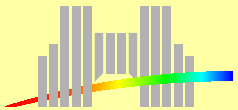
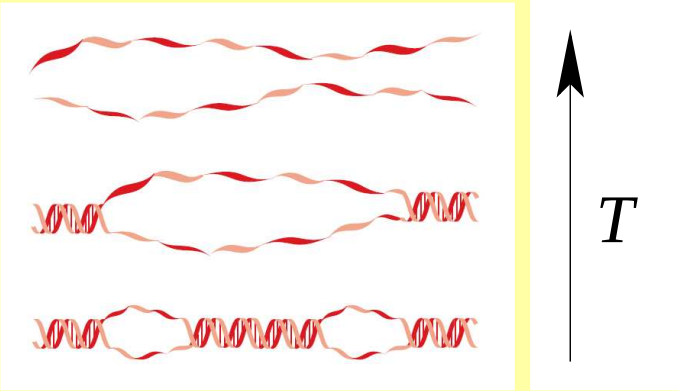

ECOLE NORMALE SUPERIEURE DE LYON

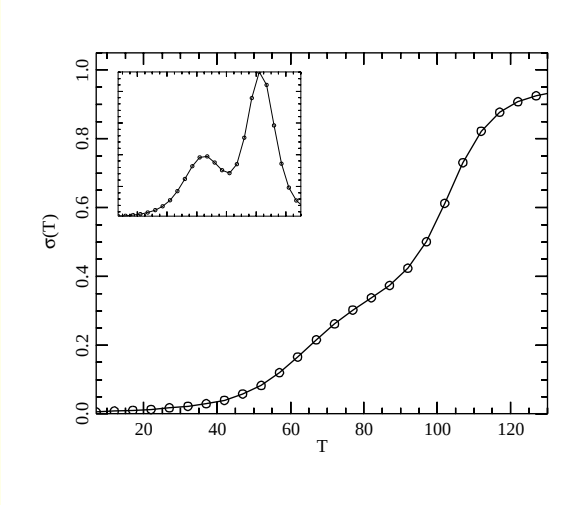
Adding a dimension to DNA melting curves

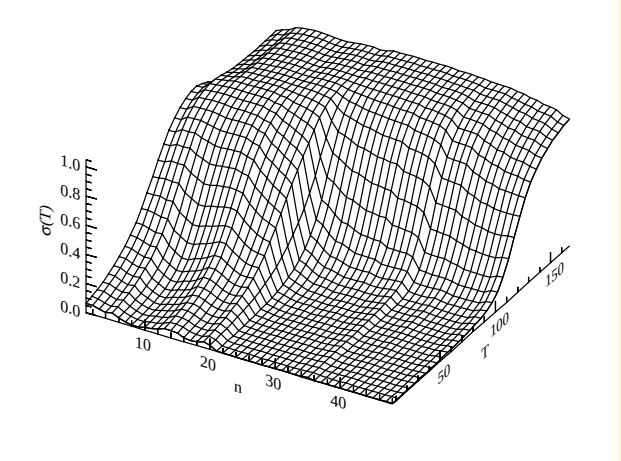
Experimental and theoretical studies of sequence effects on the fluctuations and melting of short DNA molecules

Michel Peyrard, Santiago Cuesta Lopez and Dimitar Angelov
Ecole Normale Supérieure de Lyon, France
Michel.Peyrard@ens-lyon.fr



(M. Peyrard, Nature Physics, **2**, 13-14 (2006))





ECOLE NORMALE SUPERIEURE DE LYON

- Typeset by FoilTEX

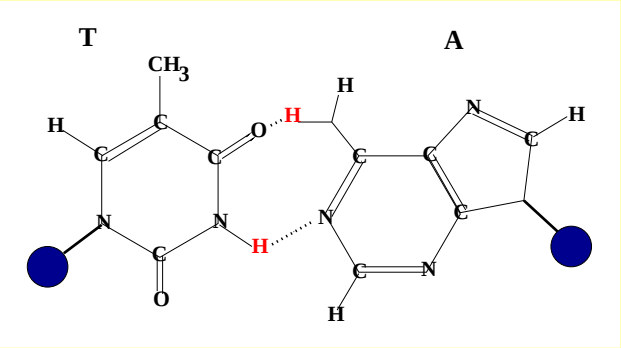
DNA is a highly dynamical entity.

“Breathing of DNA” has been known from biologists for decades.

- Observation: kinetic of imino-proton exchanges

Guéron et al. used NMR.
(Gueron et al. Nature **328**, 89 (1987))

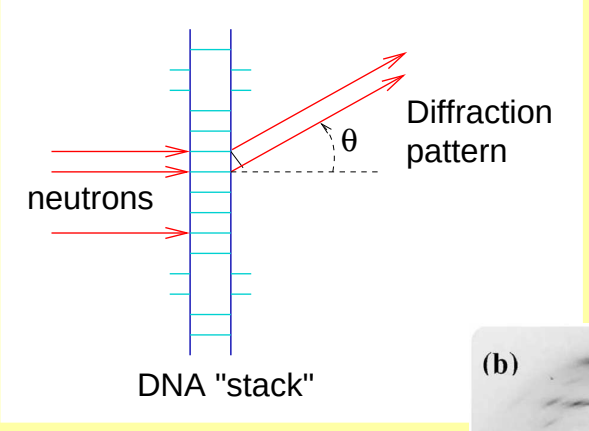
- Lifetime of a base-pair (closed time): 10 ms.
- **Individual** base pair opening (neighbours have different lifetimes)



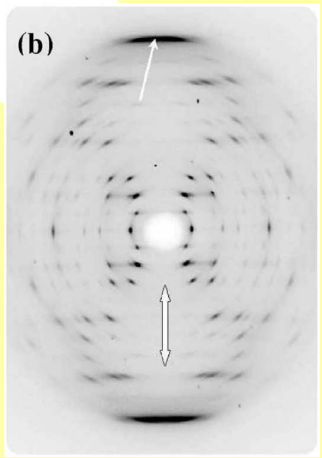
DNA fluctuations can be studied with various experimental methods

At the scale of the full molecule:

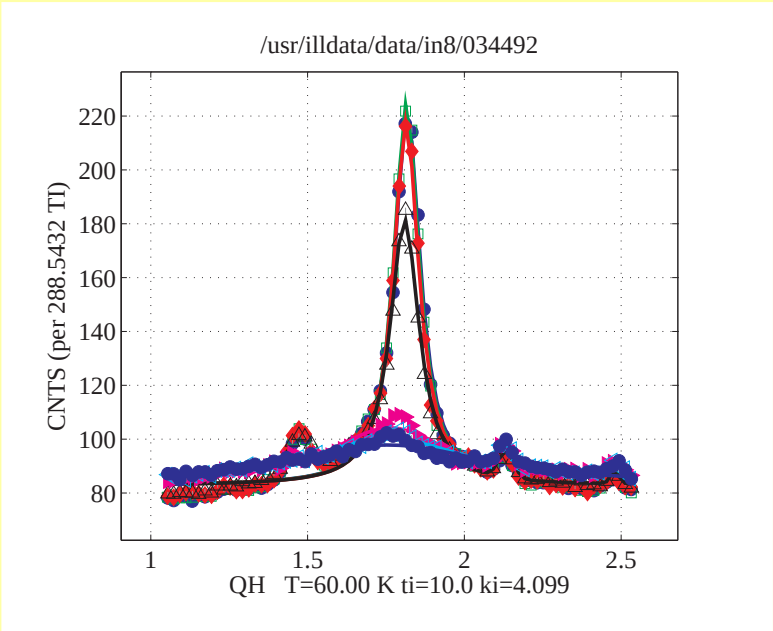
Neutron diffraction (with Andrew WILDES, ILL, Grenoble)



Measure the coherence length of the base stacking, which makes a diffraction grating.



