DNA hybridization

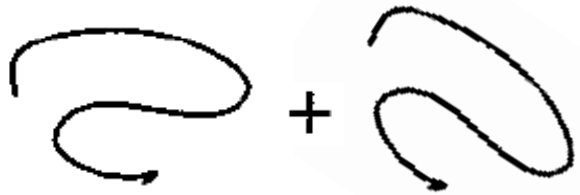
1- INTRODUCTION

Specific and non specific interactions, inter- and intra-molecular:
- competition between enthalpy and entropy (binding energies versus probabilities)

Specific interactions play an important role in the biological cell:
- Enzymes
- catalytic RNA (ribozymes)
- transcription regulation
- iRNA
- …

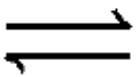
How to mesure, quantify and model the process of specific recognition between biomolecules?

Here: DNA hybridization at thermal equilibrium
The most studied and most quantitatively described example



Single-stranded DNA

Hybridization



Denaturation
Melting

Double-stranded DNA

2- Experimental and theoretical approaches

Several measurement techniques

Principle: Measure the fraction of closed base pairs as a function of temperature

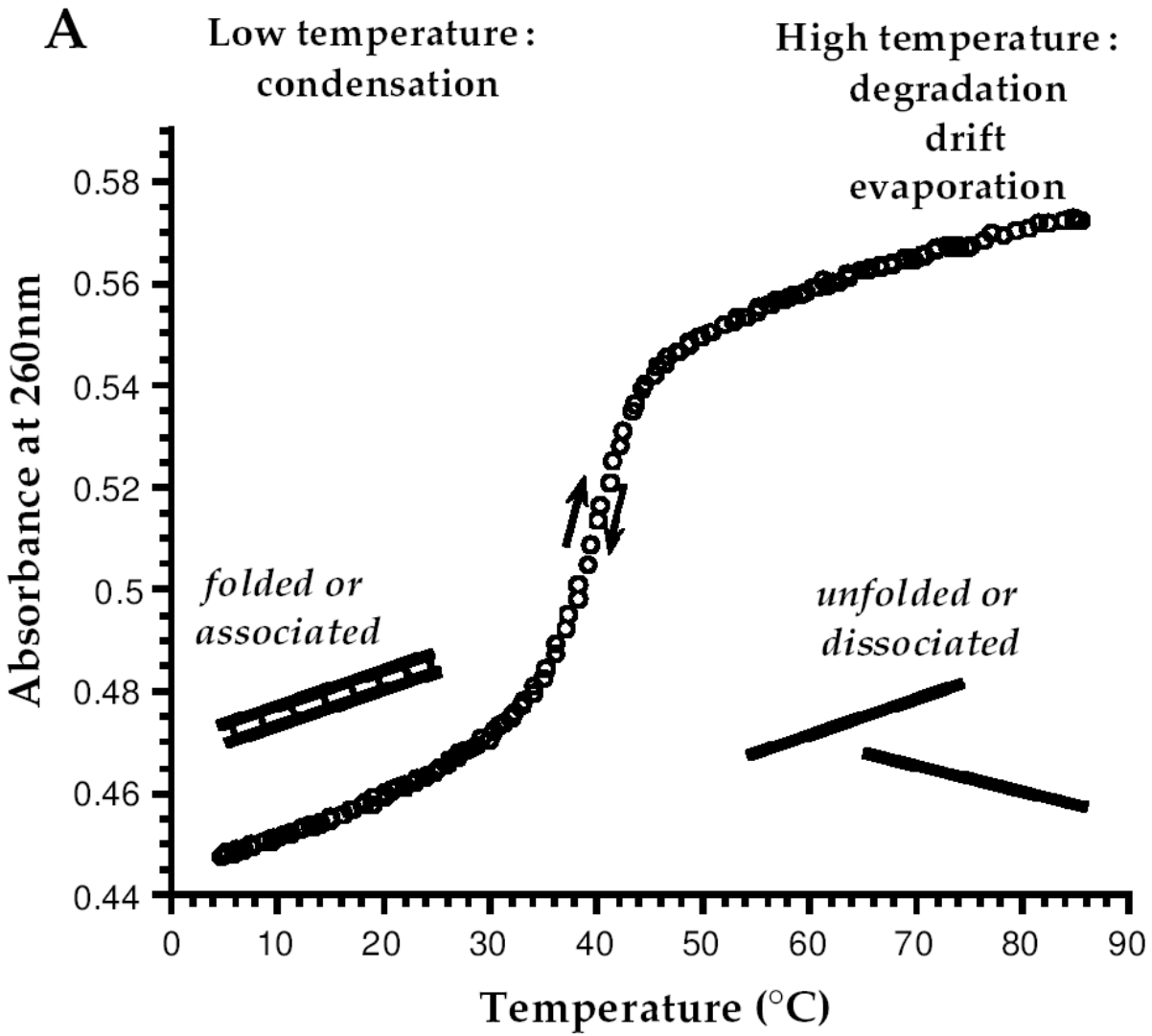
Several techniques for obtaining a signal that depends on the open or closed state:

- UV absorption (often at 260 nm)
- Fluorescence (intercalating dyes, FRET)
- NMR
- Raman diffusion

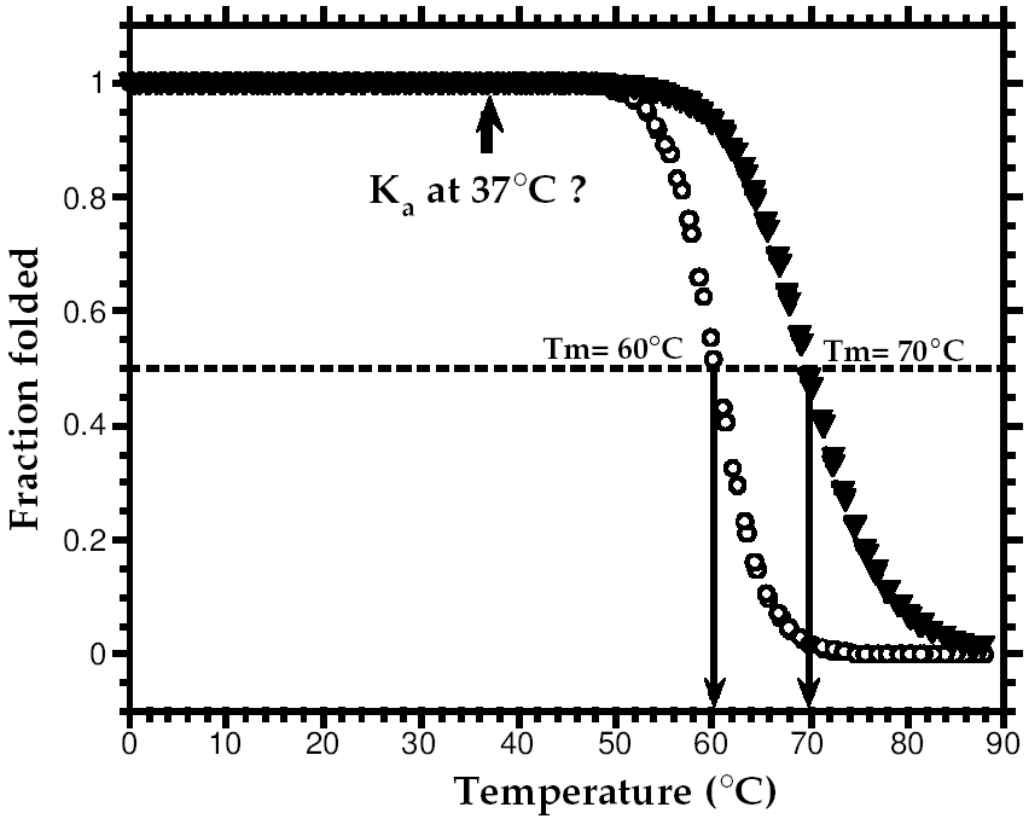
Pay attention to residual temperature dependencies

