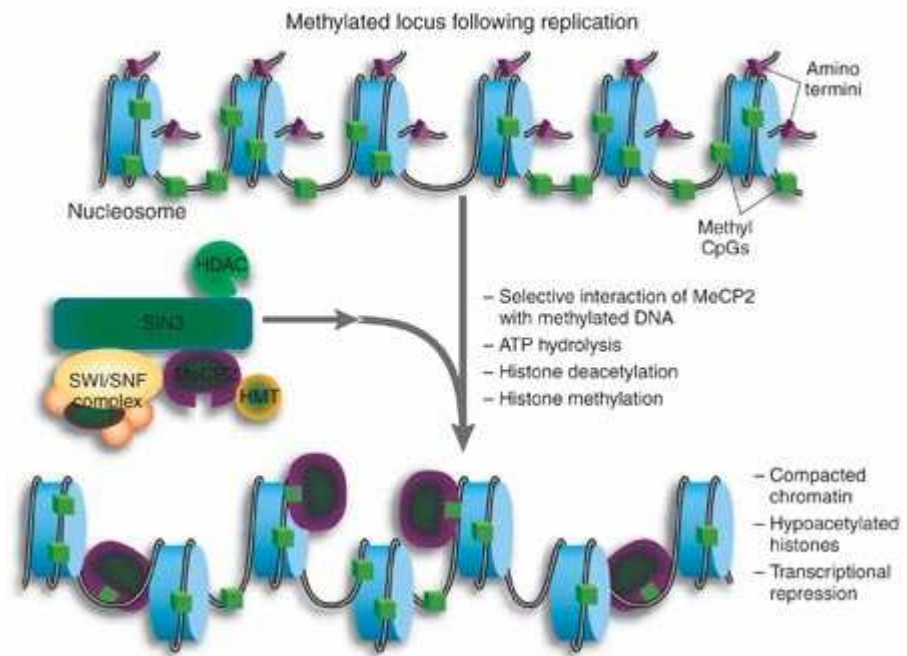
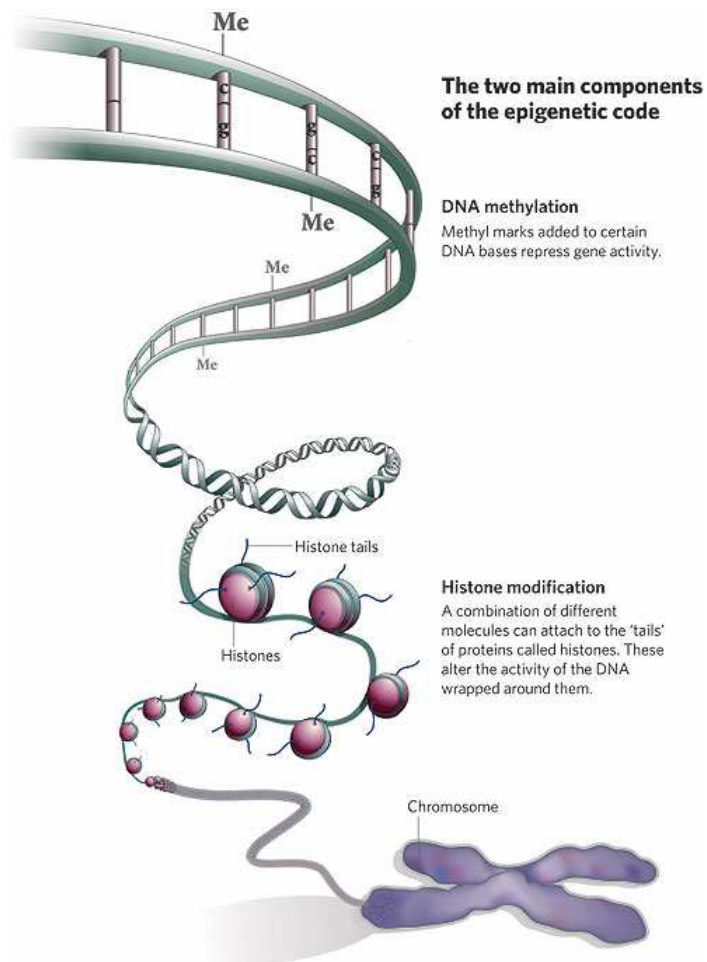
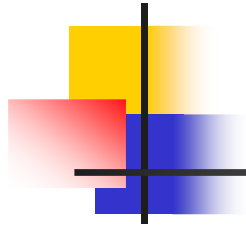
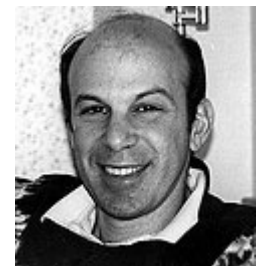


Figure 1. One potential scenario by which MeCP2 and the SWI/SNF complex might cooperate in establishment or maintenance of transcriptional repression at a methylated locus.