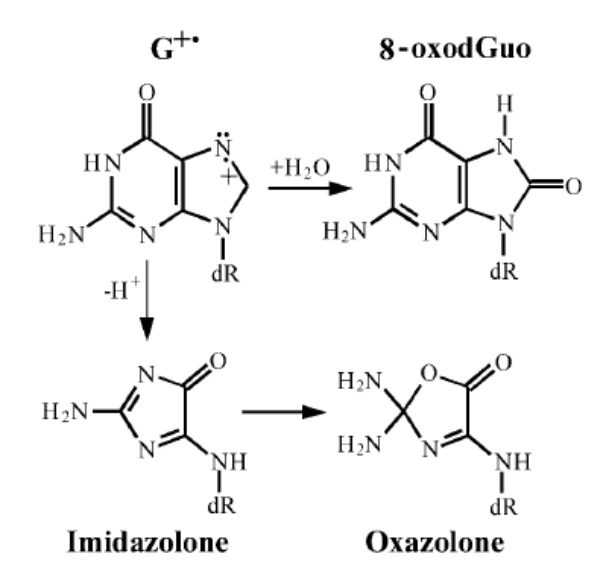
M. Krisch, et al. Phys. Rev. E **73** 061909-1-10 (2006)



$T = 60, 75, 78, 79, 80, 81\ °C$

UV Laser oxidative Guanine modifications: a probe of local stacking fluctuations.

High-intensity UV laser pulses → guanine oxidative lesions, followed by electronic redistribution strongly modulated by the local structure.



Oxozalone sensible to cleavage by piperidine glycosylase

8-oxodG sensible to cleavage by Fpg protein, and appears only in helicoidal stacking

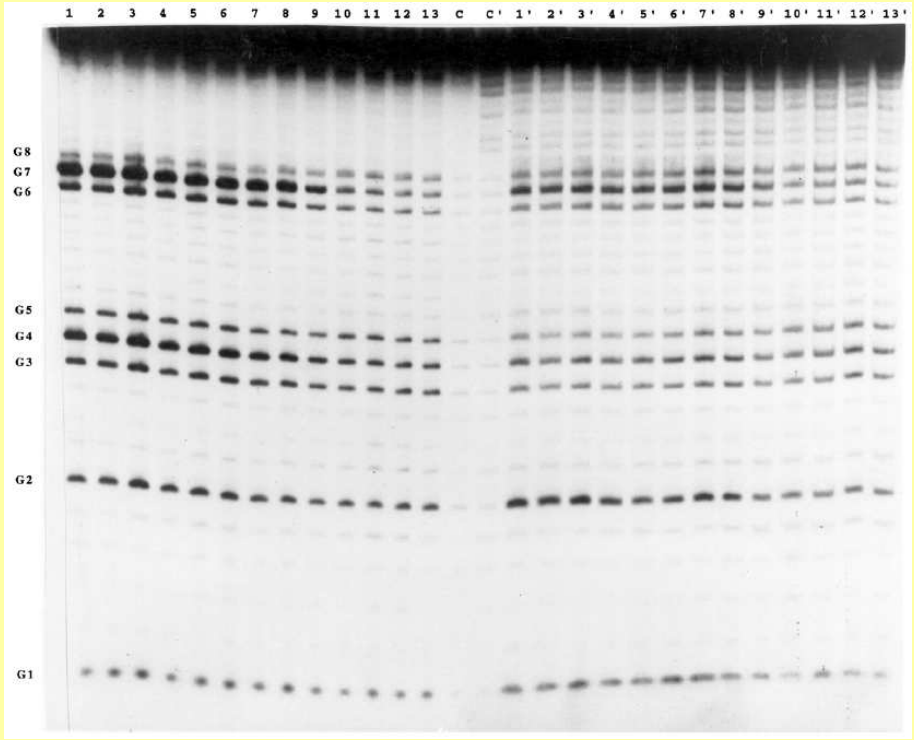
→ R_{pip}/R_{FpG} indicative of local opening

Opens the possibility to evaluate fluctuational opening versus sequence.

