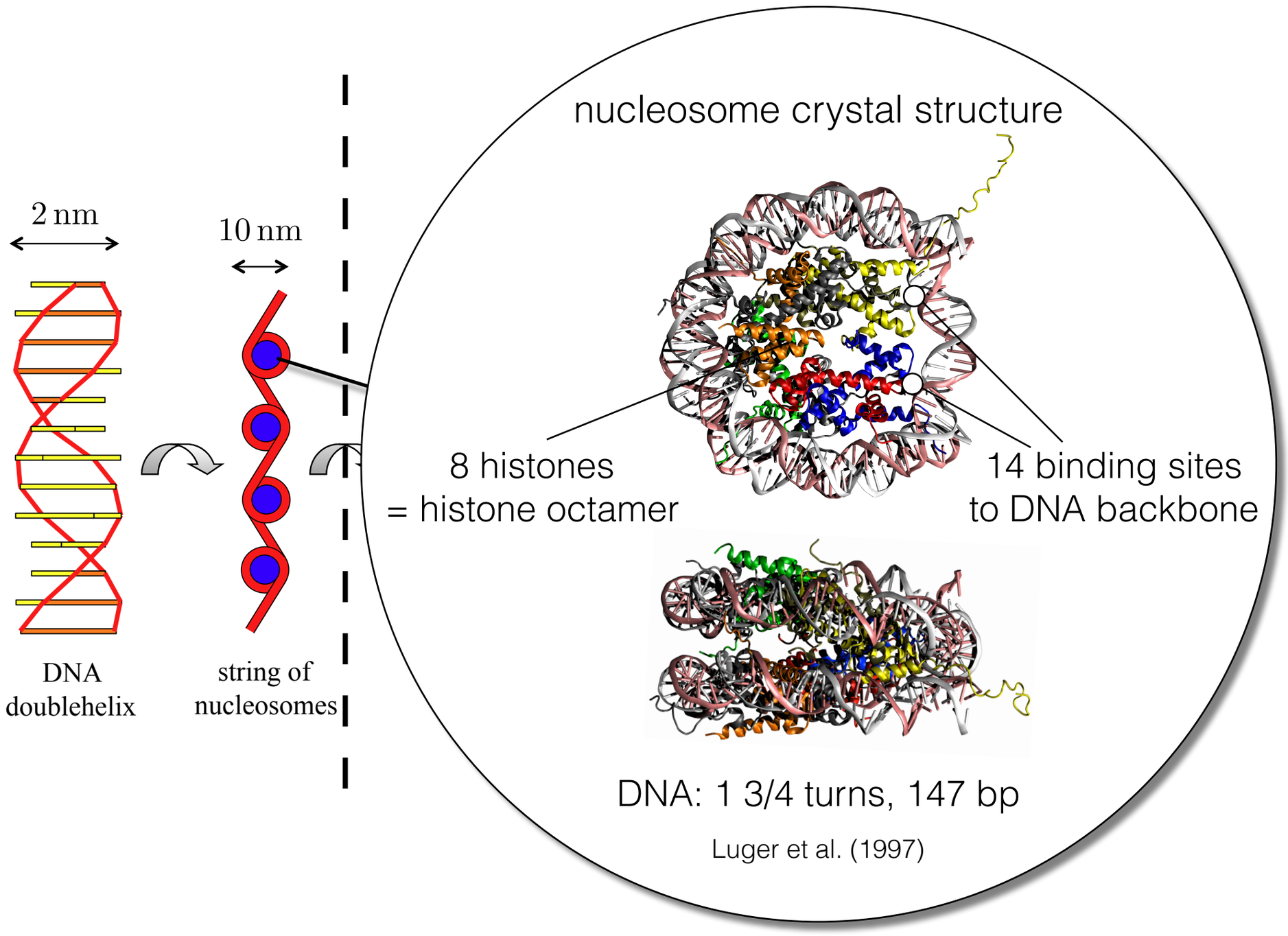
Demultiplexing information written along DNA molecules

Helmut Schiessel
Lorentz Institute
Leiden University
The Netherlands

GRUNDIG
FERRIT-ANTENNE
EIN
LW
MW
KW
UKW
SENDEREINSTELLUNG
WUNSCHKLANG
AUS
TA
LW
REGISTER
MW
KW
UKW
SPRACHE
WUNSCHKLANG
MUSIK

Chromatin



from: H. Schiessel Biophysics for Beginners (Pan Stanford Publishing 2014)

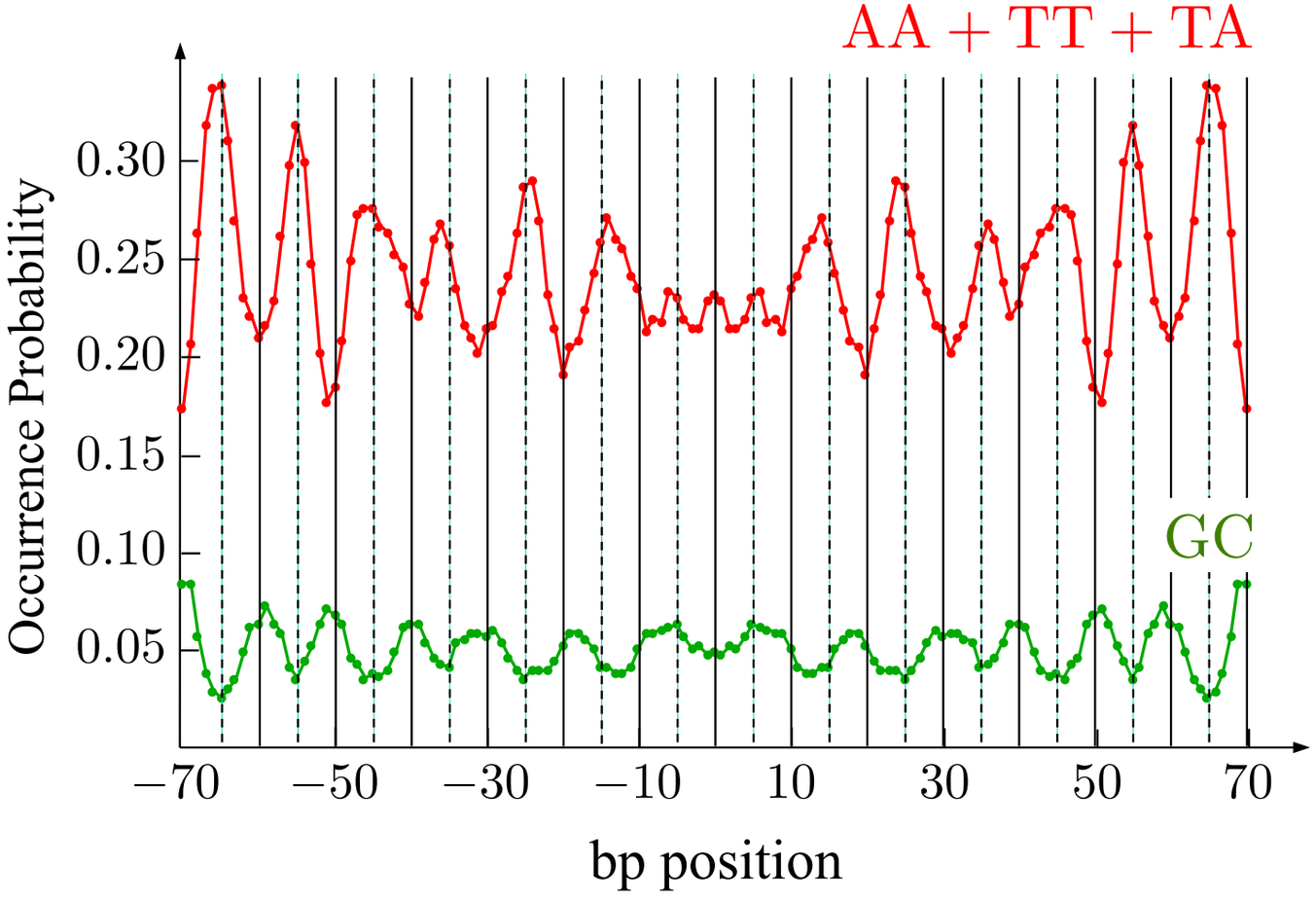
nature Vol 442|17 August 2006|doi:10.1038/nature04979

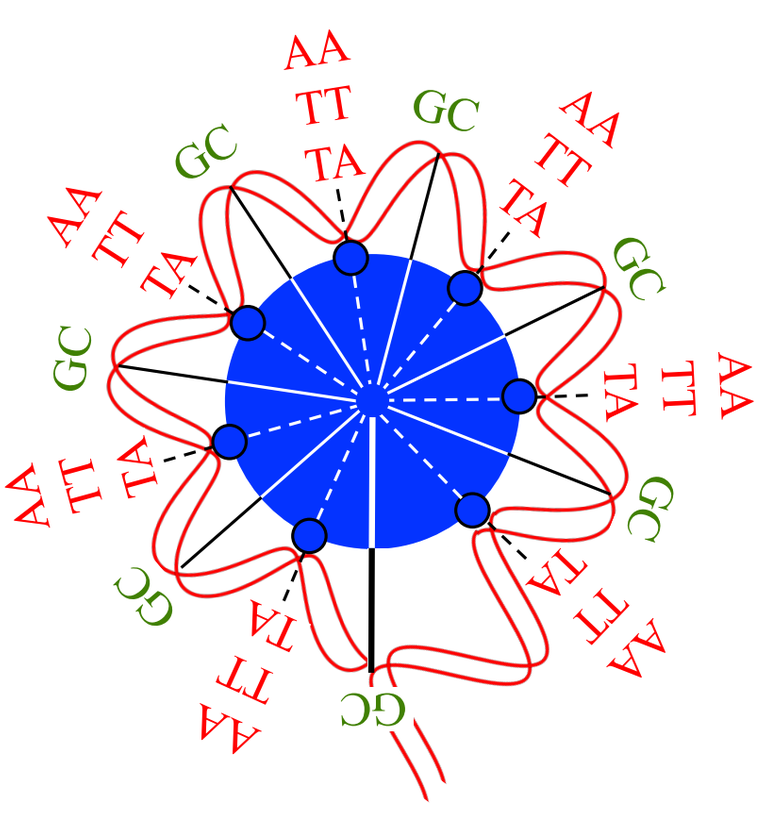
ARTICLES

A genomic code for nucleosome positioning

Eran Segal[1], Yvonne Fondufe-Mittendorf[2], Lingyi Chen[2], AnnChristine Thåström[2], Yair Field[1], Irene K. Moore[2], Ji-Ping Z. Wang[3] & Jonathan Widom[2]

Jonathan Widom
1955-2011





The nucleosomal sequence space

4^{147}

sequences can be wrapped around a nucleosome

How to explore that space?

1. Produce nucleosome maps on **genomic DNA**

   disadvantage: only small fraction of sequence space (e.g. yeast: 10^{-80})

2. Find high affinity sequences in **random pool** (Lowary & Widom 1998)

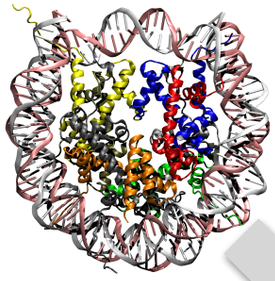
   disadvantage: experimentalists work mostly with one sequence only (‘601’)

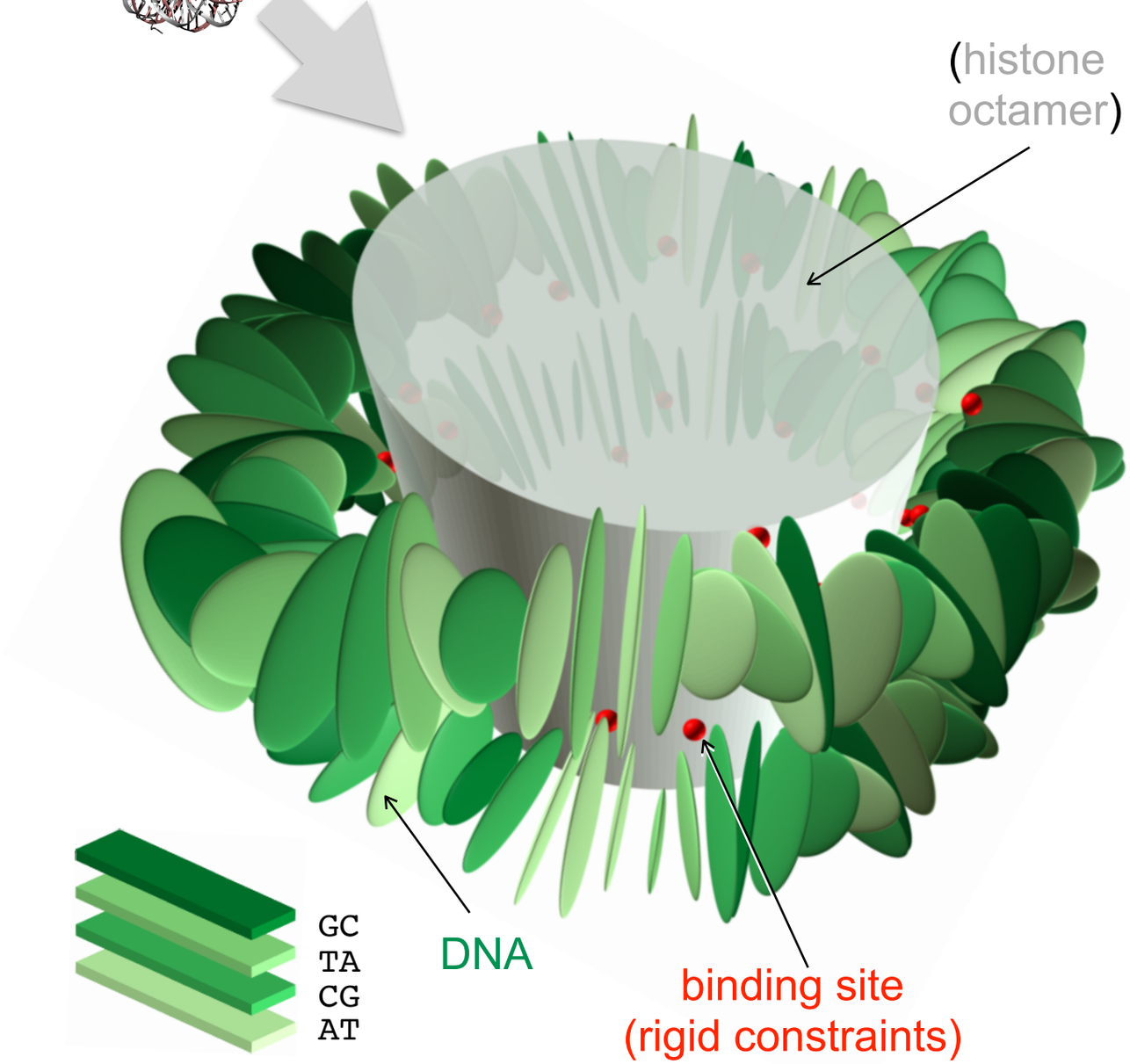
3. **Mutation Monte Carlo simulation** (Eslami-Mossallam et al., 2016)

   disadvantage: *in silico*, only as good as nucleosome computer model

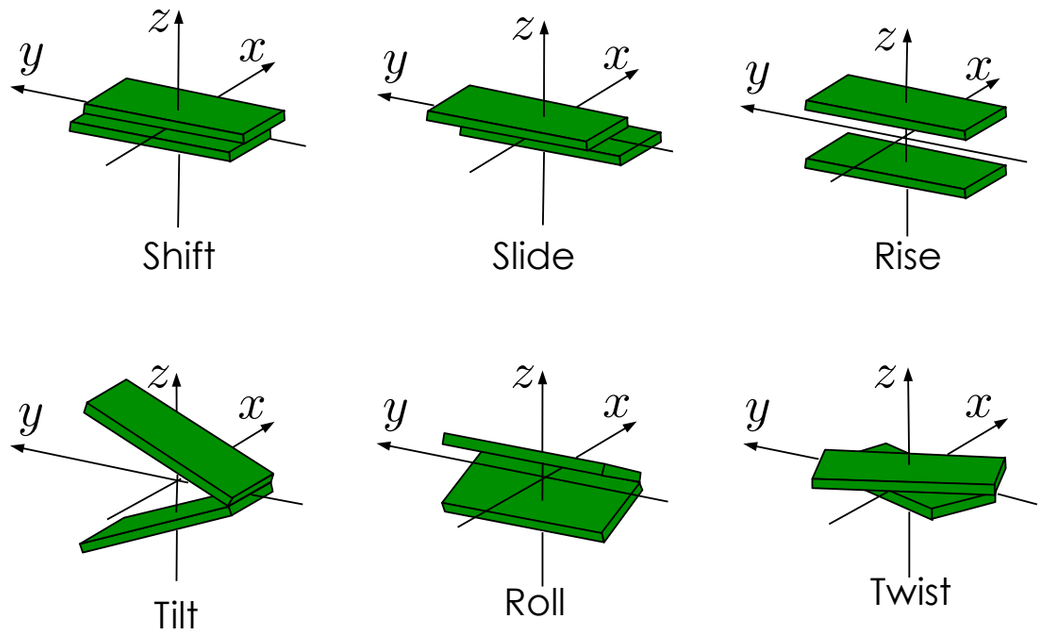
Our nucleosome model

Eslami-Mossallam, Schram, Tompitak, van Noort, Schiessel, PLoS ONE 11, e0156905 (2016)





rigid basepair model:



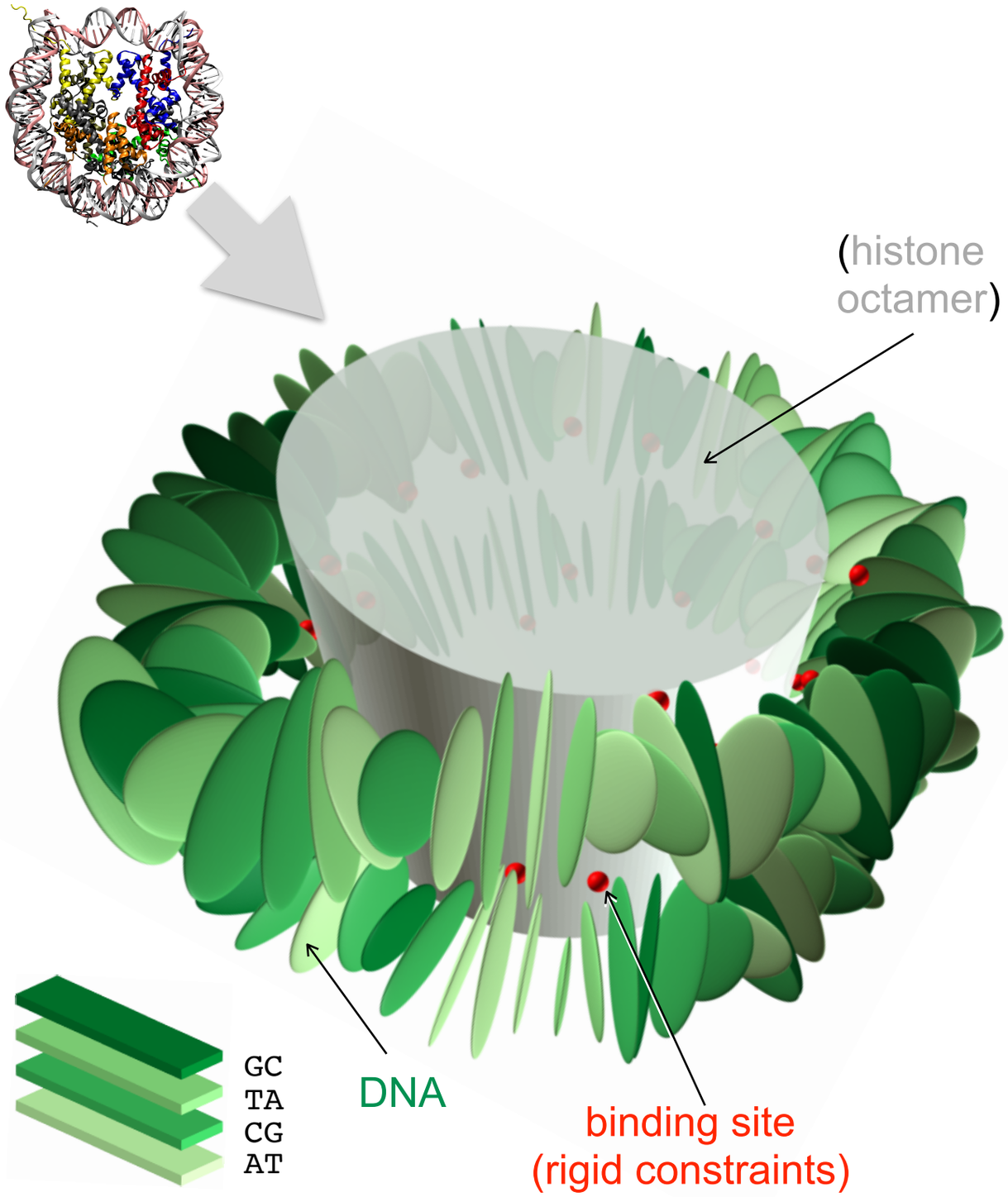
$$\vec{x} = [\text{Tilt}, \text{Roll}, \text{Twist}, \text{Shift}, \text{Slide}, \text{Rise}]$$

$$E_{\text{step}} = \frac{1}{2}\left(\vec{x} - \vec{x}^0\right)^T \mathbf{Q} \left(\vec{x} - \vec{x}^0\right)$$

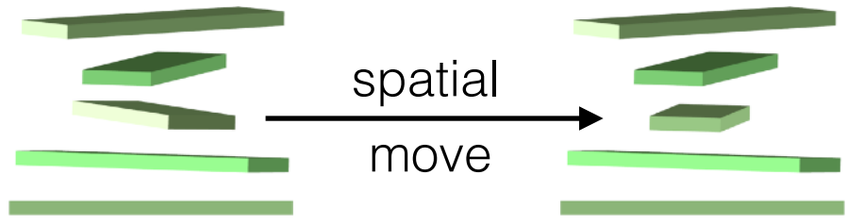
protein-DNA cocrystals
Olson et al. (1998)

all atom MD simulations
Lavery et al. (2010)

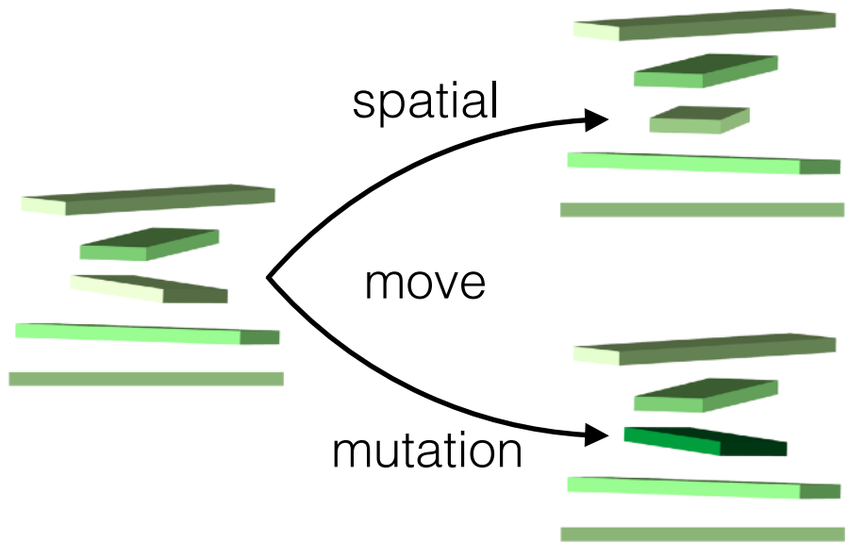
Two types of simulations



1. study fixed bp sequence:
   ordinary MC simulation



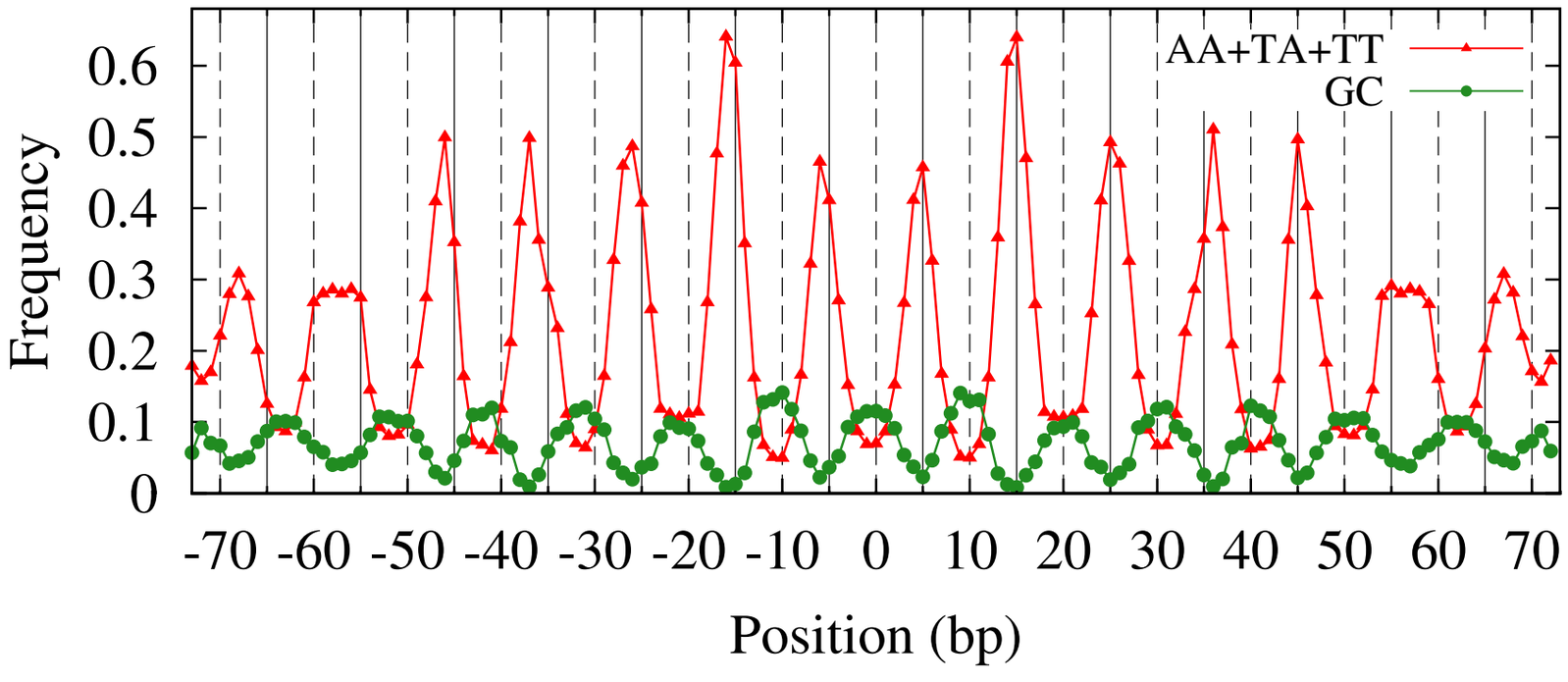
2. explore sequence space:
   mutation MC simulation

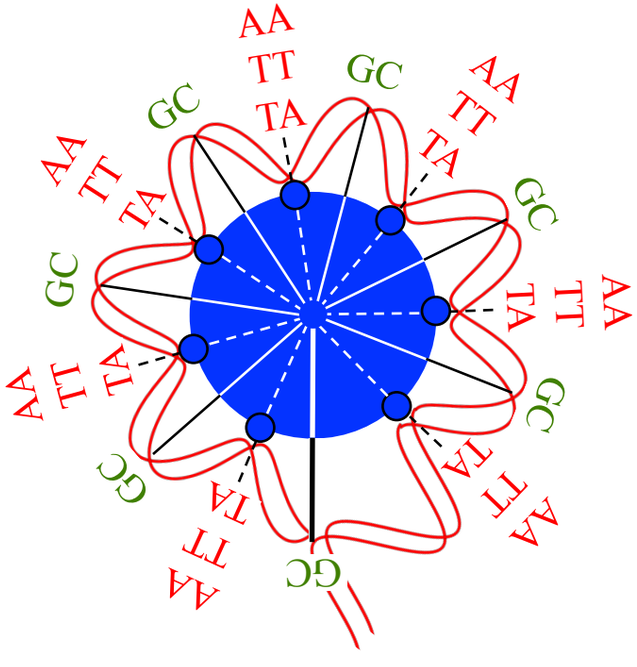


Recovering the positioning rules

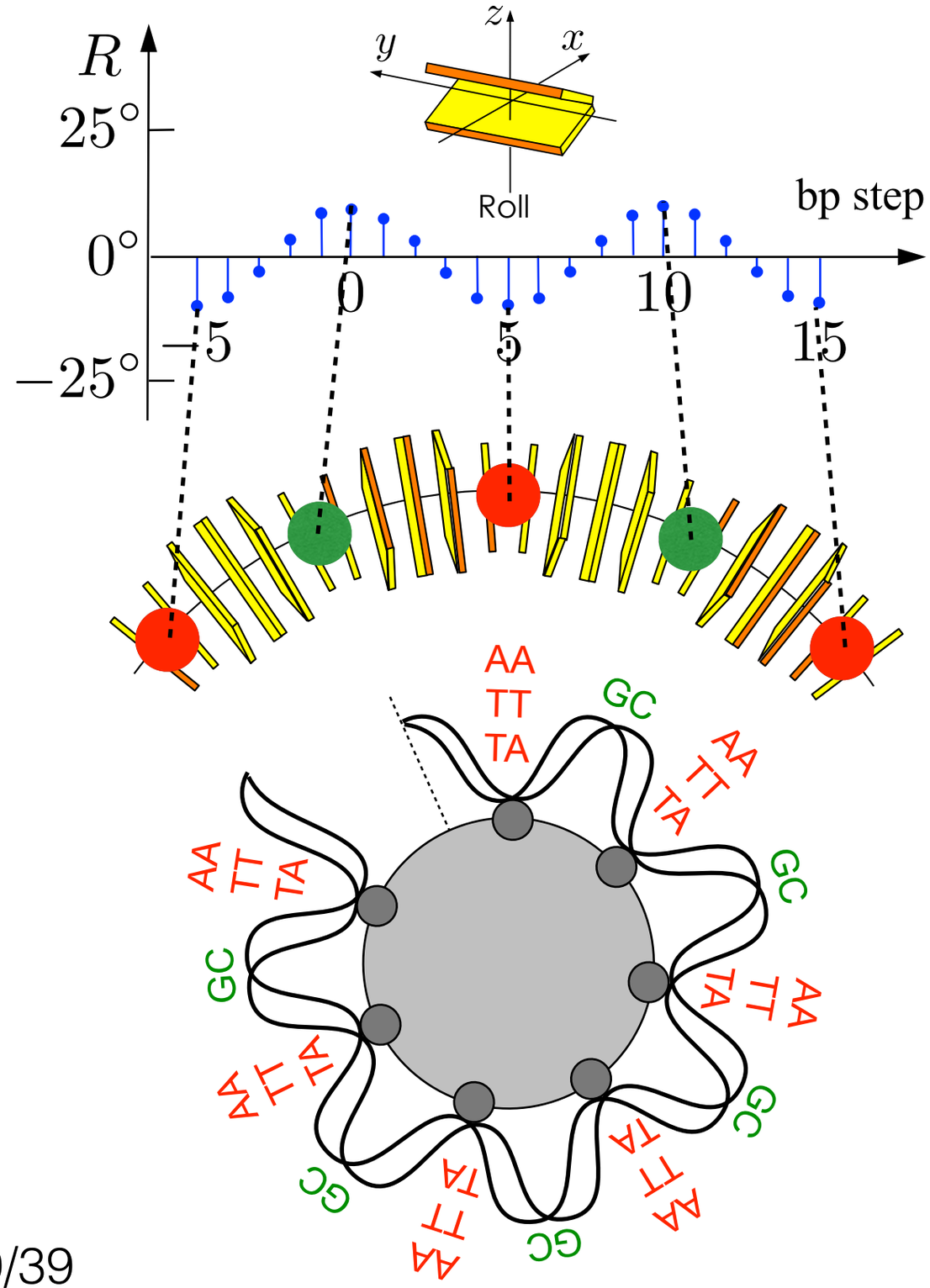
Eslami-Mossallam, Schram, Tompitak, van Noort, Schiessel, PLoS ONE 11, e0156905 (2016)

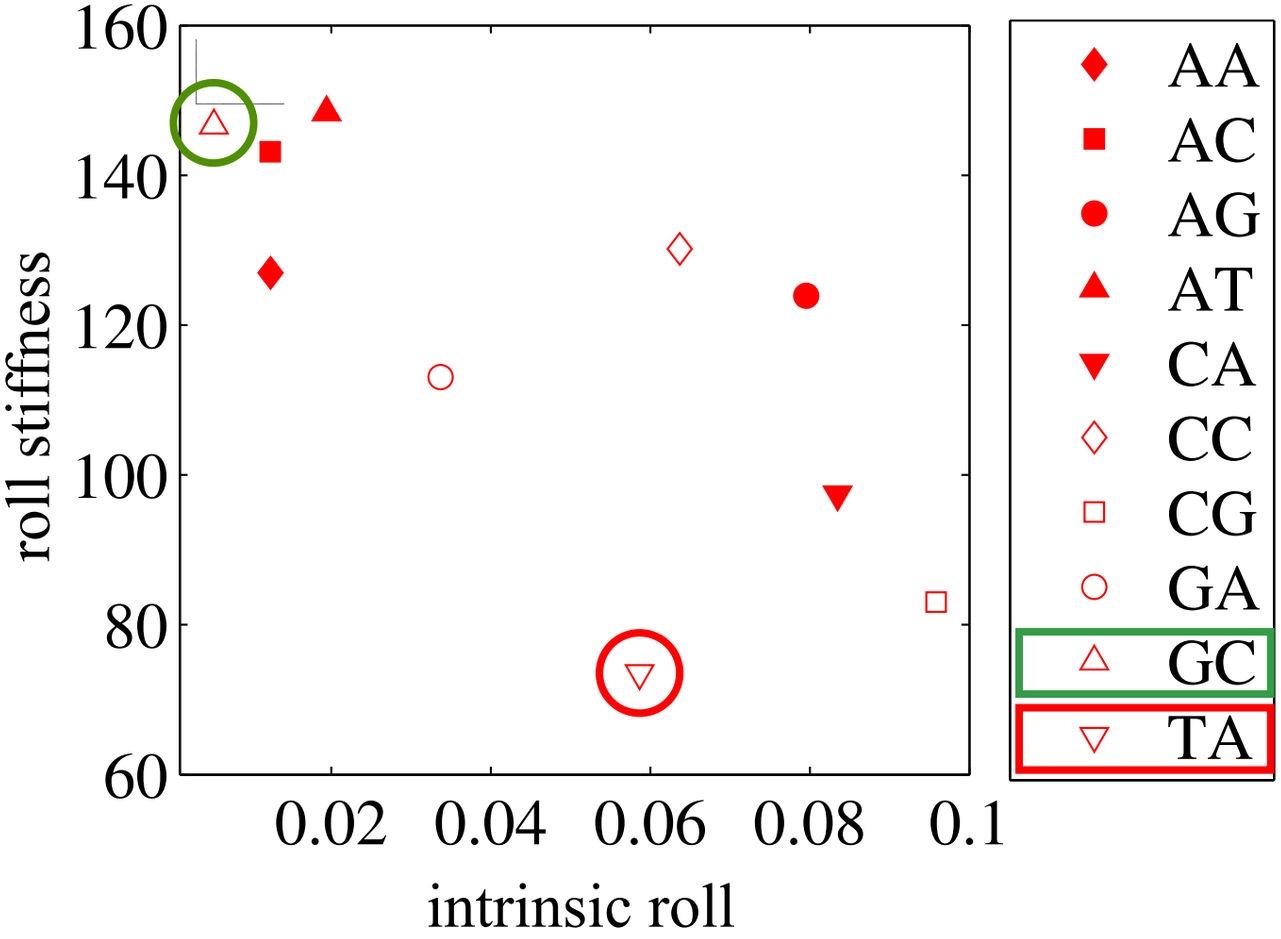
10 million sequences at 100 K





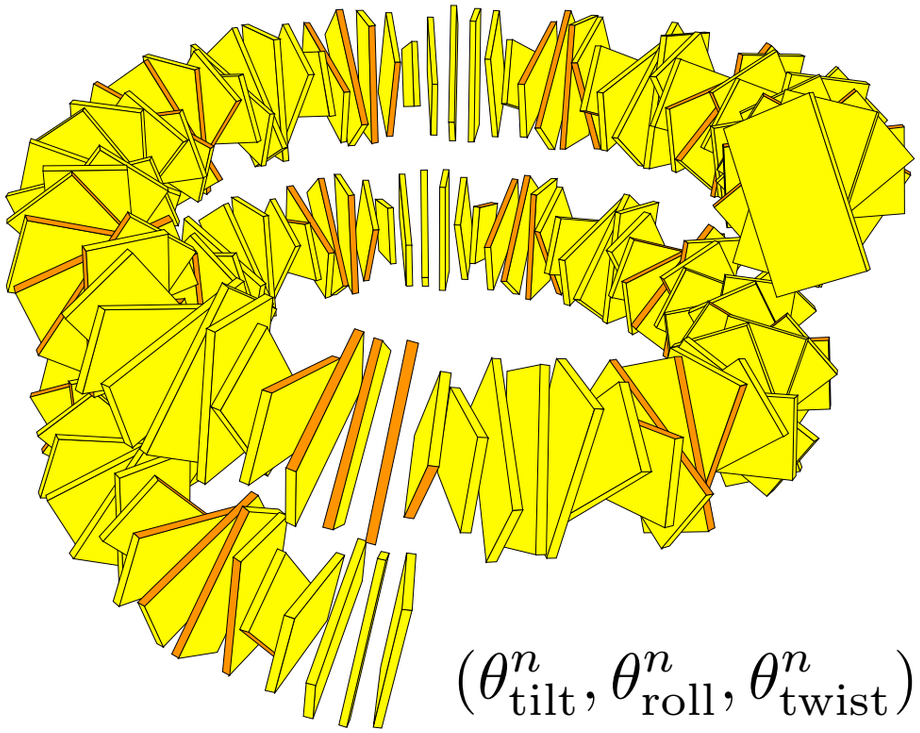
But do we now understand the positioning rules?