Changes in enthalpy ΔH can be measured by direct calorimetry

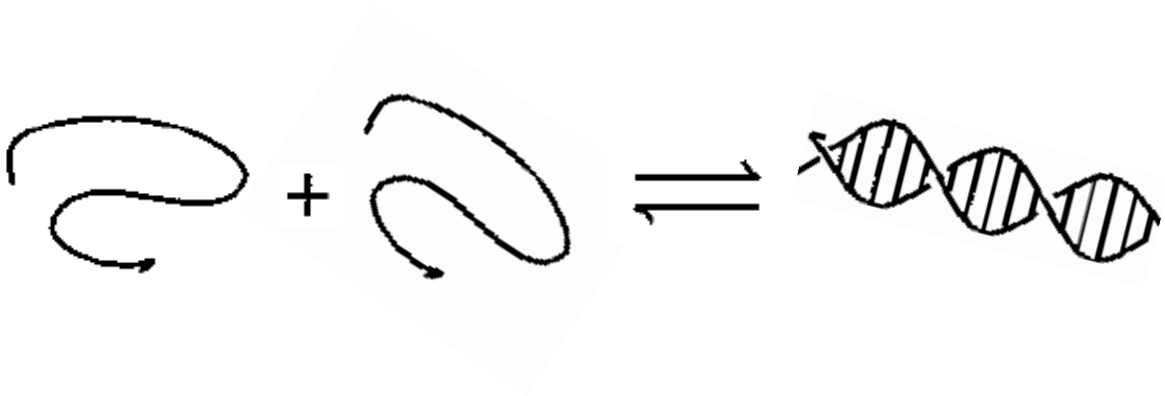
A typical measurement



The need for theoretical analysis: what duplex is more stable at 37° C?

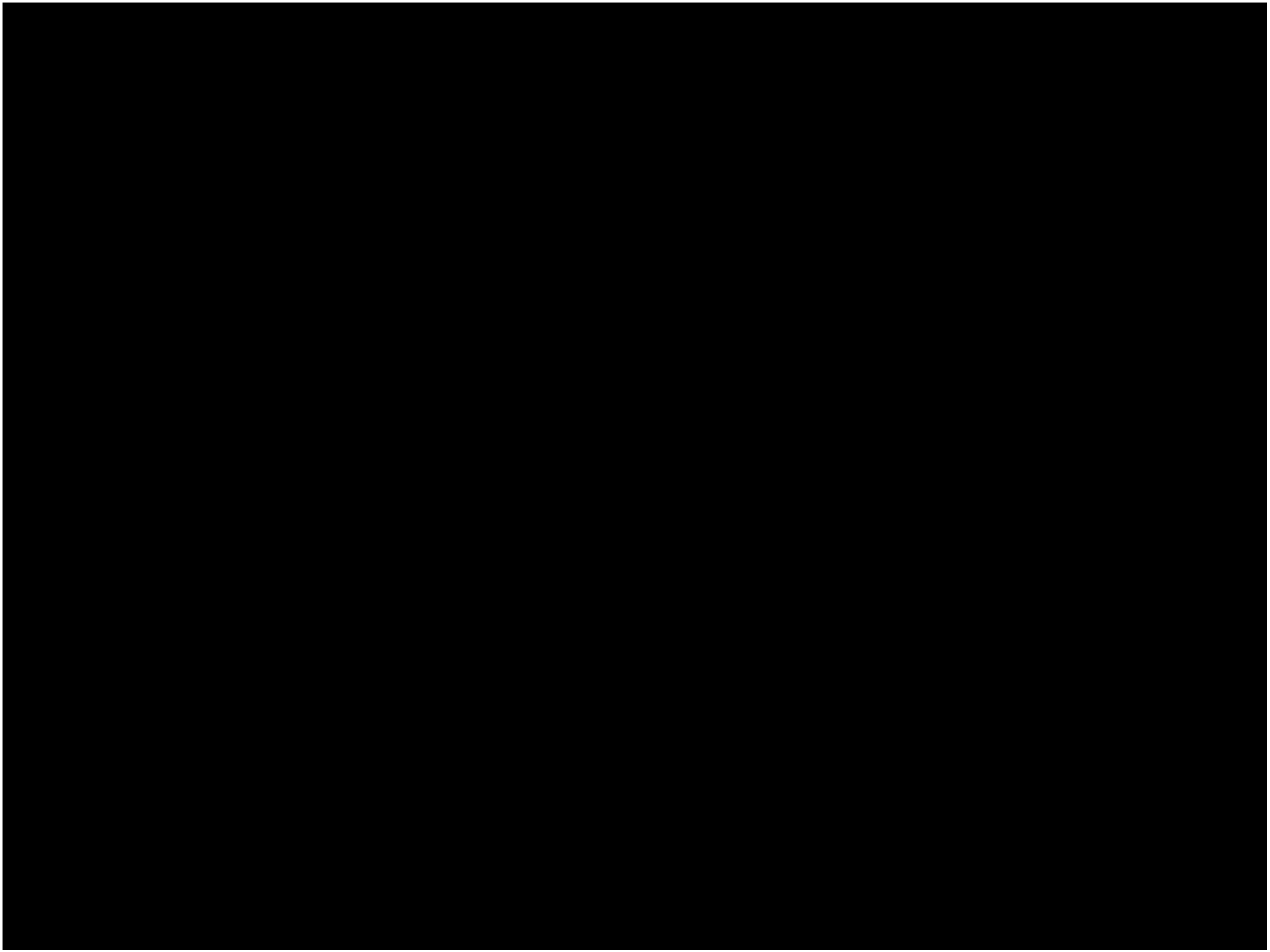


Thermodynamic description of the equilibrium between two states



Single-stranded DNA Double-stranded DNA

IN SOLUTION



SUMMARY OF THE TWO-STATE MODEL

Chemical equilibrium $K_D = e^{-\frac{\Delta H - T\Delta S}{kT}}$ ⟶ Van t'Hoff $\Delta H = kT^2 \frac{\partial}{\partial T} \ln K_D$

Bi-molecular		Intra-molecular
non-self-complementary	self-complementary	
$S_1 + S_2 \rightleftharpoons D$	$2S \rightleftharpoons D$	$S \rightleftharpoons D$
$K_D = \frac{2\theta}{c_T (1-\theta)^2}$	$\frac{\theta}{2c_T (1-\theta)^2}$	$\frac{1-\theta}{\theta}$
$T_m = \frac{\Delta H}{\Delta S + k \ln \frac{C_T}{4}}$	$\frac{\Delta H}{\Delta S + k \ln C_T}$	$\frac{\Delta H}{\Delta S}$
$\Delta H = 6kT_m^2 \left.\frac{\partial \theta}{\partial T}\right\|_{T=T_m}$		$4kT_m^2 \left.\frac{\partial \theta}{\partial T}\right\|_{T=T_m}$