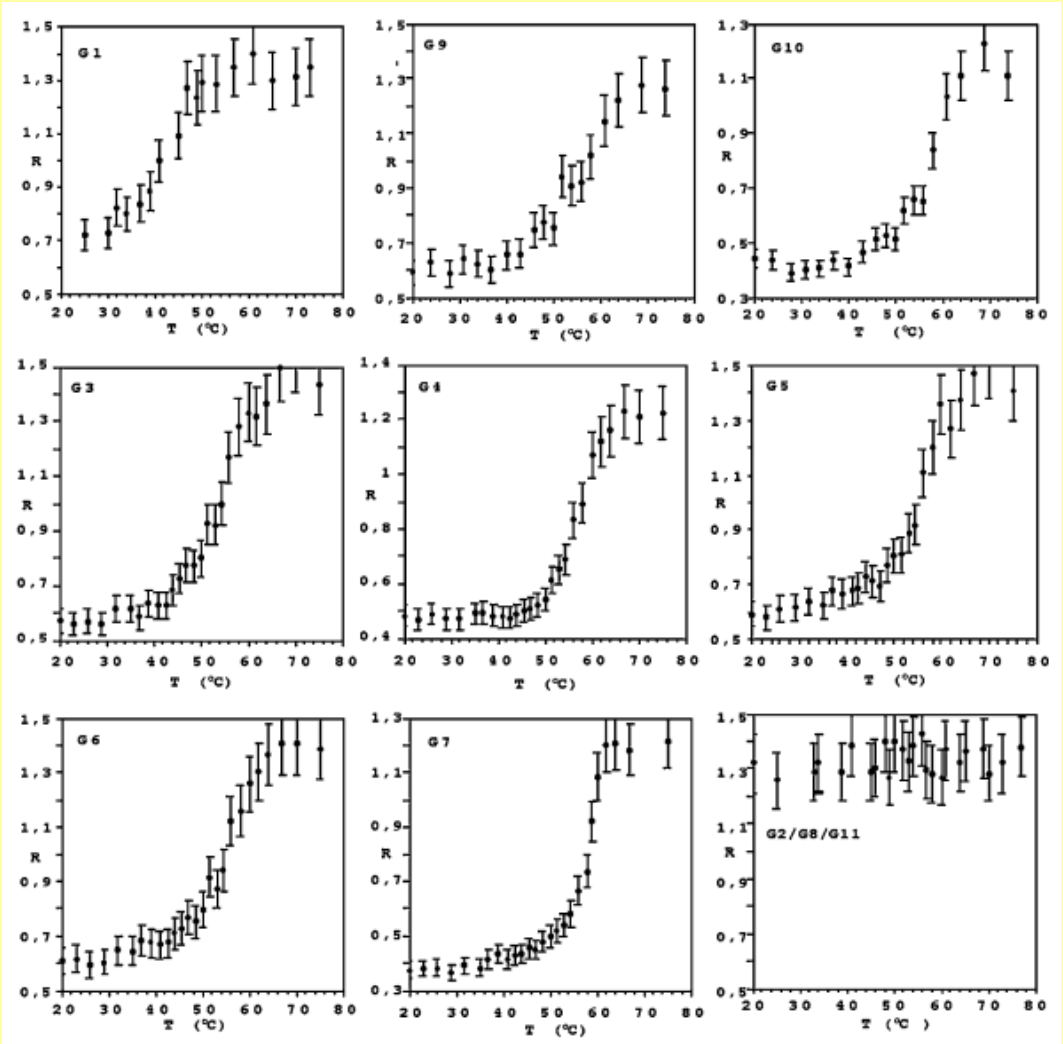
From A. Spassky and D. Angelov, J. Mol. Biol. **323** 9 (2002)

```
                1     2    345        678
5'-ATAAAAGTTACGCTAGGGATAACAGGGTAATATAACG-3'
3'-TATTTTCAATGCGATCCCTATTGTCCCATTATATTGC-5'
                  11 10          9
```

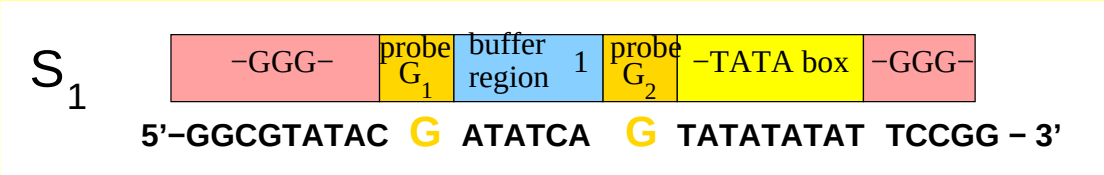
FpG protein hot piperidine

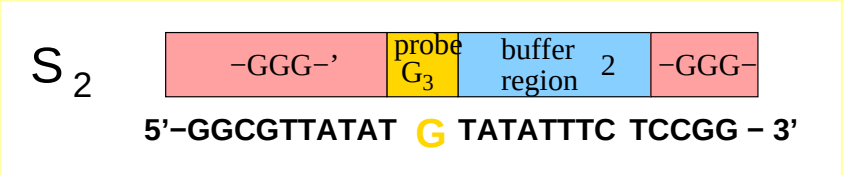


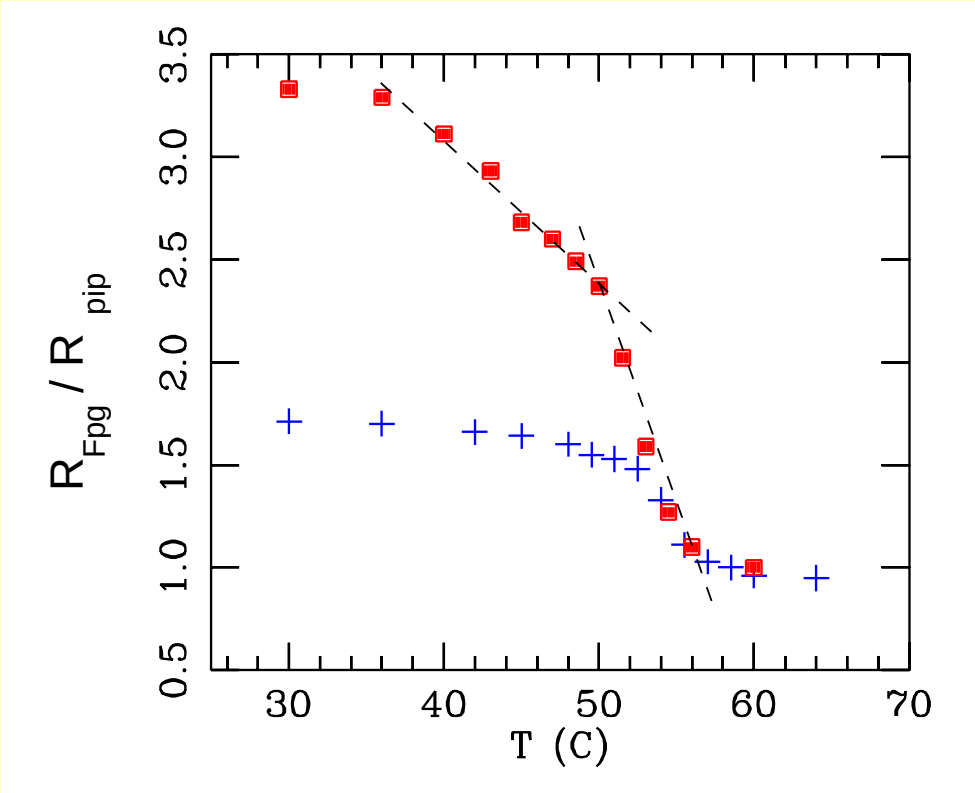
Provides data for all *GC* pairs, contrary to fluorophore–quencher methods.