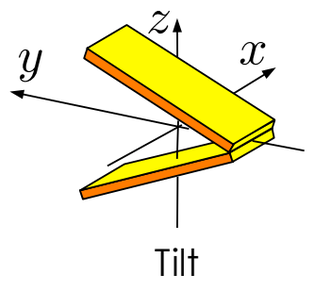


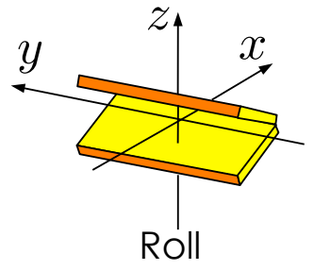


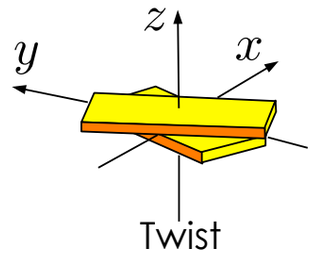
An exactly solvable model

Zuiddam, Everaers, Schiessel, Phys. Rev. E 96, 052412 (2017)









$$(\theta^n_{\text{tilt}}, \theta^n_{\text{roll}}, \theta^n_{\text{twist}}) = (\Gamma \sin(2\pi n/10 - \phi), \Gamma \cos(2\pi n/10 - \phi), \theta_{\text{twist}})$$

dinucleotide probability distribution:

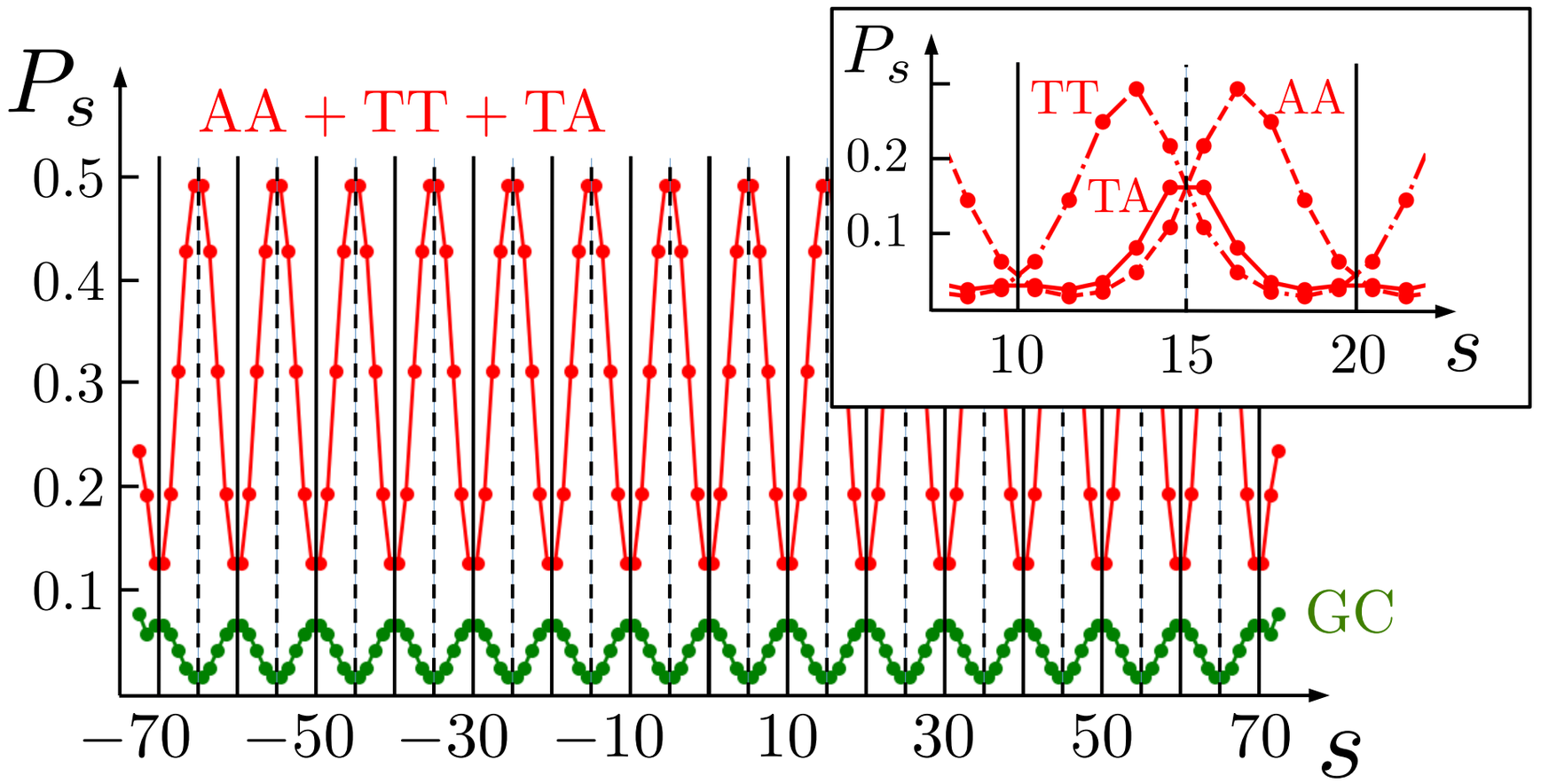
$$P_s(\mathrm{GC}) = \frac{1}{Z} \sum_{\substack{\mathrm{N}^1,\ldots,\mathrm{N}^{147} \\ \mathrm{N}^s=\mathrm{G},\mathrm{N}^{s+1}=\mathrm{C}}} \exp\left[-\beta \sum_{n=1}^{146} E^n \left(\mathrm{N}^n \mathrm{N}^{n+1}\right)\right]$$

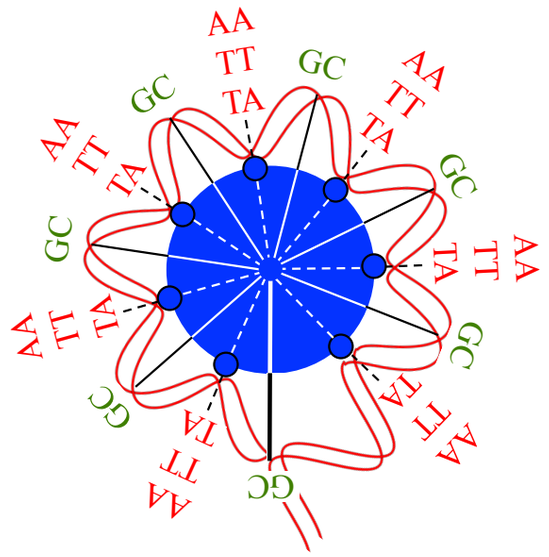
4D vectorspace with basis: $B = \{\mathrm{A}, \mathrm{T}, \mathrm{C}, \mathrm{G}\}$

transfer matrix: $$T^n_{\mathrm{GC}} = \exp\left[-\beta E^n\left(\mathrm{GC}\right)\right]$$

dinucleotide probability distribution:

$$P_s(\mathrm{GC}) = \frac{\sum\limits_{\mathrm{N}^1,\mathrm{N}^{147}} \left(T^1 ... T^{s-1}\right)_{\mathrm{N}^1\mathrm{G}} T^s_{\mathrm{GC}} \left(T^{s+1} ... T^{146}\right)_{\mathrm{CN}^{147}}}{\sum\limits_{\mathrm{N}^1,\mathrm{N}^{147}} \left(T^1 T^2 ... T^{146}\right)_{\mathrm{N}^1\mathrm{N}^{147}}}$$





But do we now understand the positioning rules?

average neighbor energy approximation:

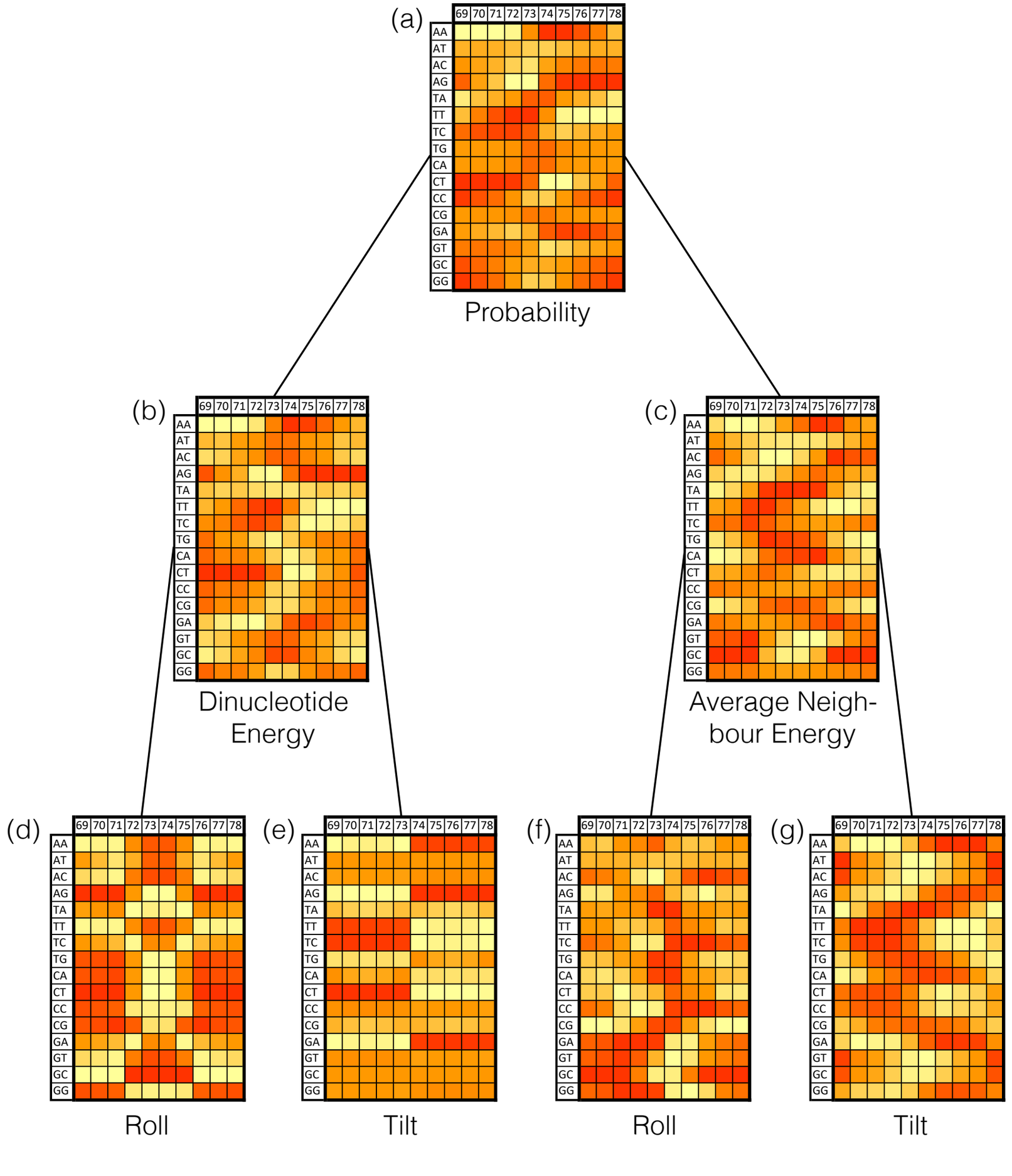
$$P_p^{(e)}(a,b) = \frac{1}{Z} \exp\left[-\beta \langle E_{p-1}(n_{p-1},a)\rangle_{n_{p-1}}\right] \exp[-\beta E_p(a,b)] \exp\left[-\beta \langle E_{p+1}(b,n_{p+2})\rangle_{n_{p+2}}\right]$$

The probability of a basepair step depends on:

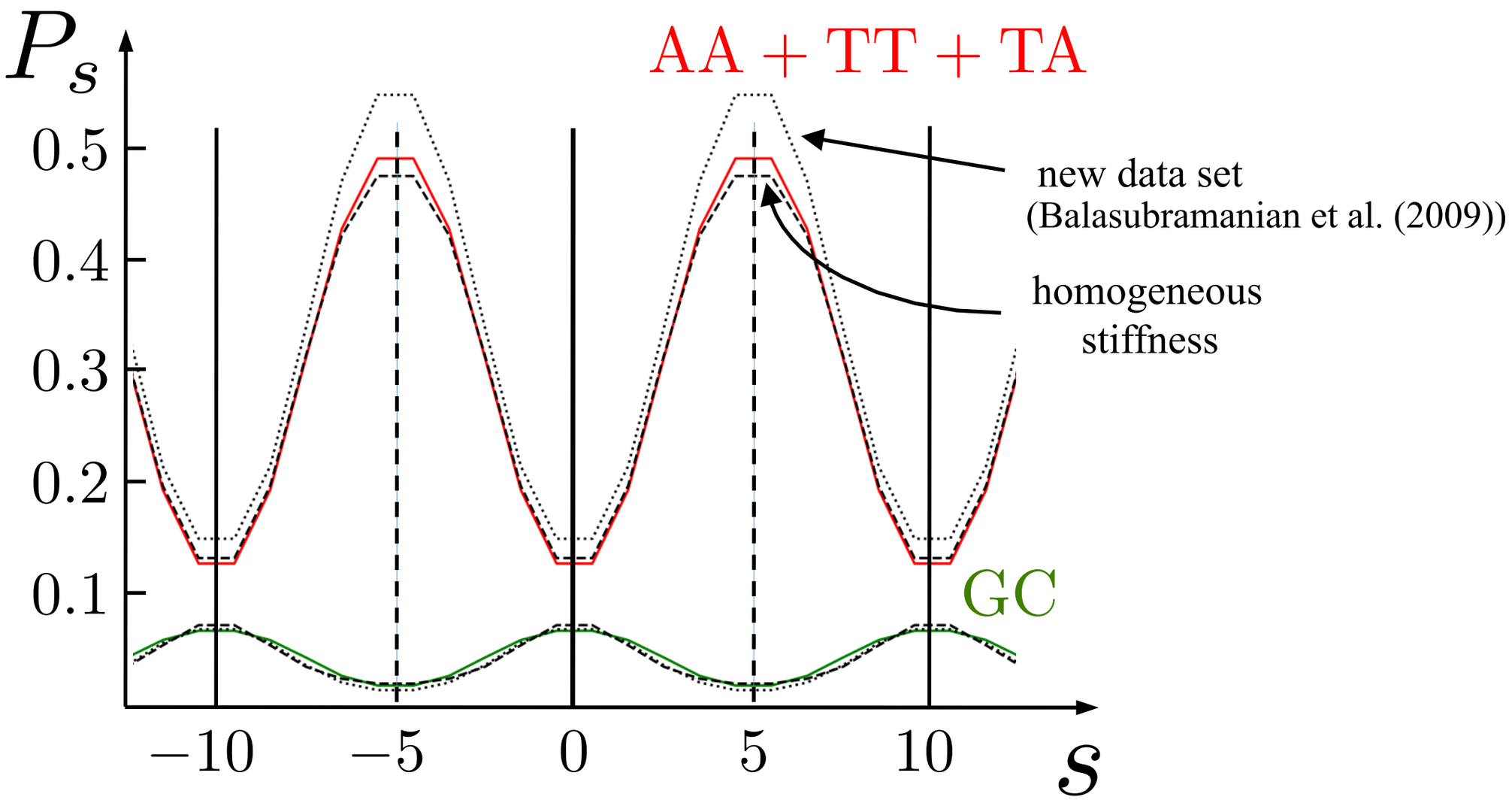
— the energy of this step

and

— the average energies of all possible steps before and after that step


(a)
69 70 71 72 73 74 75 76 77 78
AA AT AC AG TA TT TC TG CA CT CC CG GA GT GC GG
Probability
(b)
Dinucleotide Energy
(c)
Average Neighbour Energy
(d)
Roll
(e)
Tilt
(f)
Roll
(g)
Tilt

What is more important: shape or stiffness?