Paul A. Wade, *Nature Genetics* **37**, 212 - 213 (2005)



Inferring nucleosome positioning and energetics from high-throughput *in vitro* datasets



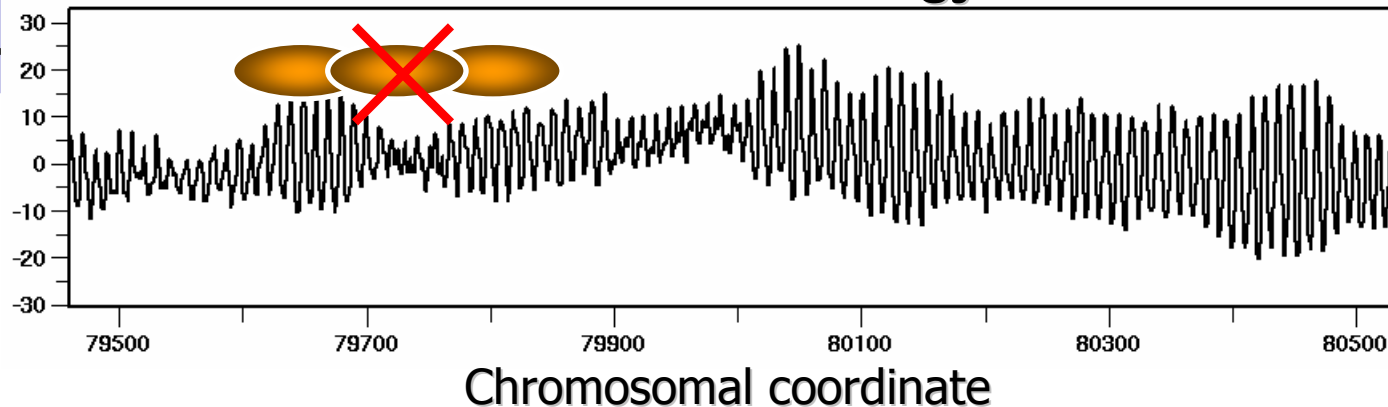
Kevin Struhl, Harvard

From nucleosome energies to probabilities and occupancies: 1D liquid of core particles



Nucleosome energy

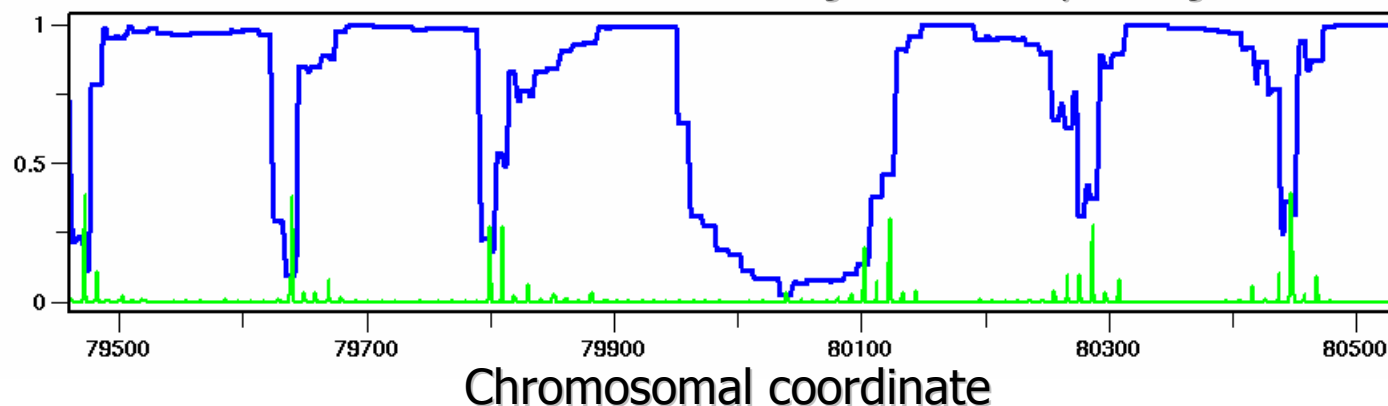
Morozov et al., Nucl. Acids Res. 2009



Use a **recursive algorithm** to find the grand canonical partition function and thus nucleosome probabilities and occupancies:

$$P_i = \frac{Z_{i-1}^f e^{-E(i)} Z_{i+L}^r}{Z}$$

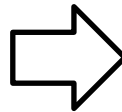
Nucleosome Probability & Occupancy



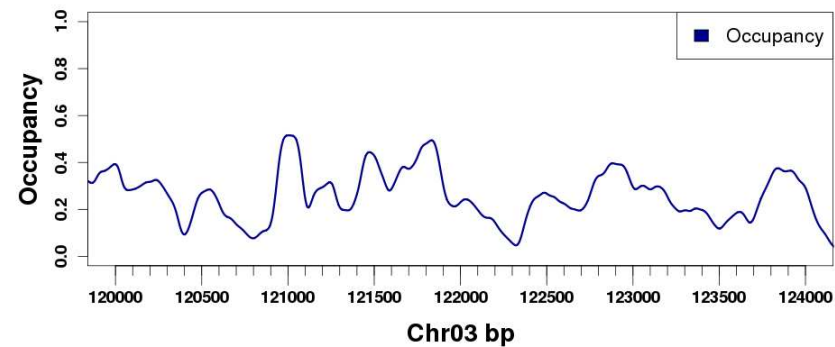
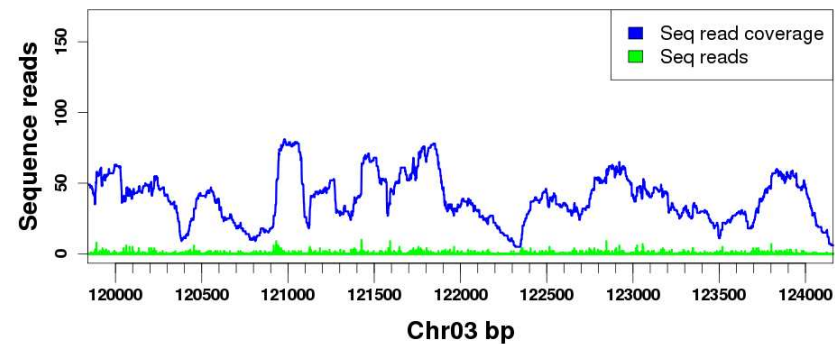
Sequence read profiles for nucleosomes positioned *in vitro* by salt dialysis