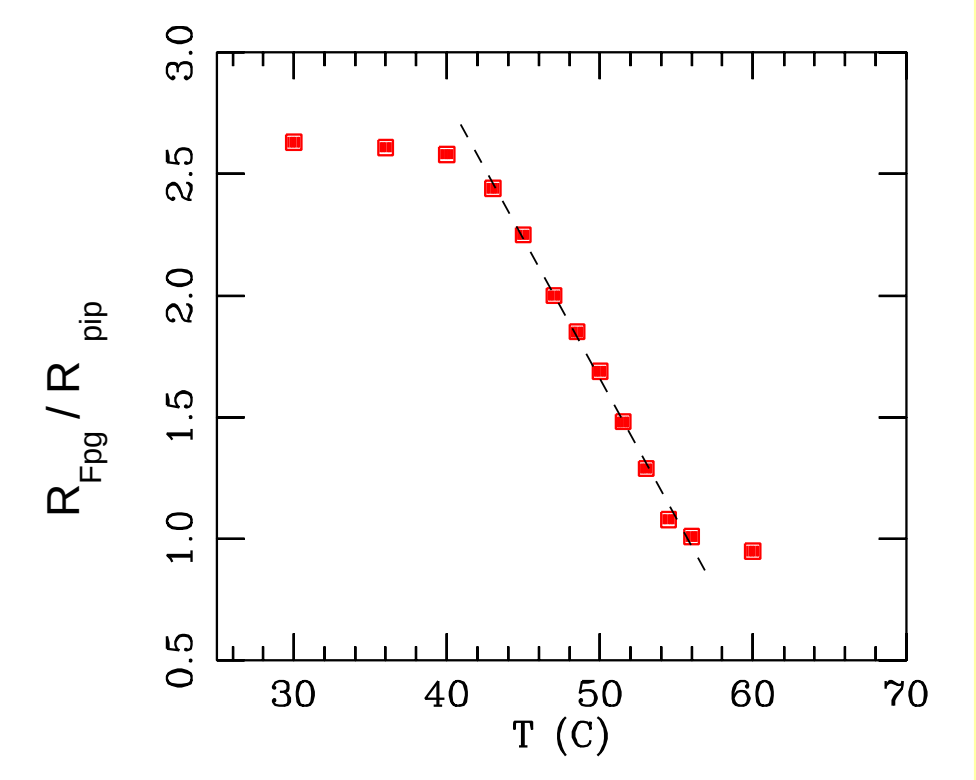


The fluctuations of a TATA box affect DNA to some distance



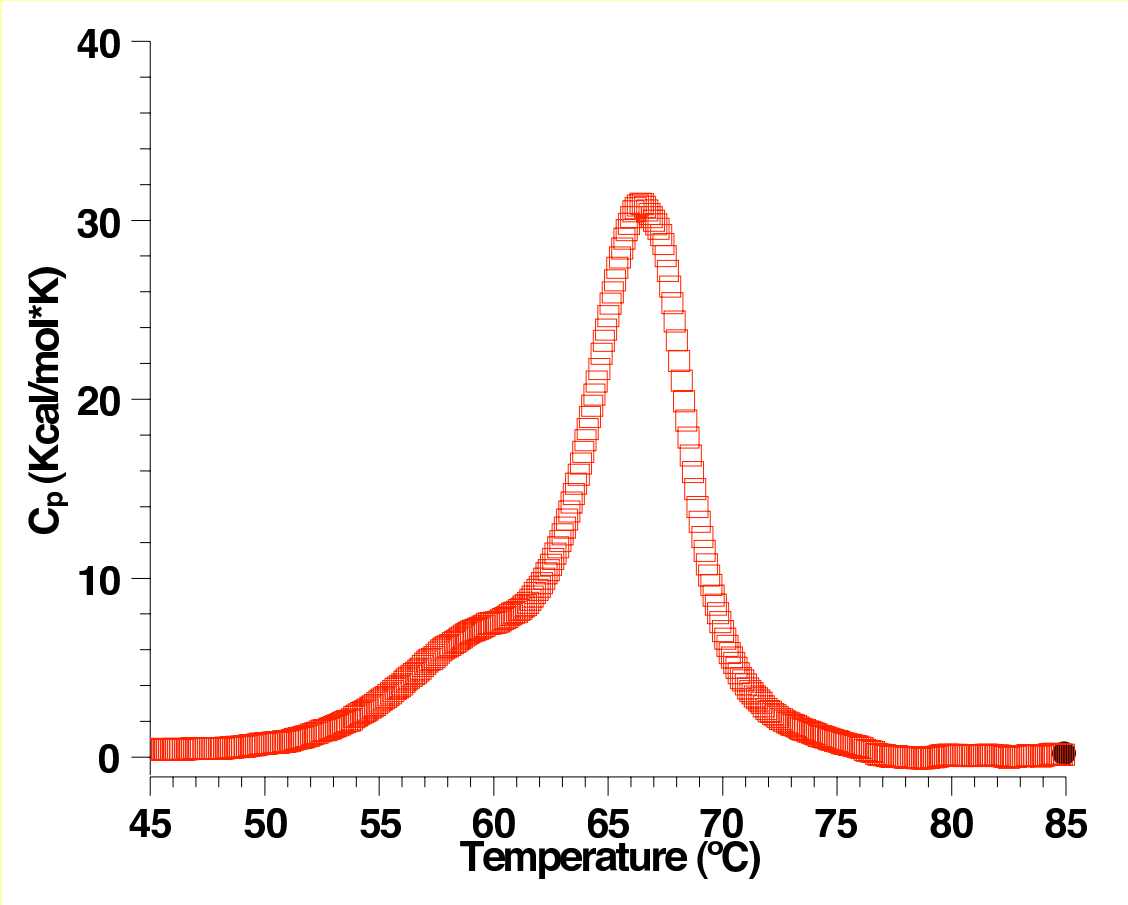






The precursor effects associated to the TATA box are also detected in calorimetry experiments