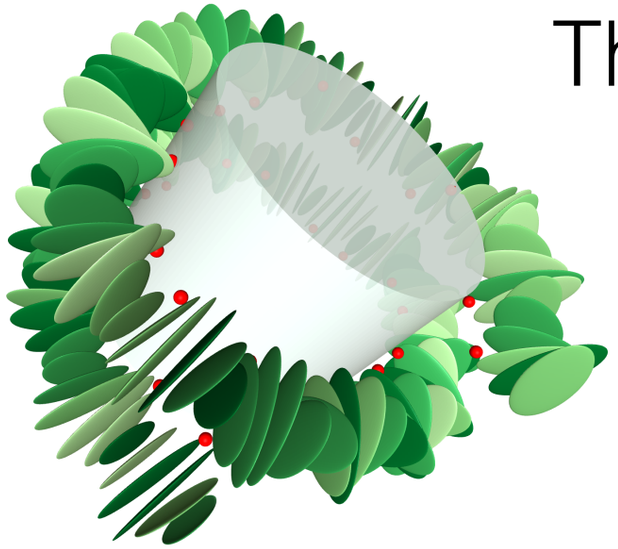


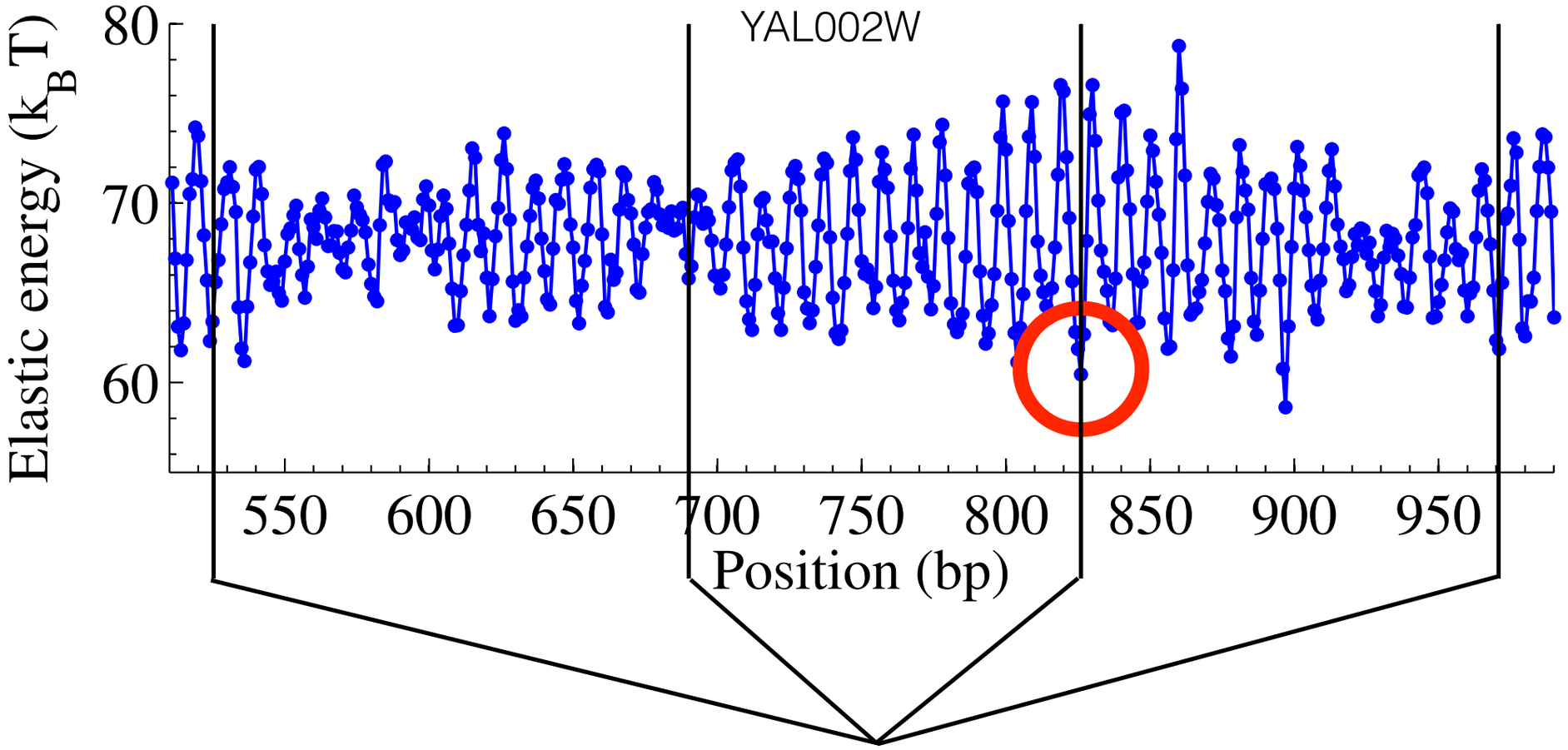
THE MECHANICAL GENOME

1. Positioning rules are mechanical
2. Multiplexing of the 2 (and even 3) layers
3. Mechanical cues on real genomes

Artwork: Amar van Leewaarde & Marco Tompitak

The energy landscape of a gene





Nucleosome mapping at basepair resolution in *S. cerevisiae*

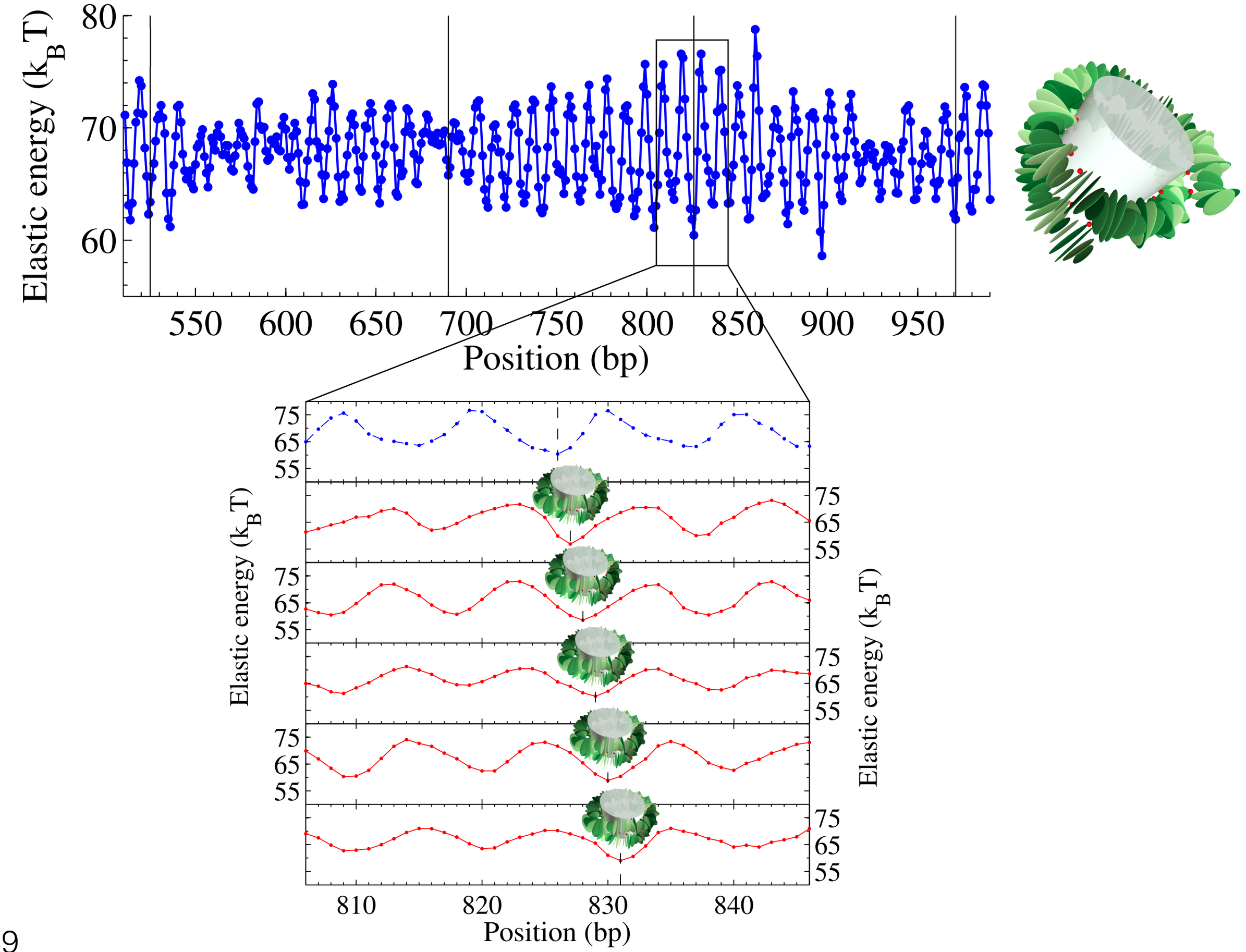
Brogaard, Xi, Wang & Widom

Nature **486** (2012) 496

Synonymous Mutation Monte Carlo

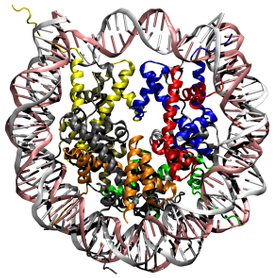
codon sequence	…	GTA	CTC	ACA	ACT	ACA	CAT	TTT	GCC	CTT	ATT	…
protein sequence	…	Val	Leu	Thr	Thr	Thr	His	Phe	Ala	Leu	Ile	…

allowed mutations										
	GTA	TTA	ACA	ACA	ACA	CAT	TTT	GCA	TTA	ATA
	GTT	TTG	ACT	ACT	ACT	CAC	TTC	GCT	TTG	ATT
	GTG	CTA	ACG	ACG	ACG			GCG	CTA	ATC
	GTC	CTT	ACC	ACC	ACC			GCC	CTT	
		CTG							CTG	
		CTC							CTC	

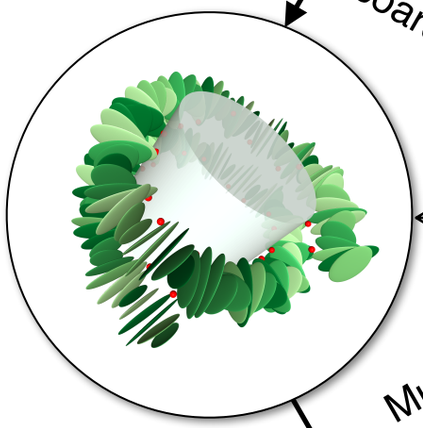

Elastic energy (k_BT)
80
70
60
550
600
650
700
750
800
850
900
950
Position (bp)
75
65
55
810
820
830
840

10^7 speed-up

crystal structure

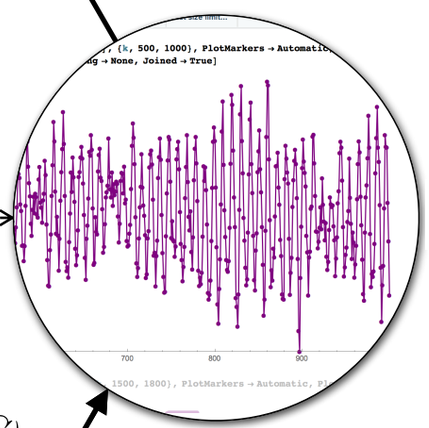


coarse graining



coarse grained model

$\sqrt{(\Delta E)^2} = 0.8\, k_B T$



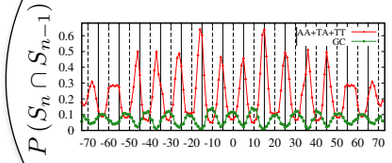
(free) energy landscape

genome-wide calculations

Mutatation Monte Carlo

$P(S_n | S_{n-1} \cap S_{n-2}) = \frac{P(S_n \cap S_{n-1} \cap S_{n-2})}{P(S_n \cap S_{n-1})}$

$-k_B T \ln(P(S))$



$P(S) = \prod_{n=1}^{147} P(S_n | S_{n-1} \cap S_{n-2})$

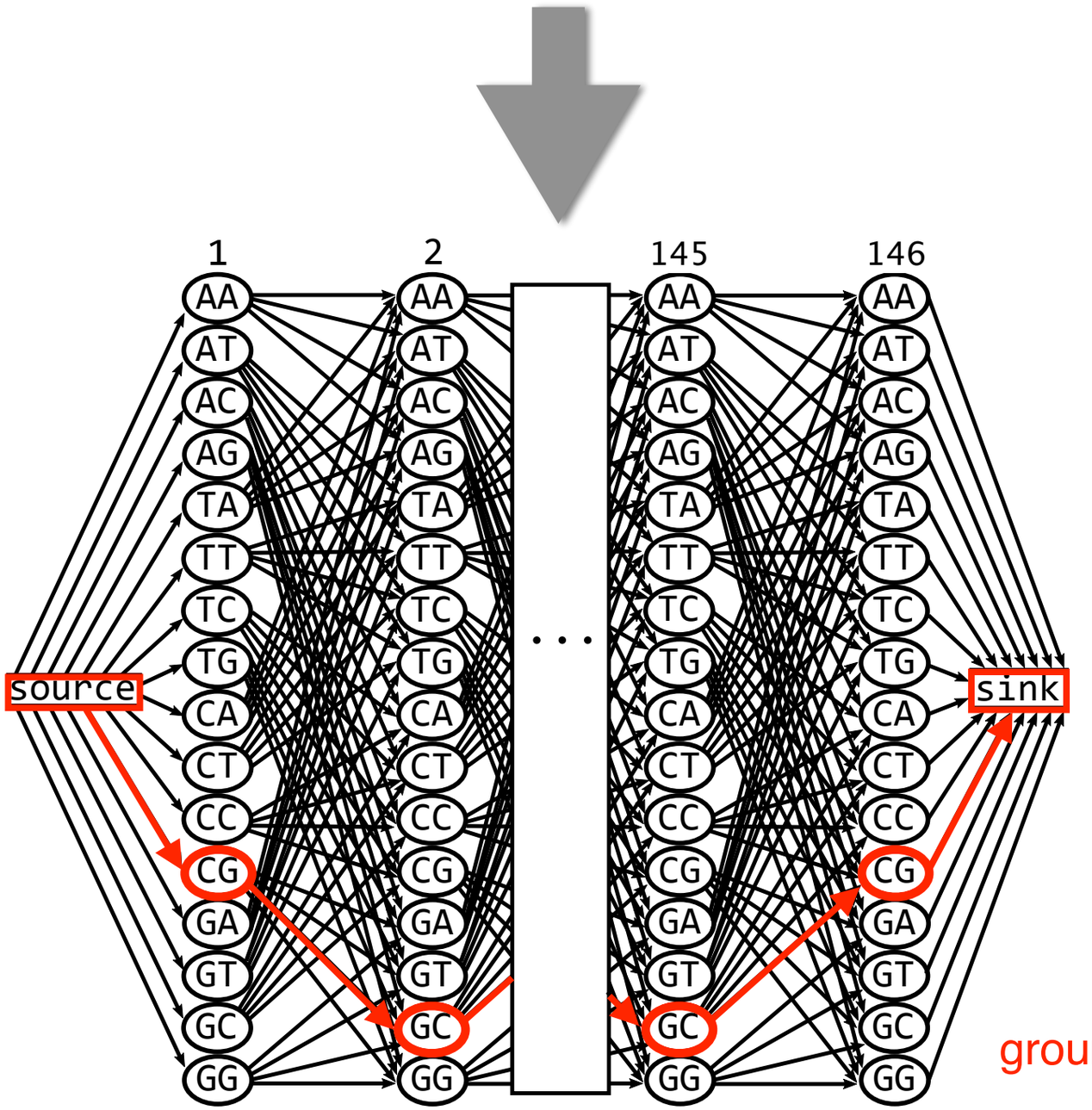
mono-, di-, trinucleo-tide probabilities

probabilistic model (2-step Markov)

Mapping of sequences on weighted graphs

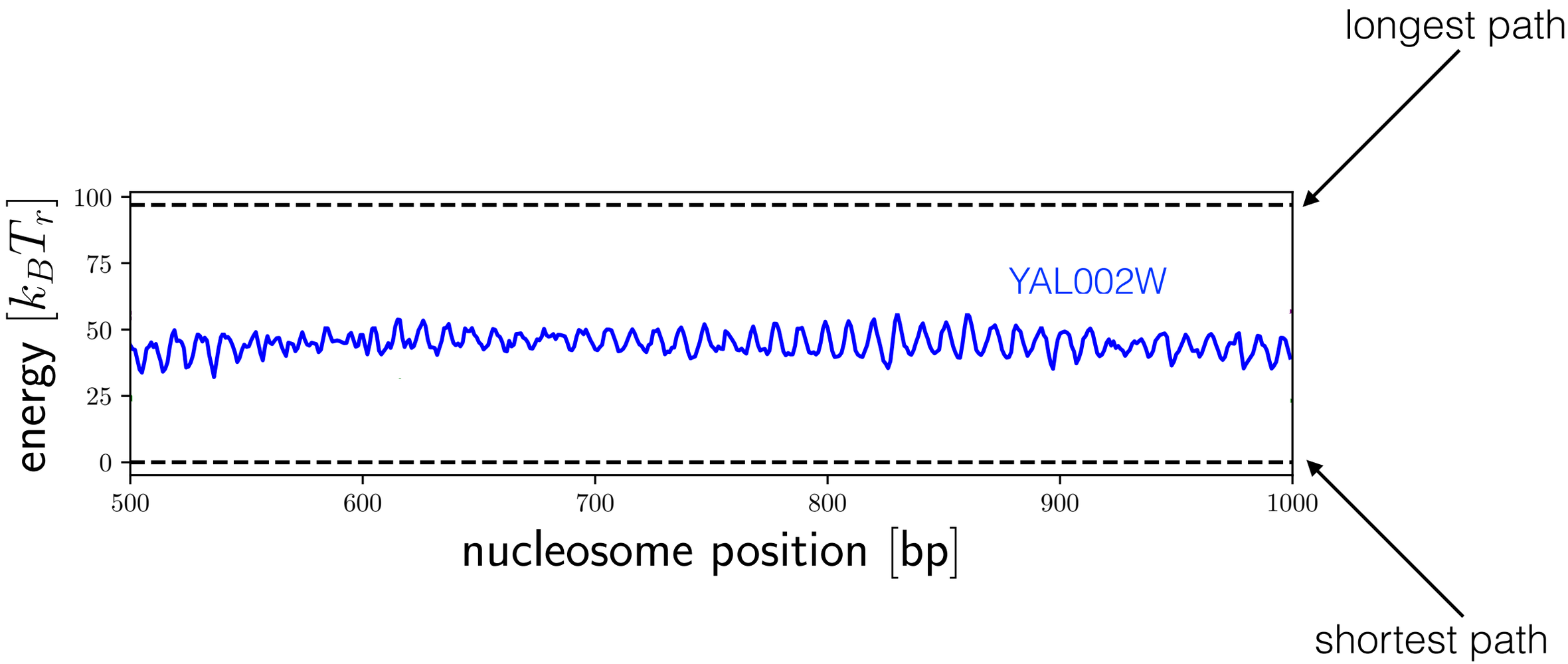
Zuiddam, Schiessel, PRE 99, 012422 (2019)