Solexa/Illumina sequencer

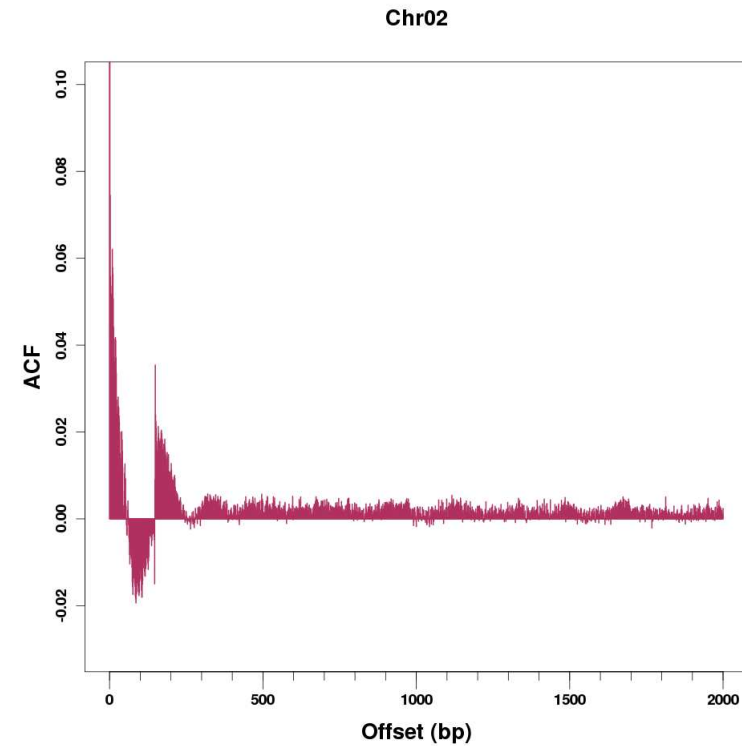
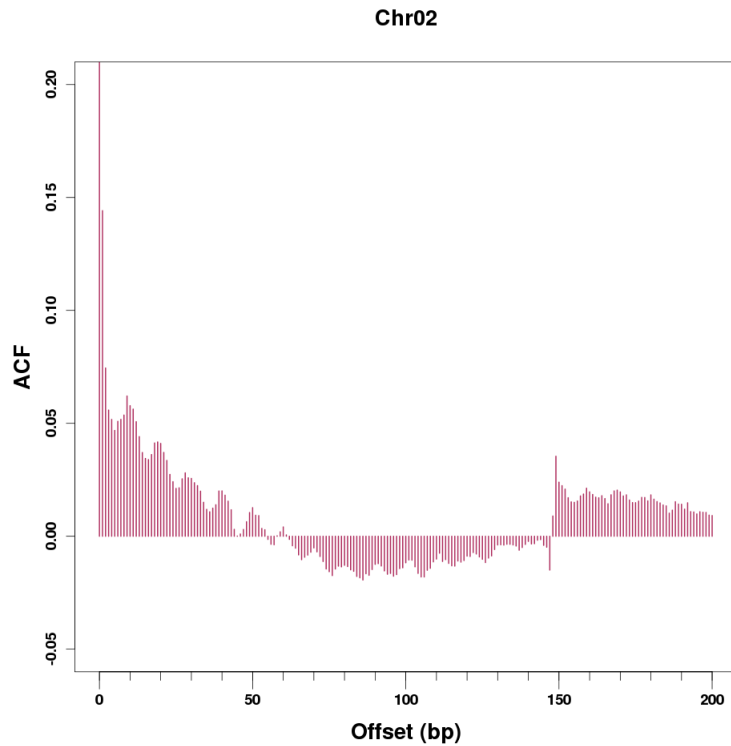


S.cerevisiae

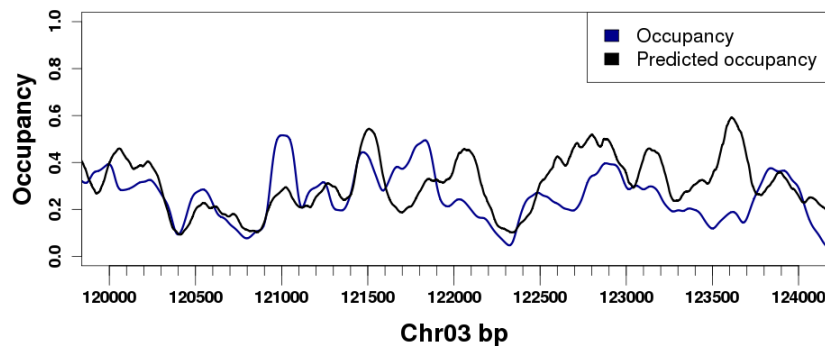
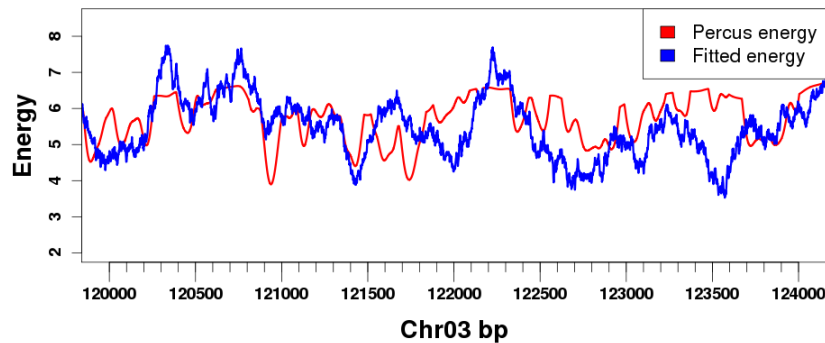