Differential scanning calorimetry on a tiny amount of DNA

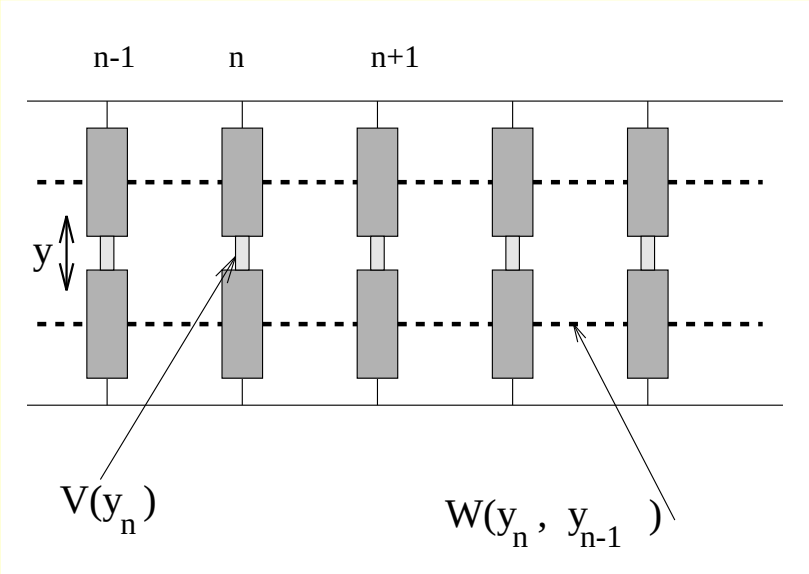


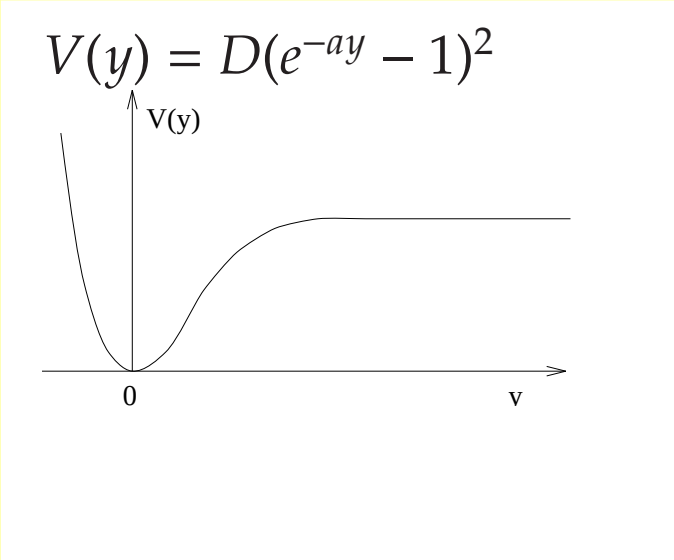
(Temperature shift due to a different ionicity)

Modelling sequence effects: a nonlinear model of DNA melting

An (old) starting point:

(M. Peyrard and A.R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989))

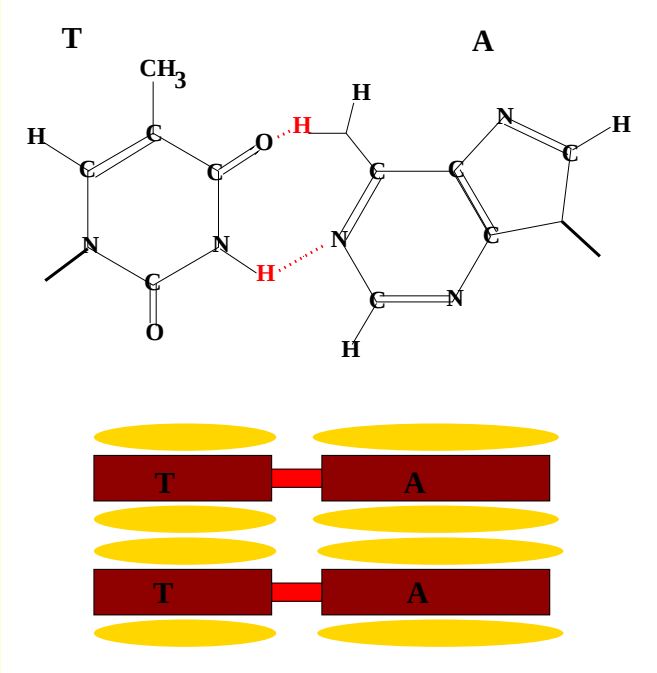




DNA structure only encoded in the potentials

$$H = \sum_n \left[\frac{1}{2m} p_n^2 + W(y_n, y_{n-1}) + V(y_n) \right] \qquad p_n = m \frac{dy_n}{dt}$$

Coupling potential W (stacking) : crucial for the statistical mechanics
(T. Dauxois, M. Peyrard and A.R. Bishop, Phys. Rev. E **47**, R44-R47 (1993))



$$W(y_n, y_{n-1}) = \frac{1}{2}K\left[1 + \rho e^{-\alpha(y_n + y_{n-1})}\right](y_n - y_{n-1})^2$$

if y_n **and** $y_{n-1} \ll 1/\alpha \Rightarrow K' \approx K(1+\rho)$
if y_n **or** $y_{n-1} \gg 1/\alpha \Rightarrow K' \approx K$

Model parameters:

Actual potential interactions → can be *estimated* from the physics of DNA and refined by comparison with experiments

Dynamics of DNA fluctuations: the model has to be improved

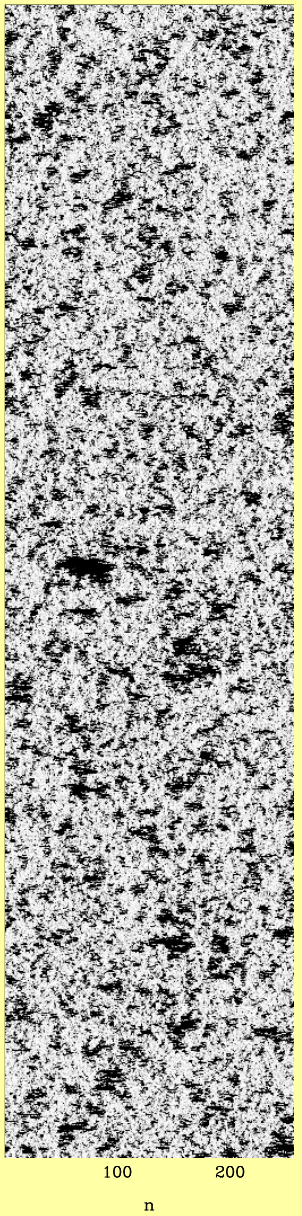
Simulation $\rightarrow$ 20 ns
Experiment:

- lifetime of an open state 30 to 300 ns
- lifetime of a closed base pair 1 to 10 ms

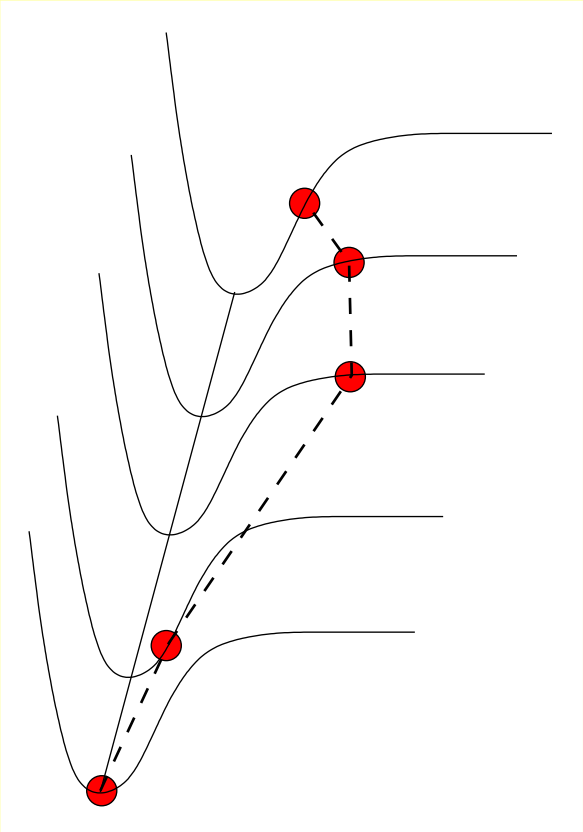
Nonlinear model: discrete breathers.

they may have a lifetime of hundreds of picoseconds (with thermal fluctuations) but opening and closing have periods of a few ps.

Disagreement by several orders of magnitude!