4^{147} sequences

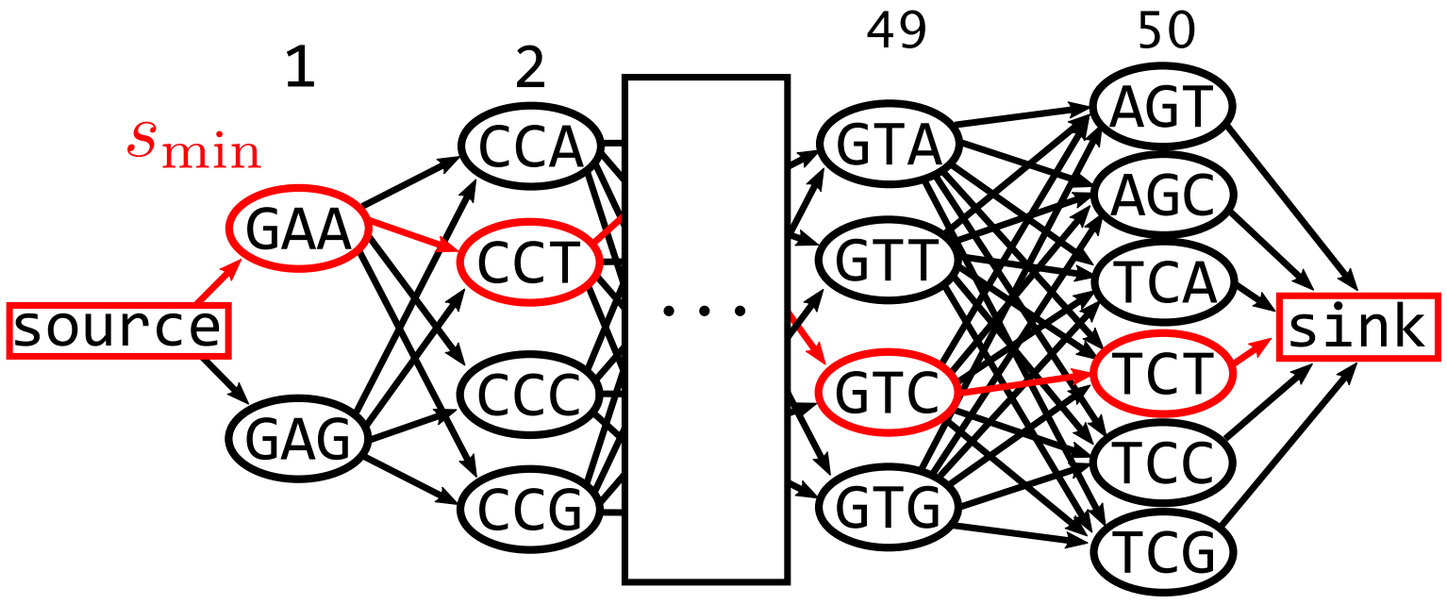


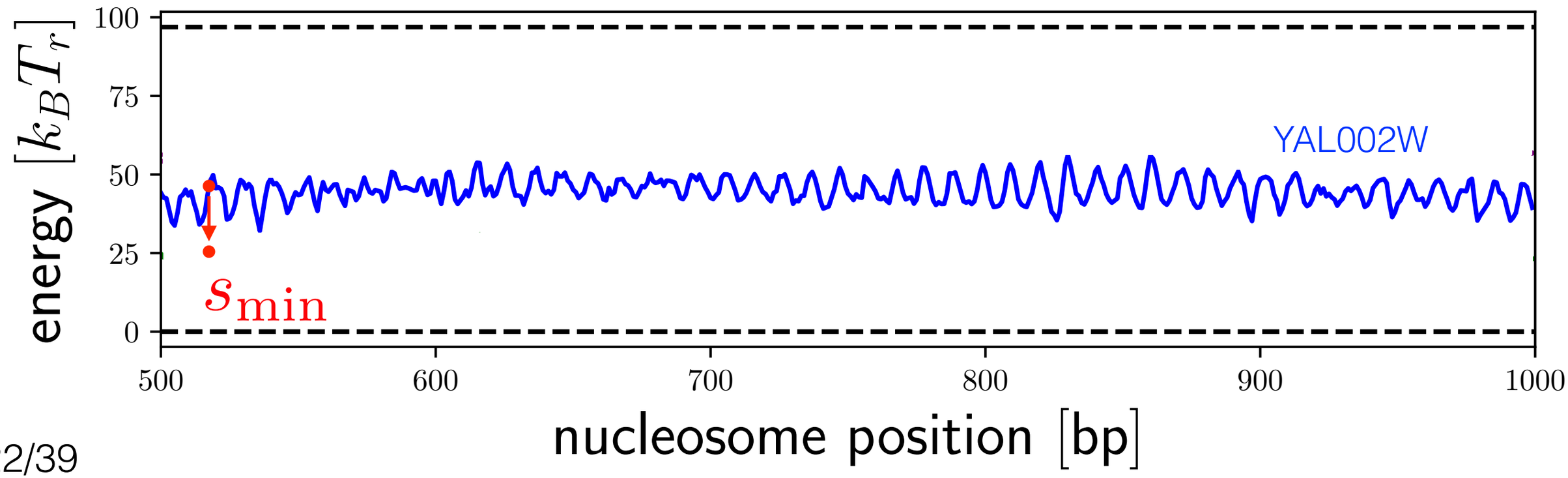
ground state sequence

The energy landscape of a gene: comparision to extreme sequences

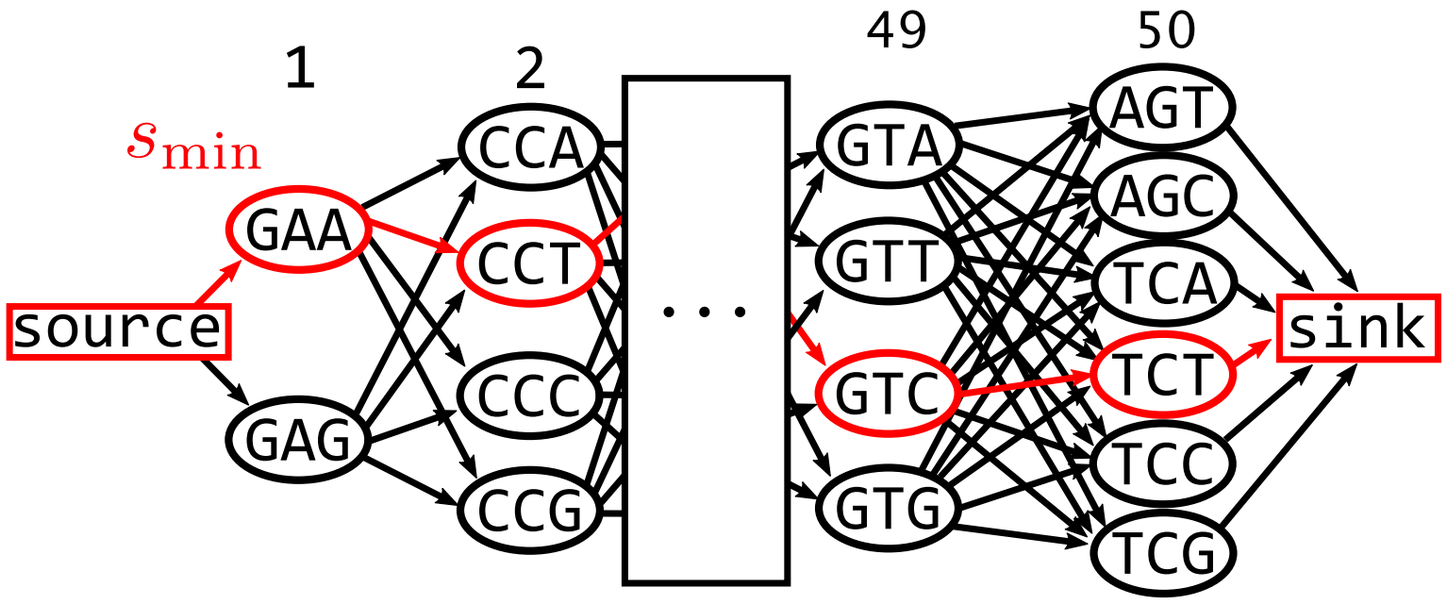


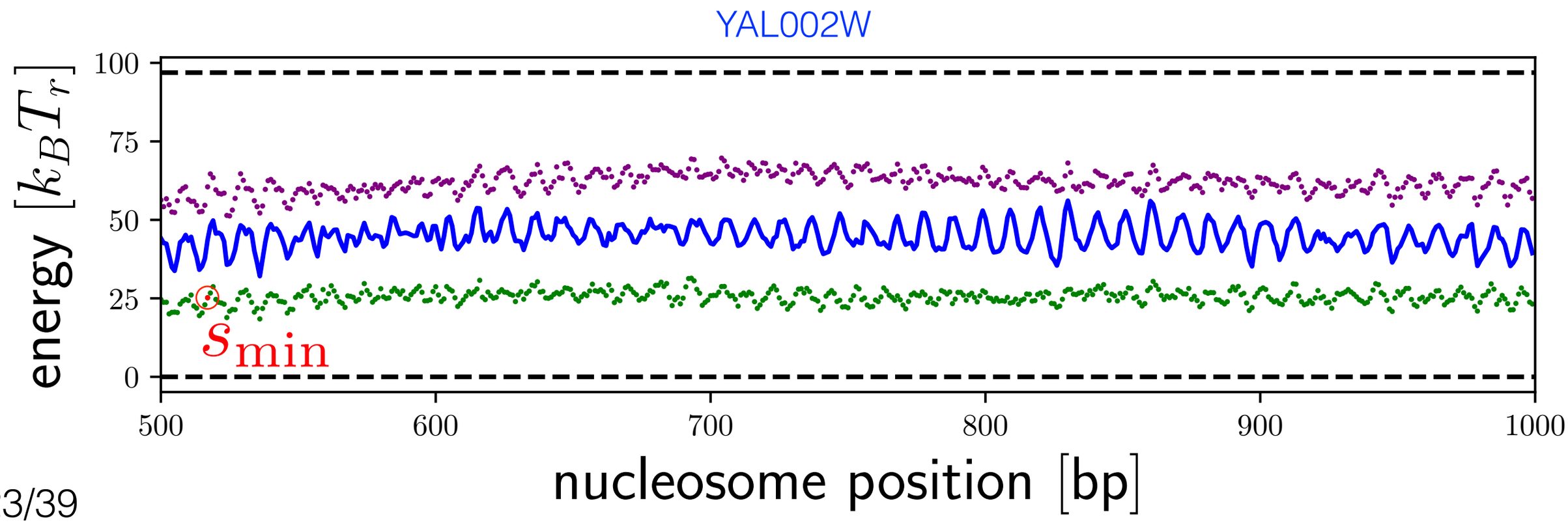
Graph on coding sequence



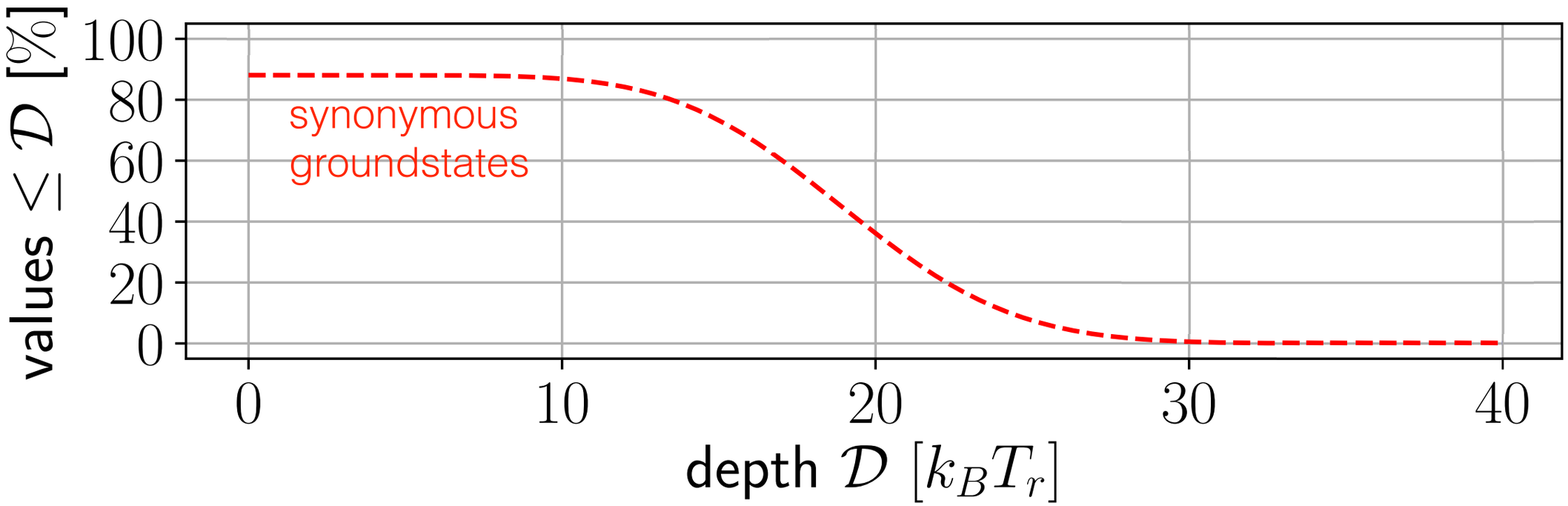


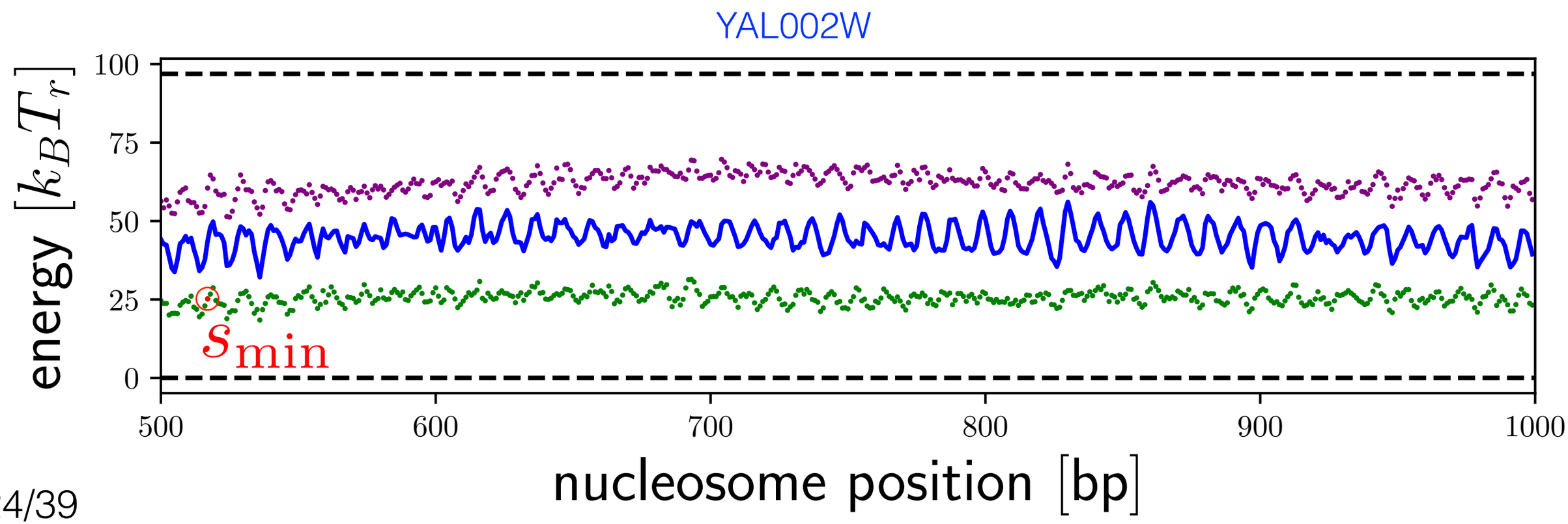
Graph on coding sequence



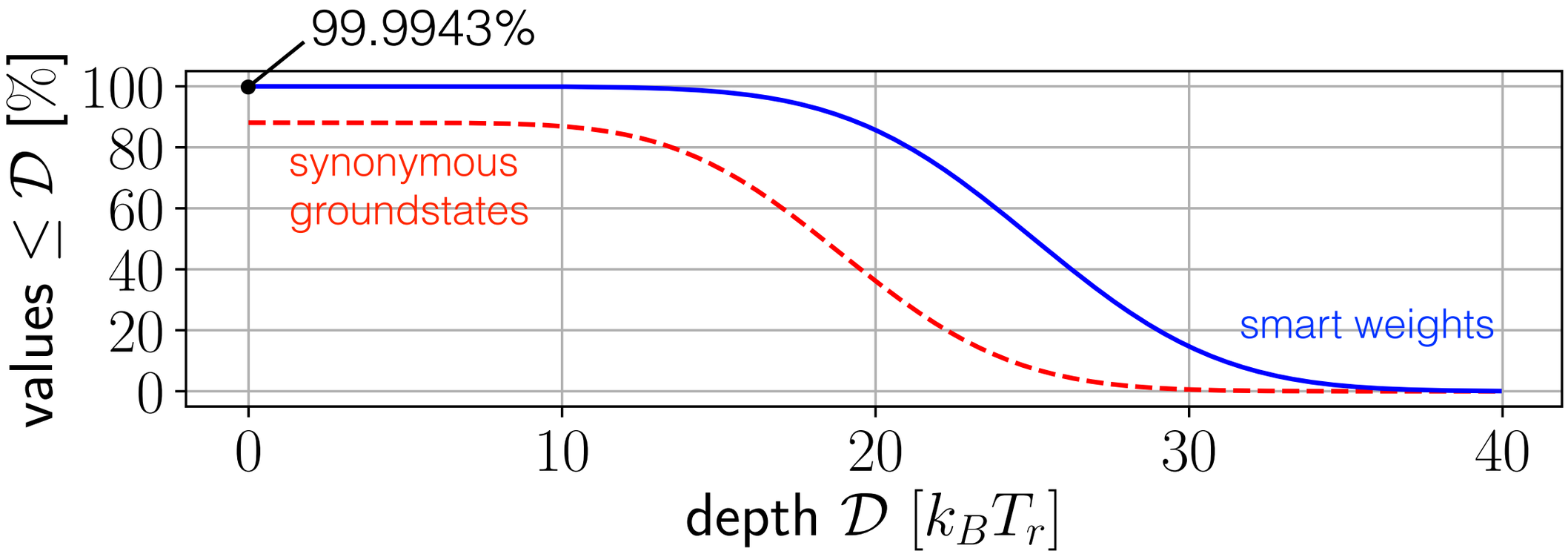


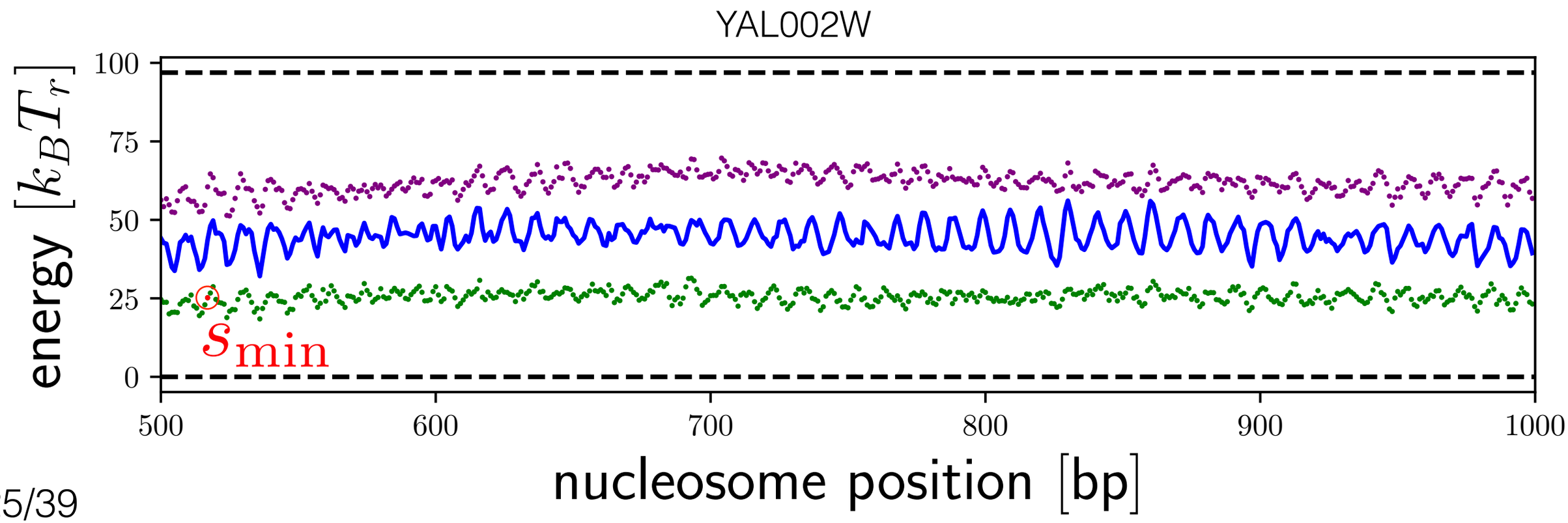
Success in minima creation on 7640994 positions