Translational and rotational positioning of *in vitro* nucleosomes

Auto-correlation function of sequence read profiles



Rigorous derivation of nucleosome energies from occupancies



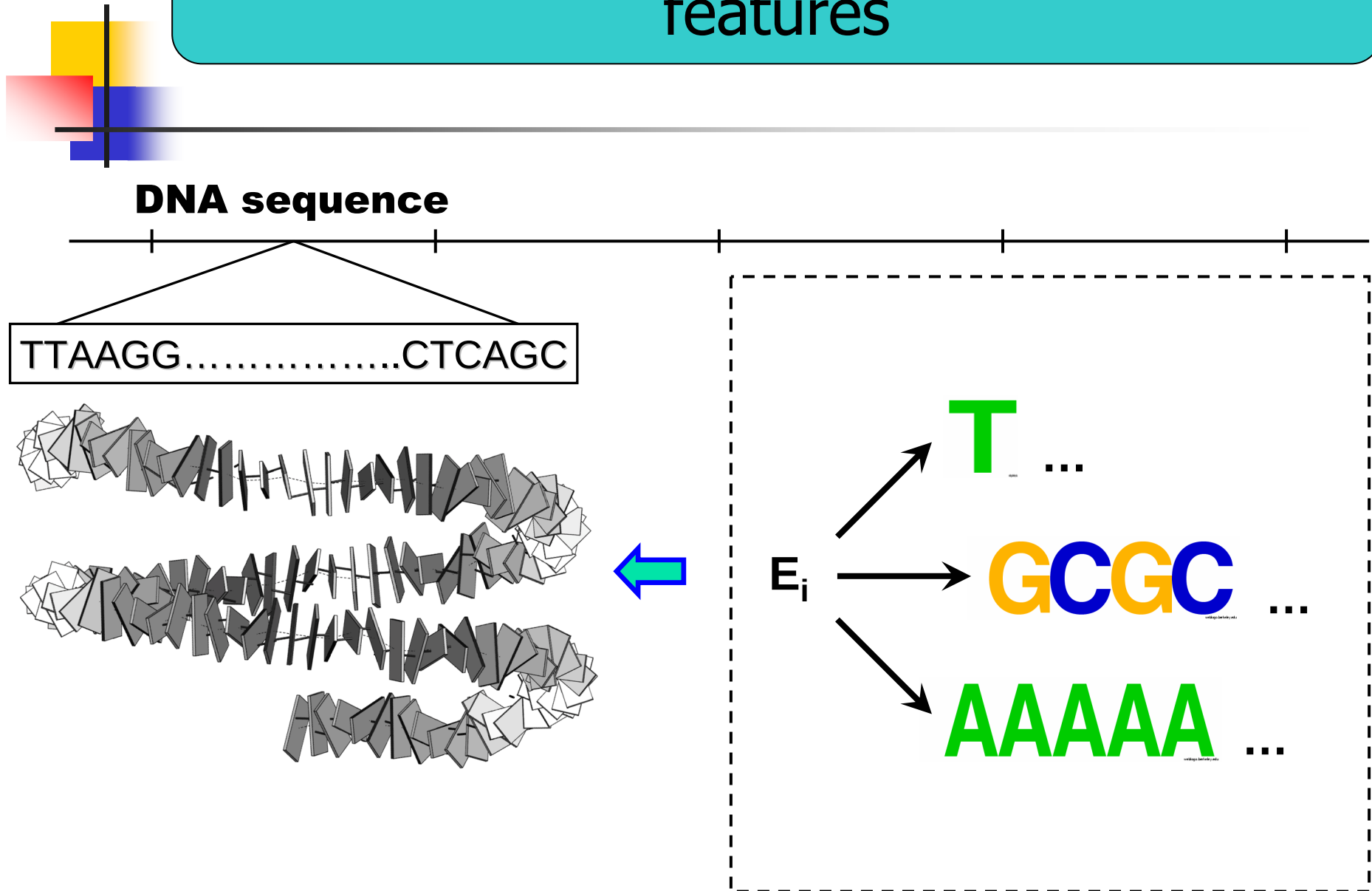
- Infer nucleosome energies directly from the occupancy profiles:

$$E_i - \mu = \log \frac{P_i + P_i^L}{P_i} + \sum_{j=i}^{i+L-1} \log \frac{P_j^L}{P_j + P_j^L}$$