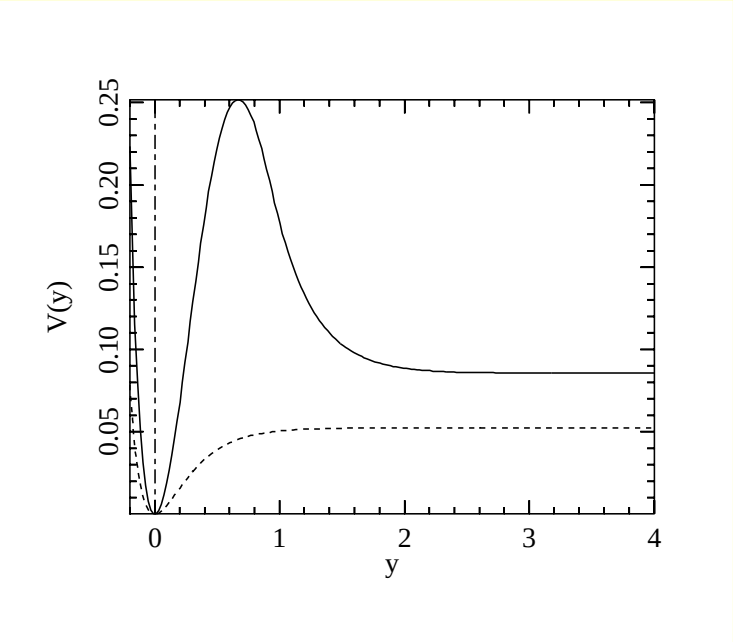


The Morse potential must be changed:



$$F = U - TS$$

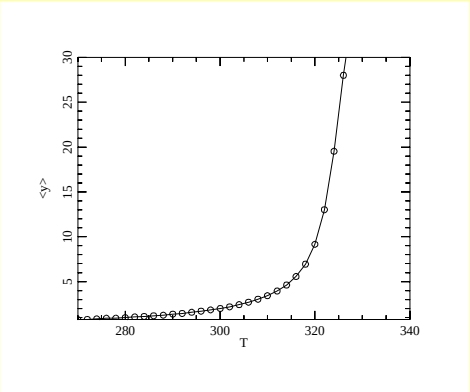
For a mesoscopic model, the “potential” is a free energy
(includes missing degrees of freedom)



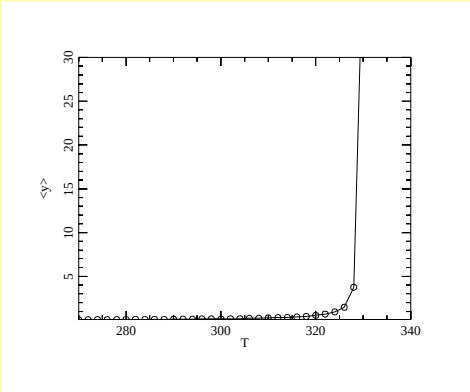
$$V_h(y) = D\left(e^{-\alpha y} - 1\right)^2 + \Theta(y)\frac{by^3}{\cosh^2[c(\alpha y - d\ \ln 2)]}$$

The equilibrium properties are hardly affected (the melting transition gets sharper)

Morse

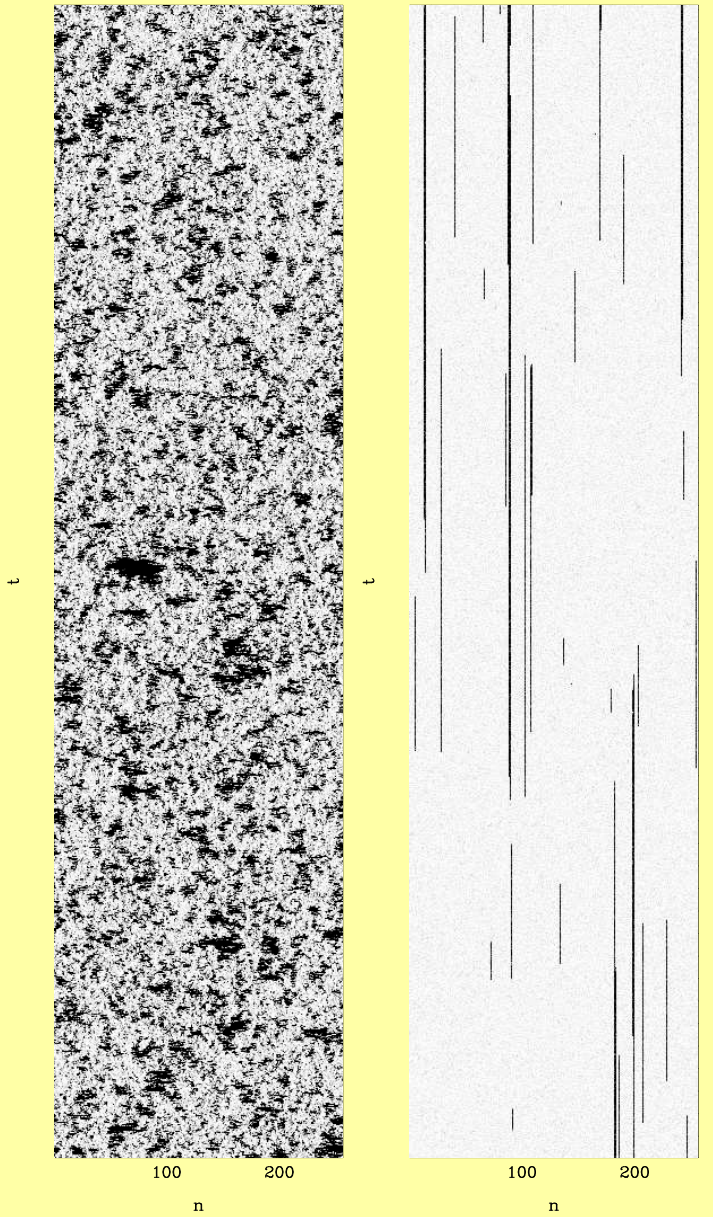


Modified Morse



The dynamics is drastically changed:

	Morse	Modified M.
Lifetime open	0.08 ns	7 ns
Lifetime closed	0.4 ns	$0.4\ \mu$s



Computing DNA melting profiles

Statistical mechanics (computing the local probability of opening) better than molecular dynamics which gives very bad statistics for large open bublles.