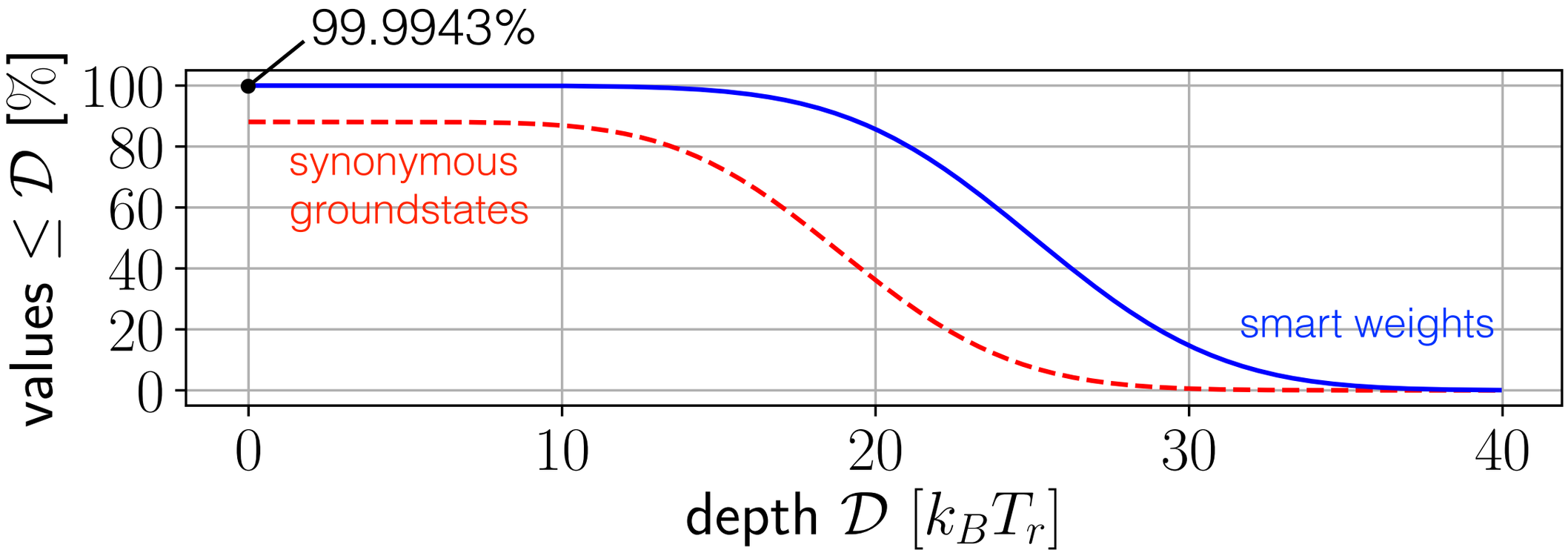


Success in minima creation on 7640994 positions

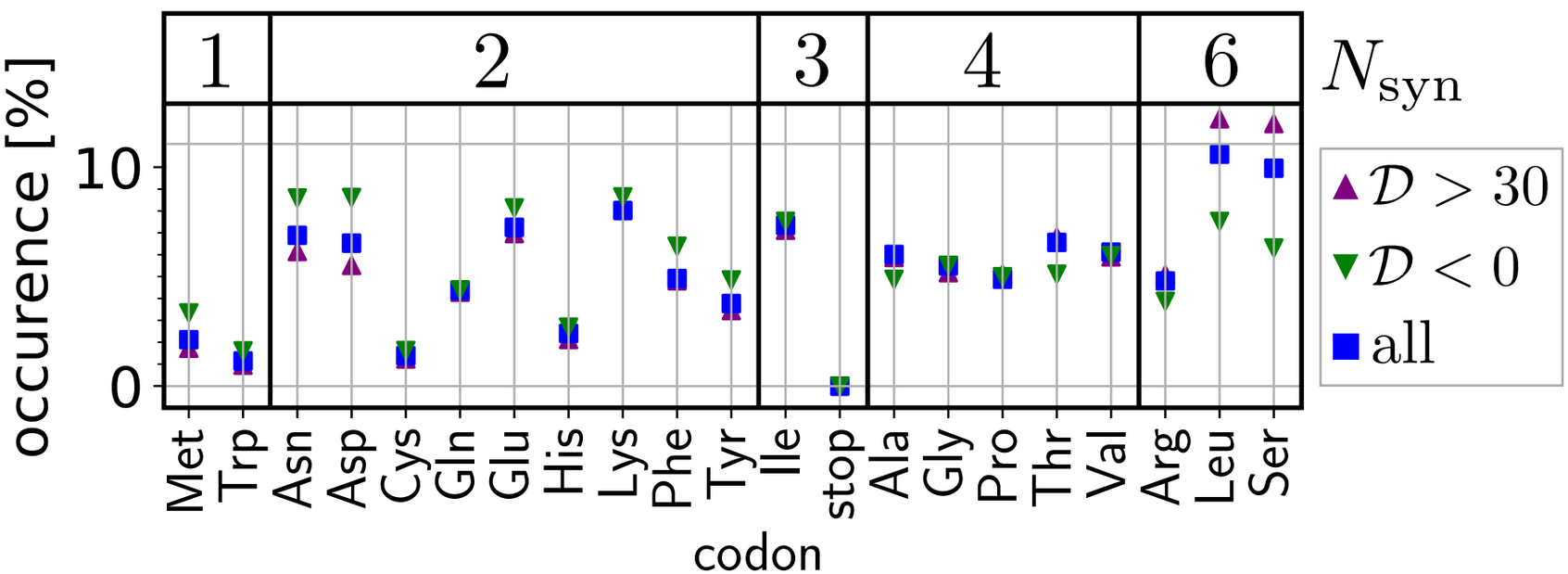




Success in minima creation on 7640994 positions

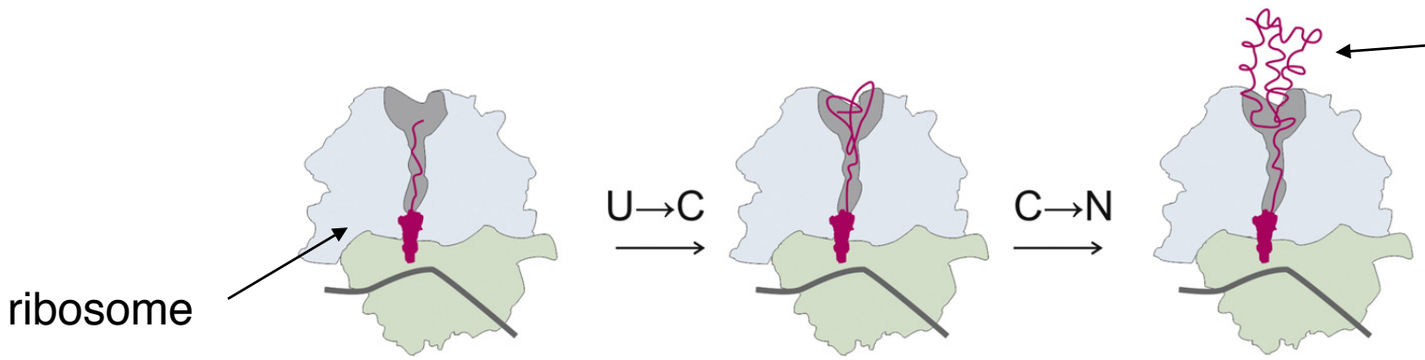


occurrence frequencies of amino acids in yeast genome:



So is there space for a third layer?

translation speed in ribosomes:

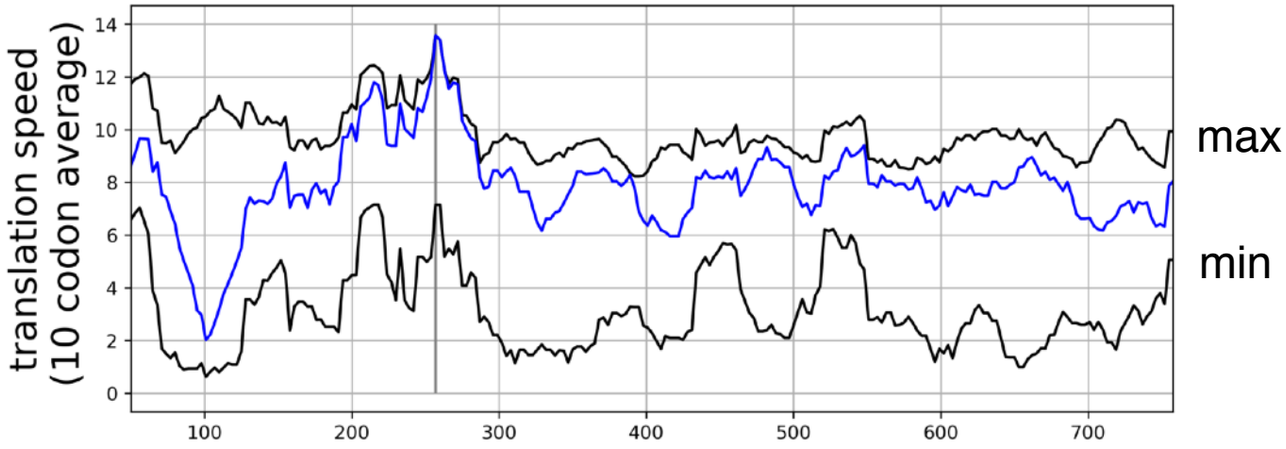


Rodnina & Wintermeyer, J. Mol. Biol. 2016

A third layer: translational speed

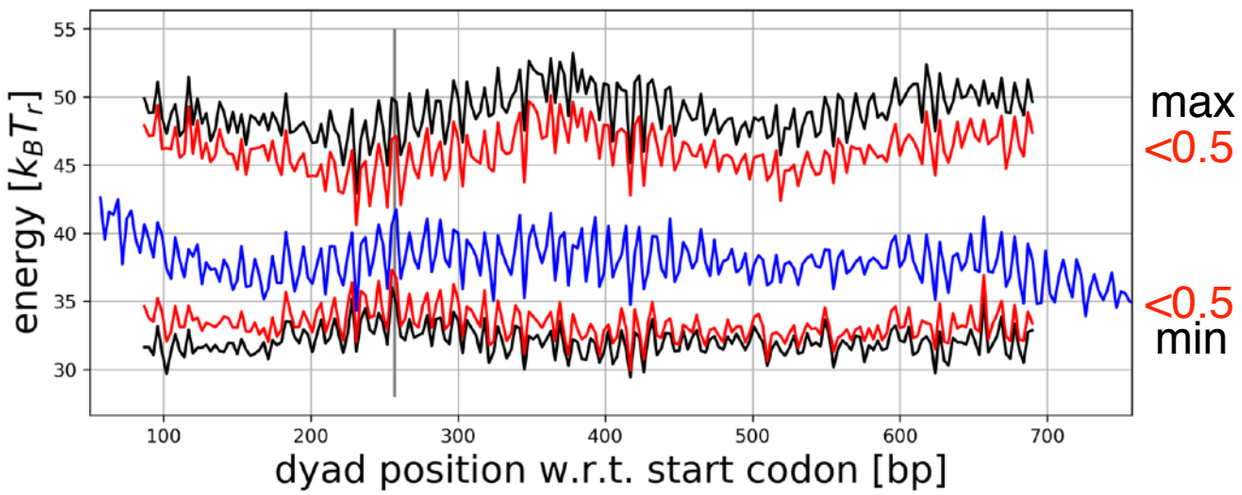
Zuiddam, Shakiba, Schiessel, in preparation

Oryza sativa



codon specific Markov model

Rudorf, …, Lipowski, PLoS Comp. Biol. 2014



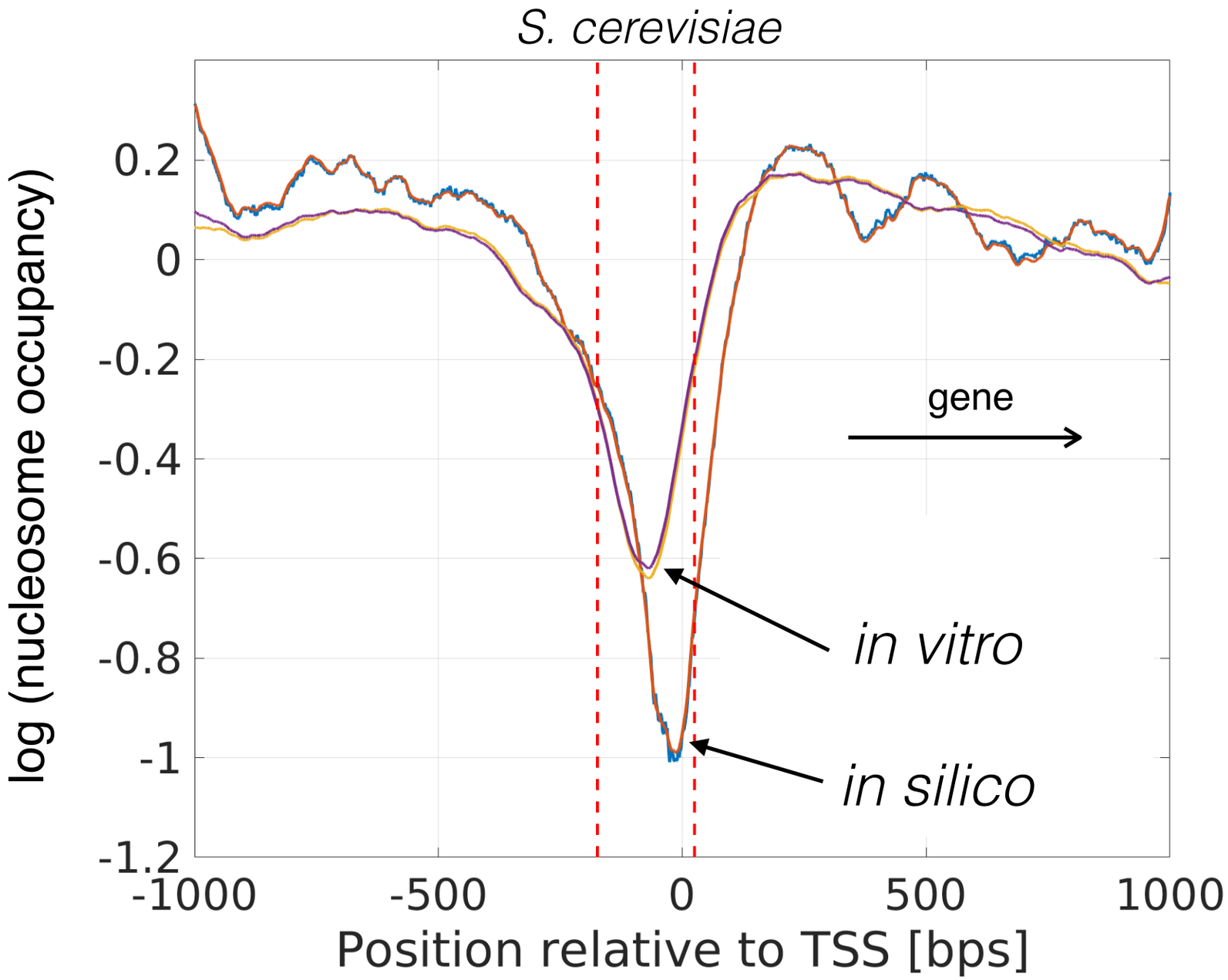
THE MECHANICAL GENOME

1. Positioning rules are mechanical
2. Multiplexing of the 2 (and even 3) layers
3. Mechanical cues on real genomes

Artwork: Amar van Leewaarde & Marco Tompitak

Nucleosome occupancy around transcription start sites

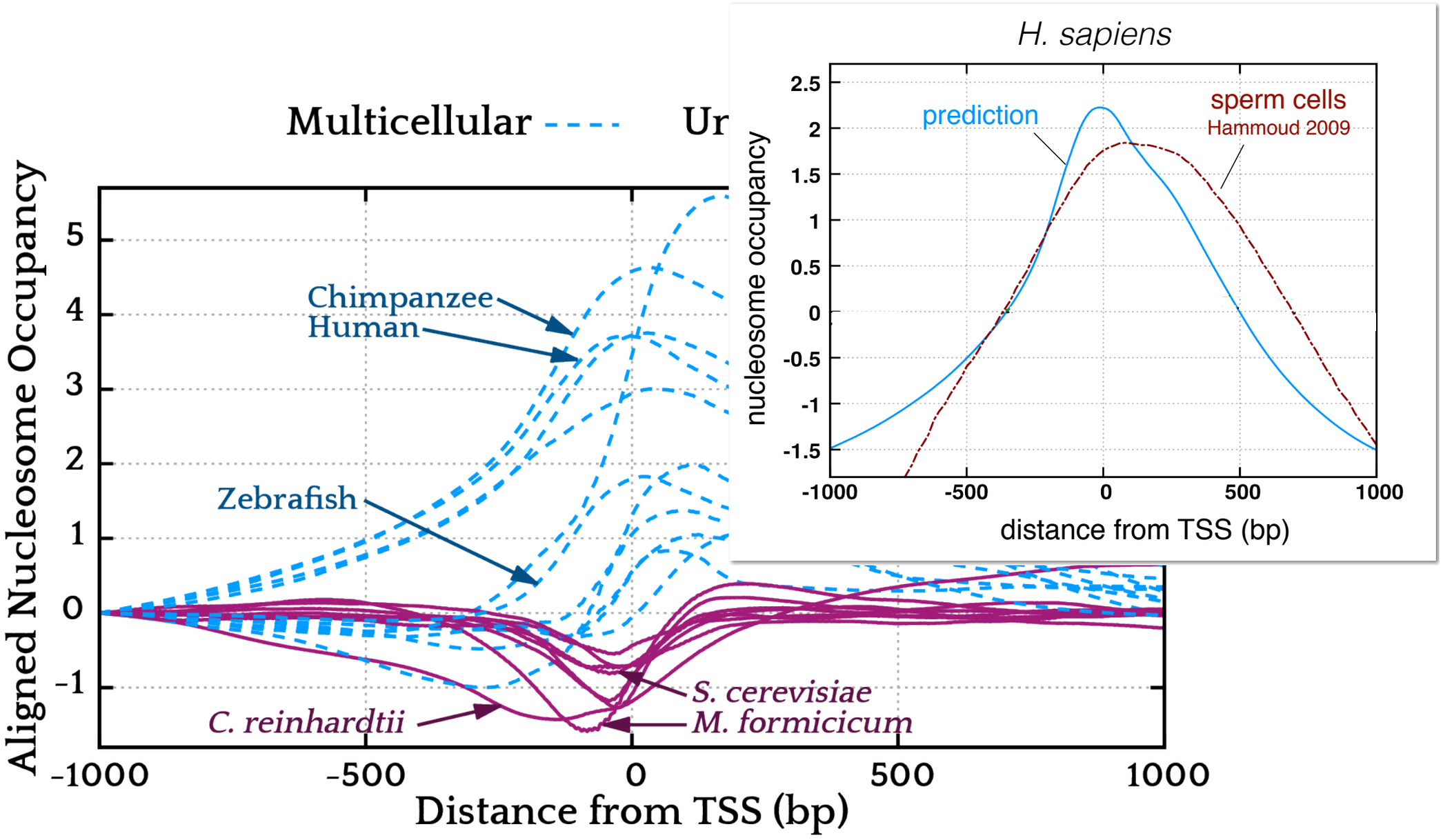
Neipel, Brandani, Schiessel, Phys. Rev. E 101, 022405 (2020)