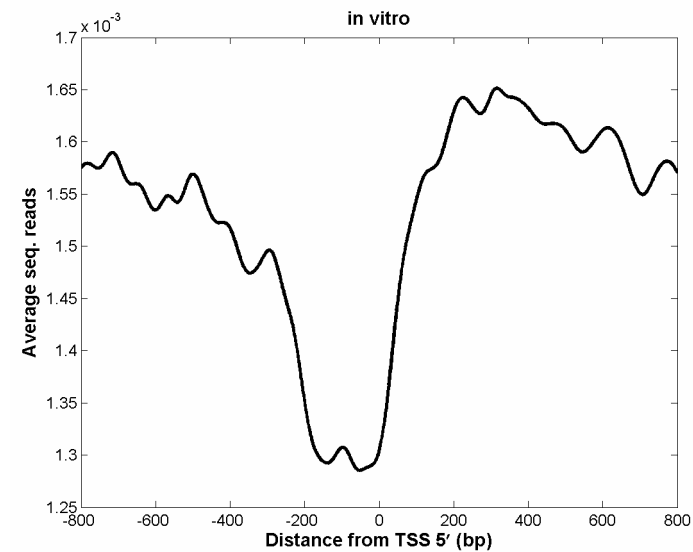
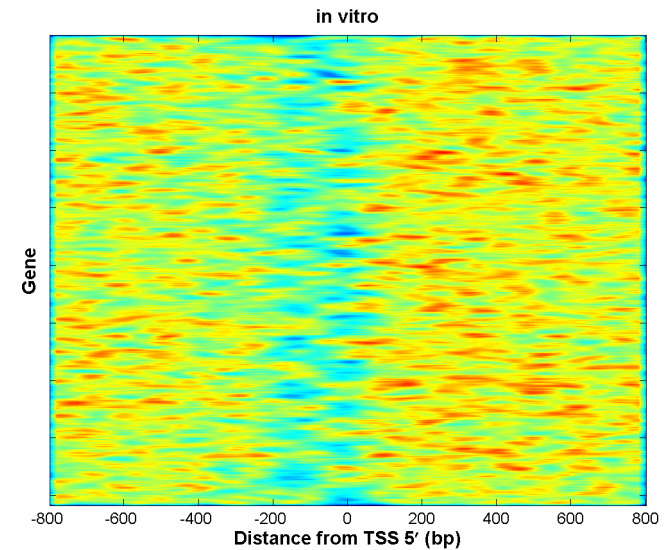
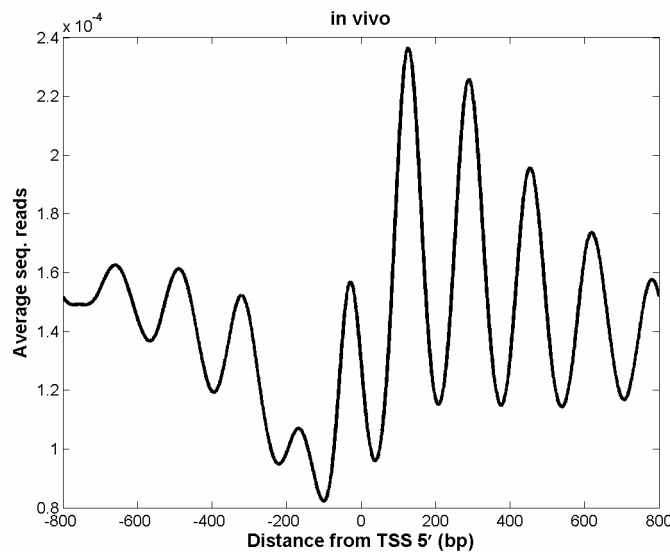
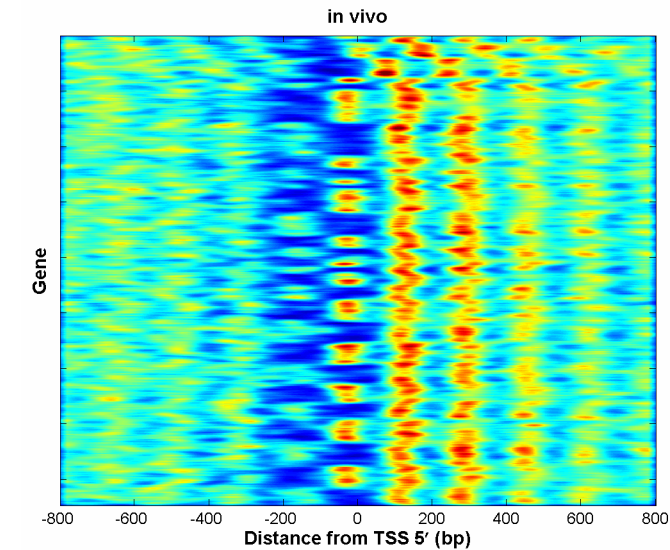
- Explain the energies in terms of sequence features through a linear model fit to the energies of individual words:

$$E_i - \mu = \sum_{j=\{\text{words}\}} n_{\text{words}}^j \epsilon_{\text{words}}^j$$

Fitting nucleosome energies to sequence features



Nucleosomes are not well-positioned with respect to TSS *in vitro* ...



Nonetheless we can extract sequence-specific positioning signals

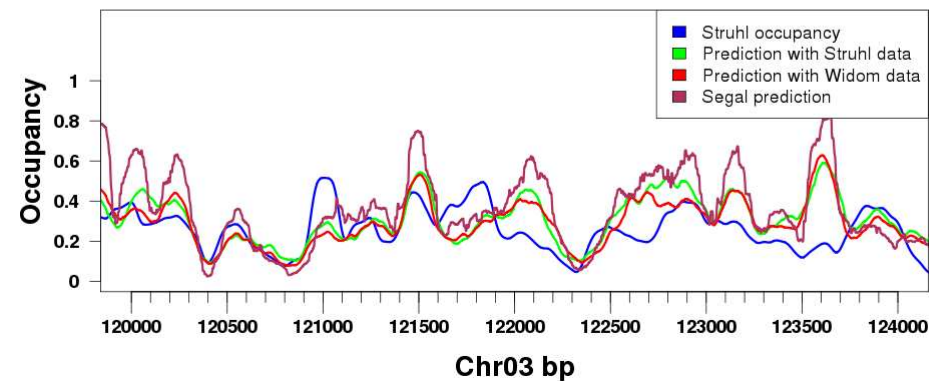
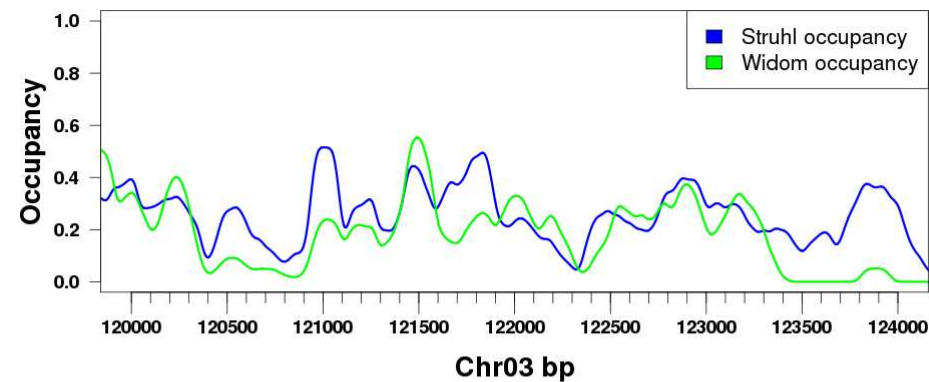
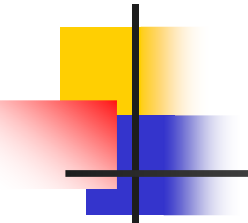


Table of correlation coefficients between various experimental & theoretical occupancy profiles:

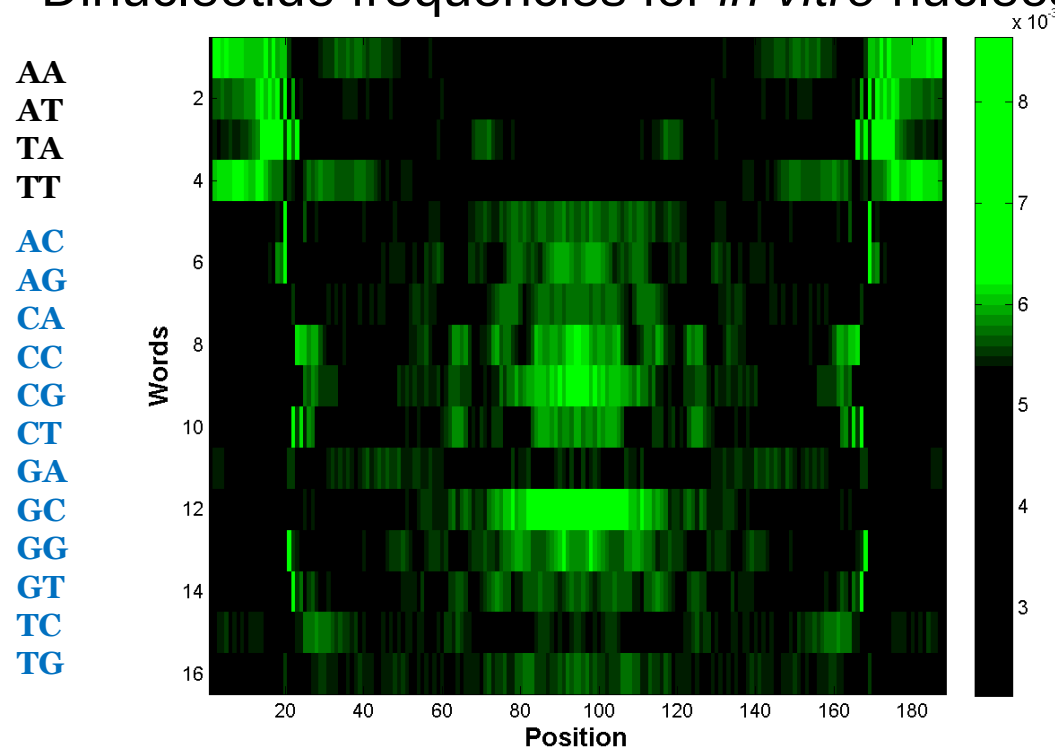


	A	B	C	D	E	F	G	H	I
A	1.0000	0.6542	0.4446	0.5903	0.2558	0.3607	0.4522	0.5993	0.3771
B	0.6542	1.0000	0.5705	0.7695	0.2400	0.4781	0.5801	0.7437	0.4489
C	0.4446	0.5705	1.0000	0.5354	0.1866	0.7879	0.9914	0.5364	0.6045
D	0.5903	0.7695	0.5354	1.0000	0.2681	0.3827	0.5614	0.8826	0.4972
E	0.2558	0.2400	0.1866	0.2681	1.0000	0.1817	0.1937	0.3141	0.2022
F	0.3607	0.4781	0.7879	0.3827	0.1817	1.0000	0.7259	0.4339	0.5073
G	0.4522	0.5801	0.9914	0.5614	0.1937	0.7259	1.0000	0.5532	0.6182
H	0.5993	0.7437	0.5364	0.8826	0.3141	0.4339	0.5532	1.0000	0.5004
I	0.3771	0.4489	0.6045	0.4972	0.2022	0.5073	0.6182	0.5004	1.0000

- A = Struhl's *in vitro* data
- B = Widom's *in vitro* data
- C = Widom's *in vivo* data (combined YPD)
- D = Segal's model
- E = Struhl's MNase data
- F = Widom's *in vivo* data (CL, YPD)
- G = Widom's *in vivo* data (no CL, YPD)
- H = Percus energy model
- I = Widom's 454 *in vivo* data (YPD)

Sequence determinants of nucleosome positioning and energetics

Dinucleotide frequencies for *in vitro* nucleosomes



People & Funding

Lab members:

- **Dr. Denis Tolkunov**
- Allan Haldane
- **George Locke**
- Julia Tsitron



Funding:

- NIH R01 HG004708
- Alfred P. Sloan Fellowship
- BioMaPS Institute for Quantitative Biology



ALFRED P. SLOAN
FOUNDATION



BioMaPS

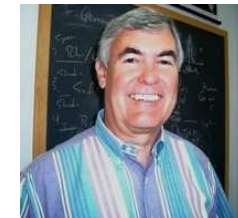
- ▣ Eric Siggia
(Rockefeller)



- ▣ Jon Widom
(Northwestern)



- ▣ James R. Broach
Karl Zawadzki
(Princeton)



- ▣ Kevin Struhl
(Harvard)



Inter-nucleosomal interactions and “quantized” linker lengths