COMPARISON TO EXPERIMENTAL DATA

Shape of Θ(T): constancy of ΔH and ΔS

5′-ATCGTCTGGA-3′
3′-TAGCAGACCT-5′

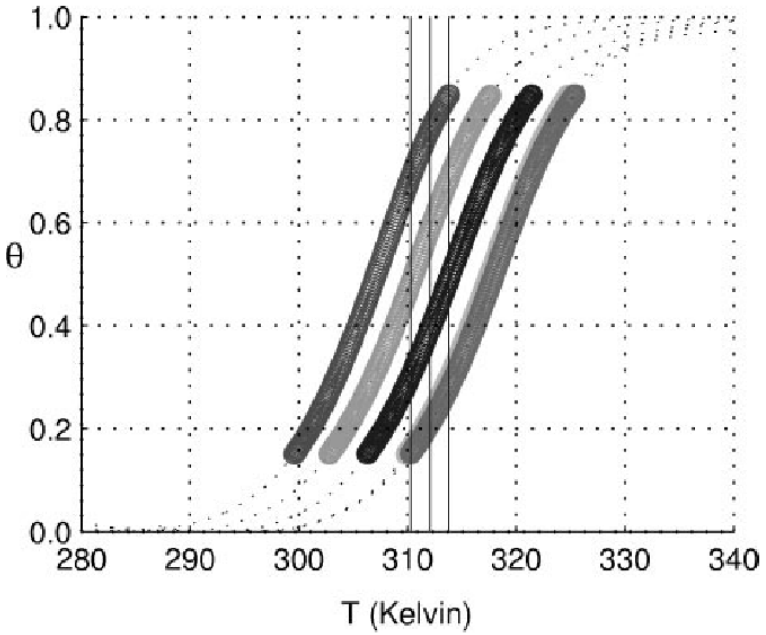


Fig. 1. Melting curves for the 10-bp DNA duplex oligomer (ODN1, $f_{GC} = 0.50$) (f_{GC}, fraction of G·C base pairs in the duplex) 5′-ATCGTCTGGA-3′ (at a C_t of 2×10^{-6} M) in solutions of sodium concentrations (left to right curves) 0.069, 0.12, 0.22, 0.62, and 1.0 M. The thick black curves span a range of θ from 0.15 to 0.85, where the data are the most accurate. The central vertical line through the curves shows the temperature chosen.

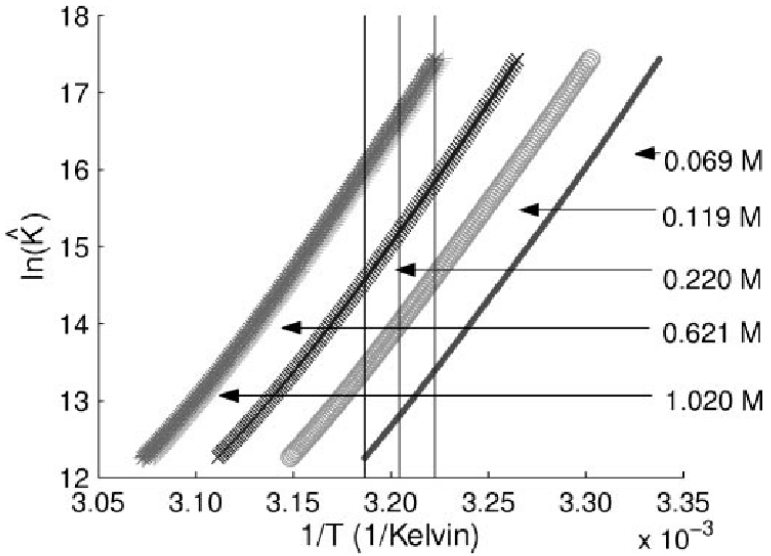
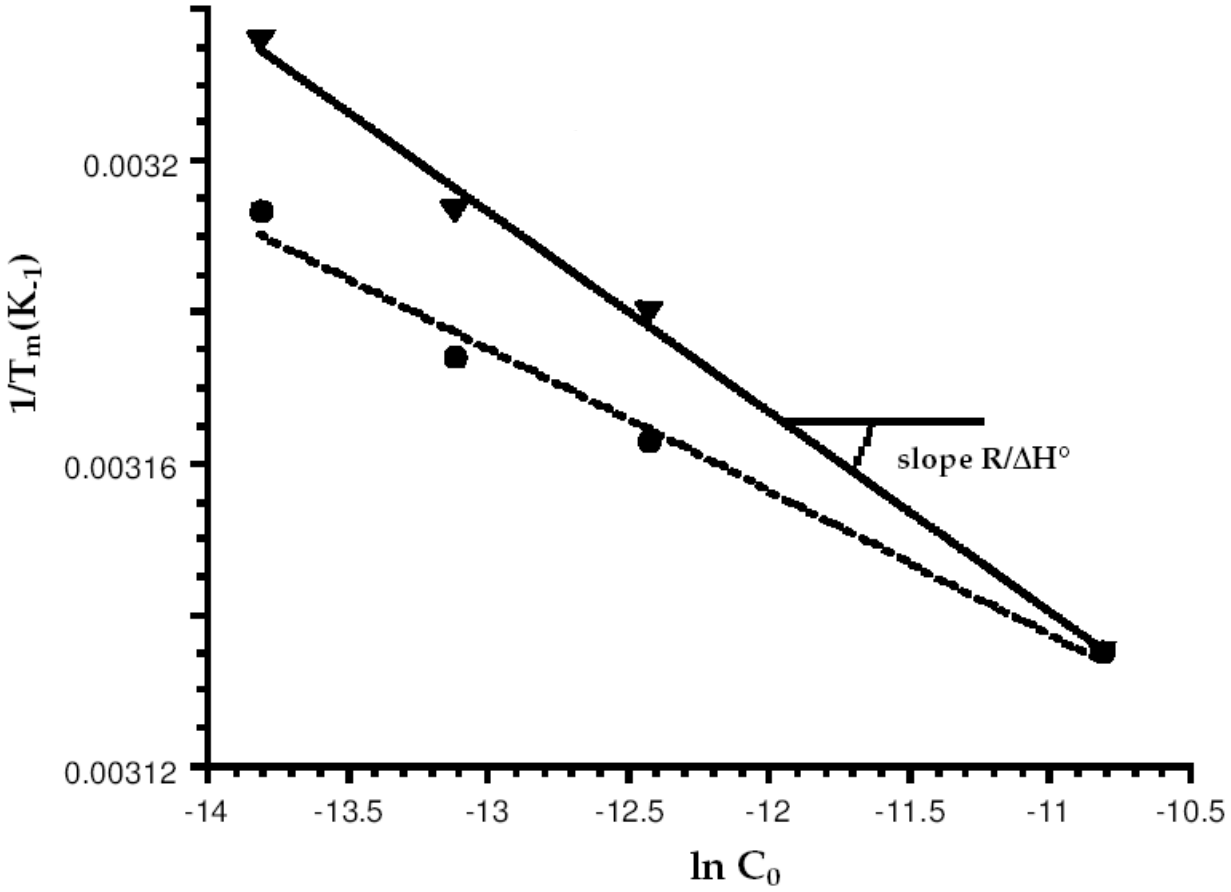


Fig. 2. Plots of ln $\hat{K}$ vs. $1/T$ for the five Na^+ concentrations (mol/liter) for ODN1.