DNA mechanics keeps promoter regions free of nucleosomes

A general rule for DNA elasticity

Tompitak, Vaillant, Schiessel, Biophys. J. 112, 505 (2017)

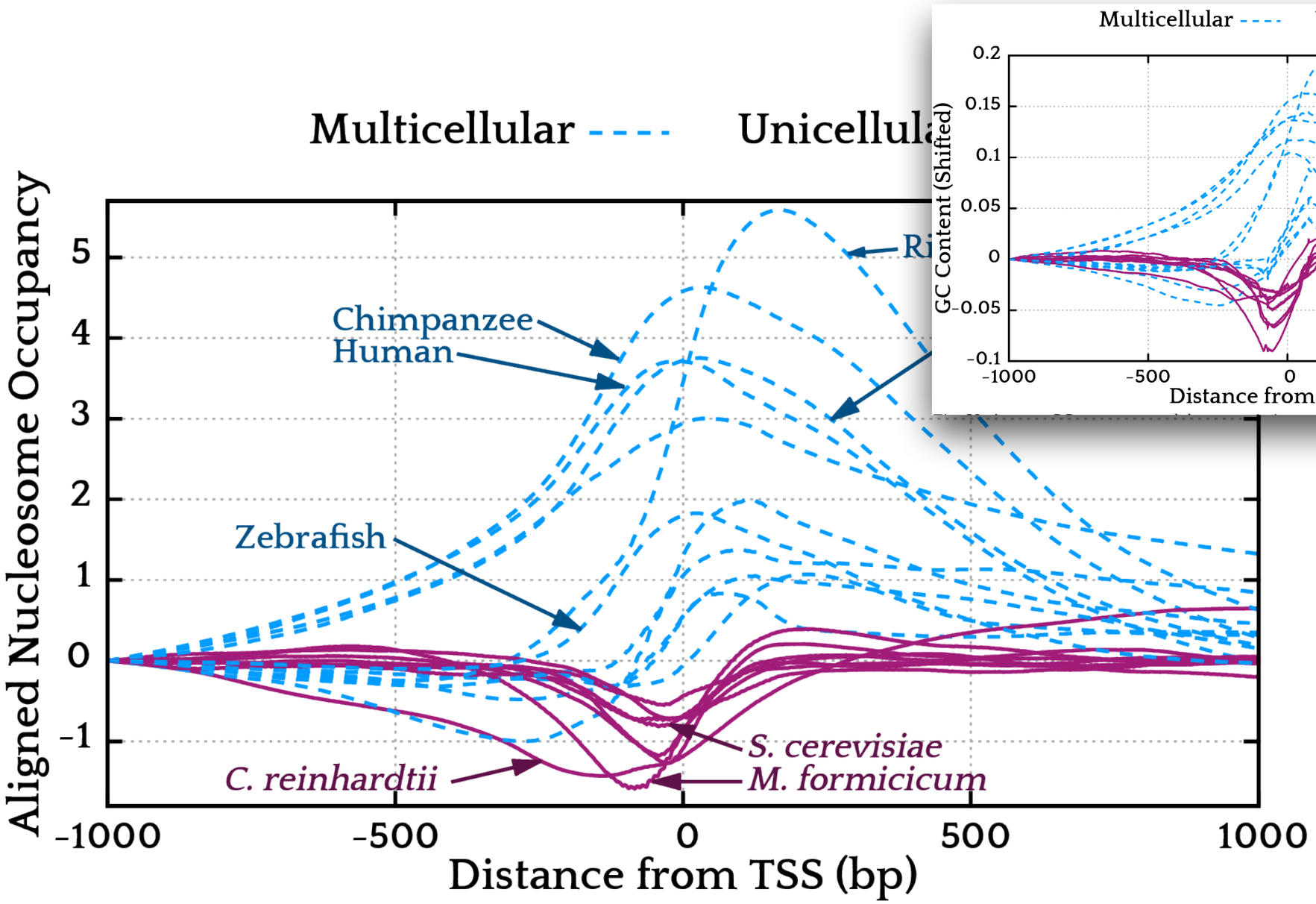


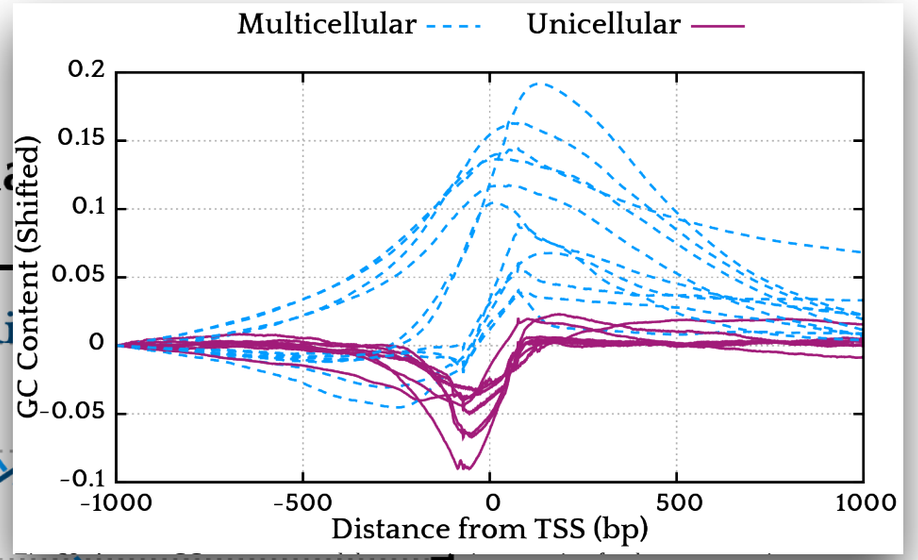
What is the function of the broad peaks in multicellular life forms?

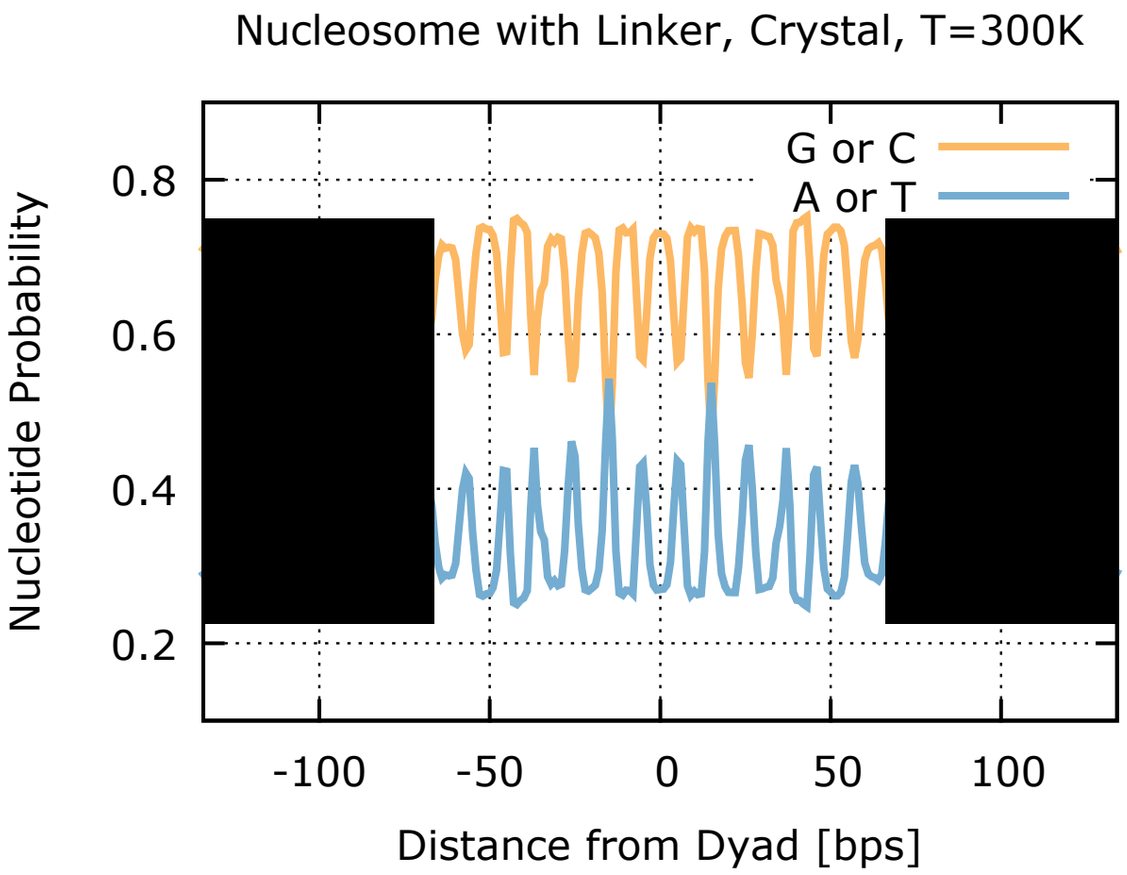
Transfer of epigenetic information via retention of nucleosomes in sperm cells.

A general rule for DNA elasticity

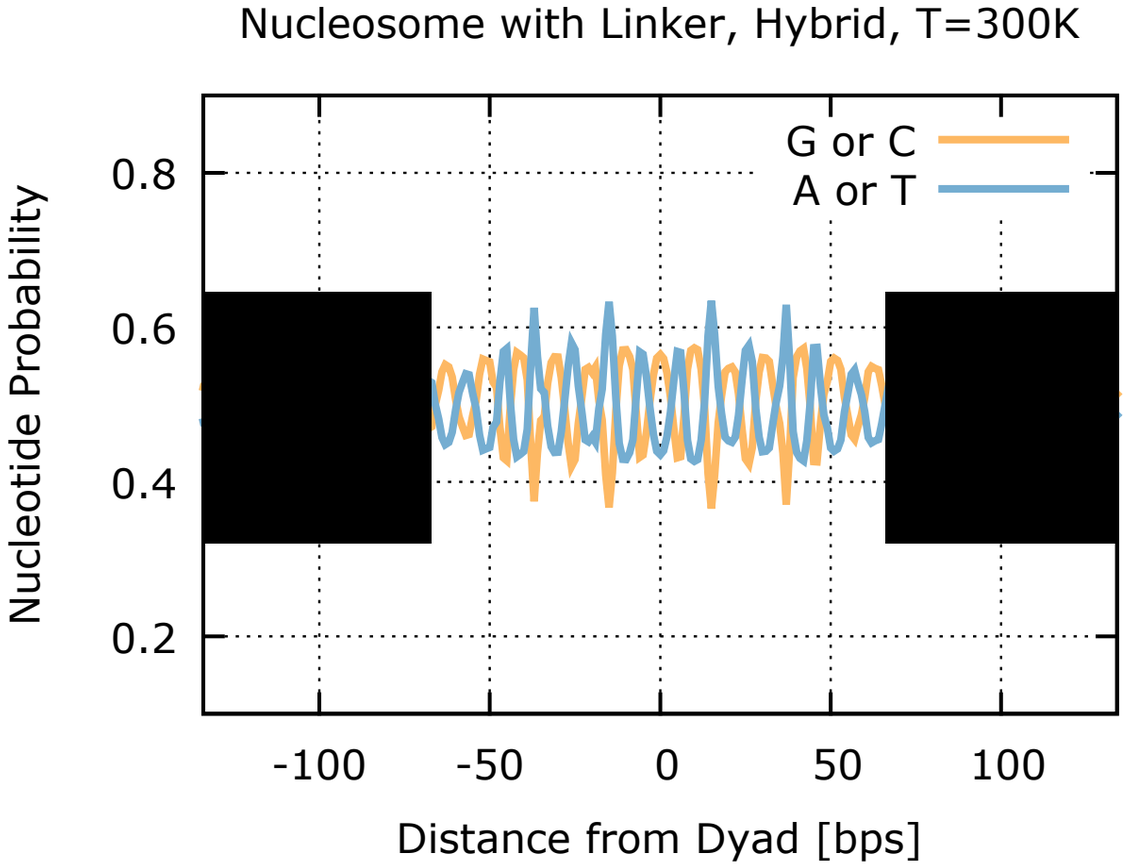
Tompitak, Vaillant, Schiessel, Biophys. J. 112, 505 (2017)



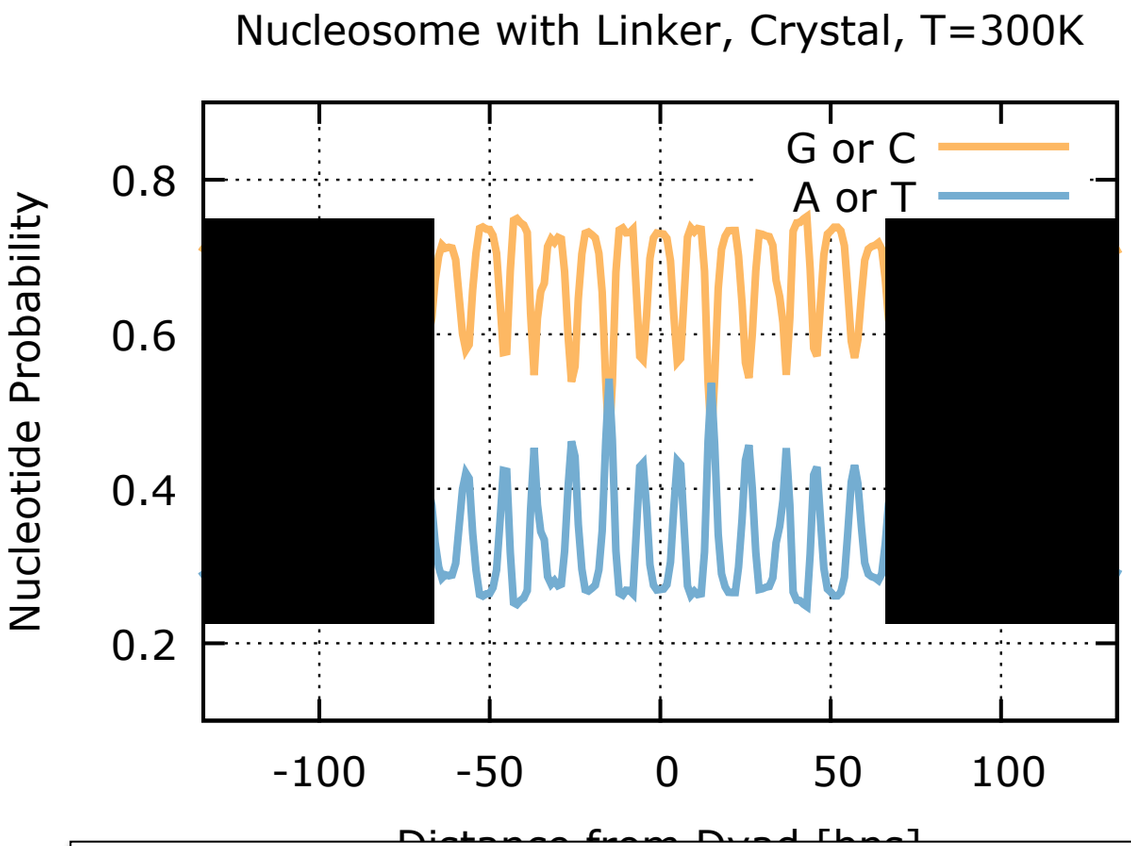


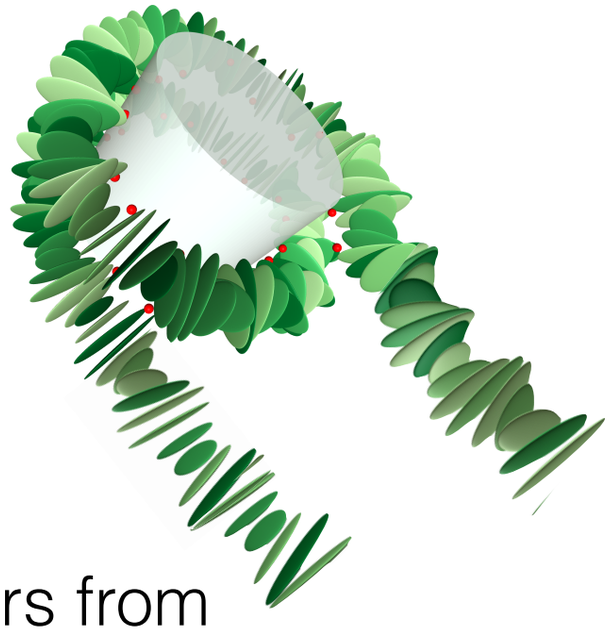


parameters from
DNA-protein cocrystals



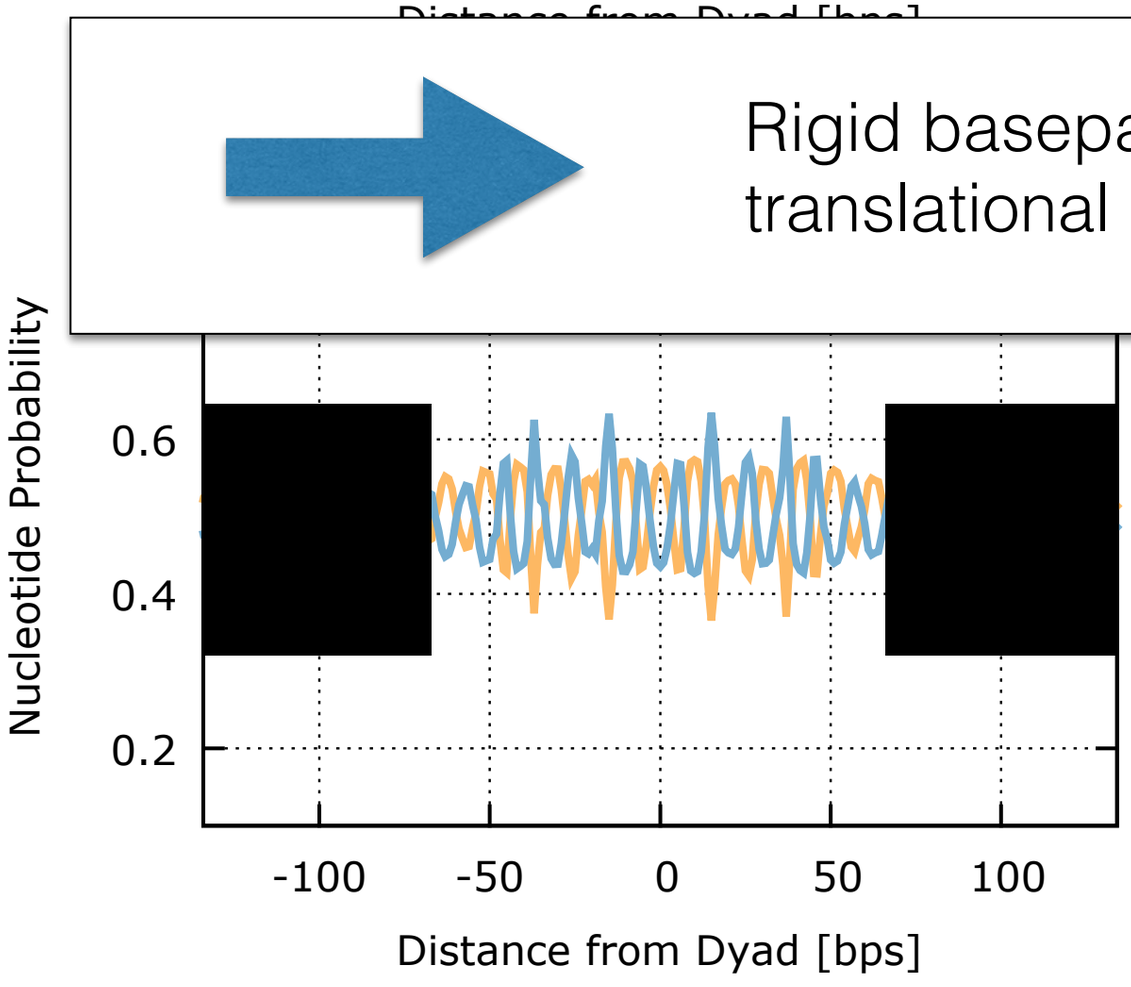
as above but MD stiffnesses
of oligonucleotides





