

2000

Dynamics of DNA Breathing and Folding for Molecular Recognition and Computation

Gregoire Bonnet

Follow this and additional works at: http://digitalcommons.rockefeller.edu/student_theses_and_dissertations



Part of the [Life Sciences Commons](#)

Recommended Citation

Bonnet, Gregoire, "Dynamics of DNA Breathing and Folding for Molecular Recognition and Computation" (2000). *Student Theses and Dissertations*. 319.
http://digitalcommons.rockefeller.edu/student_theses_and_dissertations/319

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact mcsweej@mail.rockefeller.edu.



THE LIBRARY

Rockefeller University Library
1230 York Avenue
New York, NY 10021-6399



Dynamics of DNA Breathing and Folding for Molecular Recognition and Computation

A thesis presented to the Faculty of
Rockefeller University
in partial fulfillment of the requirements
for the degree of Doctor in Philosophy

by

Grégoire Bonnet

To my parents, Roger & Monique

O yeni ıřıkta, otobüsün sađ ön camında meleđi gördüm.

Benim az ötemde, ama benden ne kadar da uzaktaydı. Gene de anladım ama: O derin, yalın ve güçlü ıřık benim için oradaydı. Magirus'un bütün hızıyla bozkırda ilerlemesine rağmen melek bana ne yaklaşıyor, ne uzaklaşıyordu. Parlak ısıđı yüzünden tamı tamına neye benzediđini de göremiyordum, ama içimde canlanan bir şaka duygusu, bir hafıflık, bir özgürlük yüzünden onu tanıdıđımı anladım.

Orhan Pamuk
Yeni Hayat

In that new light on the right side of the windshield, I beheld the angel.
The angel was so close to me and yet how far. Even so, I still knew this: the profound, plain, powerful light was there for me. Even though the Magirus hurtled through the steppe with all its might, the angel would neither draw close nor draw back. The brilliant light kept me from seeing what the angel looked like for sure, but I knew from the sense of playfulness, the sense of lightness, the sense of freedom I felt inside me that I had recognized the angel.

Orhan Pamuk
The New Life
Translated into English by Güneli Gün

Contents

Abstract	iii
Abbreviation & Notation	v
List of Figures	ix
Foreword	1
1 Introduction	5
1.1 DNA structures.	5
1.1.1 Primary structure of DNA.	5
1.1.2 Formation of B-DNA structures.	7
1.1.3 Structural deformation of B-DNA.	10
1.1.4 Single-stranded DNA.	10
1.1.5 Stacking of DNA bases.	11
1.2 Statistical Physics of DNA secondary structures.	12
1.2.1 Phenomenology of DNA hybridization.	14
1.2.2 Zipping toy-model for the coil-to-helix transition.	15
1.2.3 Nearest-neighbor model of DNA hybridization.	16
1.2.4 Single-stranded DNA loop entropy.	20
1.3 DNA conformational fluctuations.	22
1.3.1 Mechanical micromanipulation of DNA molecules.	22
1.3.2 Spectroscopic studies of DNA structures.	24
1.3.3 DNA breathing modes.	29
1.4 Fluorescence Correlation Spectroscopy.	33
1.4.1 Historical lineage of FCS.	34
1.4.2 FCS for deciphering molecular events.	34
1.4.3 Application of FCS in analytical chemistry, biophysics, and cell biology.	37
1.4.4 FCS to monitor translational mobility.	37
1.4.5 FCS to monitor intramolecular conformational fluctuations.	41
2 Material & Methods.	43
2.1 Synthesis and conjugation of the DNA samples.	43
2.1.1 Initial materials.	43
2.1.2 Reversed-Phase High-Pressure Liquid Chromatography (HPLC): our quantitative analytical purification tool.	43
2.1.3 Coupling of a succinidimyl ester onto a primary amine.	46
2.1.4 Coupling of an iodoacetamide or maleimide onto a sulfhydryl.	48
2.1.5 Complete synthesis of a molecular beacon.	50
2.1.6 Synthesis of DNA molecules to study DNA breathing.	51

2.2	Thermodynamic measurement on nucleic acid structures.	57
2.2.1	Thermal denaturation profiles.	57
2.2.2	Equilibrium analysis for the denaturation profiles of the molecular beacons. . .	57
2.2.3	Thermodynamic measurements of the hairpin to random-coil transition. . .	58
2.2.4	Thermodynamic measurements of the dissociation of probe-target duplexes. . .	59
2.3	Fluorescence Correlation Spectroscopy	59
2.3.1	Chemical-diffusion equation for the molecular beacon.	60
2.3.2	Analytical expression for the correlation function.	62
2.3.3	Data analysis for a molecular beacon.	63
2.3.4	Statistical accuracy in fluorescence correlation spectroscopy.	66
2.3.5	Experimental setup for fluorescence correlation spectroscopy.	67
3	Results.	81
3.1	Thermodynamics and Optimization of molecular recognition.	82
3.1.1	Thermodynamics of molecular beacons as DNA probes.	82
3.1.2	Optimal sensitivity of a DNA probe.	94
3.2	DNA logic gates.	102
3.2.1	OR gate.	103
3.2.2	AND gate.	104
3.2.3	Concatenation of DNA logic gates.	106
3.3	Conformational fluctuations of DNA hairpin-loops.	107
3.3.1	Measurement of the opening and closing timescales of a DNA hairpin-loop. . .	107
3.3.2	Systematic study of the hairpin-loop fluctuation dynamics.	116
3.4	Breathing modes of double-stranded DNA.	127
3.4.1	Design of the probes to study dsDNA breathing modes.	127
3.4.2	Breathing fluctuations for 18AT fluctuation box.	128
3.4.3	Breathing fluctuations for shorter AT-tracts.	134
3.4.4	Distortion of the GC clamp for different sizes of fluctuation box.	141
4	Discussion.	145
4.1	New insights on the biophysics of DNA conformational fluctuations.	145
4.1.1	Single-stranded DNA stacking and consequences.	145
4.1.2	Initiation of transcription and DNA breathing.	148
4.1.3	Entropy-enthalpy compensation in DNA hybridization.	151
4.2	DNA molecular switches and computation.	154
4.2.1	DNA and molecular code.	154
4.2.2	DNA self-assembly and nanotechnology.	157
4.2.3	DNA Molecular switches.	160
	Afterglow & Acknowledgments	165
A	DNA Computation and amplifiers.	167
A.1	DNA computation.	167
A.1.1	Quantitative PCR.	170
B	Kinetic model for the hybridization cascade.	173
C	Temperature-control device for bulk measurement.	175
D	Anodization of aluminum.	179
	Bibliography	194

Abstract

This thesis is centered on the development of the molecular beacon, as a new DNA probe for DNA genotyping, DNA computation and biophysical studies of DNA conformations.

Molecular beacons are single-stranded DNA molecules that form a stem-and-loop structure. A fluorophore and a quencher are grafted at their two ends to report their conformations: when the molecular beacon is closed, fluorophore and quencher are held in close proximity and the fluorescence is quenched; when the molecular beacon is open, fluorophore and quencher are far apart, and the fluorescence is restored. Molecular beacons are ideal DNA probes coupling conformational switch with fluorescence signal turning-ON.

We use molecular beacons to study the molecular recognition of single-stranded DNA (ssDNA) oligonucleotide. We present a thermodynamic diagram to show that structural constraints make the molecular beacon highly sensitive to the presence of mismatches in its target. We introduce a sequence sensitivity parameter to quantitatively compare different DNA probes, and propose an algorithm to optimally tune the probe's structure for enhanced sequence discrimination.

Logic gates (OR and AND gates) using molecular beacons are designed to carry most elementary molecular computations. The conformational changes associated with such computations can be used to concatenate many chemical reactions, and carry out complex molecular computations.

Molecular beacons are also ideal probes to study DNA secondary structures and their fluctuations. We develop the fluorescence correlation spectroscopy (FCS) technique to monitor the dynamics of relaxation of DNA conformational fluctuations. We first measure the opening and closing timescales of DNA hairpin-loops. Activation barriers for opening and closing for different loop lengths and sequences are analyzed to better account for the stability of DNA secondary structures. A sequence-dependent rigidity of ssDNA has been discovered, and analyzed in terms of base stacking.

We then use FCS to study the dynamics of double-stranded DNA (dsDNA) breathing modes with synthetic DNA constructs. The analysis of the basepairing fluctuation dynamics, monitored by fluorescence, unravels lifetimes of breathing modes ranging from $1\mu\text{s}$ to 1ms . Long-range distortions of the dsDNA have been unraveled for purine-rich sequences, of relevance to the specificity of transcription initiation in prokaryotes.

Abbreviation & Notation.

$\langle \dots \rangle_t$ Time average $\langle I(t) \rangle_t = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T I(u) du$

APD Avalanche Photodiode

B_0 Total concentration of molecular beacon in solution

β Amplitude of the Autocorrelation Function

c Speed of light in vacuum ($c = 3 \cdot 10^8 \text{ m/s}$)

$d(X) \bullet d(Y)$ Duplex of two DNA oligonucleotides of sequence X and Y

Da Dalton (1Da=1g/mol)

DABCYL 4-((4-(dimethylamino)phenyl)azo) benzoic acid

DNA DeoxyriboNucleic Acid

dsDNA Double-Stranded DNA

DMF DiMethyl Formamide

EDTA Ethylene Diamine Tetraacetate

ϵ Extinction coefficient (*e.g.* $\epsilon_{525nm}(Rh6G) = 1.08 \cdot 10^5 \text{ cm}^{-1} \text{ Mol}^{-1}$)

fl femtoliter 1fl=10⁻¹⁵l = (1μm)³

h Planck constant ($h = 6.63 \cdot 10^{-34} \text{ J.s}$)

IA IodoAcetamide

FCS Fluorescence Correlation Spectroscopy

FIDA Fluorescence Intensity Distribution Analysis

FRAP Fluorescence Recovery After Photobleaching

$G(\tau)$ Autocorrelation Function of the Fluorescence Intensity: $G(\tau) = \frac{\langle I(t)I(t+\tau) \rangle_t - \langle I(t) \rangle_t^2}{\langle I(t) \rangle_t^2}$

HPLC High-pressure liquid chromatography

λ Wavelength of light (typically $\lambda_{\text{laser}} = 514.5 \text{ nm}$)

MW Molecular Weight (in Dalton)

N Number of fluorescent object in the field of view

\mathcal{N}_A Avogadro number ($\mathcal{N}_A = 6.02 \cdot 10^{23} mol^{-1}$)

NN-model Nearest-Neighbor model of DNA hybridization

P Power of the excitation laser (in μW)

PCR Polymerase Chain Reaction

PID Proportional-Integral-Derivative Controller

PMT PhotoMultiplier Tube

Rh6G 5-carboxy Rhodamine 6G

RNAP RNA polymerase (from *E. coli* if unspecified)

rpm rounds *per* minute

σ Cross-section of the dye ($\sigma(\text{Rh6G}) = 1.8 \text{ \AA}^2$)

SE Succinimidyl Ester

ssDNA Single-Stranded DNA

TEpH8.0 Saline buffer, Trizma Base 0.1M, EDTA 1mM, pH8.0

τ_- Opening timescale of a DNA hairpin loop

τ_+ Closing timescale of a DNA hairpin loop

T_0 Total concentration of target oligonucleotides in solution

TTL Transistor-transistor logic signal

List of Figures

1	Schematic representation of a molecular beacon	2
1.1	Primary structure of a DNA molecule	6
1.2	Watson-Crick basepairs.	8
1.3	Structure of B-DNA	9
1.4	Structure of polyC determined by X-ray diffraction	13
1.5	Average number of basepairs in a 20-mer duplex (Zimm-Bragg zipping model).	16
1.6	Enunciation of all the possible conformations of two hybridizing oligonucleotides.	18
1.7	Single stranded domains in DNA secondary structures	20
1.8	Yield of fluorescence energy transfer	25
1.9	Review of previous studies on DNA conformations by FET	27
1.10	Single molecule fluorescence recording under confocal geometry	35
1.11	Histogram of quantum burst lag times in single molecule fluorescence recording	36
1.12	Graphical illustration of the building of the autocorrelation function	38
1.13	Application of FCS to particle counting <i>in vivo</i>	39
2.1	Chromatogram for the bioconjugation of Rh6G onto a DABCYL-oligonucleotide	45
2.2	Three bioconjugation reactions used to couple dyes onto modified oligonucleotides	47
2.3	Chromatogram of the fractions after coupling of Rh6G-SE onto a 5' amine-oligonucleotide.	49
2.4	Chromatogram of the first HPLC purification in the molecular beacon synthesis	52
2.5	Chromatogram of the second HPLC purification in the molecular beacon synthesis	53
2.6	Molecular Sketch of the stem of a molecular beacon	54
2.7	Internal tagging on modified thymidine bases	55
2.8	Constructs for DNA breathing studies	56
2.9	Sources of fluctuation in an FCS experiment with the molecular beacons.	61
2.10	Geometry of the excitation beam in an epifluorescence microscope	63
2.11	Cross-correlation function for a serial dilution of Rh6G	64
2.12	Intercept β of the cross-correlation function for a serial dilution of Rh 6G	65
2.13	Amplitude of the cross-correlation function <i>vs.</i> open probability of the molecular beacon	66
2.14	Signal/noise of the cross-correlation function for different concentrations of Rh6G	67
2.15	Signal/noise of the cross-correlation function <i>vs.</i> Acquisition time	68
2.16	Optical setup for Fluorescence Correlation Spectroscopy	69
2.17	Photograph of our FCS setup	70
2.18	Comparison of water-immersion lens and oil-immersion lens for FCS	71
2.19	Schematic ray tracing in the sample	73
2.20	Ray tracing in the confocal microscope and definition of the depth of field	74
2.21	Transmission and reflection coefficients for the optical filters	75
2.22	Eliminating artefacts from photodetector afterpulsing with the cross-correlation function	76
2.23	Jablónsky diagram for Rh6G	77

3.1	Phase transitions in solutions containing molecular beacons.	84
3.2	Determination of thermodynamic parameters for the melting of a molecular beacon.	85
3.3	Determination of thermodynamic parameters for the melting of a duplex beacon–target	86
3.4	Thermodynamic phase diagram.	88
3.5	Melting curves of a solution of the inhomogeneous molecular beacon and different targets	91
3.6	Low–temperature transition for a solution of molecular beacon and target	92
3.7	Free energy diagram of DNA hybridization.	96
3.8	Melting curves for matched and mismatched DNA duplexes (linear probe).	98
3.9	Melting curves for matched and mismatched DNA duplexes (molecular beacon probe).	99
3.10	Maximal sequence discrimination sensitivity	100
3.11	Resolution of a match and a mismatch target by hybridization of a molecular beacon	101
3.12	Sketch of the structure and function of the OR gate	103
3.13	Fluorescence signals from the OR gate	104
3.14	Sketch of the structure and function of the AND gate	105
3.15	Fluorescence signals from an AND gate	106
3.16	Sketch of the conformational thermal fluctuations of a molecular beacon	108
3.17	Data treatment for FCS of a molecular beacon	110
3.18	FCS of the chemical fluctuation of the molecular beacon at different temperatures	111
3.19	Bulk melting curve of a molecular beacon.	113
3.20	Arrhenius plot for τ_- and τ_+ for a molecular beacon with a 21–T loop.	114
3.21	Energy landscape for the fluctuation dynamics of the hairpin-loop structure.	117
3.22	Arrhenius plots of the opening and closing timescales of beacons with different lengths of T-loops.	118
3.23	Scaling laws for the closing timescale of T-loop molecular beacons as a function of the loop length	119
3.24	Closing and Opening timescales for a molecular beacon of loop T_{21} for different salt concentrations.	120
3.25	Melting curves of T-loop beacons in 0.25M NaCl	121
3.26	Melting curves of A-loop beacons in 0.25M NaCl	122
3.27	Melting Temperatures for T-loop and A-loop beacons in 0.25M NaCl	122
3.28	Closing and opening timescales for beacons of different lengths of T-loops	123
3.29	Closing and opening timescales for beacons of different lengths of A-loops	124
3.30	Barrier enthalpy for the closing of the molecular beacons as a function of the loop length	124
3.31	Closing timescales of four Poly(A) series with a single cytosine defect	126
3.32	DNA construct for dsDNA breathing study	128
3.33	Melting curve of the end-tagged and internally tagged DNA construct with a random 18AT domain	129
3.34	Conformational transitions associated with the melting transition of the breathing constructs	130
3.35	Stretched exponential for the breathing modes FCS	133
3.36	Measure of the stretching parameter	133
3.37	Arrhenius plot of the breathing timescales for 18AT–fluctuation box	134
3.38	Melting curve of the end-tagged and internally tagged DNA construct with a random 8AT domain	135
3.39	DNA construct to study the entropic constraint on the GC–clamp.	137
3.40	Comparison of the melting of 8AT–basepair–construct with 8–unmatched–bases’ construct	138
3.41	Melting curves of DNA breathing constructs with a 6AT–basepairs fluctuation box	139
3.42	FCS on 6AT breathing constructs	140

3.43	Melting curves of DNA breathing constructs with different lengths of fluctuation box	142
3.44	Melting curves of DNA breathing constructs with different lengths of fluctuation box (closup from 30°C to 80°C)	143
4.1	Double-well energy potential for dsDNA breathing and consequences.	149
4.2	Nearest-neighbor thermodynamic parameters for matched basepairs	152
4.3	Nearest-neighbor thermodynamic parameters for mismatched basepairs	153
4.4	Schematic of an hybridization cascade.	156
4.5	Simulation of the hybridization cascade.	158
4.6	Example of DNA-based memory	162
A.1	Traveling salesman problem	168
C.1	Temperature calibration of the thermistor	176
C.2	Temperature recording for a mistuned PID controller.	177
C.3	Temperature recording versus time for a PID controller set to oscillate steadily. We record the period of the oscillation ($T_u=0.66\text{min}$) and the proportional band ($PBu=1.6$) from this run to finely tune the PID controller.	178
C.4	Temperature recording versus time for a well-tuned PID controller	178

*La cigarette sans cravate
Qu'on fume à l'aube démocrate
Et le remords des cul-de-jatte
Avec la peur qui tend la patte
Le ministère de ce prêtre
Et la pitié à la fenêtre
Et le client qui n'a peut-être
Ni Dieu ni maître*

*Le fardeau blême qu'on emballe
Comme un paquet vers les étoiles
Qui tombent froides sur la dalle
Et cette rose sans pétale
Cet avocat à la serviette
Cette aube qui met la voilette
Pour des larmes qui n'ont peut-être
Ni Dieu ni maître*

*Ces bois qu'on dit de justice
Et qui poussent dans les supplices
Et pour meubler le Sacrifice
Avec le sapin de service
Cette procédure qui guette
Ceux que la Société rejette
Sous prétexte qu'ils n'ont peut-être
Ni Dieu ni maître*

*Cette parole d'évangile
Qui fait plier les imbéciles
Et qui met dans l'horreur civile
De la noblesse et puis du style
Ce cri qui n'a pas de rosette
Cette parole de prophète
Je la revendique et je vous souhaite
Ni Dieu ni maître*

Léo Ferré
Ni Dieu ni Maître

Foreword.