Owczarzy et al, PNAS 25, 14480 (2003)

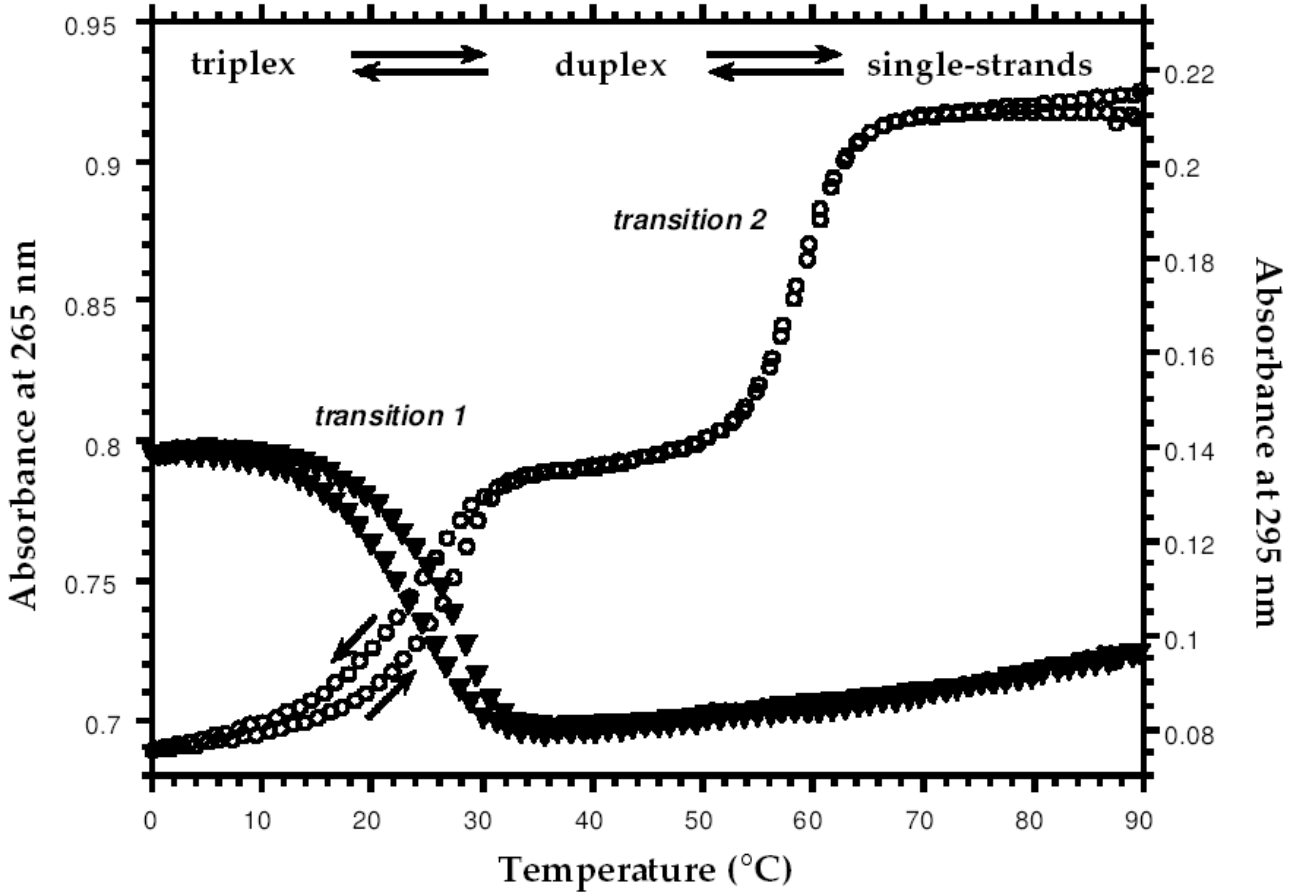
Effect of the concentration

$$S_1 + S_2 \rightleftharpoons D$$

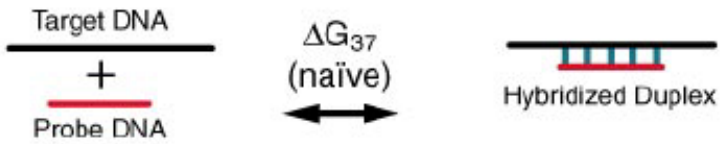


Mergny et Lacroix, Oligonculeotides 13, 515 (2003)

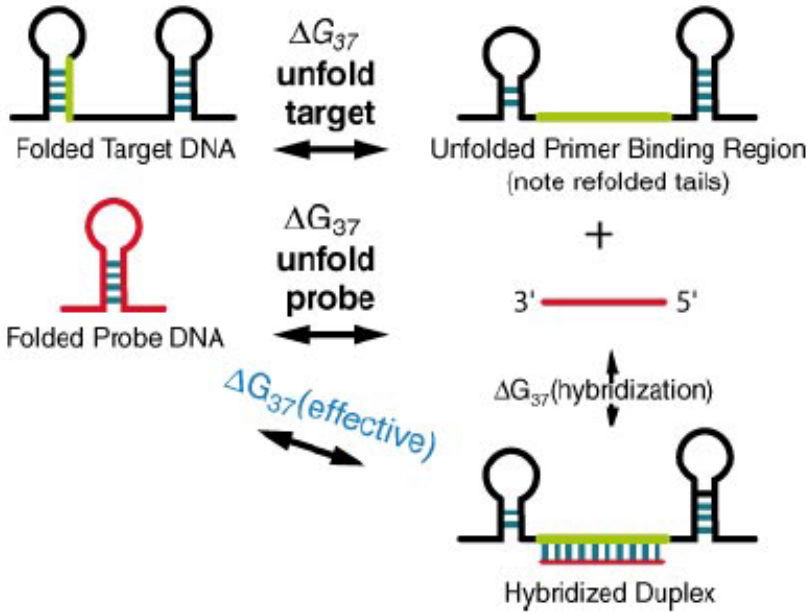
LIMITATIONS of the model: (I) multiple transitions