parameters from
DNA-protein cocrystals

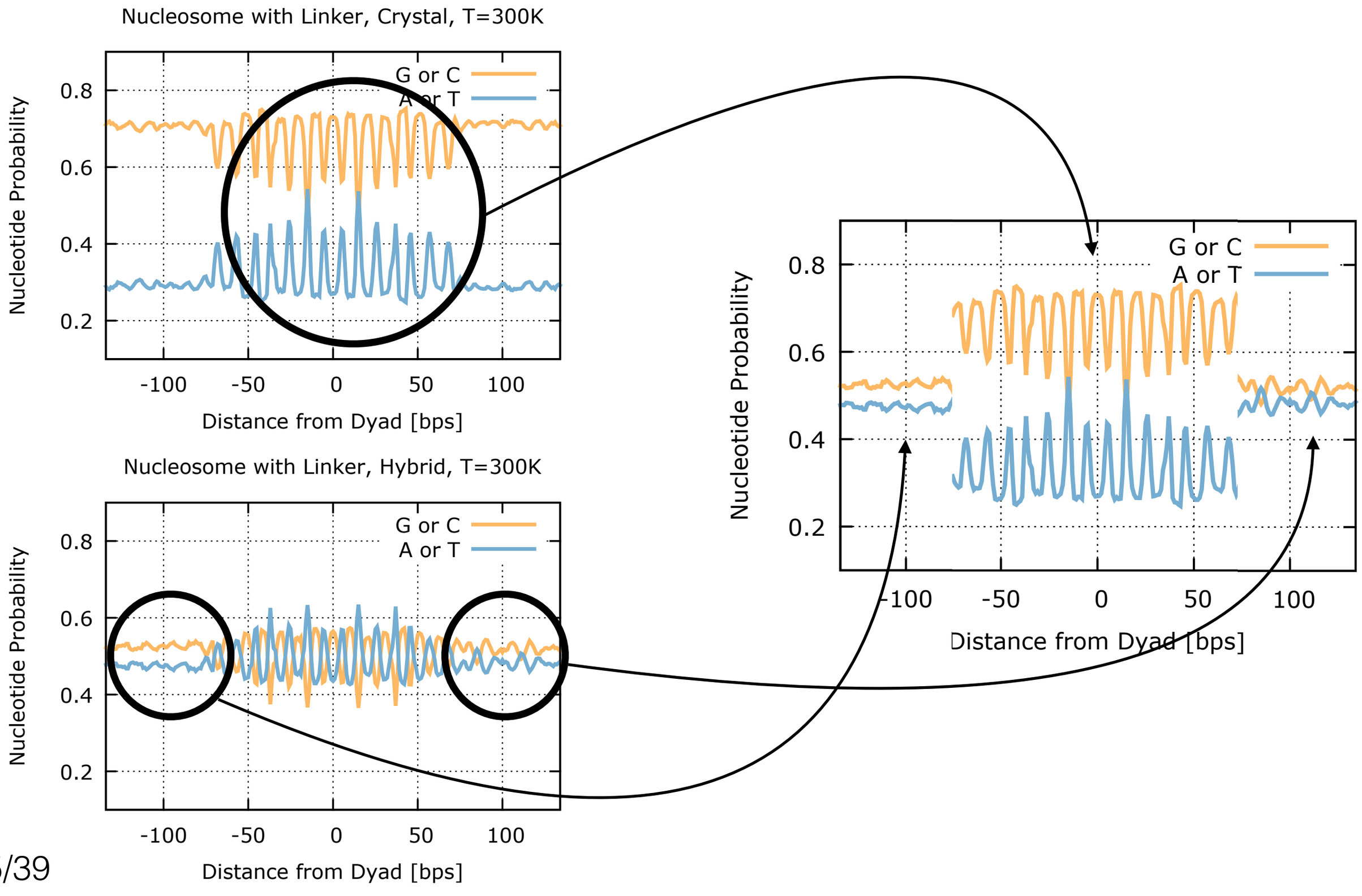
Rigid basepair model cannot predict
translational nucleosome positioning



as above but MD stiffnesses
of oligonucleotides

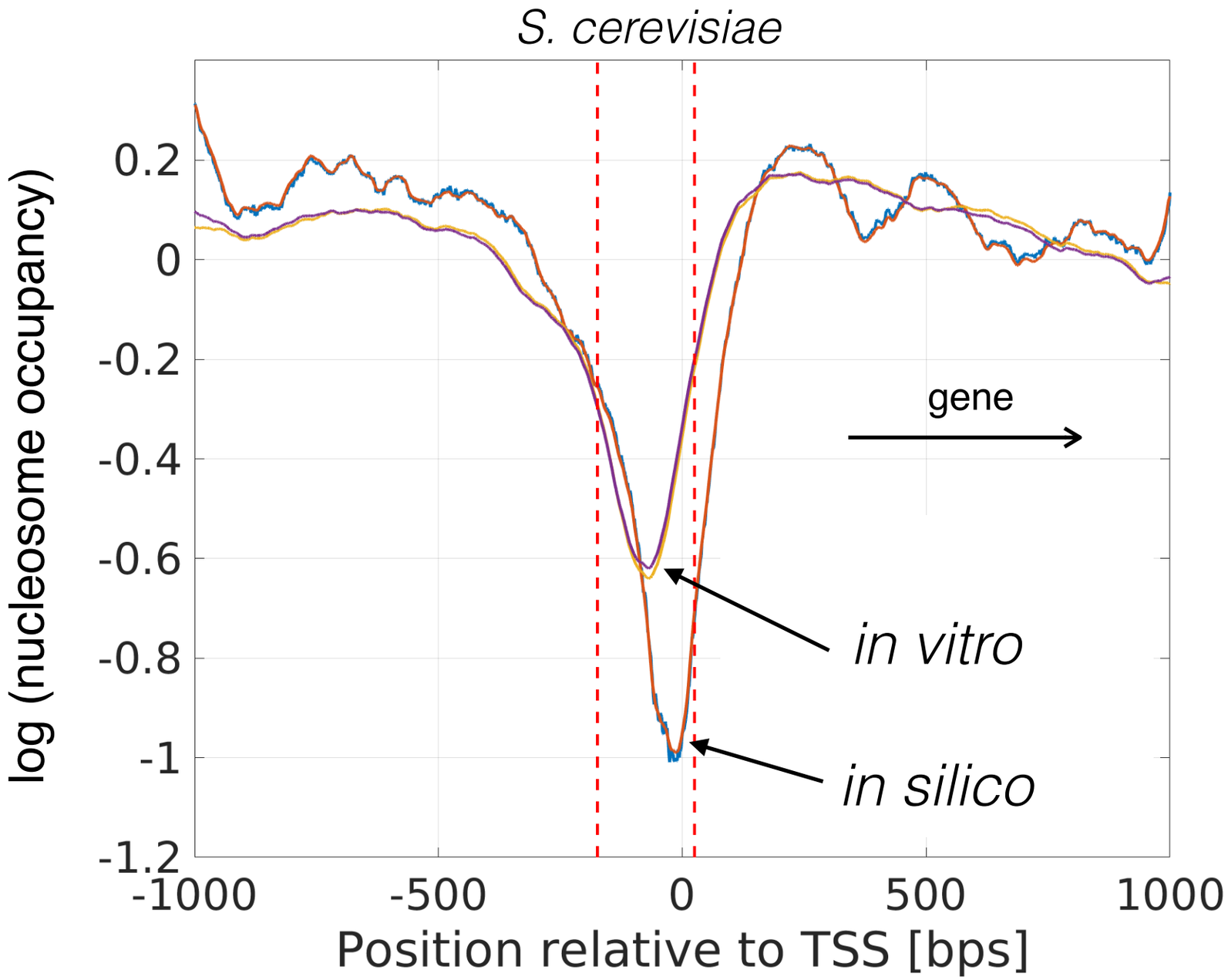
Multiharmonic model

Neipel, Brandani, Schiessel, Phys. Rev. E 101, 022405 (2020)



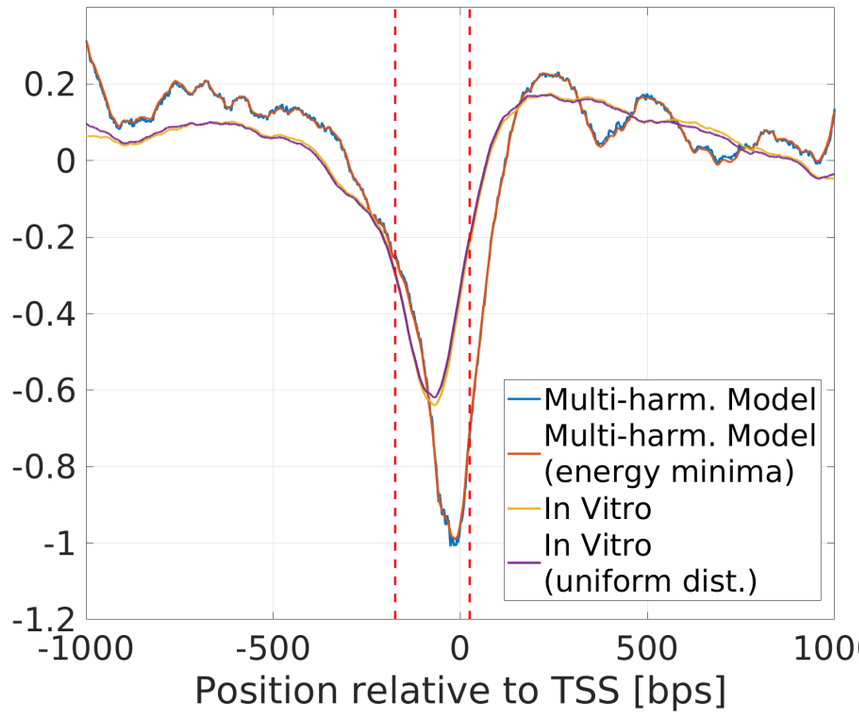
Nucleosome occupancy around transcription start sites

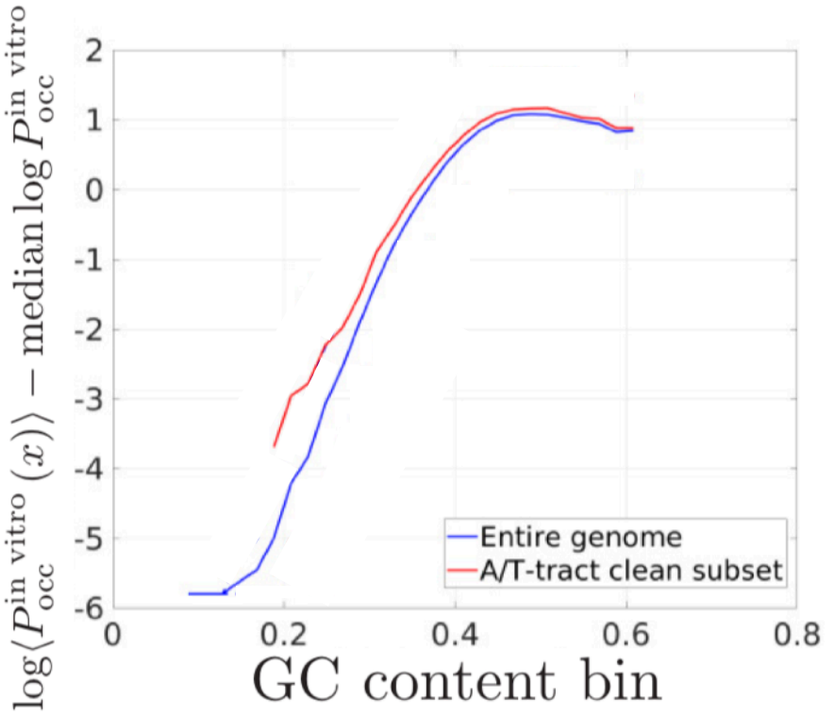
Neipel, Brandani, Schiessel, Phys. Rev. E 101, 022405 (2020)

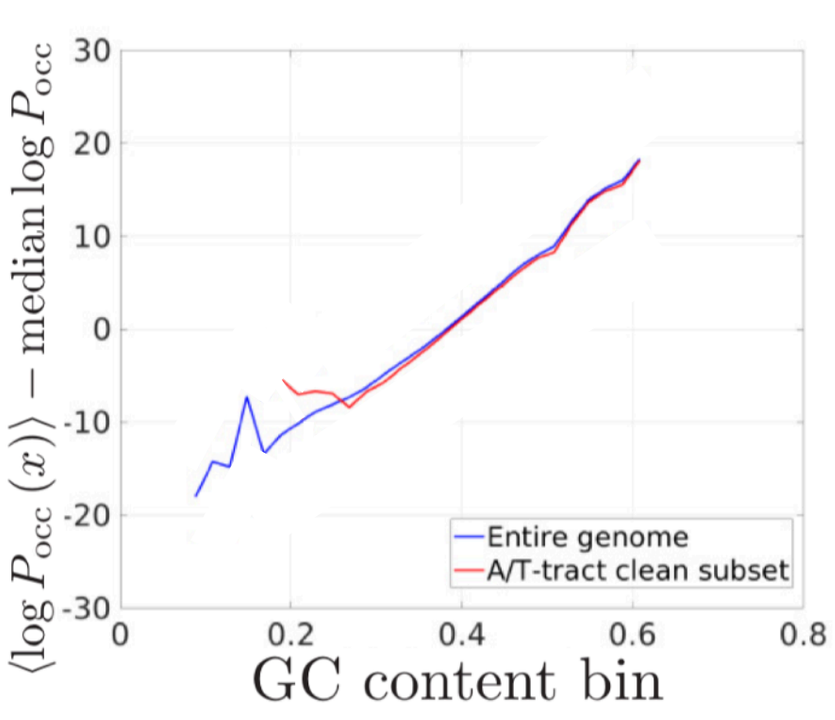


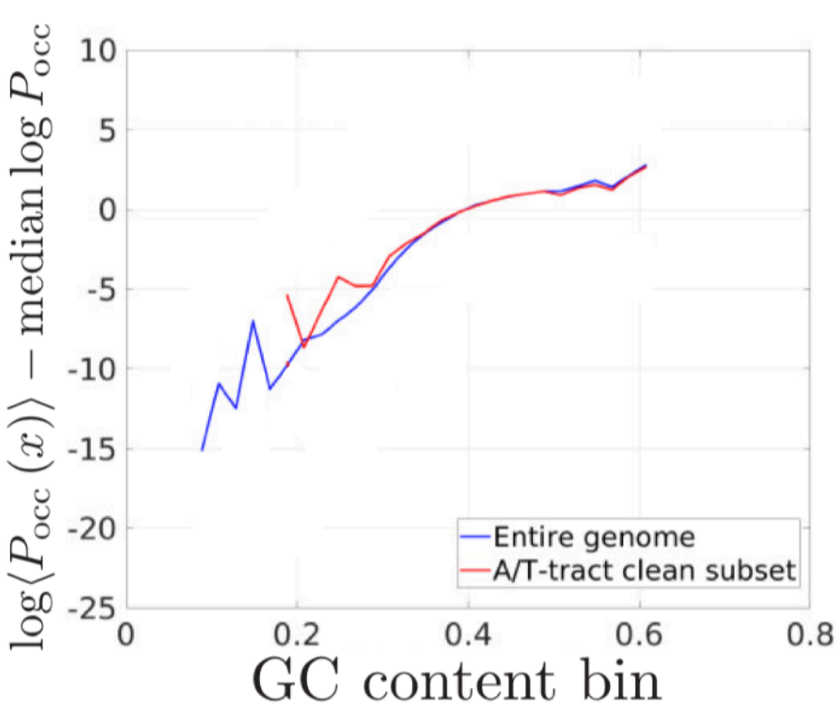
Reconstituted chromatin is not equilibrated

Neipel, Brandani, Schiessel, Phys. Rev. E 101, 022405 (2020)









in vitro model

THE MECHANICAL GENOME

1. Positioning rules are mechanical
2. Multiplexing of the 2 (and even 3) layers
3. Mechanical cues on real genomes

Artwork: Amar van Leewaarde & Marco Tompitak

Acknowledgements

Theoretical Physics of Life Processes

Current

Bahareh Shakiba

Lennart de Bruin

Martijn Zuiddam

Koen van Deelen/Adi Kiang/
Steven Lankhorst/Michael Wellens

Former

Behrouz Eslami-Mossallam

Marco Tompitak

Jonas Neipel

Raoul Schram

Joeri Wondergem

Jamie Culkin

Collaborations

John van Noort
Leiden Univ

Gerard Barkema
Utrecht Univ

Cédric Vaillant
Ralf Everaers
ENS Lyon

Ralf Blossey
Lille Univ

Giovanni Brandani
Kyoto University

Tetsuya Yamamoto
Hokkaido University

Takahira Sakaue
Aoyama Gakuin Univ.

Shelley Sazer
Baylor College Houston

Alexey Onufriev
Virginia Tech

Rob Phillips
Caltech, Pasadena