This thesis work deals with molecular beacons, as a probe to read and process DNA sequences, as well as a probe for biophysical studies of nucleic acid conformational fluctuation.

We have developed DNA probes, called molecular beacons, that become fluorescent when they bind to complementary nucleic acids [1]. These probes are single-stranded DNA molecules that form a stem-and-loop structure (Figure 1). A fluorophore is linked to one end of the molecule and a quencher is linked to the other end. Consequently, fluorescence is quenched when the probe is in the stem-and-loop conformation. However, when the probe sequence in the loop anneals to a complementary nucleic acid target sequence, the longer and stronger probe–target duplex overcomes the shorter hairpin stem, leading to a conformational reorganization. The rigidity of the probe–target helix forces the hairpin stem to unwind, resulting in the separation of the fluorophore and quencher, and fluorescence is switched on. Another mode of unwinding of the molecular beacons’ stem is the thermal melting of the stem basepairs.

We show that molecular beacons have an exceptional signal/noise: their fluorescence increases 50 to 100 fold times upon opening of the beacon stem. Thus, molecular beacons have been used as ideal DNA probes coupling their conformational switch with a fluorescence reporter signal.

This thesis work is presented in four chapters:

Chapter 1: Introduction. This chapter presents background information on the concepts and techniques we used to study the molecular beacons. We review structural data on DNA molecules (§ 1.1), the physics of secondary structure formation (§ 1.2) and the importance of conformational fluctuation in biomolecules (§ 1.3). We also review the field of fluorescence correlation spectroscopy as a tool to analyze the dynamics of DNA conformational fluctuations (§ 1.3).

Chapter 2: Material and Methods. In this chapter, we present the materials and methods we used to carry out our experiments: this includes the bioconjugation chemistry (§ 2.1) to prepare the DNA constructs, the thermodynamics analysis of the conformational change (§ 2.2). The third section (§ 2.3) presents a detailed account of our implementation of the Fluorescence Correlation Spectroscopy technique.

Chapter 3: Results. This chapter is divided into four sections.

The first section presents our measurements on the thermodynamics of the enhanced specificity of structurally–constrained DNA probes (§ 3.1.1).

The second section deals with the design of two molecular logic gates with molecular beacons: OR and AND gate. The possibility of concatenation of the output of these molecular switches is then discussed (§ 3.2).

The third section presents our results on the dynamics of DNA hairpin–loop fluctuations. The combination of fluorescence correlation spectroscopy and fluorescence quenching in the molecular beacons enable us to monitor the open–close transition kinetics. Our FCS setup is such that we can record these dynamics virtually from single molecule, in a range of timescale

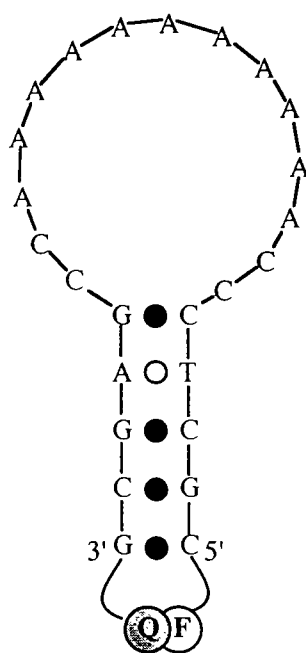


Figure 1: Schematic representation of a molecular beacon. The molecule has an polyadenosine probe sequence embedded within complementary arm sequences. The arms form a hairpin stem and the probe sequence is located in the hairpin loop. A fluorophore (F) is covalently linked to the end of one arm and a quencher (Q) is covalently linked to the end of the other arm.

from 100ns up to 10ms. We analyze the energy landscape guiding the thermal fluctuations of the molecular beacons. The effect of the loop length and sequence is presented and a surprising sequence-dependent rigidity of single-stranded DNA is analyzed.

Finally, the last section presents our experiments on DNA breathing fluctuations (§ 3.4): DNA constructs have been designed to monitor the dynamics of internal basepairing in double-stranded DNA. We show that breathing modes are fast intramolecular transitions with a distribution of timescales spanning from 100ns up to 1ms. We also unravel surprising large-scale distortion modes of double-stranded DNA, specific to small AT-rich domains (6–8 basepairs).

Chapter 4: Discussion. This chapter discusses the scope of our experimental results. First, we discuss the importance of sequence-dependent base stacking in single stranded DNA, and the relevance of singular breathing modes to the issue of the specificity of initiation of transcription in prokaryotes. Second, we come back to our thermodynamics measurements to discuss the nature of the molecular code of nucleic acids and its possible applications in DNA computation and nanotechnology.

Владимир Высоцкий.

Я не люблю

(Из спектакля "Свой остров")

Я не люблю фатального исхода.
От жизни никогда не устаю.
Я не люблю любое время года,
Когда веселых песен не пою.

Я не люблю открытого цинизма,
В восторженность не верю, и еще,
Когда чужой мои читает письма,
Заглядывая мне через плечо.

Я не люблю, когда наполовину
Или когда прервали разговор.
Я не люблю, когда стреляют в спину,
Я также против выстрелов в упор.

Я ненавижу сплетни в виде версий,
Червей сомненья, почестей иглу,
Или, когда все время против шерсти,
Или, когда железом по стеклу.

Я не люблю уверенности сытой,
Уж лучше пусть откажут тормоза!
Досадно мне, что слово "честь" забыто,
И что в чести наветы за глаза.

Когда я вижу сломанные крылья,
Нет жалости во мне и неспроста -
Я не люблю насилие и бессилье,
Вот только жаль распятого Христа.

Я не люблю себя, когда я трушу,
Обидно мне, когда невинных бьют,
Я не люблю, когда мне лезут в душу,
Тем более, когда в нее плюют.

Я не люблю манежи и арены,
На них мильон меняют по рублю,
Пусть впереди большие перемены,
Я это никогда не полюблю.

1969 г.

Chapter 1

Introduction

1.1 DNA structures.

This thesis work is centered on DNA molecules as a recognition and computational tool. DNA, and more generally nucleic acids, are the most fascinating molecular objects one can envision: linear polymers, with a combinatorial richness of the nucleotide arrangement, used as a repository of, or a key for genetic information. Central to our fascination is the ability to self-assemble into double-strands. Biological functional consequences of the discovery of the B-DNA structure are found in DNA replication, DNA transcription and DNA recognition.

Beyond static structures, the conformational polymorphism of nucleic acids enables a physico-chemical processing of the nucleic acids' code: for example, the discovery of self-splicing enzymatic activity in Group I RNA intron further enhanced the interest as one starts to envision an RNA world as a fundamental evolutionary step in the origin of living matters (where RNA nucleic acids would be the repository of the information as well as the processing machinery). Fifty-seven years after the discovery of DNA as the substrate of genetic information by Avery, McCarthy & McLeod [2] and forty-seven years after the elucidation of the B-DNA structure by Watson and Crick [3], the flurry of work related to nucleic acid is inexhaustible.

DNA molecules through their chemistry are ideal building blocks for nanotechnology and computation. We will present basic facts on DNA structures and on the physics of its self-assembly to show that DNA molecules might constitute the best material to design functional allosteric biomolecules. In other words, the structure-function relationship of DNA is both predictable and designable to stimulate applications.

In this section, we present very elementary data related to the structure of DNA molecules. We will focus on structural information on B-DNA structures, hairpin structures as well as base stacking, relevant to our study of DNA molecular recognition and fluctuation.

1.1.1 Primary structure of DNA.

The chemical structure of a single strand of deoxyribonucleic acids is presented in Figure 1.1. Many features, such as the charges, the rotation angles, and the hydrogen-bond donor and acceptor, are crucial to the understanding of the structure and dynamics of DNA molecules. Three chemical components can be distinguished: the phosphate backbone, the sugars and the bases.

The phosphate skeleton is negatively charged (one charge *per* nucleotide) which implies the electrostatic repulsion of two oligonucleotide strands unless the phosphate charges are electrostatically screened. The phosphate backbone is also highly flexible: six angular rotation angles are available (usually names $\alpha, \beta, \gamma, \delta, \epsilon, \zeta$ -cf Figure 1.1).

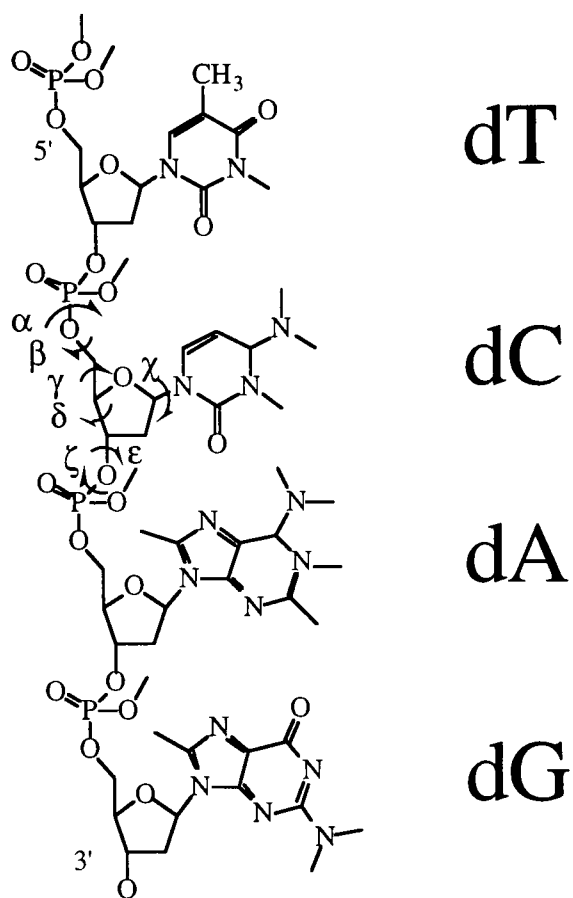


Figure 1.1: Primary structure of a DNA molecule.

The sugars (deoxyribose) are likely to pucker but they are rather inert chemically (unlike ribonucleic acids RNA- where the additional hydroxyl group on C_2 of the ribose can act as an hydrogen-bond donor). There are 4 degrees of freedom in the sugar conformational changes (five atoms are described by 5×3 Cartesian coordinates, minus 5 constraints due to the fixed lengths of the covalent bonds, minus 3 rotational constraints minus 3 translational constraints for the center of mass)

The DNA bases are the information carriers and the active components of DNA molecules. Their pattern of hydrogen bond donor and acceptor enables the sequence-specific formation of secondary structures and the template-guided replication of the DNA. The bases are rigid aromatic rings linked to the sugar by a freely-rotating glycosidic bond (this angular degree of freedom is commonly called χ). This torsional mode is crucial to characterize the helical structure of double-stranded DNA as it measures the involvement of bases in pairing of a secondary structure. Four bases are naturally present in DNA molecules: adenine and guanine (purines with two aromatic rings) and thymine and cytosine (pyrimidines with one single aromatic ring). The chain sequence is conventionally presented along the 5'-3' direction of reading.

The dimension of a single-stranded DNA is very much dependent on its conformation. However, the contour length of the sugar-phosphate backbone is classically estimated to be $5\text{\AA}/\text{nucleotide}$ [4].

1.1.2 Formation of B-DNA structures.

The fundamental property of DNA molecules is their ability of self-pairing by complementarity [3]: guanine can interact specifically with cytosine, and adenine can interact specifically with thymidine, to form basepairs of quite similar geometrical span (Figure 1.2). In fact, alternative complementarity patterns (such as Hoogsten pairs, or reversed Watson-Crick pairs) can be accommodated with the base structure, but they will not be relevant to our study of DNA secondary structures.

Many alternative conformations are compatible with the chemistry of DNA molecules, however B-DNA is the preferred conformation *in vivo*, under non-dehydrating conditions. Two complementary and anti-parallel strands spontaneously self-assemble in a B-DNA structure : it is a right-handed double-helix with an helical pitch of 34\AA . The DNA bases are positioned perpendicularly to the helical axis. The double-helix is strongly stabilized by stacking between successive bases. This base stacking wraps 10 basepairs into one helical pitch, which means a separation of 3.4\AA between bases (to be compared with the 5.0\AA -separation along the backbone). The negative repulsion of the phosphates is neutralized by the electrostatic screening by positive counterions from the buffer, as well as a snugged positioning away from each other: the width of the B-DNA helical structure is 20\AA (Figure 1.3), which is basically the distance between the phosphates of two complementary paired bases.

The interaction between complementary bases to form a Watson-Crick basepair is energetically delicate: the hydrogen bonding generates the specificity of the base-base interaction but does not stabilize strongly the structure (typically $5-12k_B T$ per hydrogen bond [5]). It is the base-stacking inside the B-DNA structure (*i.e.* a combination of hydrophobic interaction and electrostatic alignment between the bases' π -electron clouds) which brings most of the stabilizing free energy. As soon as the Watson-Crick pairing was proposed, the stability of GC basepairs (forming three hydrogen bonds) was compared to the stability of AT basepairs (with two hydrogen bonds): typically, a GC basepair has a free energy of $3k_B T$, whereas an AT basepair has a free energy $2k_B T$ (see § 1.2).

The description of the structural coordinates of a basepair requires 16 helical parameters: 2 translational displacement, shear, stretch, stagger, shift, slide, rise, inclination, tip, buckle, propeller, opening, tilt, roll and twist (these parameters have been standardized for the "local-approach" modeling of dsDNA, requiring no a-priori definition of its helical axis; other parameters could have been chosen).

The hydration of B-DNA molecules is crucial to their stability. The hydrophobic interaction of adjacent bases expulses the water out of the stack of the B-DNA structure, whereas water molecules are crucial to neutralize electronegative atoms (phosphate, oxygen of the ribose, nitrogen of the

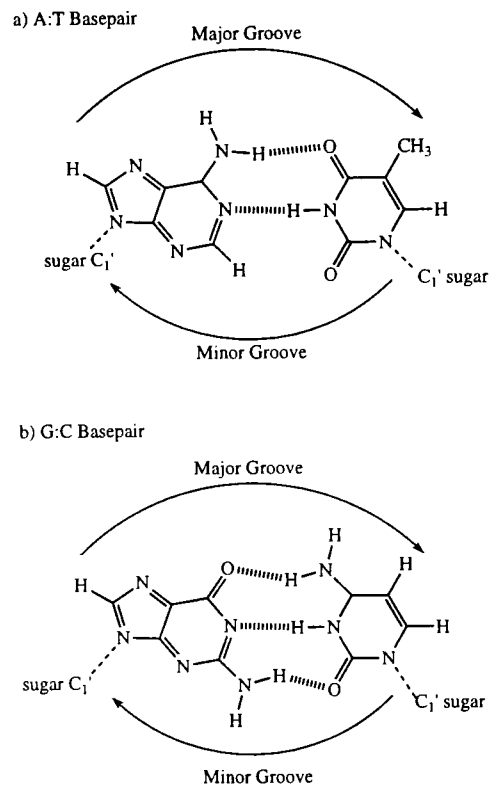


Figure 1.2: Watson-Crick basepairs.

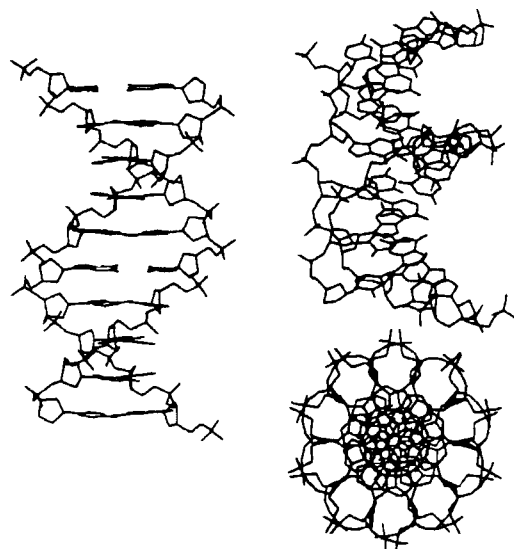


Figure 1.3: Structure of B-DNA.

bases). The question of the average negative charge carried per nucleotide in a B-DNA structure is still an open question and subject of a lot of computational evaluation [6, 7]. At lower hydration level, B-DNA is not stable and gets destabilized into A-DNA (with no known biological function). The average number of water molecules per nucleotide is estimated to 15-25, which makes the modeling of DNA structure by molecular dynamics a computing nightmare (the best simulation computes at most 1ns of the life of a DNA molecule [8]).

As a conclusion, the structure of B-DNA is a delicate balance of contradictory forces: electrostatic repulsion of the phosphate backbones, hydrophobic attraction of adjacent bases, hydrogen-bonding of complementary bases, hydrophilic stabilization of electronegative atoms

1.1.3 Structural deformation of B-DNA.

Deformations of B-DNA are crucial to its function as it introduces irregularities in the double-helix as recognition sites for DNA-binding proteins. The classical example is the sequence-specific curvature of dsDNA. First observations of these unusual structures came from gel electrophoresis whereby the mobility of certain DNA strands were reduced. The usual suspect-sequence (unraveled by point mutation in kinetoplast DNA) is a repeat of a short tract of adenosines (5-6 basepairs) in phase with the helical repeat of the B-DNA (one repeat every 10 bases). The bases in between A-tracts were shown not to influence the bending of the double helix. Structural data are still lacking, but few possible models have been presented to understand systematically the bending of dsDNA: unusual structure between A-tract and consecutive segments, or shrunk wedge of the AA dinucleotide, or unusual structure of the intervening sequences. The last possibility (commonly called non-A-tract model) seems to be in favor nowadays according to the review by Hartmann & Lavery [8].

Many deformations are expected to disturb the order of the B-DNA structure: flipped-out bases, internal loop, bulges, kinks, etc. Unfortunately, the structural analysis of nucleic acid seems to remain a technical challenge, specially in crystallography. Most DNA structures are derived from NMR studies, but the number of deposited structures (494 in March 2000) is not as impressive as the number of solved protein structures [9]. The molecular details are crucial to many published structures, where unusual basepairing, triple-strandedness, stacked loops, stabilized mismatches (such as wobble basepair) are present. A lot of effort has been produced in the field of molecular dynamics, but simulations are as hard as the protein folding problem: even though some periodicity could be taken advantage of, the degeneracy of the free energy landscape (no native structure is expected a priori) complicates the folding problem. Altogether, systematic structural understanding of DNA is still in its infancy.