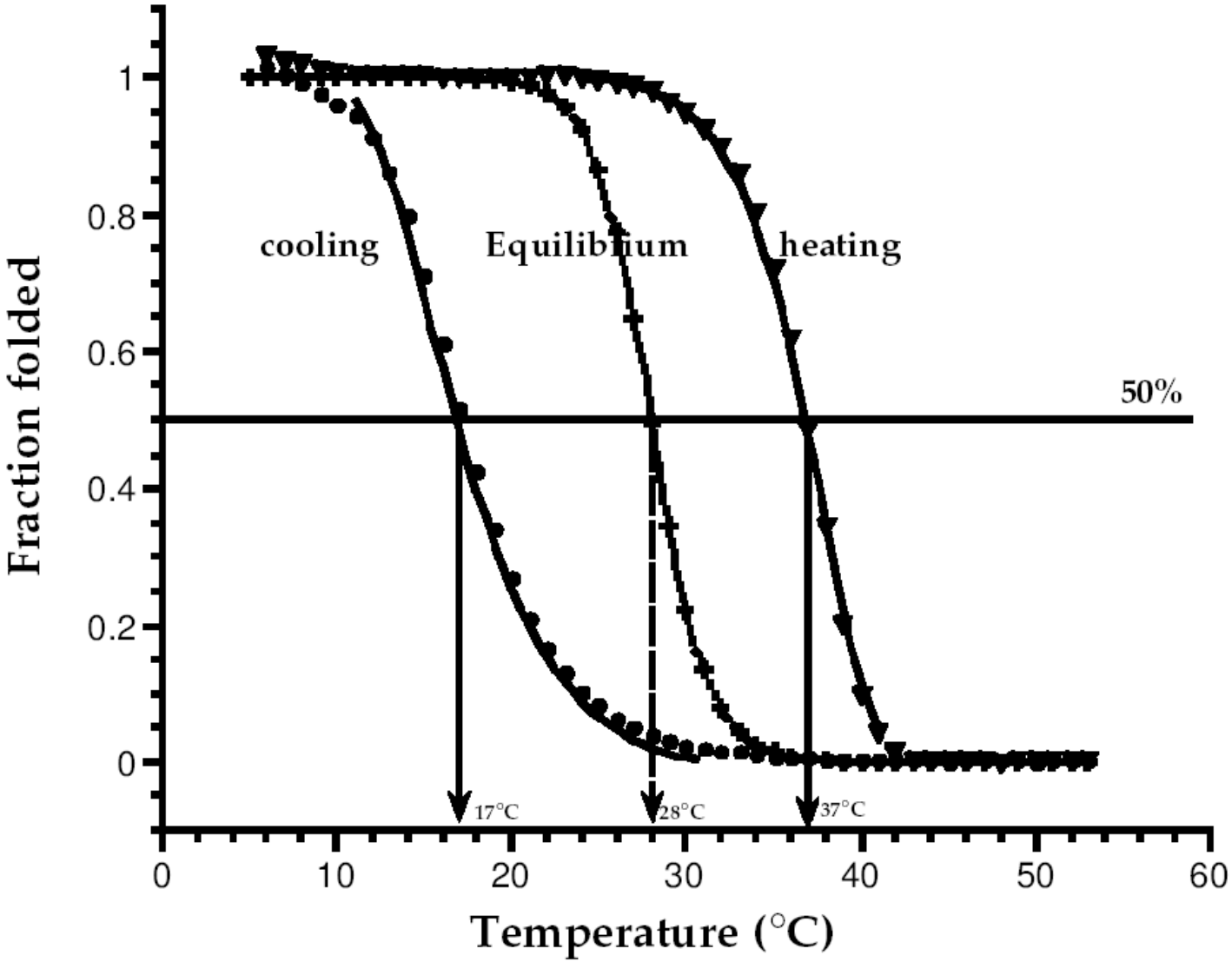
2 State Model



N-State Model *(N≥7)*



LIMITATIONS of the model: (II) kinetic effects



Predicting hybridization and secondary structures of DNA and RNA from the sequence:

MODEL WITH FIRST-NEIGHBOR INTERACTIONS

Under appropriate experimental conditions the representations

$\ln K$ versus $1/T$

$1/T_m$ versus $\ln C_T$

both allow to determine ΔH and ΔS with good precision.

Accordingly, 108 oligonucleotides of various size and sequence were analysed. 12 parameters were extracted by linear regression.

These studies allowed to derive a model to calculate ΔH et ΔS corresponding to DNA duplex formation, as a function of sequence, of DNA concentration and of the concentration of monovalent salt.

Quantitative agreement between experiment and theoretical prediction

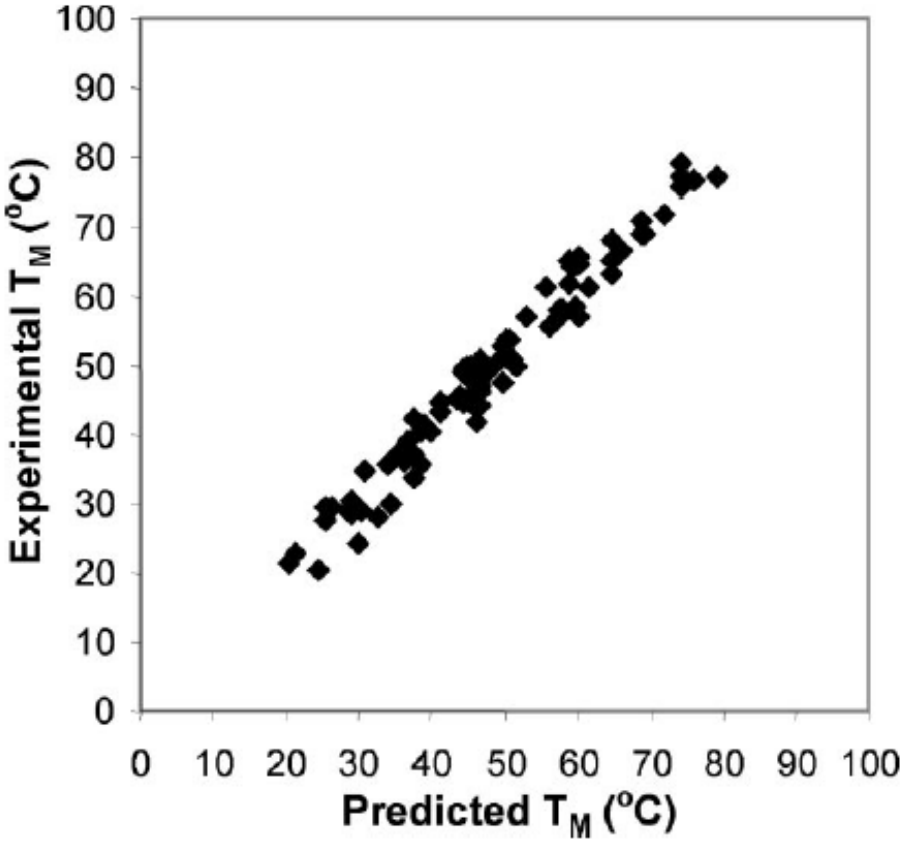
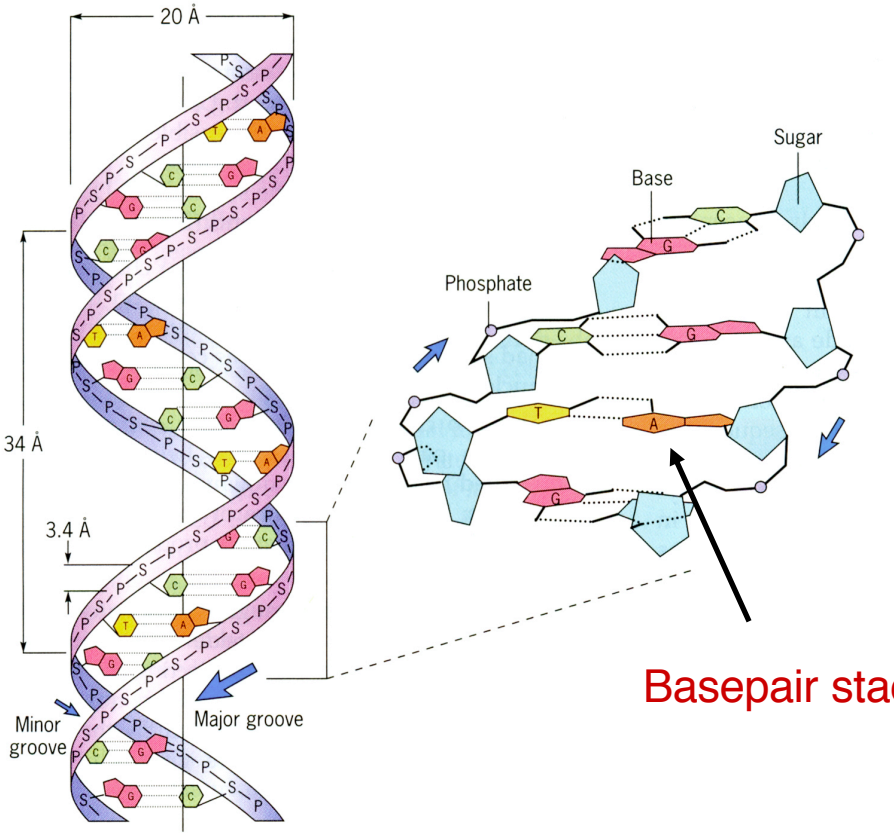