Introducing the sequence in the model:

It affects base pairing **and** base stacking

Potential V	D	α	b	c	d
AT base pair	0.09075 eV	3.0 Å^{-1}	4.00 eV	0.74 Å^{-1}	0.20
GC base pair	0.09900 eV	3.4 Å^{-1}	6.00 eV	0.74 Å^{-1}	0.20

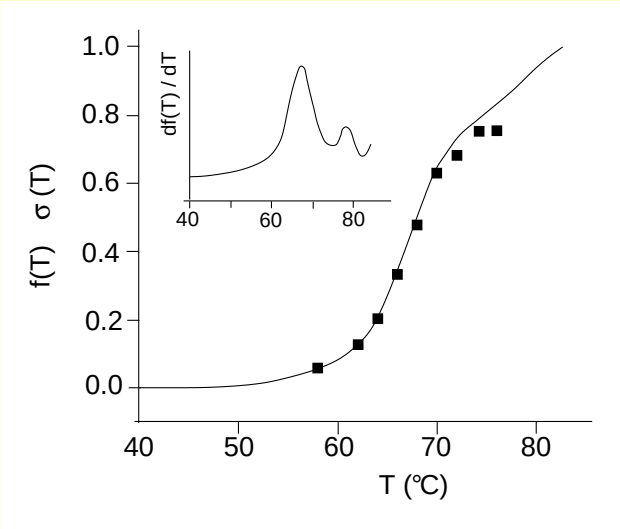
Potential W	$\rho = 25$			$\delta = 0.8$ Å^{-1}		
Dimer	A–T	A–A	T–T	G–T	A–C	T–A
K (eV Å^{-3})	0.00176	0.00418	0.00418	0.00480	0.00462	0.00506
Dimer	G–A	T–C	C–C	G–G	G–C	C–T
K (eV Å^{-3})	0.00546	0.00546	0.00810	0.00810	0.00865	0.00865
Dimer	A–G	C–A	T–G	C–G		
K (eV Å^{-3})	0.00865	0.01140	0.01140	0.01690		

Some sequences are particularly challenging !

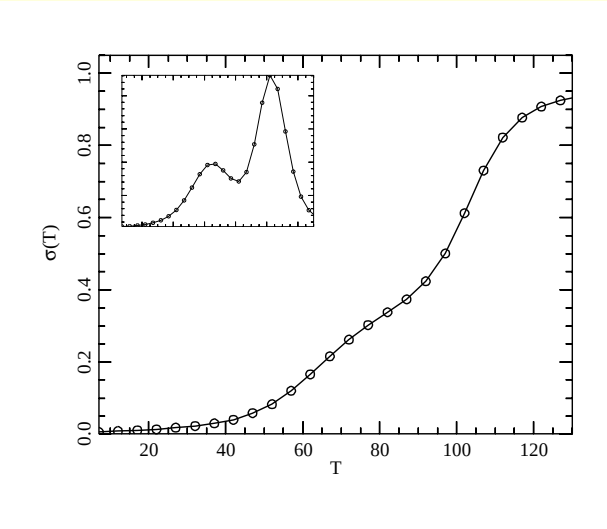
1 7 10 21 30 35 40 42 48

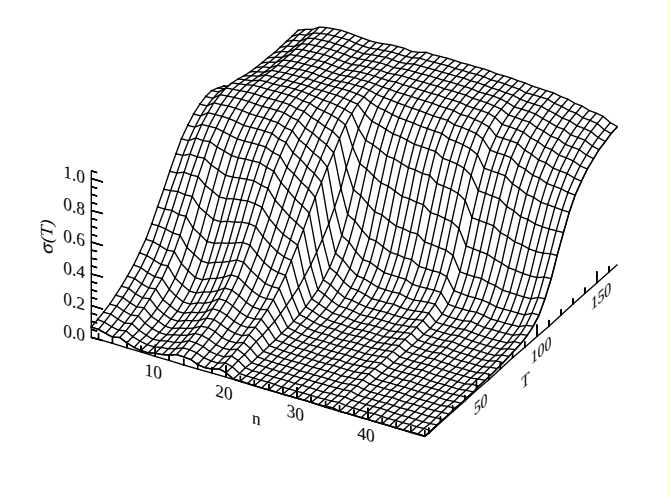
CA TAAT ACT TTATATTTAATT GGCGGCG CAC GG GAC CC GTG CGCCGCC

LAS48, from A. Montrichok? G Gruner, G. Zocchi, Europhys. Lett **62**, 452 (2003)



Experiment





Theory

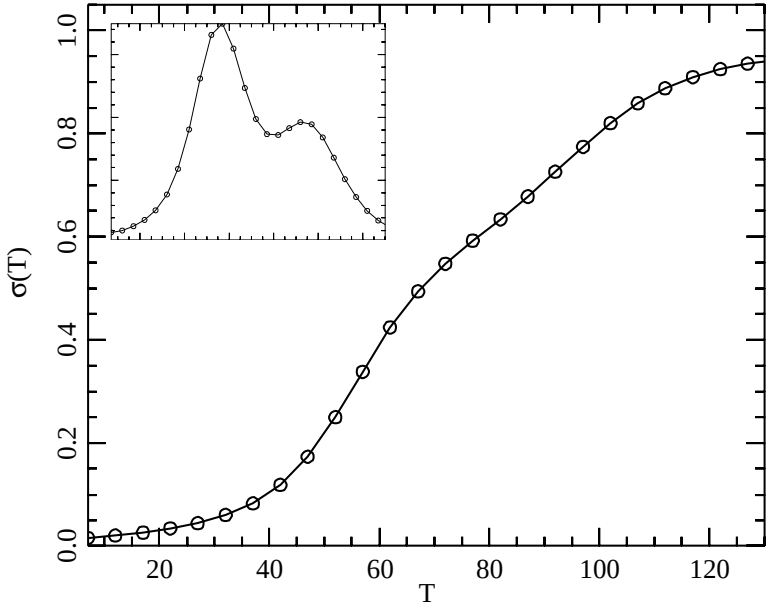
Solving the discrepancy between theory and experiment: Introducing non local effects

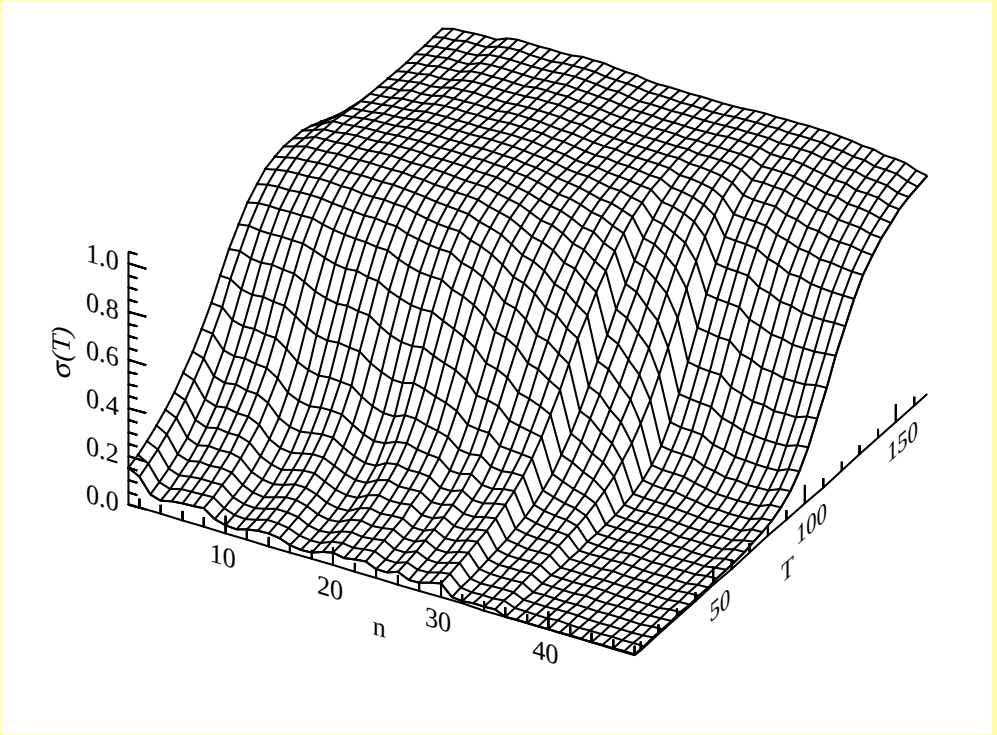
Stacking interactions are softened in the vicinity of an AT-rich region