1.1.4 Single-stranded DNA.

Classical considerations on the primary structure of ssDNA infer that a ssDNA molecule, as a polymer, should behave as a freely-jointed chain. Thus, ssDNA is always regarded as a feature-less highly-flexible polymer. The problem of the flexibility of a ssDNA chain has been addressed mostly to better control the reptation conditions of ssDNA coil in gel electrophoresis, as well as to better predict the stability of large DNA secondary structures.

A classical parameter used to characterize the intrinsic flexibility of a polymer is its persistence length L_p . A polymer conformation can be characterized by the tangent vector $\vec{t}(s)$ to the polymer contour as a function of the curvilinear coordinate s . Then, by definition, the persistence length L_p is the correlation length on the polymer contour under thermal fluctuation:

$$\langle \vec{t}(s) \cdot \vec{t}(s+l) \rangle_s = e^{-l/L_p}$$

One can show that the correlation function is the length scale onto which the polymer bends under $k_B T$, in other words the entropic Young modulus of the polymer coil is $k_B T L_p$.

Three experimental techniques have been used to measure the persistence length of single-stranded DNA (ssDNA), based on analyzing the entropic conformation of ssDNA coils.

Smith *et al.* [10] used micromanipulation of individual denatured λ DNA to stretch the molecule, and measure its entropic-elastic compliance. The force-extension curve is classically fitted with a worm-like chain model [11] yielding the persistence length of the long ssDNA strands: $L_p = 7.5 \text{ \AA}$. in 0.15M NaCl (pH8.0)¹. Note that this measurement is made on very large molecules (λ DNA of 48000 bases), and requires large deformation of the molecule out of the entropic regime.

The second technique (used by Tinland *et al.* [12]) consists in measuring the coefficient of diffusion of large single-stranded plasmids by Fluorescence Recovery After Photobleaching (FRAP). The coefficient of diffusion of the coils are inversely proportional to the radius of gyration, which itself scales with the square root of the size of the polymer measured in terms of persistence lengths. The results show that the persistence length L_p of ssDNA is independent of the salt concentration for $[Na^+] > 0.1M$, and $8 \text{ \AA} < L_p < 13 \text{ \AA}$.

The third technique, implemented by Mills *et al.* [13], consists in measuring the transient electric birefringence where the relaxation time yield information on the rotation of the probed molecules. Duplex DNA and gapped duplex DNA with single-stranded domain are compared to estimate a persistence length of poly(dT) to 20-30 \AA with an helix rise of 5-7 \AA /nucleotide, whereas poly(dA), at 4°C, has a persistence length of 78 \AA , with a helix rise of 3.2 \AA /nucleotide (buffer: 45mM NaCl, TEB, pH8.0).

The two first measurements are quite consistent in presenting ssDNA as essentially a freely-jointed chain (its persistence length is basically one nucleotide). Mills *et al.* points out the intrinsic rigidity of poly(dA) ssDNA, compared to poly(dT) ssDNA.

These measurements were performed on “large” polymers, where the concept of persistence length is indeed relevant. However, for most practical applications (folding and electrophoretic mobility of ssDNA), one would like to directly measure the dynamics of local fluctuation of ssDNA, instead of inferring this dynamics from measurements on “large-scale” polymer. We will present our results on the fluctuations of hairpin-loops, studied by fluorescence correlation spectroscopy, to infer quantitative measures on the rigidity of ssDNA.

1.1.5 Stacking of DNA bases.

Our results on ssDNA intrinsic rigidity will emphasize the higher flexibility of Thymines (T) bases compared to Adenine (A) bases. This effect could have been conjectured from a known phenomena of ssDNA in solution: base stacking.

The base stacking in B-DNA structure is well-documented as it is crucial to the stability of the double-helix. A first description of the energetics of base stacking was presented by Sinanoğlu *et al.* and Abdunur in 1964 [14]: they describe quantitatively the ejection of solvent molecules out of a stack of bases as a surface tension effect. Water having a very high surface tension, the DNA bases tend to self-organize and create a single hydrophobic cavity to minimize its contact with the water molecules. This argument, though quite phenomenological, describes accurately the higher stability of dsDNA in water compared to lower-surface tension solvent (ethanol, formamide, glycerol...).

In fact, electrostatic, hydrophobic, Van der Waals and even hydrogen bonds are likely to contribute to the stabilization of this base stacking [15]. A quantitative understanding of base-base stacking in solution, or inside B-DNA structure is still lacking as it involves a complete description of the water molecules and counterions around the bases.

Base stacking is also crucial (though largely underestimated) to explain the structures of ssDNA in neutral-pH solution: the absorbance of a poly(dA) (recorded as the extinction coefficient at 260nm) was shown to increase by 25% when switching from 0°C to 80°C, as a signature of the melting of the base stacking with a recovery of the hypochromicity of the DNA bases. Dewey & Turner [16] carried

¹Note that the same measurement has been made on double-stranded λ DNA to yield a persistence length of 53nm or 155basepairs in 0.15M NaCl (pH8.0): B-DNA is thus 21 times harder to bend than ssDNA.

out a spectroscopic study of the stacking melting to prove that the base-stacking in poly(dA) is an intramolecular process, occurring on fast timescales (~ 100 ns). They measured the thermodynamics parameters associated with the melting transition of the base stacking: $\Delta H^\circ = 9$ kcal/mol, and $\Delta H^\circ = 26.5$ e.u. per mol of stacks (1 e.u.=1 cal.mol⁻¹.K⁻¹), for the unstacking reaction (buffer: 0.05M of Sodium Cacodylate).

As far as structure is concerned, circular dichroism measurements showed that the bases of a poly(A) are stacked parallel to the axis of the helix, with an angle of 30°-45° between bases [17]. The helix rise has been estimated to 3.4Å per nucleotide by low-angle X-ray scattering (a value similar to the helical rise of B-DNA). Finally, Arnott *et al.* [18] presented a structure of stacked ssDNA for poly(dC) obtained by X-ray diffraction: the poly(C) has an helical structure with a 6-base pitch of stacked bases (Figure 1.4). All these studies have been carried out at low salt concentration (< 0.1 Mol) to enhance the base-stacking effect in ssDNA. Note that all these structural data have been obtained with polyribonucleic acid.

The ability of DNA molecules to self-assemble into a double-helical B-DNA structure satisfies a delicate balance of repulsive (electrostatic and entropic) against attractive (hydrophobic and hydrogen-bond) forces. Conformational changes of a DNA molecule are drastic: from a highly flexible polymer (with a persistence length of ~ 2 bases), it becomes a very rigid polymer upon base-pairing (with a persistence length of ~ 150 bases).

Few information are available concerning the structures, thermodynamics and dynamics of single-stranded DNA. However, as we will show in the next section, the importance of ssDNA domains is crucial to the prediction of DNA secondary structures. So far, the contribution of ssDNA domains to the overall stability of secondary structures has been generally overlooked. One of the questions, addressed by our thesis work, relates to the dynamics of conformational fluctuations of DNA; this study will lead us to emphasize the dynamic consequence of ssDNA base stacking as a competing structure to B-DNA, as well as an elastic constraint regulating the stability of DNA secondary structures.

1.2 Statistical Physics of DNA secondary structures.

As soon as the structure of B-DNA had been proposed by Watson& Crick [3], understanding the physico-chemistry of the formation of DNA secondary structures became of fundamental importance. The core issue is to understand the specific interaction of complementary strands of DNA, or formation of duplex DNA, and extrapolate this information for the prediction of DNA nucleic acid structure. This problem is ubiquitous to many techniques in molecular biology (Northern blotting, Southern blotting, Polymerase Chain Reaction, DNA chips,...) under the denomination of DNA hybridization. Although inappropriate, this name conveys the striking picture of interaction of two distinct molecules to form one single duplex object (the word and the techniques were introduced by Hall & Spiegelman in 1961 to describe the formation of duplexes between ³²P-labelled RNA and complementary DNA, after melting and cooling down of the mixture: the formation of the DNA-RNA hybrid was then detected by density-gradient centrifugation [20]).

In this section, we will review statistical physics models of double-helix formation, specially the "nearest-neighbor" (NN) model, introduced to describe oligonucleotide hybridization. Our work on molecular recognition of ssDNA sequences by molecular beacons relies heavily on the classical analysis of hybridization experiment by the NN-model. We will also present theoretical considerations on the conformational fluctuations of ssDNA domains inside large secondary structures, to emphasize the need of experimental validation.

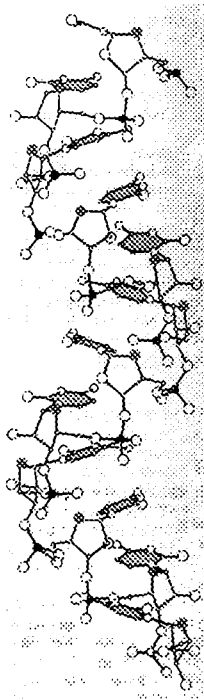


Figure 1.4: Structure of polyC determined by X-ray diffraction (from [18], cited in [19]). The C-bases have been shaded: to show their self-organization into a single-stranded stacked helix.

1.2.1 Phenomenology of DNA hybridization.

DNA hybridization was first observed by Marmur & Doty in 1962 [21] by UV spectroscopy. The experiment consisted in monitoring DNA UV-absorption (at 260nm) at low temperature and at high temperature. The melting of the dsDNA into ssDNA with specific hyperchromicity was shown to be reversible upon recooling of the solution. Thus, dsDNA was shown to assemble reversibly by hybridization of two complementary ssDNAs'

As soon as this fascinating transition was observed, UV spectroscopy techniques became ubiquitous to the monitoring of the coil-to-helix transition: essentially, helical DNA structures were shown to have a lower extinction coefficient at 260nm than ssDNA (this hyperchromicity due to the base stacking in the helix can be used to monitor the degree of hybridization), the birefringence of a DNA solution is also altered by the DNA's helical structure hence the use of circular dichroism to monitor the helix-to-coil transition. Alternatively, NMR of the imino proton exchange can be used to measure the involvement of hydrogen atoms in hydrogen bonds. Finally, Raman spectroscopy provides quantitative measurements of vibrational spectra of all the bonds in the dsDNA.

Many "cooking" recipes are routinely used, in molecular biology lab, to predict the stability of oligonucleotides duplex. These empiric "laws" mainly predict the melting temperature T_m of the homologous DNA duplex (T_m is, by definition, the temperature at which all of the DNA is single-stranded and half is double-stranded). For example, for the dimer $X(n)_2$, where $X(n) = 5' - A_n T_n - 3'$ [19],

$$\begin{aligned} 2 X(n) &\rightleftharpoons X(n)_2 \\ \frac{1}{T_m} &= A + \frac{B}{n} \text{ (for } n < 15). \end{aligned} \quad (1.1)$$

This law can be justified qualitatively: the formation of a duplex of two complementary DNAs' costs an entropy linear with the length of the oligonucleotide $\Delta S(n) = \Delta S_{nuc} + n\Delta S_{bp}$ (where ΔS_{nuc} is the nucleation entropy for bringing together two separated objects, and ΔS_{bp} the entropy *per* formed basepair), and gain an enthalpy linear with the chain of the oligonucleotide $\Delta H(n) = n\Delta H_{bp}$ (ΔH_{bp} is the enthalpy *per* formed basepair). Thus, the melting temperature of the structure (corresponding to a null free energy for the transition $\Delta G(T_m) = \Delta H - T_m\Delta S = 0$) is:

$$T_m(n) = \frac{\Delta H(n)}{\Delta S(n)} = \frac{n\Delta H_{bp}}{\Delta S_{nuc} + n\Delta S_{bp}} \Rightarrow \frac{1}{T_m(n)} = \frac{\Delta S_{bp}}{\Delta H_{bp}} + \frac{\Delta S_{nuc}}{\Delta H_{bp}} \frac{1}{n}. \quad (1.2)$$

(compare equations 1.1& 1.2).

The physics of this qualitative law takes into account the nucleation price vs. the basepairing gain to predict the stability of a duplex DNA. For long DNA duplexes (large n), the nucleation becomes negligible and the melting temperature is then essentially independent of the length of the duplex:

$$T_m(\text{large } n) \sim \frac{\Delta H_{bp}}{\Delta S_{bp}} \sim 350K \text{ in } 0.1M \text{ NaCl}$$

Another law is classically presented to take into account the effect of the DNA concentration on the melting DNA duplex. One expect intuitively that, for high concentration of DNA, when one strand unbinds from its complementary strand, the probability to interact with another strand is very high, the overall probability of one single strand to be left unhybridized becomes negligible. This is purely an entropic effect of counting the number of single strands available for hybridization, thus:

$$\frac{1}{T_m} = A + C \ln([DNA]). \quad (1.3)$$

where $A > 0$ and $C < 0$. This formula (classically used in the Van't Hoff method for determining thermodynamic parameters –see § 2.2 for a full derivation) captures the increase in the melting temperature of the duplex DNA, as one increases the concentration of DNA.

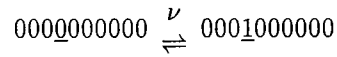
The laws (1.1) & (1.3) are coarse as they neglect any sequence-dependence effect and other “details”. A statistical-physics description of the DNA hybridization refine the prediction of the stability of duplex stability.

1.2.2 Zipping toy-model for the coil-to-helix transition.

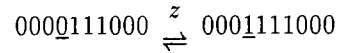
Zimm & Bragg introduced a simplified model of helix formation to capture the essential features of the thermodynamic transition helix-to-coil [22]. The initial model was developed to describe the formation of α -helices in polypeptide, but the formalism completely holds to the case of helix-to-coil transition in nucleic acids.

The nucleic chain is described as a binary word, whereby 0 corresponds to an unpaired base, and 1 corresponds to a paired base. Two type of thermodynamic equilibria are needed to describe the whole thermodynamics of the transition:

1. Basepair nucleation: when a basepair is formed without any surrounding preformed basepair



2. Helix propagation or zipping: when a basepair is added at the extremities of an already formed domain of basepairs



Classically, ν has been related to the zipping constant z with a parameter μ : $\nu = \mu z$. One can anticipate that $0 < \mu < 1$, as the nucleation of a basepair must be less favorable thermodynamically than the zipping. With these definitions, one can assign an equilibrium constant K for the melting of any configuration of basepairs between two complementary strands into two random coils. For example, for the word 001100011110100, there are six interfaces paired-unpaired (01 or 10), and four neighboring basepairs (11), thus :

$$K(001100011110100) = \nu^6 z^4 = \mu^6 z^{10}.$$

This description of the helix formation is similar to an one-dimensional-Ising model for ferromagnetism. It enables us to predict the melting temperature (at which the transition occurs), the width of the transition, and other thermodynamic characteristics.

All the thermodynamics of the hybridization system can be derived from the partition function, which can be fully expressed analytically [22]:

$$\mathcal{Z} = \sum_i K_i = 1 + \sum_{l=1}^{E(\frac{n-2}{2})} \mu z^l \sum_{k=1}^{n-l-2} \frac{(k-1)!(n-k-2)!z^k}{l!(l-1)!(k-l)!(n-k-l-2)!}, \quad (1.4)$$

where $E(x)$ is the floor integer of x . Although this expression seems indigest, it can be easily approximated with the Kramer's transfer matrix method [23, 24]

$$\mathcal{Z} \sim \left(\frac{1 + z + \sqrt{(1-z)^2 + 4\mu z}}{2} \right)^n$$

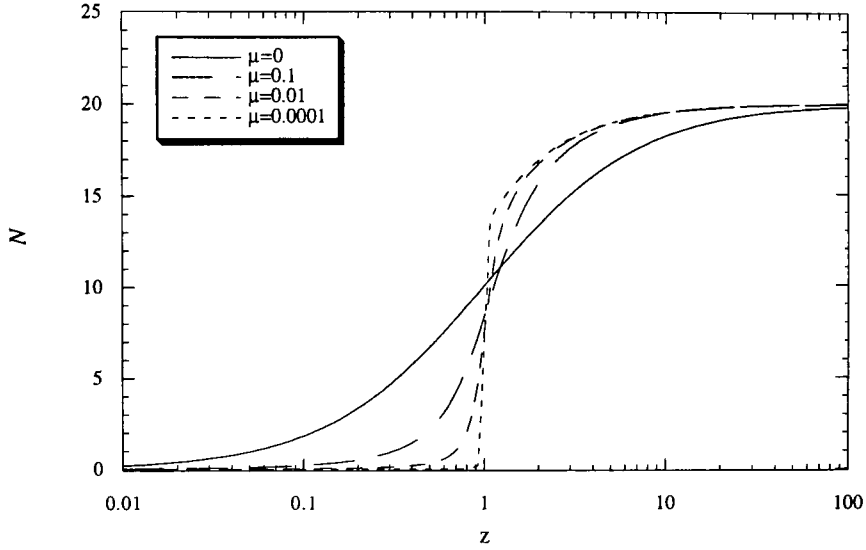


Figure 1.5: Average number \mathcal{N} of basepairs in a 20-mer duplex *vs.* z (zipping probability)

This expression is an approximation as the smaller eigenvalue of the Transfer Matrix has been neglected: this is valid mostly in the case of very large n . Once the partition function has been evaluated, it is easy to derive the average number \mathcal{N} of bonds formed in the structure:

$$\mathcal{N} = \frac{\partial \ln \mathcal{Z}}{\partial \ln z}$$

$$\Rightarrow \mathcal{N} \sim n \frac{z(1 - \frac{1-z}{1+z+\sqrt{(1-z)^2+4\mu z}})}{1+z+\sqrt{(1-z)^2+4\mu z}}.$$

This expression shows that, with the minimal assumption of the zipping model, Zimm & Bragg could characterize the phase transition between helix and coiled DNAs' and show the competition between end-effects and basepair stacking. On Figure 1.5, we plot the average number of formed basepairs between two n -mer oligonucleotides. This graph shows that the phase transition (or melting temperature of the duplex DNA) corresponds to $z = 1$.

Note that, for $\mu = 1$, there is no interaction between successive segments of formed basepairs, and z is simply the equilibrium constant for the formation of a basepairs. The average number of formed basepairs is then:

$$\mathcal{N} = n \frac{z}{z+1}.$$

On the other hand, when μ decreases, the probability of completing the basepairing of a segment is much higher than the probability of initiating (or nucleating) the basepairing in another region. Thus, the sharpness of the transition increases: for $\mu = 0$, basepaired domains grow as soon as nucleated, the transition occurs at $z = 1$, with an infinite cooperativity (the width of the transition is null).

1.2.3 Nearest-neighbor model of DNA hybridization.

Zimm & Bragg's zipping model carries all the basic features of the transition coil-to-helix in DNA hybridization (inefficient nucleation, efficient zipping of the basepaired domain) However, it oversim-

plifies the details of the molecular interactions between bases to make the partition function computable analytically: all sequence-dependent effects are lost in the zipping toy-model. We present here a more accurate model called “nearest-neighbor model”, refining the Zimm & Bragg model.