Figure 3 Experimental T_M versus predicted T_M for 81 duplexes 6 to 24 bp in length in solutions ranging from 0.01 to 0.5 M NaCl. Linear regression gives a slope of 1.02, intercept of 0.11, and $R^2 = 0.97$. The average absolute deviation is 2.3°C.

SantaLucia et al, Annu. Rev. Biomol. Struct. 33, 415 (2004)

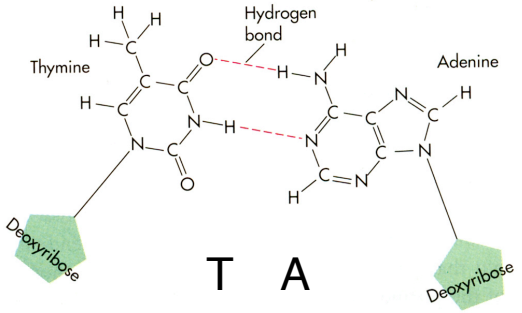
DNA double helix



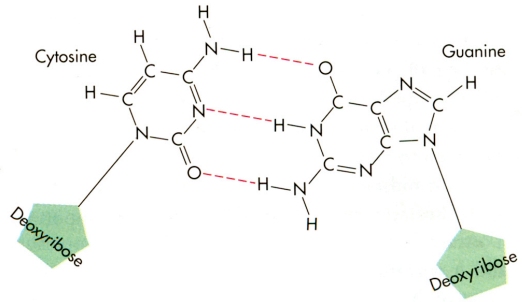
Basepair stacking

Hydrogen bonds between bases

Pyrimidine - Purine



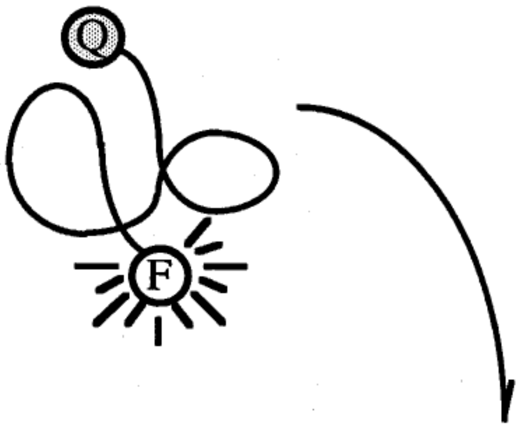
T A



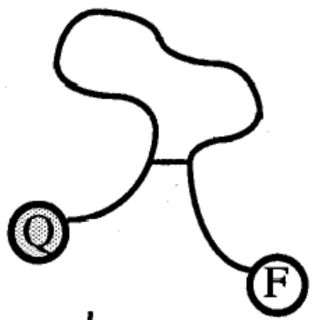
C G

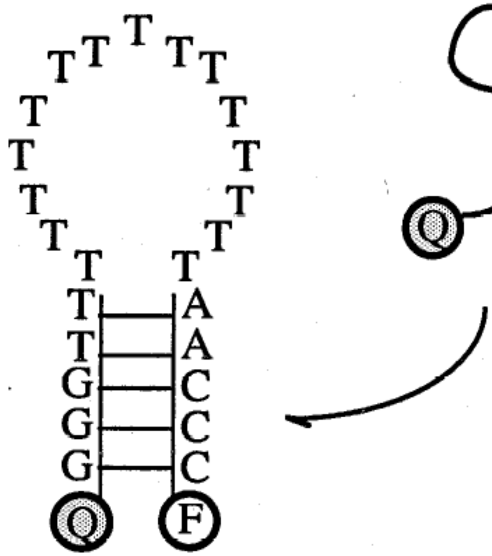
Genetic information: …GTCAGTAAC…

Main ideas behind the model



Closing the first basepair:
INITIATION
Formation of hydrogen bonds
(2 for A/T, 3 for G/C)





Closing the chain:
PROPAGATION
Hydrogen bonding + hydrophobic interaction

(depends on the 2 basepairs at the fork)

TABLE 1 Nearest-neighbor thermodynamic parameters for DNA Watson-Crick pairs in 1 M NaCl[a]

Propagation sequence	ΔH° (kcal mol^{-1})	ΔS° (e.u.) = 10^{-3} kcal/(mol K)	ΔG°_{37} (kcal mol^{-1})
AA/TT	−7.6	−21.3	−1.00
AT/TA	−7.2	−20.4	−0.88
TA/AT	−7.2	−21.3	−0.58
CA/GT	−8.5	−22.7	−1.45
GT/CA	−8.4	−22.4	−1.44
CT/GA	−7.8	−21.0	−1.28
GA/CT	−8.2	−22.2	−1.30
CG/GC	−10.6	−27.2	−2.17
GC/CG	−9.8	−24.4	−2.24
GG/CC	−8.0	−19.9	−1.84
Initiation	+0.2	−5.7	+1.96
Terminal AT penalty	+2.2	+6.9	+0.05
Symmetry correction	0.0	−1.4	+0.43

[a]The slash indicates the sequences are given in antiparallel orientation. (e.g., AC/TG means 5′-AC-3′ is Watson-Crick base paired with 3′-TG-5′). The symmetry correction applies to only self-complementary duplexes. The terminal AT penalty is applied for each end of a duplex that has a terminal AT (a duplex with both end closed by AT pairs would have a penalty of +0.1 kcal/mol for ΔG°_{37}).