Presently set of empirical rules (valid for *any* sequence)

- Only AT-rich regions which exceed a given size of k consecutive AT pairs are taken into account. *(minimum size)*
- Inside the AT-rich domain, the parameters are not modified.
- for all stacking interactions which are less than n_1 bonds from the AT-rich region, the coupling constant K is multiplied by a factor $c_0 < 1$, *(softening)*
- for all interactions which are n bonds away from the AT-rich region, $n_1 > n > n_2$, K is multiplied by $c_1 = c_0 + (1 - c_0)(n - n1)/(n_2 - n_1)$ in order to reach smoothly the value 1 n_2 bonds away from the AT-rich region.
- for all stacking interactions which are at least n_2 bonds away from the AT-rich region, the parameters of the interaction potential are not modified. *(maximum range)*
- if a stacking bond is affected by more than one AT-rich region, which could be the case for the stacking interactions which are between two AT-rich domains, only the largest correction is considered.

Result for LS48 sequence



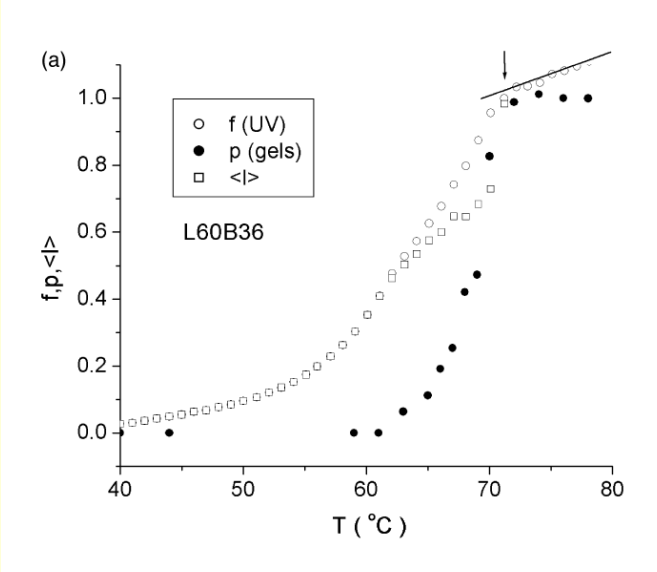


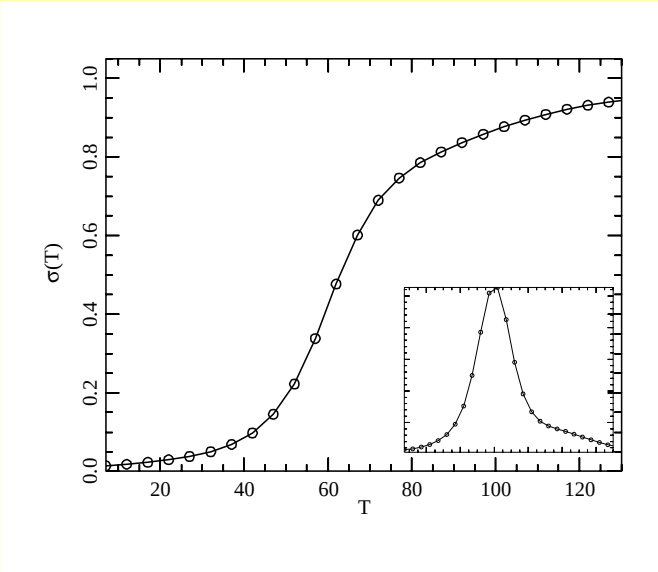
Qualitative, and semi-quantitative agreement with experiment (transition too broad, entropy effects not properly described yet)

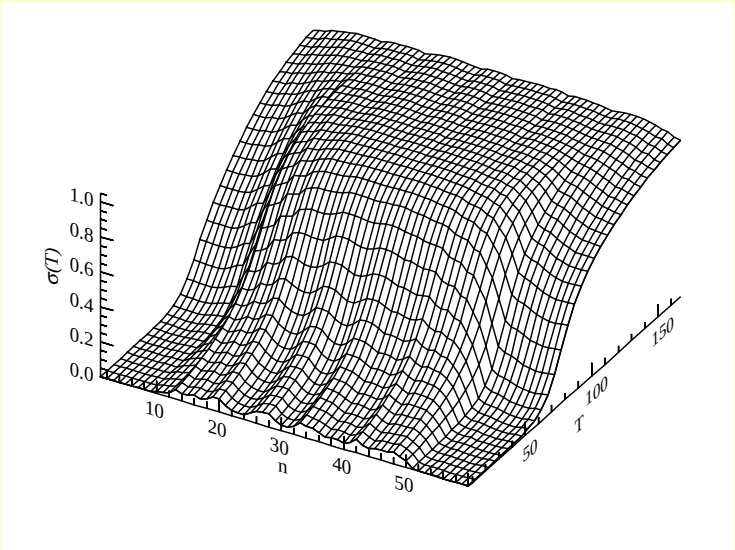
Other sequences

Bubble in the middle

C C G C C A G C G G C G T T A T T A C A T T T A A T T C T T A A G T A T T A T A A G T A A T A T G G C C G C T G C G C C







Same parameter set also valid for other sequences

Conclusion

- Using UV bi-photonic excitations and selective cleavage, it is possible to map the opening along the sequence, at the G sites

- Experiments indicate that the large fluctuations at the TATA box affect base pairs up to 10 base pairs away

- A theoretical model can be designed to describe DNA fluctuations for short segments (a few tens of base pairs)

- For sequences containig a large TATA box, a correct qualitative results can only be obtained when "non-local" effects are taken into account.

  $\Rightarrow$ both experiments and theory converge, but the nature of these "long range" effects has yet to be clarified