To predict theoretically the stability a duplex of oligonucleotides, one must enunciate all the possible states of hybridization ($\{i = 1, 2, 3 \dots\}$), and assign their probability p_i of occurrence. Classical Statistical Physics give an evaluation of p_i as a Boltzmann weight of the free energy G_i of the state i :

$$p_i = e^{-\frac{G_i}{k_B T}}.$$

All the thermodynamics of the hybridization system can be derived from the partition function \mathcal{Z} :

$$\mathcal{Z} = \sum_i p_i = \sum_i e^{-\frac{G_i}{k_B T}} \quad (1.5)$$

In fact, the relevance of the partition function is valid up to a normalization factor, such as classic handling of the free energy in the DNA hybridization case can be replaced by the difference between the free energy of the conformation i , and the free energy of randomly-coiled conformation:

$$q_B = \mathcal{Z} * e^{+\frac{G_{coil}}{k_B T}} = \sum_i e^{-\frac{\Delta G_i}{k_B T}} \quad \text{where} \quad \Delta G_i = G_i - G_{coil}, \quad (1.6)$$

q_B has been classically studied instead of \mathcal{Z} , as single-stranded properties were believed to be independent of temperature, salt, and concentration [19]. Thus, the only relevant structures in this model are basepairing and double-strand stacking.

On Figure 1.6, we sketch all the possible conformations to be taken into account to evaluate the partition function. The probability of nucleating the first basepair is named ν ; the probability to nucleate a basepair without neighboring basepair is different from ν when the two complementary oligonucleotides are already held together, and it is named ν_j (the subscript indicates that this nucleation weight depends a lot on the conformation of the bubble of j bases left between the two domains of paired bases); the probability of zipping an additional basepair, next to an already formed basepair, is named $\zeta_{i,i+1}$: this weight is very sequence-dependent on the identity of the basepairs $< i, i + 1 >$ to be stacked onto each other.

For example, for the conformation i circled in Figure 1.6, the equilibrium constant K_i for the transition between random coils and conformation i is:

$$K_i = \nu \prod_{l=1}^{n-1} \nu_l \prod_{\text{domain}} \zeta_{i,i+1} \quad (1.7)$$

The difference between the Nearest-Neighbor model and the Zimm & Bragg zipping model consists in weighting the formation of adjacent basepairs with a more realistic sequence-dependent weight $\zeta_{i,i+1}$. Once again, an approximation is taken to neglect the long-range and hard to compute effect of the second nucleation weight; the hybridization of two short oligonucleotides is assumed to be a one-nucleation process, and only two states dominates the partition function: fully hybridized oligonucleotides or fully coiled. The partition function for the hybridization of two complementary short oligonucleotides is thus:

$$\mathcal{Z} = 1 + \nu \prod_{i=1}^{n-1} \zeta_{i,i+1} \quad (1.8)$$

Experimental studies on melting transitions of short oligonucleotides duplexes validate this All-or-None approximation of the partition function: the melting transition (as monitored by spectroscopy

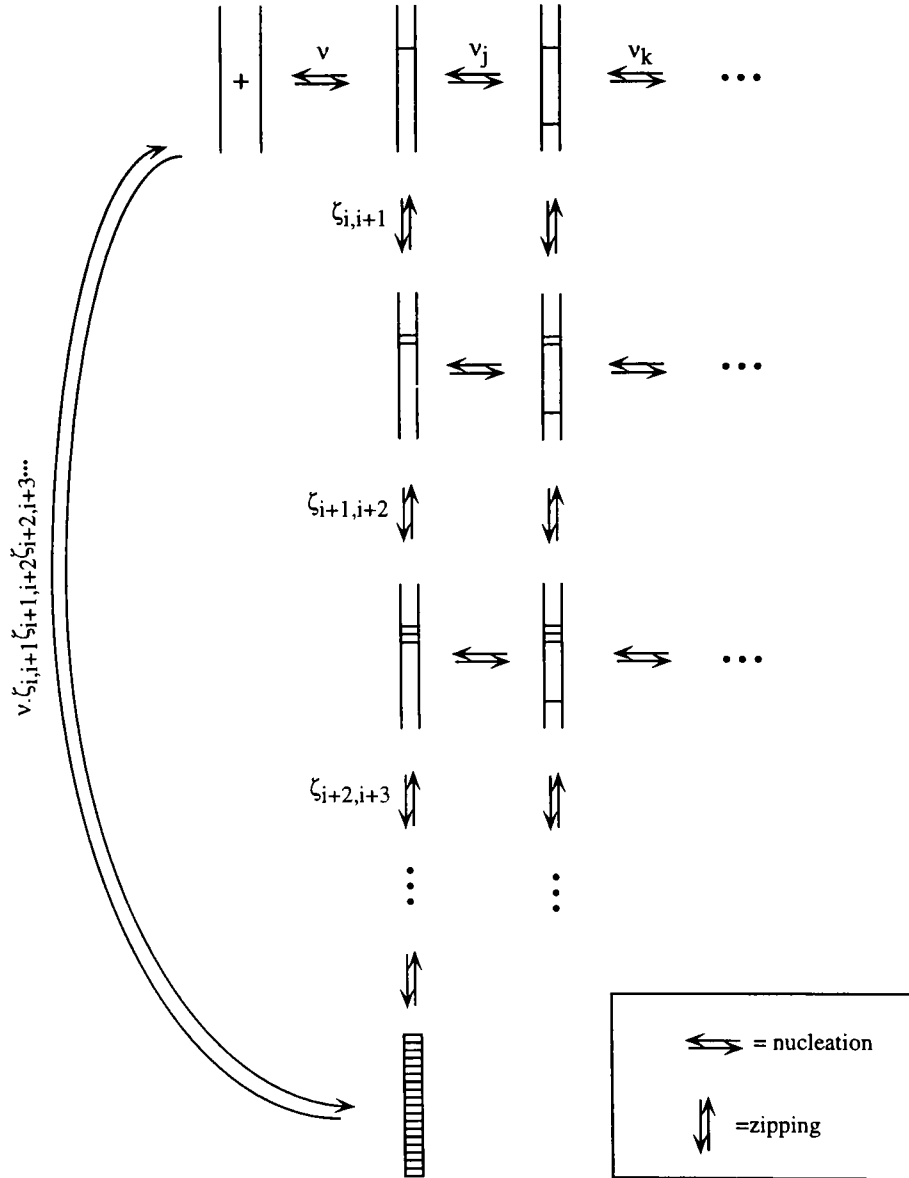


Figure 1.6: Enunciation of all the possible conformations of two hybridizing oligonucleotides. We classify the configuration by the number of formed basepair, and present the equilibrium constant for each transition. To compute the probability of a given configuration, one must multiply the equilibrium constant to go back to the randomly-coiled state.

Nearest-Neighbor Thermodynamic weights for DNA hybridization in 1M NaCl

Basepair Sequence	ΔG^0 (kcal/mol)	ΔH^0 (kcal/mol)	ΔS^0 (cal/mol/K)
AA/TT	-1.00	-7.9	-22.2
AT/TA	-0.88	-7.2	-20.4
TA/AT	-0.58	-7.2	-21.3
CA/GT	-1.45	-8.5	-22.7
GT/CA	-1.44	-8.4	-22.4
CT/GA	-1.28	-7.8	-21.0
GA/CT	-1.30	-8.2	-22.0
CG/GC	-2.17	-10.6	-27.2
GC/CG	-2.24	-9.8	-24.4
GG/CC	-1.84	-8.0	-19.9
Initiation term G.C	+0.98	0.1	-2.8
Initiation term A.T	+1.03	2.3	4.1
Symmetry correction	0	0	-1.4

or by calorimetry) can be fitted with a one-step equilibrium constant, and the products of the nearest-neighbor energies $\zeta_{i,i+1}$ can be deconvoluted from sets of melting experiments [25]. Ultimately, results on DNA hybridization are compiled into a table of 10 thermodynamic weights ΔG_{ij} to assess the stability of each basepair [26]. Table 1.2.3 presents the nearest-neighbor thermodynamic weights compiled by SantaLucia in [25].

The application of the Nearest-Neighbor Model for DNA hybridization prediction is straightforward: one must sum up individual contributions of stacked basepairs in the duplex phase to evaluate the free energy of that duplex:

$$\Delta G^0(\text{total}) = \sum_{i=1}^{n-1} \Delta G^0(i, i+1) + \Delta G^0(\text{initiation}) + \Delta G^0(\text{symmetry}).$$

All these values being given for normal conditions (1M of DNA complementary oligonucleotides, extrapolated from dilute regimes). One must add an entropic correction to the entropy to compute the free energy for a specific setting. For a solution containing a concentration C_X of the oligonucleotide X , and $C_{\bar{X}}$ of the complementary oligonucleotide \bar{X} , the entropy of coil-to-helix transition is:

$$\Delta S = \Delta S^0 + R \ln \left(\frac{C_X - C_{\bar{X}}}{2} \right), \text{ for } C_X > C_{\bar{X}}$$

and the melting temperature T_m of the duplex (defined by $\Delta G(T_m) = 0$) is:

$$T_m = \frac{\Delta H}{\Delta S} = \frac{\Delta H^0}{\Delta S^0 + R \ln \left(\frac{C_X - C_{\bar{X}}}{2} \right)}. \quad (1.9)$$

Equation (1.9) will be crucial to our analysis of the thermodynamics of DNA hybridization (see Chapter 2).

Experimentally, there are many techniques to monitor the formation of the DNA duplex (UV hyperchromicity, circular dichroism, calorimetry...): one can then measure a melting curve of an oligonucleotide duplex by monitoring the amount of duplex formation as a function of temperature. The melting temperature T_m is experimentally determined as the mid-point of the melting curve (50% of the oligonucleotides are hybridized, 50% of the oligonucleotides are free in solution). By plotting the inverse of the melting temperature $1/T_m$ as a function of the concentration of oligonucleotides (in log scale), one gets a linear plot to be fitted to yield the entropy ΔS^0 and enthalpy

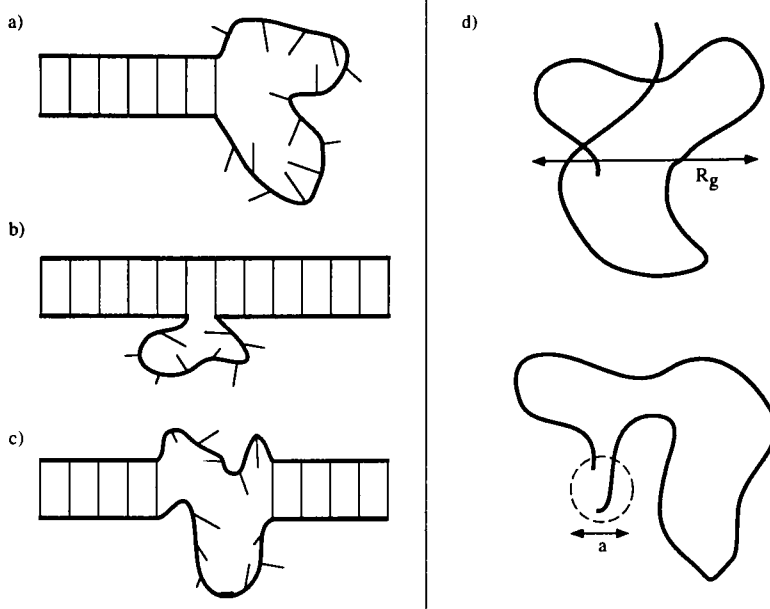


Figure 1.7: Single stranded domains in DNA secondary structures: a) hairpin loops; b) bulges; c) bubbles; d) conformations of a random coil of ssDNA, before and after cyclization.

ΔH^0 of the hybridization. This method (commonly called Van't Hoff method [27]) is very robust as one needs only to determine the inflection point of the melting curve. A direct fit of the melting curve could yield directly ΔH^0 and ΔS^0 , but this fit would rely heavily on the determination of the baseline (low temperature) and plateau (high temperature) of the melting curve. For most hybridization experiment, the Van't Hoff approach is the most robust way to determine the thermodynamics parameters.

1.2.4 Single-stranded DNA loop entropy.

Most DNA secondary structures include single-stranded domains of unpaired DNA, such as hairpin loops, bulges, bubbles or double-strand loops (Figure 1.7). To accurately predict the stability of these secondary structures, one must be able to assess the free energies associated with these single-stranded domains. In this paragraph, we will focus on the single-stranded loops of the hairpin structures.

Very few quantitative data and structures are available to predict the stability of single-stranded loops, but minimal assumptions can be taken: single-stranded DNA is expected to behave as a pure random coiled polymer such as ssDNA domains in DNA secondary structure can be described as a freely-jointed chain, with geometrical constraints (cyclization) imposed by the double-stranded domains [4, 28]. The free energy contribution of the single-stranded domains is thus purely entropic. One must be able to evaluate the total number of conformations Ω in the loop and in the coil to estimate the entropy of the loop ΔS :

$$\Delta S = R \ln \left(\frac{\Omega_{loop}}{\Omega_{coil}} \right) = R \ln p_{cyclization}.$$

A straightforward argument tells us that the cyclization probability $p_{cyclization}$ (probability to

find the two polymer ends within a volume a^3 , where a is a nucleation length) must scale with the volume of the random coil (R_g^3 , where R_g is the gyration radius of the coil):

$$p_{cyclization} = \left(\frac{a}{R_g}\right)^3 = \frac{\Omega_{loop}}{\Omega_{coil}}.$$

One must then estimate how R_g scales with the loop length N_{loop} . For a purely random coil, the polymer chain is described with a random walk of N_{loop} steps of length b , thus:

$$R_g = b\sqrt{N_{loop}},$$

$$\text{thus } \Delta S_{loop} = -\frac{3}{2}R \ln N_{loop} + 3R \ln\left(\frac{a}{b}\right).$$

The nucleation length a is somewhat arbitrary but must be a molecular scale comparable to $b \sim 5\text{\AA}$. Thus the constant term is classically neglected [29].

The model of ssDNA as a random coil is underestimating excluding-volume interaction along the DNA chain, which tends to swell the polymer. Classical argument gives the scaling law for the radius of gyration of a self-avoiding polymer in good solvent:

$$R_g = bN_{loop}^{0.6}$$

This new scaling exponent implies a stronger dependence of the loop entropy with the loop length:

$$\Delta S_{loop} = -1.8R \ln N_{loop}.$$

The free energy of the loop is then:

$$\Delta G_{loop} = 1.8RT \ln N + cste. \quad (1.10)$$

Note that this evaluation of the free energy of the loop relies heavily on the model of the hairpin loop as a pure random coil. This approximation is valid mostly for large loops, where enthalpic effects are classically negligible. Note as well that this implies that the free energy of a ssDNA loop is sequence-independent. In our work, we address the issue of the free energy of ssDNA loops, and challenge the assumed random-coil conformations, to get a more accurate model of ssDNA domains among secondary structures.

In conclusion, the competition between nucleation/zippering of basepairs and the entropy of unpaired domains determines the stability of DNA secondary structure. Both phenomena can be described with minimal assumption in a simple model: the nearest-neighbor model. We will use this classical thermodynamic analysis to study quantitatively the molecular recognition of ssDNA by molecular beacons, as well as to address issues of DNA folding.

Our work deals also with conformational switches in molecular beacons: as Watson-Crick basepairing sets simple physico-chemical rules to the folding of nucleic acids, DNA molecules constitute for us the building block of choice to design molecular switches. In other words, we are going to use the simple sequence-secondary structure relationship of DNAs' to generate a toolbox of molecular objects whose allosteric changes are fully predictable, tunable and functional.

1.3 DNA conformational fluctuations.

Fluctuation modes of DNA molecules drive many biologically-relevant function of DNA-binding proteins. The general paradigm is that most DNA-binding proteins do not induce mechanically conformational changes (bending, coiling, stretching, unzipping...) in double-stranded DNAs', but rather interact and stabilize pre-existing fluctuations in the structure of the dsDNA. The relevance of this paradigm lies in the understanding of the sequence specificity of DNA fluctuations modes, in order to understand the specificity of DNA-protein interaction. In this thesis, we will focus on the fluctuation of secondary structures of DNA to better understand biologically relevant problems: hairpin-loops will be shown to spontaneously fluctuate between open and closed configurations, with a dynamics very much dependent on their sequence; dsDNA will display breathing modes whose shape and dynamics challenge classical understanding of the role of RNA polymerase in the initiation of transcription.

Let us present a quick example illustrating the possible biological relevance of DNA conformational fluctuations. The TATA box is a stretch of 8 basepairs of sequence 5'-TATATAAA-3' upstream of transcription sites in eukaryotes. The TATA binding protein (TBP) interacts very specifically with the TATA box, and the structure of the DNA-protein complex unravels a bending angle of 80° towards the major groove in the classical B-DNA structure of the double-stranded DNA [30, 31]. Independent studies showed that prebending the TATA box towards the major groove increases TBP affinity compared to linear TATA box or prebent TATA box towards the minor groove [32]. Thus, TBP can be suspected not to induce bending after binding the TATA box, but rather to interact specifically with a prebent stretch of dsDNA. However, the bending angle of the bare dsDNA does not match *ab initio* the structural requirement of TBP: the angular match might be reached by thermal fluctuation of the already prebent DNA (instead of bending by TBP). Complete experimental evidence for that model are still lacking but initial attempts (further discussed in § 1.3.2) do not find any time separation between the binding and the bending events. This experimental evidence is crucial to understand the sequence specificity of TBP: the transcription factor does not have to scan blindly all the DNA sequence in the chromosome, but limits its interaction to structurally-prebent DNA domains. The reading of the DNA sequence is thus more discriminatory as it involves "aberrant" structures of DNA.

In the following section, we will review the experimental studies of these fluctuation-driven DNA-protein interaction. The experimental problem consists in unraveling those "flimsy" modes of DNA and address their relevance in terms of DNA-protein interaction.

Two experimental approaches have been developed. In mechanical studies, DNA micromanipulation "distorts" the DNA and stabilizes the fluctuation modes under investigation to ultimately show enhanced binding. In spectroscopic studies, the transition state has a specific fluorescent signature (fluorescence energy transfer), and the binding limiting step can be directly observed, molecule by molecule. We will show how the fluorescence energy transfer technique, applied to nucleic acid molecules, can unravel local conformational changes within large DNA. This will introduce the use of molecular beacons to monitor conformational fluctuations in DNA secondary structures. Specifically, we will review the current understanding on breathing modes in dsDNA, and their relevance to the initiation of transcription in prokaryotes.