12 parameters

10 different pairs

11

12

23 kcal/mol = 1 eV = 40 kT_{300K}

SantaLucia et al, Annu. Rev. Biomol. Struct. 33, 415 (2004)

Exemple: calculation of $\Delta G=\Delta H-T\Delta S$ at 37° C

$$\Delta G^{\circ}_{37}(\text{total}) = \Delta G^{\circ}_{37\,\text{initiation}} + \Delta G^{\circ}_{37\,\text{symmetry}} + \Sigma \Delta G^{\circ}_{37\,\text{stack}} + \Delta G^{\circ}_{\text{AT terminal}}$$

$$\begin{array}{l} 5'\text{-CGTTGA-}3' \\ 3'\text{-GCAACT-}5' \end{array} = \Delta G^{\circ}_{37\,\text{initiation}} + \Delta G^{\circ}_{37\,\text{symmetry}} + \underset{GC}{CG} + \underset{CA}{GT} + \underset{AA}{TT} + \underset{AC}{TG} + \underset{CT}{GA} + AT_{\text{terminal}}$$

$$\Delta G^{\circ}_{37}\,(\text{predicted}) = 1.96 + 0 - 2.17 - 1.44 - 1.00 - 1.45 - 1.30 + 0.05$$

$$\Delta G^{\circ}_{37}\,(\text{predicted}) = -5.35\ \text{kcal mol}^{-1}.$$

23 kcal/mol = 1 eV = 40 kT_{300K}

Effect of mispairing

$$
\begin{array}{l}
5'\text{-GGACTGACG-}3' \\
3'\text{-CCTGGCTGC-}5'
\end{array}
= \text{initiation} + \text{symmetry} + \underset{CC}{GG} + \underset{CT}{GA} + \underset{TG}{AC}
$$

$$
+ \underset{GG}{CT} + \underset{GT}{CG} + \underset{CT}{GA} + \underset{TG}{AC} + \underset{GC}{CG}
$$

$$
\begin{aligned}
\Delta G^{\circ}_{37}\,(\text{predicted}) &= +1.96 + 0 - 1.84 - 1.30 - 1.44 - 0.32 - 0.47 \\
&\quad - 1.30 - 1.44 - 2.17 \\
&= -8.32\ \text{kcal mol}^{-1}.
\end{aligned}
$$

TABLE 2 Nearest-neighbor ΔG°_{37} increments (kcal mol^{-1}) for internal single mismatches next to Watson-Crick pairs in 1 M NaCl[a]

Propagation sequence	X	Y			
		A	C	G	T
GX/CY	A	0.17	0.81	−0.25	WC
	C	0.47	0.79	WC	0.62
	G	−0.52	WC	−1.11	0.08
	T	WC	0.98	−0.59	0.45
CX/GY	A	0.43	0.75	0.03	WC
	C	0.79	0.70	WC	0.62
	G	0.11	WC	−0.11	−0.47
	T	WC	0.40	−0.32	−0.12
AX/TY	A	0.61	0.88	0.14	WC
	C	0.77	1.33	WC	0.64
	G	0.02	WC	−0.13	0.71
	T	WC	0.73	0.07	0.69
TX/AY	A	0.69	0.92	0.42	WC
	C	1.33	1.05	WC	0.97
	G	0.74	WC	0.44	0.43
	T	WC	0.75	0.34	0.68

[a]WC indicates a Watson-Crick pair, which is given in Table 1. Error bars and ΔH° and ΔS° parameters are provided in the original references.

Parameters for single-strand loops

TABLE 4 ΔG°_{37} increments (kcal mol^{-1}) for length dependence of loop motifs in 1 M NaCl[a]

Loop size[b]	Internal loops[c]	Bulge loops[d]	Hairpin loops[e]
1	—	4.0	—
2	(f)	2.9	—
3	3.2	3.1	3.5
4	3.6	3.2	3.5
5	4.0	3.3	3.3
6	4.4	3.5	4.0
7	4.6	3.7	4.2
8	4.8	3.9	4.3
9	4.9	4.1	4.5
10	4.9	4.3	4.6
12	5.2	4.5	5.0
14	5.4	4.8	5.1
16	5.6	5.0	5.3
18	5.8	5.2	5.5
20	5.9	5.3	5.7
25	6.3	5.6	6.1
30	6.6	5.9	6.3

[a]A dash indicates that the loop length is not allowed. All loop ΔH° parameters are assumed to equal zero. The loop ΔS° increment may be calculated from: $\Delta S^{\circ} = \Delta G^{\circ}_{37} \times 1000/310.15$.

[b]The increments for loop lengths not shown may be calculated with Equation 7 (see text).

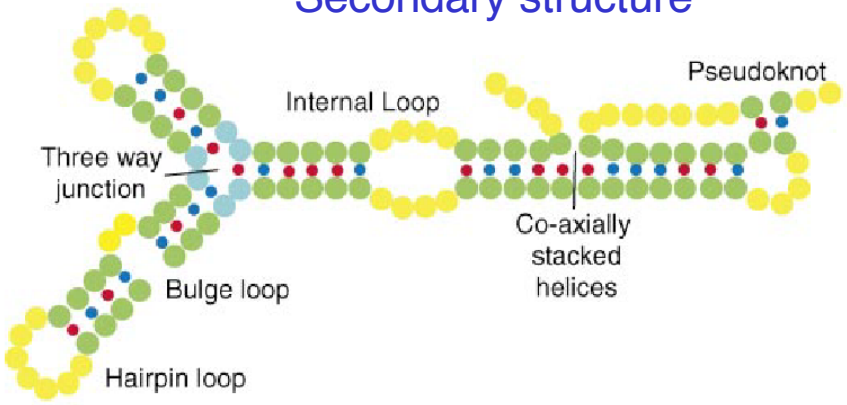
[c]For asymmetric internal loops an additional correction must be applied (see text).

[d]For bulge loops with one nucleotide, the intervening base pair stack must be added.

[e]For hairpin loops of length 3 or 4, special sequence dependent triloop and tetraloop corrections must be applied (see supplementary material).

[f]Internal loops of two are calculated using the mismatch nearest neighbor parameters (see Table 2).