1.3.1 Mechanical micromanipulation of DNA molecules.

The first direct experimental proof of fluctuation-driven interaction of proteins with DNA comes from a study of the polymerization of RecA onto dsDNA by the Chatenay group [33]. RecA is a 38kDa protein involved in strand-exchange and recombinaison of homologous strands of DNA. Electron microscopy studies show that RecA binds and stretches dsDNA by a factor of 1.5 (the

helical pitch is stretched from 34Å to 51Å [34]). Single-molecule micromanipulation of long dsDNA unraveled a new phase of dsDNA (called stretched DNA, or S-DNA) [35], confirmed by molecular dynamics simulation. The helical pitch of the S-DNA is measured to be 58Å *i.e.* 1.7 times larger than the helical pitch of B-DNA. The predicted structure of dsDNA in RecA-DNA filament has a lot of similarity with the predicted structure of S-DNA [36]: notably, their helical paths are comparable (51Å for RecA-DNA and 58Å for S-DNA, against 34Å for B-DNA). This observation led to the proposed model that RecA does not directly stretch the B-DNA to bind, but rather binds specifically to domains of S-DNA in the double helix.

The experimental evidence proving this model relies on single-molecule stretching of DNA. Léger *et al.* monitored the kinetics of RecA binding on a pre-stretched λ -DNA [33]. RecA induces a stretching or shrinking of dsDNA (depending if the DNA is originally mostly in the B-phase or in the S-phase), thus RecA binding onto the dsDNA can be monitored by measuring the length extension of the molecule as a function of time (for each experiment, a different constant loading force is applied onto the DNA). As the loading force is increased, the probability of finding stretched domains of S-DNA in the B-DNA increases. The probability of finding domains with an helical pitch matching that of the RecA-bound DNA structure increases as well. The typical binding timescale for RecA was found to be reduced for stretched DNA, consistently with the amount of pre-stretched DNA. In fact the prerequisite pre-stretching of B-DNA into S-DNA can be achieved by thermal fluctuations as it requires only $3 k_B T$ per basepair [33].

This single-molecule manipulation of DNA molecules is crucial to address the validity of fluctuation-driven binding of RecA. Stretching the dsDNA molecule enables the tuning of its initial state to match better the binding pocket of RecA. At equilibrium, this state would be accessible only through rare thermal fluctuations of the B-DNA structure. The mechanical distortion of the DNA makes these fluctuation states the dominant phase in the dsDNA, and accelerates the RecA binding.

Many other experiments have unraveled similar fluctuation-driven molecular interactions. Allemand *et al.* [37] showed that dsDNA under stretching (>3 pN) and positive supercoiling ($\sigma > +0.037$) adopts an inside-out structure whereby the phosphate backbones are buried inside the structure and the bases are sticking outside. This new phase (called P-DNA, and similar to the ill-fated structure of dsDNA predicted by Pauling *et al.*) is relevant to dsDNA packaging in Pfl bacteriophages.

Strick *et al.* [38] were able to apply a negative supercoiling to dsDNA ($\sigma < -0.036$) while stretching a λ -DNA with a force around 2pN. They showed that this mechanical distortion of the dsDNA induced its denaturation by presenting complementary ssDNA: when the mechanical distortion is released, the DNA molecule does not go back to its resting topological conformation, as the ssDNA probe remains bound to the denatured domain in the dsDNA. The authors noticed also that the domains which denature first are AT-rich, which is consistent with models of supercoiling-induced regulation of transcription initiation or recombinational repair.

Another application of single-manipulation of large DNA molecules is the monitoring of chemical reactions. Shivashankar *et al.* [39] presented an experimental study of the assembly of RecA filament onto dsDNA from λ -bacteriophage. Single λ DNA micromanipulation enables the extension and measure of its contour length. RecA binding induces the stretching by a factor 1.5 as described before, thus it can be directly monitored by plotting the DNA contour length as a function of time. The kinetics of RecA assembly-disassembly were studied and a model of nucleation-growth presented: the authors could distinguish between single-nucleus growth in the case of ATP-driven assembly of RecA, and multi-nucleus growth in the case of non-hydrolyzable ATP γ S-driven assembly. In that study, the single-molecule micromanipulation of DNA molecules bypasses the synchronization problem of more classical enzymology techniques.

New DNA structures have thus been discovered by micromanipulation of single DNA molecules, as well as new modes of protein DNA-binding, controlled by thermal fluctuation. Almost explicitly, micromanipulation provides a handle to the experimenter to control DNA conformations.

1.3.2 Spectroscopic studies of DNA structures.

A whole different approach, based on spectroscopic tools, can be used to analyze DNA's conformational fluctuations and further address the issue of their relevance for DNA-binding proteins. Instead of perturbing mechanically the DNA, the experimenter can visualize the spontaneous thermal fluctuations of the DNA molecules, and probe the phase space of the available conformations.

Single-molecule spectroscopy is also a field which showed great developments during the 90s' [40]. The most ubiquitous tool used in these studies is Fluorescence Energy Transfer (FET) between two fluorophores. If two properly-chosen fluorophores are in vicinity (typically less than few nanometers), an electromagnetic coupling can be established and the excitation energy can be transferred from a donor fluorophore to the acceptor fluorophore. The efficiency of the coupling is very distance-dependent, such as FET can be used as a very quantitative "spectroscopic ruler". In this section, we will introduce the classical theory of FET and present its application in spectroscopic studies of DNA conformational fluctuations.

Theory of Fluorescence Energy Transfer.

Fluorescence Energy Transfer is quantum in its nature. However, a classical description of the electromagnetic coupling can be used to evaluate the efficiency of the coupling.

When a fluorophore is pumped into an excited state, it adopts a permanent dipole (excitation dipole). The common pathway of relaxation consists in emitting a photon (of lower energy, *i.e.* higher wavelength). When another fluorophore (the acceptor) is presented close to the already excited fluorophore (the donor), a dipole-induced dipole interaction sets, and the excitation energy is transferred from donor to acceptor. The conditions for efficient coupling are then very intuitive: donor and acceptor must be close to each other, the donor emission spectrum must match the acceptor excitation spectrum, and the orientations of the two fluorophores must be such, that the emission dipole of the donor is aligned with the excitation dipole of the acceptor.

Förster predicted the efficiency of the FET coupling to be [41]:

$$\eta = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6}, \quad (1.11)$$

$$\text{where } R_0 = 0.211 \left[\kappa^2 n^{-4} Q_D \left(\int_0^\infty F_D(\lambda) \epsilon_A(\lambda) \lambda^4 d\lambda \right) \right]^{\frac{1}{6}} \text{ in } \text{\AA}, \quad (1.12)$$

where $F_D(\lambda)$ is the normalized fluorescence intensity emitted by the donor, $\epsilon_A(\lambda)$ is the extinction coefficient of the acceptor, Q_D the quantum efficiency of the donor fluorophore, n the refraction index of the surrounding media, κ^2 a geometrical parameter evaluating the relative alignment of the donor and acceptor dipoles (for any practical purpose, in the case of randomly-oriented dyes, $\kappa^2 = 2/3$). The characteristic distance R_0 for FET is commonly called the Förster distance. It varies from one donor-acceptor pair to another: typically, for fluorescein and rhodamine, $R_0 = 54 \text{\AA}$, while, for EDANS and DABCYL, $R_0 = 33 \text{\AA}$. The resulting theoretical distance-dependence for the FET yield is plotted on Figure 1.8.

Theoretically, there is not limit on the spectral separation one can choose between donor emission and acceptor absorption: a large spectral separation would simply imply a small overlap of the two spectra, thus a short Förster distance (*cf* equation 1.11). A good yield of FET would be preserved if the spatial separation between acceptor and donor is smaller than the Förster distance. However, a short range quenching between the two dyes can occur when two dyes are less than 10\AA apart (this effect, similar to collisional quenching, is poorly documented). Consequently, we must impose a minimum separation between donor and acceptor to avoid collisional quenching, then choose an overlapping pair of dyes to set an appropriate Förster distance and achieve a proper coupling.

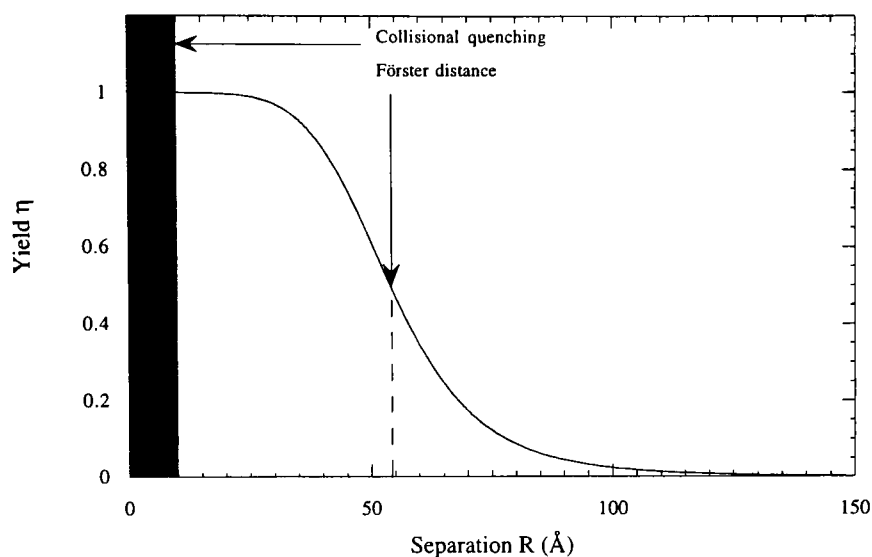


Figure 1.8: Theoretical yield of fluorescence energy transfer, as a function of the acceptor-donor separation R (here $R_0 = 54\text{\AA}$, for a fluorescein-rhodamine pair). At very short separations, donor and acceptor fluorophores interact and quench each other by collisional quenching: the FET yield collapses below $\sim 10\text{\AA}$.

Applications of FET to study DNA conformational fluctuations.

As soon as the electromagnetic coupling in FET was theoretically understood, biologists used the FET mechanism as a “spectroscopic” ruler to study conformational change in proteins [42]. Stryer and Haugland’s textbook experiment consisted in tagging oligopeptides of poly-L-proline with α -naphthyl (the donor) at the N terminal and with dansyl (the acceptor) at the C terminal. Different length of oligomers (from 10\AA to 45\AA , *i.e.* 1 to 12 residues) were synthesized and the efficiency of energy transfer measured: the agreement with the $1/R^6$ was very accurately checked (4 decades), and proved the usefulness of FRET as a “spectroscopic ruler”.

Application of FET to DNA molecules became feasible only at the beginning of the 90s’ once the chemistry of bioconjugation of the dyes had been worked out. DNA molecules lack the flexibility of protein chemistry *i.e.* thiols and primary amines are not naturally present in DNA (in proteins, they are naturally available as cysteines or histidines). Historically, the chemistry was worked out to develop dye-termination method of DNA sequencing. Today, the DNA chemistry is so accessible that nucleic acids have become the choice molecules for physico-chemist to study biomolecules’ conformational dynamics.

Let us review the results of FET studies on DNA conformational fluctuations.

In 1993, Morrison & Stols [43] modified specifically 5’ and 3’ ends of DNA oligonucleotides to further couple fluorescein or rhodamine on the ends of two complementary oligonucleotides. They used these modified molecules to measure the thermodynamics and kinetics of their hybridization (Figure 1.9A). They showed that the sensitivity of the FET detection of DNA hybridization (6 fold increase of FET upon hybridization) is beyond the signal/noise of any other spectroscopic detection scheme. Moreover, the fluorescence signal can be resolved at 1000-fold lower concentration of tagged DNA than the concentrations accessible by UV spectroscopy. Morrison & Stols also compared quantitatively the stability of DNA duplexes with and without grafted fluorophores (comparison of melting curves acquired by UV hyperchromicity or by FET): even though fluorophores have been suspected to intercalate and stack between dsDNA bases, they stabilize the double-stranded DNA only by -1.5kcal/mol in ΔG_{298K}^0 (to be compared to -40kcal/mol for the formation of the duplex

20-mer)

In 1994, Mergny *et al.* [44] used other dye-coupled DNA constructs to monitor the formation of secondary structures in single-stranded DNA (Figure 1.9B). They showed that the FET detection scheme creates a very local measurement of base-base proximity, by opposition to standard spectroscopic methods whereby a global measure of basepair formation is monitored.

In 1995, Parkhurst & Parkhurst [45] worked out the chemistry to tag specifically one 16-mer oligonucleotide with two different dyes (fluorescein and rhodamine). They then compared the FET efficiency of the two dyes when the oligonucleotide is coiled and when the oligonucleotide is “rigidified” by hybridization to its complement: in the coiled configuration, fluorescein and rhodamine are on average 51.5 ± 10.0 Å apart, the FET efficiency is 70%; in the hybridized configuration, the two dyes are 68.5 ± 6.4 Å apart, and the FET efficiency drops to 30% (Figure 1.9C): note that the decrease of the FET efficiency as well as the decrease of the standard deviation of the separation value for duplex DNA is compatible with the rigidity of that structure. Thus, the ability to tag specifically a single oligonucleotide with two FET-coupled dyes enables the monitoring of its conformational fluctuation: the design of this DNA probe is very close to that of the molecular beacon by Tyagi and Kramer [1].

Parkhurst *et al.* used this “proto”-molecular beacon to study the hybridization of the short DNA probe onto a large single-stranded plasmid DNA (M13mp18 phage). Kinetics of the hybridization of such a small DNA probe onto a large yields some information concerning the local secondary structure state of the M13 particle: the binding follows second-order kinetics and the hybridization rate is $5.7 \cdot 10^4 \text{M}^{-1} \text{s}^{-1}$ for M13mp18 (to be compared with $5.7 \cdot 10^5 \text{M}^{-1} \text{s}^{-1}$ for the complementary short oligonucleotide). The 10-fold decrease is investigated to prove that it corresponds to secondary structure static hindrance towards binding of the short oligonucleotide. Thus, the use of this FET oligonucleotide probe enabled Parkhurst *et al.* to probe the strength of local secondary structures inside a large ssDNA.

Other DNA structures have been studied by FET spectroscopy such as four-way DNA junction [46], triple-strand intermediates in RecA-catalyzed strand-exchange [47], B-Z DNA transition [48]. These studies provide detailed structural information on the distance separation between two DNA sites: in all cases, FET is used as a spectroscopic tool to measure distance separation as well as monitor the rearrangement of DNA structures.

Few studies have yet used FET to the study of DNA-protein interactions. Parkhurst *et al.* [49] used their doubly-labeled oligonucleotide to study the kinetics of binding and bending of the TATA-box to the TATA binding protein. A 16-mer double-stranded oligonucleotide is doubly labeled with a fluorescein and a rhodamine at the two ends. The analysis of the fluorescein lifetime shows that the FET efficiency increases upon TBP binding: the average separation between the two dyes decrease from 56.5 Å to 46.8 Å. This decrease is associated with the bending of the dsDNA by TBP. The analysis of the binding kinetics shows that the association follows second-order kinetics with no distinguishable events between binding and bending ($k_{\text{association}} = 2.4 \cdot 10^6 \text{M}^{-1} \cdot \text{s}^{-1}$ at 30°C). This rate is smaller than diffusion-limited rates ($10^8 \text{M}^{-1} \cdot \text{s}^{-1}$), which strongly suggests that the pre-bending of the TATA box is the limiting step in the association of TBP. However, the bulk measurements performed by Parkhurst *et al.* [49] is still inconclusive as bending fluctuations might occur at an unresolvable timescale.

Another elegant application of FET on nucleic acids is the validation of the Förster equation (1.11). Double-stranded B-DNA structures are well known (§ 1.1), and the crystallographic spacing between nucleotides has been documented with good resolution (typically, $1\text{--}2$ Å). The structure of B-DNA is also known to be very rigid with a persistence length of 150 basepairs in 0.1M Na^+ [11], such as one expects its structure in solution to be similar to the one found in crystals. Thus, one can use the B-DNA structure as a rigid ladder onto which donor and acceptor dyes are grafted at very precise distance separation. Different groups synthesized dsDNA oligonucleotides of varying length with two dyes at different separations [45, 46, 50, 51, 52]. The efficiency of the fluorescence energy transfer can then be monitored and compared to the Förster prediction in equation (1.11):

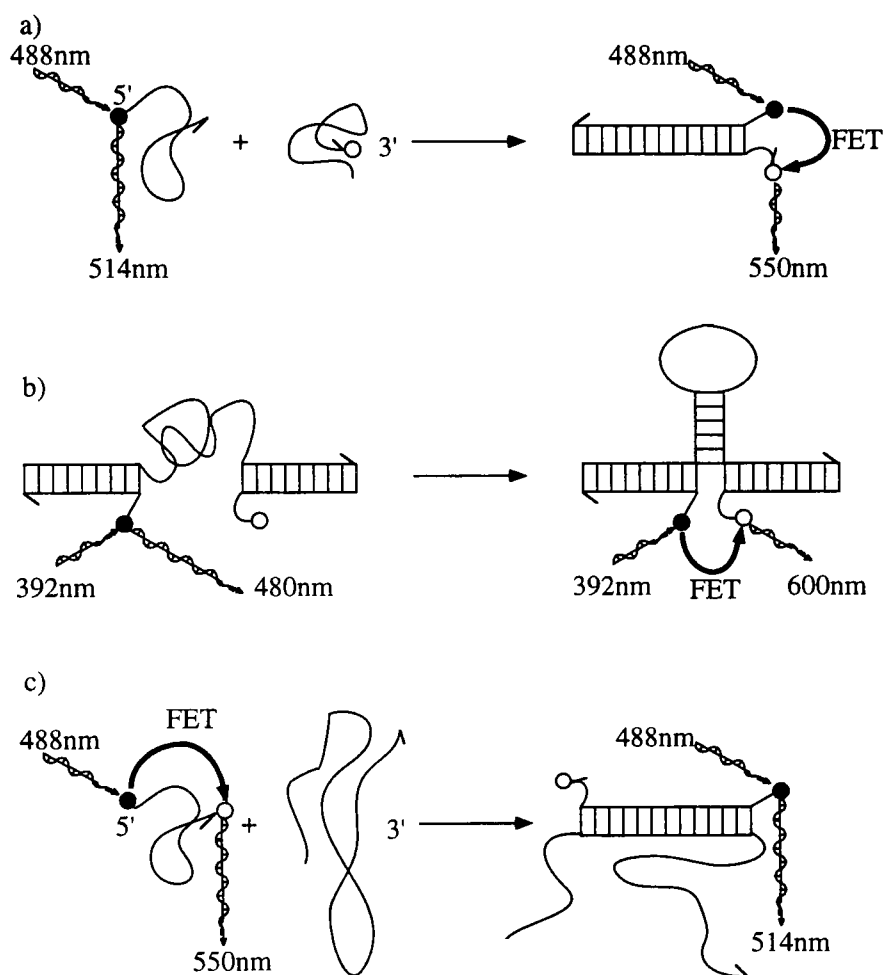


Figure 1.9: Review of previous studies on DNA conformations by Fluorescence Energy Transfer: A) Monitoring DNA hybridization of two complementary oligonucleotides (donor: fluorescein; acceptor: rhodamine) [43]; B) Monitoring the formation of an internal hairpin (donor: coumarin; acceptor: ethidium bromide) [44]; C) Monitoring DNA hybridization of a small oligonucleotide onto a large ssDNA (donor: fluorescein; acceptor: rhodamine) [45].

a fair agreement with the $1/R^6$ decay of FET has been recorded, but some systematic discrepancy has been noticed. Indeed, Förster's theory applies for two dyes in solution, the grafting of donor and acceptor fluorophores onto dsDNA brings aromatic bases close to the dyes and affects their spectroscopic surrounding [53]. Moreover, the angular separation between the donor emission dipole and the acceptor absorption dipole is hardly rigidly fixed, and introduced some uncertainty in the Förster distance evaluation. As far as the Förster's theory validation is concerned, the use of nucleic acid structures to spatially position a donor fluorophore and an acceptor fluorophore yields better-controlled results than the classical peptide approach introduced by Stryer and Haugland [42].

Collisional quenching.

Most relevant to the design of the molecular beacon is the collisional quenching of fluorophores by the DABCYL dye. Many chemical pathways can justify the quenching of a dye (excited-state reactions, charge coupling, energy transfer, molecular rearrangement), but the quenching of practically any fluorescent dye by DABCYL was a serendipitous discovery by S. Tyagi during the development of the molecular beacons [54]. Two experimental evidence for a charge-transfer quenching by DABCYL have been proposed:

- DABCYL's absorption peak centers at 450nm. Standard Förster's theory would predict that only EDANS and other UV-excited dyes would have the required spectral overlap to FET-couple to DABCYL. Tyagi *et al.* synthesized different molecular beacons with a DABCYL at the 3'end of the stem and different dyes at the 5'end. These dyes were chosen to span the whole visible range of excitation-emission (EDANS, Lucifer Yellow, Fluorescein, Rhodamine6G, Tetra Methyl Rhodamine, Texas Red). All these molecular beacons had very good signal-to-noise which constitutes a proof that all the dyes could be quenched efficiently by DABCYL at close range, and dequench when the molecular beacon opens. The quenching process could not be an energy transfer as the emission peak of Texas Red does not overlap at all with the absorption peak of DABCYL.
- the absorption spectrum of closed beacons (where the fluorescent dye and DABCYL are within 20Å) is very different from the one of open beacons. This proves the existence of a spectroscopic coupling between DABCYL and any dye, when in close proximity.

There is no direct evidence of collisional quenching between DABCYL and fluorescent dye. However, many experimental facts argue for it: DABCYL is very hydrophobic (it is hardly soluble in water, even when diluted first in dimethyl formamide), and it is a huge electron acceptor. Thus, one can rationale that DABCYL stacks on any aromatic dyes (such as fluorescein or rhodamine), and pump out and dissipate the energy of the fluorescent excited state.

For our studies, the quality of the DABCYL quenching of rhodamine6G was crucial to obtain good signal-to-noise and wide dynamic range of equilibrium constant measurement.

The discovery of universal quenching by DABCYL is one of the reason of the success of molecular beacon: these DNA probes became available for standard optical setup (confocal microscope or PCR light cycler). As far as our experiment is concerned, the digital nature of the collisional quenching simplified the analysis of conformational changes: the fluorophore is efficiently quenched at close range, and completely free to fluoresce when DABCYL and fluorophore are pulled apart, no intermediate is to be taken into account.

Fluorescence Energy Transfer (FET) studies on nucleic acids bring experimental validation of crystallographic structures, and offer a flexible tool to monitor locally DNA conformational states. The Förster's theory enables a

precise determination of the distance separation between two sites along the DNA, and has been used to measure structural characteristics of DNA structures in solution. Ingenious tagging with a collisional quencher enables the experimenter to create a digital signal (easier to analyze) related to specific DNA conformational switching: the molecular beacons and our construct to study DNA breathing constitute good examples of such use of fluorescence quenching.

1.3.3 DNA breathing modes.

The opening of basepairs of double stranded DNA (basepair breathing) is a prerequisite conformational change for many biologically relevant events: replication and transcription involve the stabilization and displacement of an open region (DNA bubble) in a double-stranded DNA. Breathing modes are also relevant to the understanding of the conformational fluctuations of dsDNA: single-stranded DNA is very flexible (so much that a freely-jointed chain model constitutes a good description), it can then release all the structural rigidity stored in dsDNA, such as bending or supercoiling.

DNA breathing is also relevant to many recognition events. DNA-binding proteins have been shown structurally to recognize specific sequences “indirectly”: the sequence specificity is imposed by the structure of the B-DNA where bases make hydrogen bond acceptors and donors accessible according to their identity (for example, a guanosine sticks out). The sequence recognition is indirect as the pattern of hydrogen bonds are not the one involved in the classical Watson-Crick basepairing. Chemists have been struggling with this specificity problem in their quest for specific markers of dsDNA. The triple-strand approach [55] has been dropped and the best solution so far consists of a pyridazole-imidazole polymer [55] which invades and expands the minor-groove of dsDNA. This experimental tour-de-force is quoted to show how sequence-recognition of dsDNA seems impossible without distortion of the B-DNA structure. Understanding the dynamics and sequence specificity of the dsDNA breathing modes will unravel stretches of DNA which are accessible to direct sequence reading (thanks to their single-strandedness).

Steady-state measurements of DNA breathing.

Steady states of open DNA bubbles in large genomic DNA have been visualized via specific digestion of ssDNA (by endonuclease R [56]), or specific cleavage of ssDNA (by orthophenantroline [57]) or specific chemical tagging of N-H groups in unpaired bases (by formaldehyde [58]), followed by electron microscopy or gel electrophoresis. These techniques draw the classical conclusion that AT-rich region of dsDNA are more likely to open up than GC-rich regions, as expected by comparison of their relative free-energies of basepairing (typically, an AT rich basepair is stabilized by $1.0k_B T$, whereas a GC basepair is $3.4k_B T$ more stable²).

Imino proton exchange measurement of DNA basepair breathing.

In the mid-eighties, the École Polytechnique group worked out the analysis of the imino-proton exchange technique to analyze the dynamics of DNA basepair opening. The important feature of this NMR application is that imino and amino proton exchange can readily be distinguished by titration of an exchange catalyst (solvent water or rather ammonia). In the following paragraph, we will review the details of this NMR technique and the results drawn from it.

Principle and theory of NMR measurement of proton exchange.

The dynamics of the basepair opening is measured by monitoring the imino proton exchange, catalyzed by proton acceptors (OH^- , H_2O , or chemical bases such as NH_3). It is crucial to notice

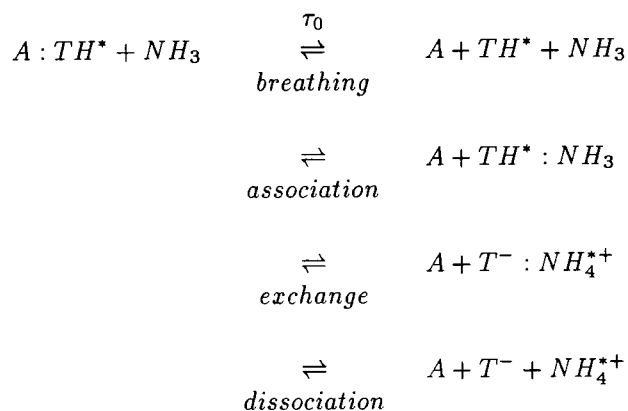
²Data extracted from nearest-neighbor model of DNA hybridization at 37°C , in 1M NaCl, 10mM Cacodylate, 1mM EDTA pH 7.0. (cf table on page 19).

that this catalyzed exchange can occur if and only if the basepair is open. Thus, two timescales are involved in the exchange lifetime: the opening timescales (intramolecular), and the collision timescale of the proton acceptor with the imino proton for an unpaired base.

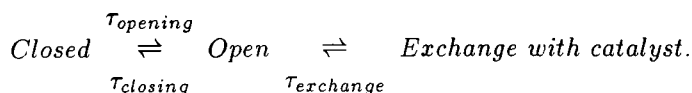
If the concentration of proton acceptor is large enough, then the probability of collision for an open basepair is one, and the only rate-limiting step in the imino-proton relaxation is the lifetime of the basepair. To the contrary, when the concentration of proton acceptor (NH_3 classically) is low, the exchange time τ_{ex} is determined by the association rate. The general formula for the exchange time is thus [59]:

$$\tau_{ex} = \tau_0 + \frac{1 + 10^{pK(\text{imino}) - pK(\text{NH}_3)}}{k_{\text{association}}} \frac{1}{[\text{NH}_3]} \left(1 + \frac{1}{K_{\text{dissociation}}}\right), \quad (1.13)$$

for the following equilibria:



(a full derivation of the different equilibria is presented in the review by Guéron & Leroy [60]). For our purpose, only three states of the imino proton are relevant to the breathing modes (with only three limiting steps):



At low temperature, for good salt conditions ($\approx 0.1\text{M}$ monovalent ions), one expects the concentration of open basepair to reach a stationary state (this corresponds to $\tau_{\text{opening}} \ll \tau_{\text{closing}}$, or $K = \tau_{\text{closing}}/\tau_{\text{opening}} \gg 1$). These kinetic equations can be solved to yield the timescale associated the relaxation mode of exchange of the imino proton.

$$(1.13) \Rightarrow \tau_{\text{relaxation}} = \tau_{\text{opening}} + \frac{\beta}{[\text{NH}_3]}, \text{ with } \beta = \frac{1}{k_{\text{exchange}}} \frac{\tau_{\text{opening}}}{\tau_{\text{closing}}} = \frac{\tau_{\text{exchange}}}{K}$$

thus, extrapolations of the exchange time to infinite concentrations of the proton acceptor NH_3 yields to the basepair opening timescale τ_{opening} .

We can estimate the slope of this extrapolation. For the imino proton of the thymidine base, the pK is 9.6 [60]. The pK for the ammonia is 9.3. If the association of the ammonia with the base is purely diffusion limited, and we can estimate the rate of association to $k_{\text{exchange}} = 10^9 \text{M}^{-1} \cdot \text{s}^{-1}$, the equilibrium constant of the opening-closing of the basepair can be estimated with another bulk NMR measurement (typically, $K = 10^{-5}$ for an AT basepair and $K = 10^{-6}$ for a GC basepair), then:

$$\beta = \frac{1 + 10^{9.6-9.3}}{10^9 10^{-5}} \sim 10^{-4} s.Mol^{-1}$$

The value of this slope explains the range of basepair opening timescale accessible by this titration of the exchange timescale. The highest concentration of ammonia one can achieve is 0.3M thus $\beta/[NH_3] \sim 0.3ms$, which should be smaller than the prefactor $\tau_{opening}$. In practice, imino proton has been used to measure opening timescale in the ms range. For example, the basepair lifetime of the AT basepair in the duplex $5' - CGCGATCGCG - 3' \bullet 5' - CGCGATCGCG - 3'$ is $2 \pm 0.7 \mu s$ at $25^\circ C$ (buffer: 0.1M NaCl pH8.8). By measuring the temperature dependence of this lifetime, the enthalpic barrier associated with the unzipping of the basepair has been estimated to be 23kcal/mol [61]: this huge barrier corresponds to the full disruption of the structure stabilizing the basepair (base-stacking and hydrogen bonds). This measure thus confirms that the measured lifetime is accurately attributed to the closed-basepair's lifetime.

The value of β explains also how site-specific the measure of the exchange rate is. The pK of other proton in other bases is much lower (typically 4 for the adenosines or cytosines), and the accessible timescale for reasonable concentrations of ammonia becomes much larger. The fact that the relaxation peaks can be accredited to specific bases is a plus of the technique.

Exchange times for the imino proton can thus range from days down to 1ms. For long timescales (above 2s), a simple bulk measurement of the real-time exchange is possible (the molecules are tritiated with protonated water, and suddenly mixed with an excess of deuterated water: the exchange spectrum is then measured as a steady-state). For shorter timescales, NMR measurements must be carried on and the titration with ammonia is the experimental method to be used. Let us insist that the timescale measured by this NMR technique is the lifetime of the basepair inside dsDNA. To the contrary, we developed the technique of fluorescence correlation spectroscopy to measure the lifetime of the "open basepair", and give us direct insights into the dynamics of accessibility of DNA bubbles inside dsDNA.

Study of dsDNA breathing modes by Raman Spectroscopy.

Recently, the resolution of Raman spectroscopy measurements improved drastically to such an extent that details on the premelting of large dsDNA could be monitored [62]. The principle of a Raman spectroscopy measurement is as follows: one excites a solution of biomolecules with a monochromatic light (typically with a wavelength of 514nm), and records the quasi-elastic scattering of the excitation photons at higher wavelength. The shifted peak of scattering (at a specific wavelength measured in wave number) corresponds to specific vibration modes in the biomolecules.

Raman spectroscopy of dsDNA fluctuations can discriminate between the vibration of the hydrogen bonds, the fluctuation of the phosphate backbone. For example, Raman bands at 728 (dA) 1236 (dT) and 1301 cm^{-1} (dA) are determined by basestacking in these nucleotides; Raman bands at 1182 (dT) and 1512 cm^{-1} (dA) measure the unpairing of the AT basepair; Raman bands at 792 and 842 cm^{-1} are characteristic of the O-P-O stretch bonds, leaving 819 cm^{-1} as a characteristic of the backbone vibrational modes; The amplitude of the Raman band at 1481 cm^{-1} yields the amount of hydrogen bonding on the atome N1 of the adenine bases, etc. [62, 63]. A systematic titration of each Raman peak as a function of temperature enables Movileanu *et al.* to measure a premelting of poly(dA-dT) duplexes between 10 and $66^\circ C$, before complete melting between 66 and $75^\circ C$. They show that the premelting transition is associated with a perturbation of the B-DNA structure without unpairing (increased motions are recorded on the phosphate backbone), whereas the melting implies a complete separation of the two complementary strands [63].

Raman spectroscopy is thus a powerful tool to analyze, chemical bond by chemical bond, the melting transition in double-stranded DNA. However, it is a steady-state measurement, yielding average values of the amplitude of the chemical bonds' vibrations, with no insight on the dynamics of the fluctuation and melting. Once again, the technique of fluorescence correlation spectroscopy

we developed to study nucleic acid conformational fluctuations will be a useful tool to analyze the premelting (breathing) modes of AT-rich domains in dsDNA.

DNA breathing modes and the initiation of transcription in prokaryotes.

In this section, we present the problem of the initiation of transcription in prokaryotes, and show how a systematic study on the dynamics of dsDNA breathing can challenge standard model on the action of the RNA polymerase (RNAP).

If RNAP was to start transcription at random, it would obviously be poorly efficient as most transcripts would be junk mRNA. Moreover, if RNAP was to miss the start site of transcription by one or two bases, then the whole transcript would be out of phase from the coding sequence, thus functionless. Consequently, the specificity of the initiation of transcription is a crucial problem in molecular biology: how does RNAP recognize accurately the start site of transcription ?

In the case of eukaryotes, RNAP can rely on transcription factors to dock precisely at the start site (the problem of specificity must then be addressed in terms of the specificity of the DNA-binding specificity of the transcription factor [64]). In the case of prokaryotes, RNAP without the help of ancillary proteins is enough to generate accurate transcripts. In 1975, in an experimental *tour de force*, Pribnow [65] was the first to discover a well-conserved binding site of RNAP at an early T7 promoter. First, it proved that the specific initiation of transcription is associated with a tight binary complex between the dsDNA and RNAP. Second, the binding sequence (subsequently called "Pribnow box") was found to be a very well conserved sequence,

$$5' - TATAAT - 3',$$

placed 10 basepairs upstream of the start site of transcription. The AT-richness of this box yielded Pribnow to hypothesize that the formation of the tight binary complex involves the local melting of the dsDNA strands.

Siebenlist [66] sequenced the 3 promoters of bacteriophage T7 to discover a conserved sequence 35 bases upstream of the transcript first nucleotide. This region (called -35 region) was shown to be involved with the initial recognition by RNAP, whereas the Pribnow box was proven to be the melting region. It was later shown that -35 region participates in promoter function but is not essential (it does not define the transcription start point [67]). Most early studies [68, 69, 70] involved differential cytosine methylation between ssDNA and dsDNA to map the single-stranded regions in duplex DNA: RNAP was shown to melt a region starting from the Pribnow box down to the initiation start site (12 base sequence). However, the mechanism by which the open region is unwound, propagated down the helix and stabilized by the RNAP was still unclear.

NMR study of the Pribnow box region unraveled sequence-specific premelting of the promoter regions [71, 72] at 55°C, in 0.1M phosphate buffer: the Pribnow sequence was suspected early to display a rate-limiting helix opening for the initiation of transcription. In fact, the NMR measurement of the imino proton exchange lifetime proves that the melting in an AT-rich domain is not cooperative at the core of the duplex [72] but becomes cooperative on larger scale (10 basepairs). The sequence dependence of the isomerization rate was found: 5' - TATAGT - 3' does not premelt, whereas 5' - TATAAT - 3' premelts relatively rapidly [71]. It was argued that the presence of TAA or TAT steps is crucial to the destabilization of the DNA. Thus the details of the AT-domain seems crucial to explain the premelting of a promoter region.

Further NMR experiments correlated the sequence-specific base-pair kinetics with dsDNA curvature [61]. Anomalous long lifetimes for the AT basepairs were recorded in DNA tracts of contiguous AT basepairs (at least 4 adenines, with maybe one 5' - AT - 3' step but no 5' - TA - 3' stay closed 50ms at 15°C pH8.8 (a 5' - TA - 3' step brings the lifetime of the basepair down to 1ms). It is then argued that four or more AT basepairs must adopt a structure deviating from classical B-DNA, such as B'-DNA (large propeller twist, good stacking, bifurcated additional hydrogen bonds and narrow minor groove).

These studies did not register any change in the proton chemical shift of the base protons of the non-terminal dG–dC basepair next to the Pribnow box. One must then conclude that isomerization of the Pribnow box is limited to its core of AT basepairs (6 bases): that is short of 7 bases to open a transcription bubble at the start site. The development of the orthophenanthroline cuprous complex to map the accessibility of DNA sites [73, 74] further established the correlation between Pribnow box breathing and formation of the complex RNAP–unwound DNA. Thus, local unwinding of the dsDNA helix seemed limited to the Pribnow box.

Today, one big debate about the initiation of transcription concerns the mechanism for the strand separation in the promoter region: does RNAP transfer a torsional stress onto the DNA upon binding, to force the unwinding of dsDNA, or does the RNA Polymerase simply stabilize the already melted region of the DNA promoter ? The second hypothesis belongs to the general paradigm of fluctuation–induced binding interaction in DNA–protein molecular recognition

Many tools have been used to map the formation of the open complex between RNAP and the promoter region: DNaseI digestion, methylation, orthophenanthroline cleavage, permanganate oxydation of thymine bases, NMR assay on the imino proton exchange. All these techniques measure bulk average formation of ssDNA domains in the promoter sequence. The precision of the mapping is also questionable as it relies on the accessibility of third–party cutters. In our work, we introduced new DNA constructs with internal fluorescent tagging to monitor the dynamic fluctuations of dsDNA, at the single–molecule level. This approach, although limited to the bare fluctuations of dsDNA without RNAP, unravels unknown large–scale breathing modes of dsDNA, of relevance to the problem of the initiation of transcription.