Secondary structure

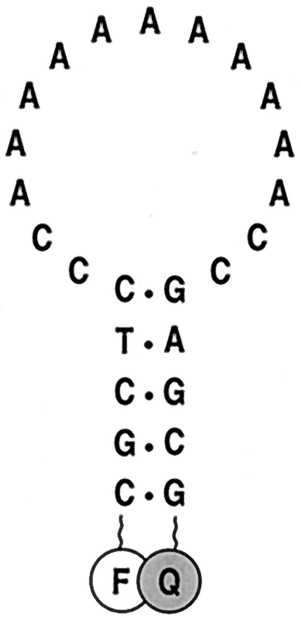


RNA: catalytic activities, regulation

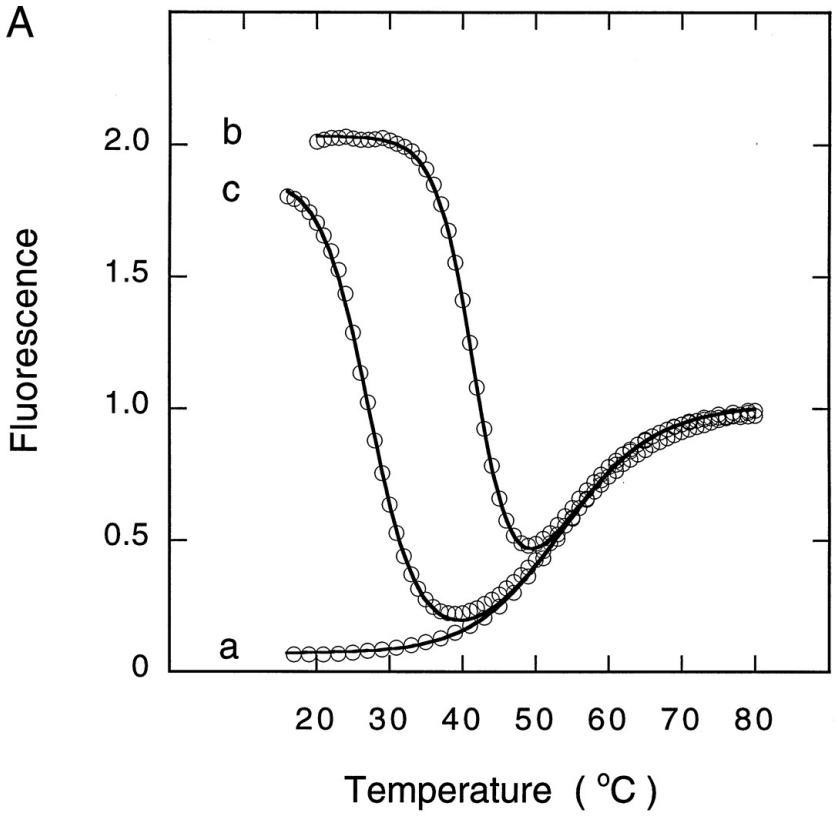
WEB servers for calculating
ΔH, ΔS, T_m and secondary structures:

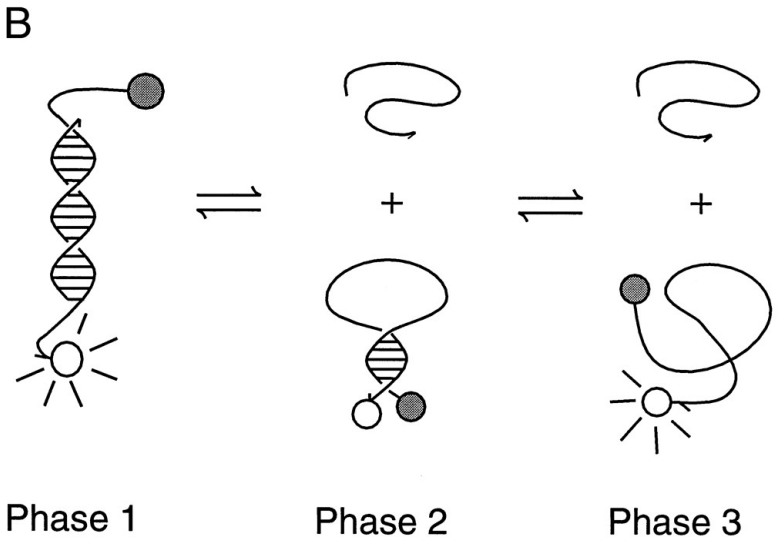
mfold, hyther, vienna

« Enhanced specificity of molecular beacon probes »

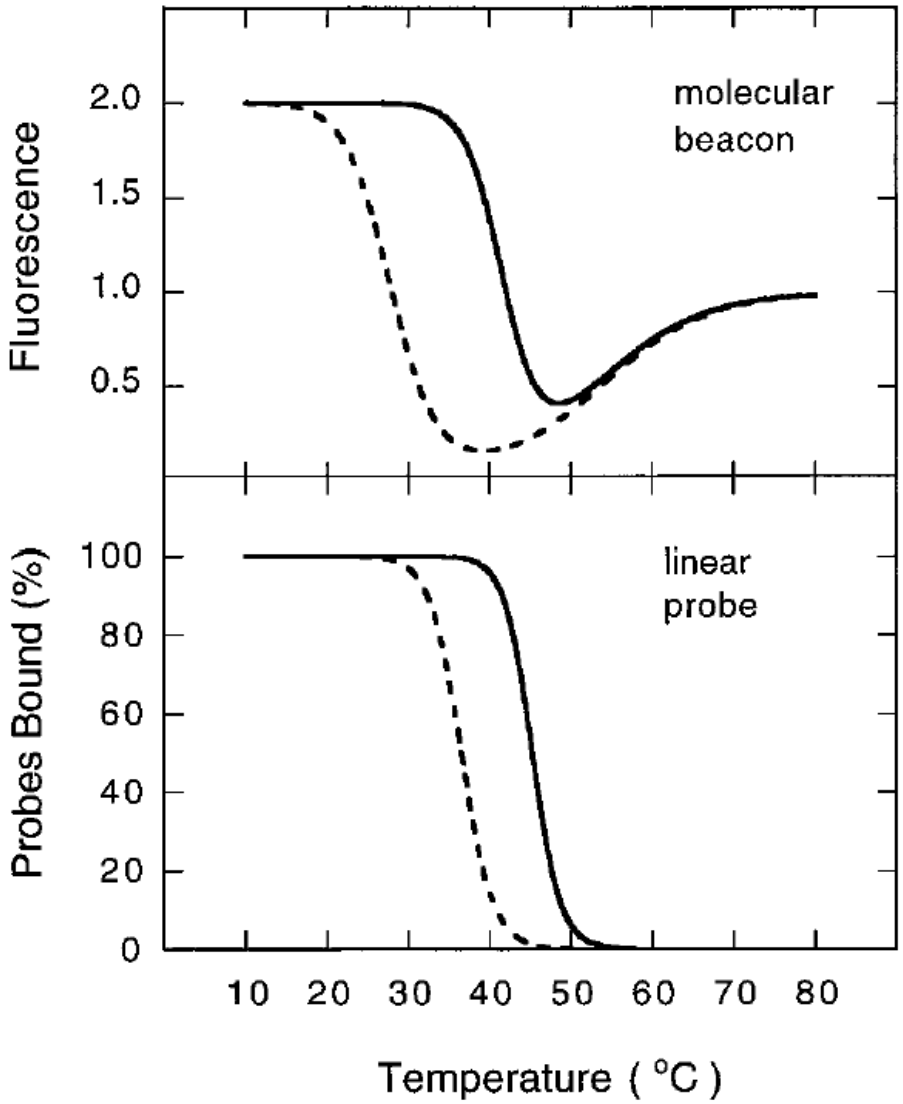


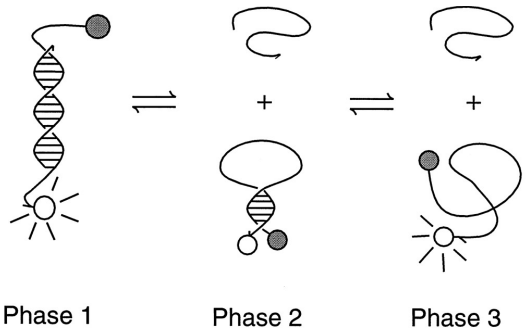
Bonnet, Tyagi, Libchaber and Kramer, PNAS 96, 6171 (1999)





General result: the molecular beacon exhibits higher specificity.
Why?





$$\Delta H_{1p} < \Delta H_{1m} < \Delta H_2 < \Delta H_3$$

$$\Delta S_{1p} < \Delta S_{1m} < \Delta S_2 < \Delta S_3$$

Zero-energy state.

Energy diagram derived from measured thermodynamic parameters:

$$\Delta G = \Delta H - T\Delta S$$