1.4 Fluorescence Correlation Spectroscopy.

To study the conformational fluctuation dynamics of our DNA constructs (molecular beacons and DNA breathing constructs), we used a technique combining fast–time resolution with rapid processing time: Fluorescence Correlation Spectroscopy (FCS). In this section, we review the historic lineage, the basic theory and the classical applications of FCS in analytical chemistry, biophysics and cell biology.

Fluorescence Correlation Spectroscopy is a statistical-physics based tool introduced to decipher molecular information. It was first introduced by Magde *et al.* [75] to measure the diffusion and binding kinetics of Ethidium Bromide onto double-stranded DNA. A flurry of experimental studies followed on the study of rotational diffusion [76], translational diffusion [77], chemical reaction [78] and others... However, the technique fell out of fashion in the 80s' as it remained cumbersome and inapplicable to *in vivo* studies.

FCS got revived at the beginning of the 90s' with the introduction of the confocal microscopy by Rigler & Eigen's groups [79, 80]. FCS is a good technique to monitor the relaxation of thermal fluctuations of fluorescent particles, and decipher molecular events such as motility or conformational change. As was pointed out early, one needs to reduce the number of monitored molecules to enlarge the amplitude of the thermal fluctuations. Thus, a confocal illumination was clearly the solution to enhance the sensitivity of FCS as it reduces the sampling volumes to small dimensions (typically $0.4\mu\text{m}\times 0.4\mu\text{m}\times 2\mu\text{m}\sim 0.3\text{fl}$), and brings down the detection limit to individual molecules. Today, the FCS technique has matured to be commercially available (the Zeiss & Evotec companies released

the first model of a confocal microscope for FCS applications in 1996 [81]). In our work, we used FCS as a tool to measure the dynamics of fluctuations in nucleic acid structure.

1.4.1 Historical lineage of FCS.

The understanding of fluctuation-relaxation in thermodynamics has been a great achievement of statistical physics. The theory of Brownian movement, presented by Einstein in one of his famous 1905 papers [82], not only established a macroscopic understanding of the consequence of the atom existence, but also opened up a whole new area of research related to the study of systems near-equilibrium. Einstein's most revolutionary idea in this paper is the deep connection between equilibrium statistics and fluctuation phenomena: this is the fluctuation-relaxation theorem.

Perrin's observation of Brownian particles of mastic under a microscope [83] brings a supremely-elegant understanding and experimental support for this atomistic theory. Other experiments are worth mentioning such as the Zeeman's observation of the Brownian torsional vibrations of a galvanometer, and Henri's measurement of the diffusion law followed by caoutchouc particles in solution [84]. It is humbling to point out that, right at the beginning of the XXth century, these experimentalists had already figured out the importance of reducing the sample size and using ultramicroscopy to unravel atomic wonders.

Fluorescence Correlation Spectroscopy is a technique which very much builds upon the intuition that thermal noise, usually a source of annoyance in an experimental measurement, can be used in our profit to glean some information on the system studied. The explicit formulation of the fluctuation-dissipation theorem by Onsager states that the dynamics governing the relaxation of a system out-of-equilibrium are embedded in the equilibrium statistics [85]. Experimentally, this theory has been applied by M. Eigen and followers whereupon thermal perturbations of a system are applied and the relaxation monitored to decipher the dynamics of the biomolecules (this technique is commonly called the T-jump). Less invasive techniques such as quasi-elastic light scattering (QELS) are also used to monitor fluctuations of refraction index and get dynamics information [86]. However, these techniques have the disadvantage to be perturbative (T-jump) or poorly sensitive (QELS) when molecular dynamics are to be monitored.

1.4.2 FCS for deciphering molecular events.

Fluorescence Correlation Spectroscopy corrects the shortcomings of the two techniques mentioned above, as it monitors the relaxation of fluctuations around the equilibrium state in a non-invasive fashion. It relies on the robust and specific signal provided by fluorescent particles to analyze the dynamics of these particles. We will present the principle of the FCS technique, and then briefly review the applications of FCS in analytical chemistry, biophysics and cell biology.

Fluorescence at a single molecule level.

On Figure 1.10, we present a time trace of the TTL signal recorded by a photodiode in our confocal setup. We will present the details of the measurement in the following chapter. We present this measurement to show the stochastic nature of the fluorescent signals emitted by any fluorescent particle. In this experiment, fluorophores are diffusing in and out of the sampling volume, and, once inside the sampling volume, are emitting stochastically. This example points out the possibility of gaining some submicroscopic information from single molecule spectroscopy, but it also emphasizes the importance of the statistical processing need to compress the stochastic noise of the measurement.

Statistical analysis of the stochastic fluorescence.

Two simple statistical analysis can be carried out from this single fluorophore measurement. First, one can plot the histogram of the lagtimes between TTL signals (*i.e.* detected photons): this his-

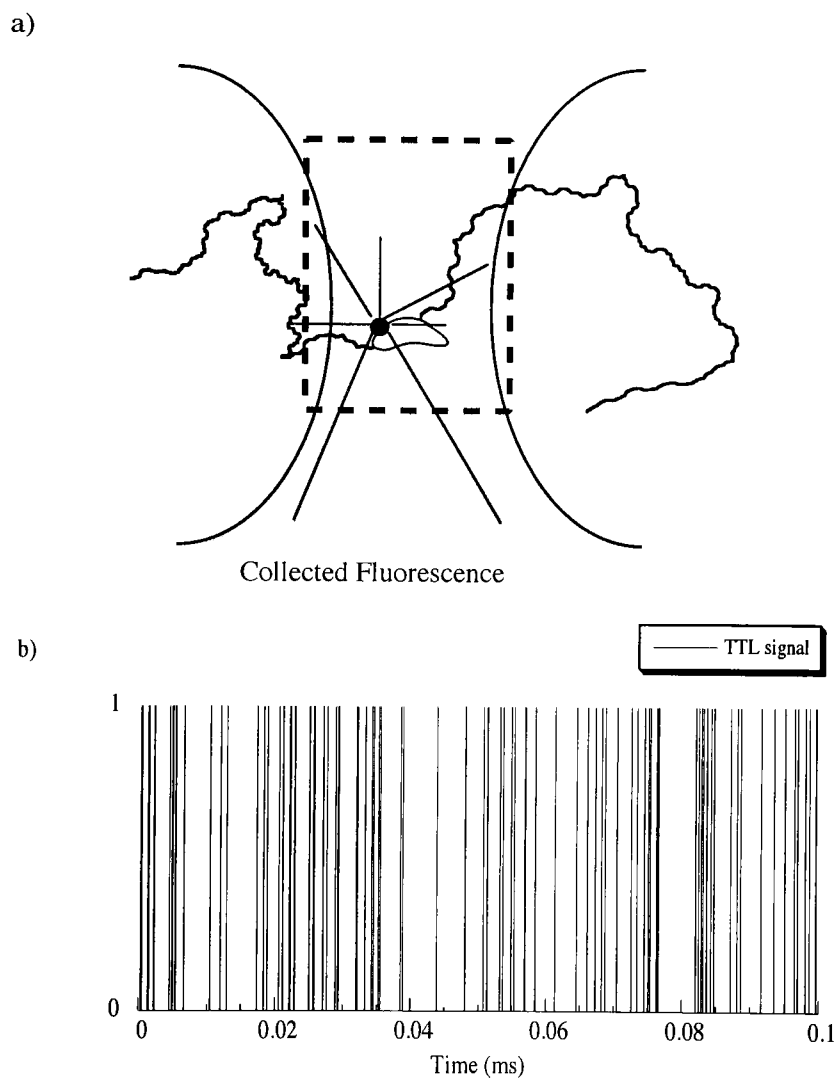


Figure 1.10: a) Sketch of the confocal geometry for single molecule spectroscopy. b) Time recording of the quantum bursts from a solution of dilute Rh6G. This figure shows the stochastic nature of any fluorescence signal when collected from small sampling volume (~ 1 fl), and from dilute solutions.

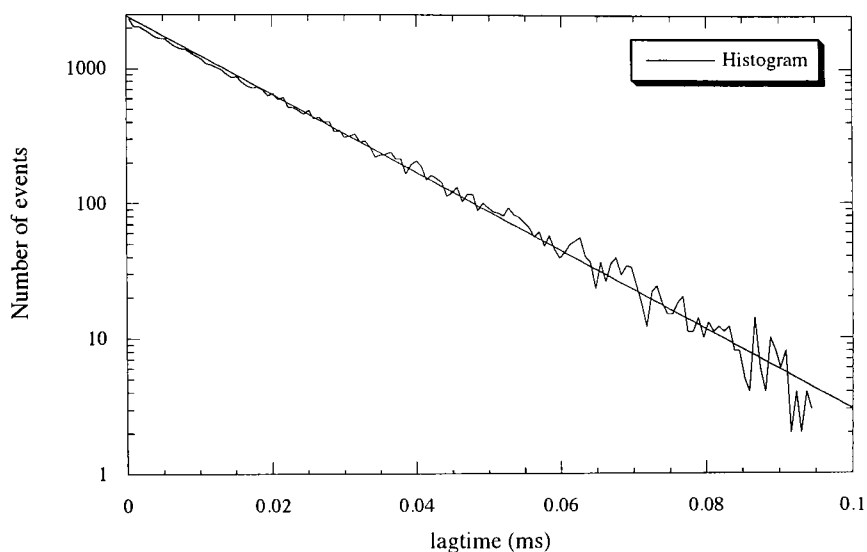


Figure 1.11: Histogram $P(\Delta t)$ of the TTL lagtimes Δt measured in the previous figure. This histogram is a monoexponential corresponding to the Poissonian distribution of the emission of photons by a single fluorescent particles: $P(\Delta t) = \frac{1}{f}e^{-f\Delta t}$, where f is the brightness of the dye (*i.e.* total number of collected photons per dye). The fit yields a brightness $f = 67kHz$ or 67000 photons are collected *per* individual fluorophores *per* second.

togram 1.11 can be fitted by a single exponential, as one would expect the distribution of lagtimes between emitted photons to follow a Poissonian distribution³. This fit can be used to measure the inner brightness of individual fluorescent particles. This analysis has been introduced by Qian *et al.* [87, 88], and further developed by Chen *et al.* [89, 90, 91] and Kask *et al.* [92] to monitor molecular aggregation of fluorescent particles. We will not dwell onto this kind of analysis, and present it more thoroughly in the Discussion Chapter.

The second statistical treatment of a time trace $I(t)$ as presented in Figure 1.10 is the computation of the autocorrelation function $G(\tau)$ of the intensity signal:

$$G(\tau) = \frac{\langle I(t)I(t+\tau) \rangle_t - \langle I(t) \rangle_t^2}{\langle I(t) \rangle_t^2} \quad (1.14)$$

$G(\tau)$ can be seen as the Fourier transform of the power spectrum of the collected fluorescence. On Figure 1.12, we illustrate the way the autocorrelation function is built, and the reason why it unravels dynamic information from the collected fluorescence. Moreover, this figure points out what kind of timescale one can extract from a stochastic signal with the autocorrelation function.

In the case of a fluorescent signal switching between a high and a low signal (think of a molecular beacon switching between quenched-closed conformation and fluorescent-open conformation), two timescales are relevant to the fluctuation dynamics: the lifetime of the fluorescent state τ_F , and the lifetime of the quenched state τ_Q . In the case of Figure 1.12, $\tau_F \ll \tau_Q$: the molecule is more often quenched than fluorescent. Note that the stochastic nature of the signal comes from the Poissonian distribution of switching time between the two states. When translating the fluorescence signal $I(t)$ by a lagtime τ shorter than τ_F , the initial signal $I(t)$ and the translated signal $I(t+\tau)$ overlap and their product is large. However, when one translates by a lagtime larger than τ_F , one completely decorrelates signal and translated signal, the product of the two signals is essentially 0. Thus, the

³The distribution of lagtimes will be a convolution of Poissonian distribution when the sample is less dilute and more than one particle are in the sampling volume

autocorrelation function is a function decaying with a typical timescale τ_F (see Figure 1.12). This observation will naturally show up in the analytical derivation of the autocorrelation function (see the Material&Methods Chapter), but it is important to understand now that the only relevant timescale, as far as the autocorrelation function is concerned, is the fast timescale of the fluctuation (in this particular case, τ_F). We presented this illustration to stress the usefulness of the autocorrelation function in extracting fluctuation timescale out of a stochastic signal, as well as to point out that this function unravels only the fast timescale in the fluctuation.

1.4.3 Application of FCS in analytical chemistry, biophysics, and cell biology.

Fluorescence Correlation Spectroscopy has been used in many experimental setting to extract molecular information, specially since its refinement in 1993. We will review some of the applications of FCS

FCS for particle counting.

The amplitude of the autocorrelation function is inversely proportional to the absolute number N of fluorescent particles in the field of view. Indeed,

$$G(0) = \frac{\langle \delta I(t)^2 \rangle_t}{\langle I(t) \rangle_t^2} = \frac{\langle \delta N^2 \rangle}{\langle N \rangle^2},$$

and for a Gaussian variable such as N , the variance is the square root of the average, $\langle \delta N^2 \rangle = N$, and:

$$G(0) = \frac{\langle \delta N^2 \rangle}{\langle N \rangle^2} = \frac{N}{N^2} = \frac{1}{N}.$$

pH environment and electron-acceptor quenchers (Oxygen, aromatic compounds) can affect the fluorescence of fluorescent particles *in vivo*. One can not rely on a fluorescent measurement to assess the concentration of fluorescent particles. FCS offers a measure of the absolute number of fluorescent particles inside the sampling volume, whatever the environment of the fluorescent particles.

Many groups have been applying FCS to the measure of an absolute concentration since 1972. An elegant example of application of FCS towards quantitating fluorescence *in vivo* was achieved by Cluzel *et al.* [93]. The experimental issue was the relation between the tumbling rate of the EColi chemotaxis motor, and the concentration of the signaling kinase CheY involved in the regulation of the chemotaxis. To titrate the response of the motor *versus* a change of concentration at the level of an individual cell, a fusion protein GFP-CheY-P has been cloned (this clone retains the kinase activity of CheY while being tagged fluorescently by GFP), and a latex bead has been coupled to the motor flagellum (Figure 1.13A). The absolute concentration of CheY-GFP was then measured by FCS (the sampling volume being calibrated before hand to be able to convert the absolute number of the FCS measurement into a concentration), while the tumbling rate of the motor was monitored by tracking the rotation of the bead. Ultimately, a titration curve of the bias motor *versus* the absolute concentration of CheY-P was built by inducing the expression of CheY at different concentrations.

This titration curve (Figure 1.13B) is an example of enzymology correlated with the phenotype, at the level of a single cell. FCS was crucial to this measurement to provide a non-invasive method to count the GFP-fused enzymes *in vivo*.

1.4.4 FCS to monitor translational mobility.

The shape of the autocorrelation function of an FCS experiment can be fitted to yield some information on the diffusion constant hence the size of the fluorescent particles. The experiment corresponds

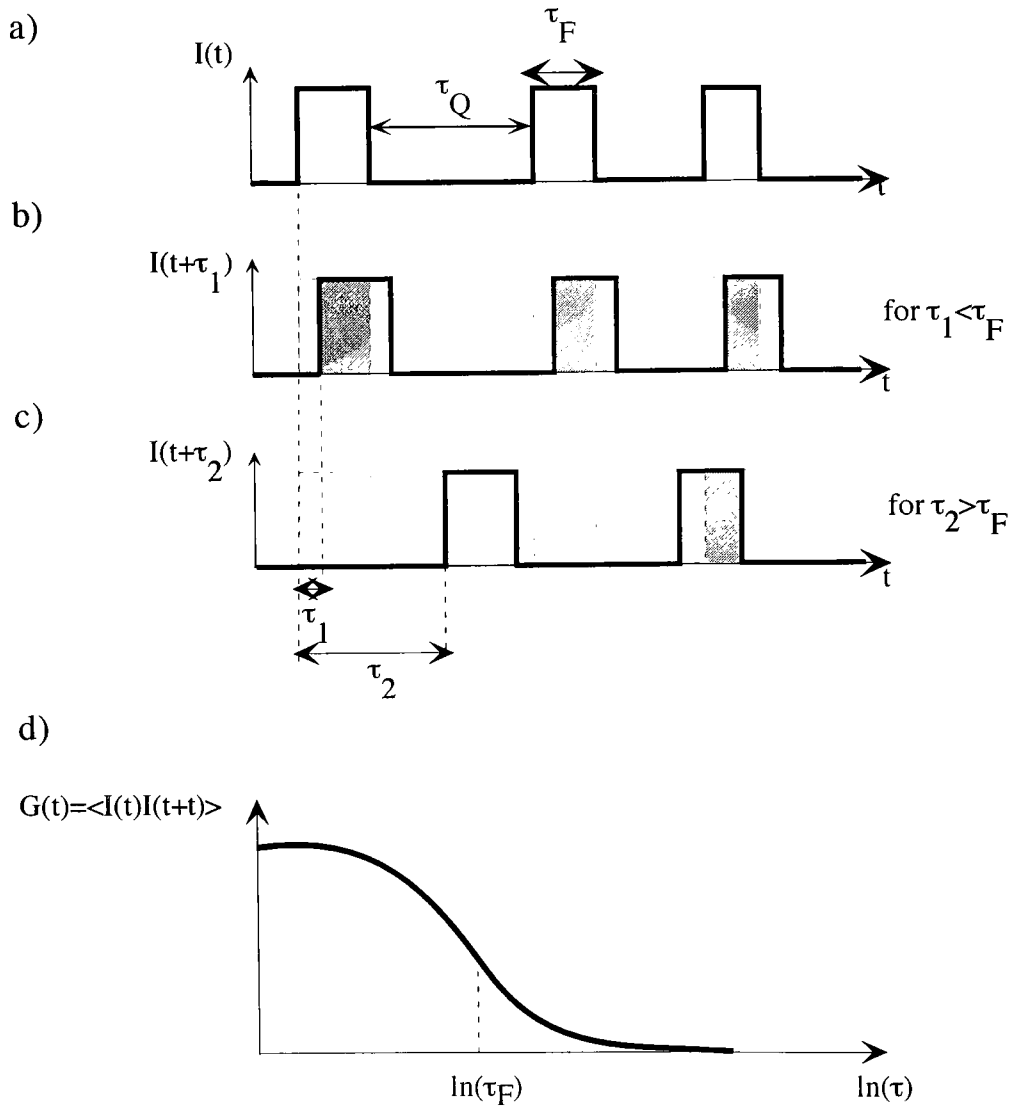


Figure 1.12: Graphical illustration of the building of the autocorrelation function $G(\tau)$ of a fluorescent signal $I(t)$. This illustration is done with an analog signal for clarity purpose, a) Initial signal $I(t)$: note the characteristic timescale τ_F of the events; b) Translated signal by a lag time τ_1 smaller than τ_F ; c) Translated signal by a lag time τ_2 larger than τ_F ; d) Autocorrelation function $G(\tau)$.

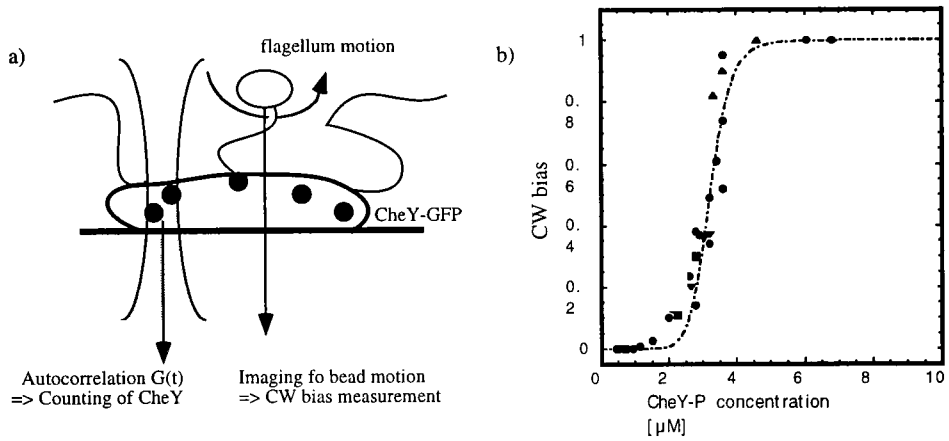


Figure 1.13: a) Sketch of Cluzel & Leibler's experiment [93]: a latex bead is antibody-coupled onto the flagellum of an E.Coli. Its rotation bias can be monitored by video microscopy. On the same cell, a FCS measurement is carried on to evaluate the absolute concentration of CheY (tagged with the Green Fluorescent Protein GFP). b) Resulting titration curve for the motor bias *versus* the concentration of CheY-P. This figure illustrates the power of FCS for counting fluorescent proteins *in vivo*

to the setup on Figure 1.10: fluorescent particles are diffusing in and out of a confocal volume. There are two typical dwell times for the fluorophore in this volume: $\tau_{\perp} = \frac{w^2}{4D}$ and $\tau_{\parallel} = \frac{Z^2}{4D}$, where w is the radius of the beam waist, Z is the longitudinal beam waist and D is the coefficient of diffusion of the fluorescent particle. To measure these timescales, one needs to fit the autocorrelation function

$$G(\tau) = \frac{1}{N} \frac{1}{1 + \frac{\tau}{\tau_{\perp}}} \sqrt{\frac{1}{1 + \frac{\tau}{\tau_{\parallel}}}}.$$

Thus, the temporal decay of the autocorrelation function can be used to measure the diffusion coefficients of fluorescent particles [94, 79]. It is a less invasive technique than the Fluorescence Recovery After Photobleaching (FRAP) technique [95], as it does not require the irreversible photobleaching of the fluorescent particles. It is also more convenient than classical quasi-elastic light scattering (QELS), since a fluorescence signal is more specific and more robust than a change in index of refraction.

Measuring a diffusion coefficient by FCS can be carried out in Analytical Chemistry, to be able to characterize specially-tagged molecules. For example, Kinjo and Rigler[96] have measured the hybridization of a short fluorescently-tagged oligonucleotides onto large single-stranded M13 phage DNA: the hybridization can be monitored by measuring the autocorrelation function of the fluorescence. The free oligonucleotides diffuse fast ($\tau_{free} \sim 180\mu s$), whereas the hybridized probes diffuse slower ($\tau_{bound} \sim 3.6ms$). In that case, the autocorrelation function can be fitted with two diffusing contributions:

$$G(\tau) = \frac{1}{N} \left(\frac{p}{1 + \frac{\tau}{\tau_{free}}} + \frac{1-p}{1 + \frac{\tau}{\tau_{bound}}} \right),$$

where p is the fraction of unhybridized probes. Kinjo and Rigler simply recorded the fitted variable p as a function of the concentration of added M13-phage DNA: this titration curve was then fitted to yield the binding affinity of the short oligonucleotides for the long ssDNA (typical). They also showed that they could monitor the kinetics of hybridization by acquiring the autocorrelation function at different time intervals, and fitting p as a function of time: the results were shown to follow classical Arrhenius second-order kinetics with an activation enthalpy of 38.8kcal/mol (this high barrier could be interpreted as a quantitative estimate of the strength of the secondary structure inside the M13-phage DNA, to be disrupted to allow the oligonucleotide hybridization). Kinjo & Rigler's main result was a proof of feasibility of such a FCS measurement.

Beyond analytical chemistry applications, FCS is becoming an exciting tool for cell biology to glean on "molecular events" *in vivo*. Perdeson's group was the first to carry out FCS measurement *in vivo* to assess the mobility of fluorescently-tagged oligonucleotides [95, 97]. Fast diffusion rates around ($4 \cdot 10^{-7} cm^2/s$) (comparable to the ones measured in solution) were measured inside the nucleus. However, for specific sequences of the oligonucleotides (dT_{21}), the authors measured a slower diffusion rate, compatible with an interaction with ribonucleotideprotein complexes and A-tract of mRNA ($< 10^{-7} cm^2/s$). These results show that the FCS measurement of oligonucleotides mobility in cells' nucleus can constitute a valuable tool to assess the presence of specific nucleic acid sequences *in vivo*. Jacobson *et al.* [95] also compared their FCS measurements with FRAP measurements, and the two techniques were found in good agreement.

Jovin's group in Göttingen has pushed FCS for further *in vivo* applications [98, 99, 100]. They pointed out that the combination of traditional confocal microscopy with FCS provides very quantitative measurements with very good spatial resolution of molecular mobility in a cellular context. Their experiment consists in monitoring Epithelial Growth Factor Receptors (EGFR), tagged with Green Fluorescent Protein (GFP). The receptor is expected to be present in three different cellular locations: endoplasmic reticulum (ER), cytoplasm, or plasma membrane, where it becomes fully functional. Classical confocal microscopy does not allow a definite distinction between these three

locations. However, with FCS, one can measure three characteristic timescales for the fluctuations of GFP fluorescence: the fast timescale ($\sim 100\mu\text{s}$) is characteristic of intramolecular rearrangement (see § 1.4.5), the medium timescale ($\sim 1\text{ms}$) is characteristic of freely diffusing proteins, and the slow timescale ($\sim 1\text{s}$) is characteristic of photobleaching. These three processes are very well time-separated in time, thus the fitting of the fluorescence correlation function with three timescales is very reliable. What Brock *et al.* noticed is that the weight of each fluctuation timescale in the autocorrelation function is very dependent of the localization of the GFP-EGFR: for cytoplasmic and endoplasmic reticulum GFP-EGFR, the biggest contribution is for the medium timescale (at $190\mu\text{s}$), whereas, for GFP-EGFR in the plasma membrane, the biggest contribution is for a slower timescale ($810\mu\text{s}$). This is consistent with the slower diffusion of GFP-EGFR when inserted in the plasma membrane (enhancing the photobleaching) compared to freely diffusing in the cytoplasm, or partially synthesized in the ER. Thus, FCS provides a fast, sensitive and reliable tool to measure the distribution of localization of EGFR: Brock *et al.* offers this analytical tool for *in vivo* drug screening, as well as for mobility measurement in cell biology.

In conclusion, FCS can be carried out in Analytical Chemistry and in Biophysics to monitor the translational mobility of fluorescent molecules and make diagnosis on their molecular arrangement.

1.4.5 FCS to monitor intramolecular conformational fluctuations.

This application of FCS in biophysics embodies our work on the conformational fluctuations of nucleic acid. Concomitantly, Two groups used FCS to analyze the photochemistry of the fluorescence of a Green Fluorescent Protein mutant (EGFP) [101, 102]. At stakes was a better understanding of the GFP fluorescence flickering, due to intramolecular rearrangement and protonation of the protein.

The FCS experiment consists in recording the autocorrelation function of the emitted fluorescence by EGFP, in different buffers. EGFP's chromophore contains a tyrosine residue whose degree of protonation influence the quantum yield of the fluorophore. Haupts *et al.* [101] noticed that three timescales were sufficient to fit the autocorrelation function: one diffusion timescale, one pH-independent slow timescale ($350\mu\text{s}$), and one pH-dependent fast timescale ($300\mu\text{s}$ at pH7.0, $45\mu\text{s}$ at pH5.0). The quantitative titration of these timescales with the buffer pH enabled Haupts *et al.* to fully analyze and assign chemical rates to the intramolecular rearrangement of the EGFP [101]. The slow timescale is characteristic of the protonation transfer inside the proton-network of the EGFP, whereas the fast timescale corresponds to the protonation exchange of the Tyrosine with the acid-base in the buffer. FCS revealed crucial in unraveling the three states of the EGFP chromophore.

The bonus of this study is that one can envision using EGFP *in vivo* to map the pH as well as the viscosity of intracellular compartments. Indeed, as shown earlier, FCS enables a quick measurement of EGFP molecules' mobility *in vivo*, to assess the local viscosity of the compartment where EGFP resides. The same FCS measurements could be used to assess the flickering fast kinetics (in the $100\mu\text{s}$ range), deduce the average protonation state of the EGFP, and thus deduce the pH of the surrounding buffer. This application of FCS is very appealing as it is non-invasive: EGFP can be endogeneously expressed and localized by fusion with targeting proteins, and the FCS measurement is carried out to assess *in situ* the intramolecular dynamics of the chromophore (diffusion and protonation). A pH and viscosity map of the cell could then be drawn in cell compartment by cell compartment.

Note one pitfall of the FCS technique. Although straightforward, the processing of the collected fluorescence into an autocorrelation function forces the experimenter to analyze the signal in a convoluted space. In other words, the autocorrelation function is more related to the power spectrum of the fluorescence signal (it is its Fourier transform), than to the signal itself. Without model, no mercy: the autocorrelation function will be simply a featureless curves. In fact, many FCS experiments are worthless because they rely on questionable models, or on unreliable fits (basically, fits with more than four free parameters are only published by dare-devils). In our experiments with nucleic acids, we had the chance to be able to construct control molecules to reduce the complexity of the fit on our FCS measurements. Special care was also taken to sieve out the spurious

sources of fluorescence fluctuation (quenched triplet state, partial quenching). We will describe our implementation of FCS for the study of DNA conformational fluctuation in the following chapter.

In conclusion, Fluorescence Correlation Spectroscopy is a technique which has found many applications since its revival at the beginning of the 90s'. The processing of a fluorescence signal into an autocorrelation function yields all the information on the relaxation of fluctuation in the collected fluorescence from a sample. Hence, one can glean information on the absolute number of fluorescent objects in a field of view, as well as the mobility, the photochemistry, or the conformation of the fluorescent particles. The difficulty of the technique consists in developing, for each experiment, an analytically-tractable kinetic model to fit the autocorrelation function. In the case of our DNA constructs, FCS will be used to analyze quantitatively the dynamics of their conformational fluctuations at thermal equilibrium.

יא

וַיְהִי כִלְהָאָרֶץ שָׁפָה אֶחָת וּדְבָרִים אֲחֻדִּים: וַיְהִי בְנֹסְעִים אֶמְקָרָם וַיִּמָּצְאוּ בַקָּעָה בָאָרֶץ שְׁנַעַר וַיֵּשְׁבוּ שָׁם: וַיֹּאמְרוּ אִישׁ אֶל־רֵעֵהוּ הִבָּה נִלְבְּנָה לְבָנִים וְנִשְׂרָפָה לְשִׂרְפָּה וְתָהִי לָהֶם הַלְבֵנָה לְאַבֵּן וְתַחֲמֹר הִיָּה לָהֶם לַחֲמֹר: וַיֹּאמְרוּ הִבָּה | נִבְנֶה־לָּנוּ עִיר וּמִגְדָּל וְרֹאשׁוֹ בַשָּׁמַיִם וְנַעֲשֶׂה־לָּנוּ שֵׁם פָּדִנְפוּץ עַל־פְּנֵי כָל־הָאָרֶץ: וַיֵּרֶד יְהוָה הַלֹּאֲת אֶת־הָעִיר וְאֶת־הַמִּגְדָּל אֲשֶׁר בָּנוּ בָּנוּ הָאָדָם: וַיֹּאמֶר יְהוָה הֵן עַם אֶחָד וְשָׁפָה אֶחָת לְכָל־ם וְזֶה הַחֲלָס לַעֲשׂוֹת וְעַתָּה לֹא־יִבְצֹר מֵהֶם כָּל אֲשֶׁר יִזְמוּ לַעֲשׂוֹת: הִבָּה נִרְדָּה וְנִבְלָה שֵׁם שְׁפַתָּם אֲשֶׁר לֹא יִשְׁמְעוּ אִישׁ שִׁפְתֵי רֵעֵהוּ: וַיִּפֹּץ יְהוָה אֹתָם מִשָּׁם עַל־פְּנֵי כָל־הָאָרֶץ וַיַּחֲדְלוּ לִבְנֹת הָעִיר: עַל־כֵּן קָרָא שְׁמֹהּ בָּבֶל כִּי־שָׁם בָּלַל יְהוָה שִׁפְתֵי כָל־הָאָרֶץ וּמִשָּׁם הִפִּיצָם יְהוָה עַל־פְּנֵי כָל־הָאָרֶץ:

The whole earth had a common language and a common vocabulary. When the people moved eastward, they found a plain in Shinar and settled there. Then they said to one another, Come, lets make bricks and bake them thoroughly. (They had brick instead of stone and tar instead of mortar.) Then they said, Come, lets build ourselves a city and a tower with its top in the heavens so that we may make a name for ourselves. Otherwise we will be scattered across the face of the entire earth.

But God came down to see the city and the tower that the people had started building. And God said, If as one people all sharing a common language they have begun to do this, then nothing they plan to do will be beyond them. Come, lets go down and confuse their language so they wont be able to understand each other.

So God scattered them from there across the face of the entire earth, and they stopped building the city. That is why its name was called Babelbecause there Cod confused the language of the entire world, and from there God scattered them across the face of the entire earth.

Bible (Genesis 11:1–9)

New English Translation (NET Bible)

Chapter 2

Material & Methods.

2.1 Synthesis and conjugation of the DNA samples.

All our experiments involve short strands of synthetic DNA oligonucleotides. Most of these materials were further derivatized to incorporate different dyes (fluorescein, Oregon Green, rhodamine, Texas Red, Cy5 or DABCYL). In this section, we present the different bioconjugation and purification protocols used to prepare molecular beacons and other dye-coupled oligonucleotides.

2.1.1 Initial materials.

The oligodeoxynucleotides were synthesized using standard cyanoethyl phosphoramidite chemistry. After removal of the protecting groups by hydrolysis (2 hours in concentrated ammonia at 75°C), the products were desalted by gel filtration chromatography on Sephadex G25 (Pharmacia), and dried in a speed-vacuum device. The oligonucleotides were used directly as such from the ammonium salt. Modifying groups (such as 5' primary amine, 3' sulfhydryl, internal amines on dT bases...), as well as dyes (internal fluorescein on dT base, terminal fluorescein on the 5' end, or terminal DABCYL on the 3' end...) were introduced directly during synthesis when possible. Most oligonucleotides were directly ordered from synthesis laboratories (Midland Certified Reagent Company, Texas, or IDT DNA, Iowa), or synthesized on an oligonucleotide synthesizer (Perkin Elmer, California).

2.1.2 Reversed-Phase High-Pressure Liquid Chromatography (HPLC): our quantitative analytical purification tool.

Reversed-Phase HPLC was a crucial tool we used to purify the products of our bioconjugation reactions. It enables an efficient separation and characterization of bioconjugation products, according to their hydrophobicity and absorption spectrum. In this section, we present the principle of Reversed-Phase HPLC and one typical chromatography we carried on our HPLC system (Gold system with double pump # 126 & detector #168 from Beckman, California).

The HPLC column is the Bio-Sil C18 HL 90-3S (Biorad, California): it is packed with alkane-functionalized porous silica beads (diameter: 3 μ m). The C18 surname stands for the length of the alkyl chains (18 carbons), that sets the range of hydrophobicity resolvable by this column.

The principle of Reversed-Phase HPLC is straightforward: molecules (in our case, ssDNA oligonucleotides and organic dyes) are loaded onto the C18-column in an aqueous running buffer. These molecules are highly hydrophobic: the DNA, because of its purine-pyrimidine bases, and the dyes because of their aromatic rings. Thus, in an aqueous buffer, the loaded molecules interact strongly and bind to the alkyl-packed matrix to screen their hydrophobicity. As one increases the organic content of the running buffer (hence the surname "Reversed-Phase"), the hydrophobic interactions

between molecules and alkyl-matrix become negligible, the molecules unbind from the column, are flushed by the running buffer and detach from the column. Thus, the retention time of a fraction is mostly determined by its hydrophobicity; its typical size also matters (the porosity of the packing material acts as a fractal sieve, inducing a size-exclusion effect: small molecules get trapped in the small pores, whereas larger molecules do not see the porosity and have a larger mobility). Note as well that the retention time of a fraction is not fully determined by the hydrophobicity and size of this sole fraction: a Reversed-Phase HPLC is more a competitive “race” between fractions (whereby the fractions modify the chemistry of the matrix during the run, and alter the retention time of other fractions), rather than an individualistic “race”.

A spectrophotometer at the column output analyzes the released fractions. We will present characteristic chromatograms for each type of bioconjugation we carried out, but, in this section, we will thoroughly present the outcome of a typical reversed-phase HPLC purification. For this experiment, a Rh6G dye has been coupled onto a 60-mer oligonucleotides with a 3' DABCYL, and one expects three possible fractions: {DNA with DABCYL}, Rh6G alone and {DNA with DABCYL and Rh6G} (details of the bioconjugation reaction will be presented in the section §2.1.3).

The purification method goes as follows: the product materials of the bioconjugation reaction are loaded onto the C18-column in 0.1M TriethylamineAcetate pH6.5 (buffer A). The HPLC system is set to run at a flow rate of 1ml/min under a pressure of typically 100bar. The mobile phase is a binary phase of Acetonitrile (methyl cyanide) in buffer A: the ratio is a gradient from 0 to 50% of acetonitrile in 25min. After 25 min, the column is flushed with 75% Acetonitrile/25% Buffer A for 10min, and, then, reequilibrated with 100% Buffer A for the next purification.

The chromatogram for this HPLC purification is presented in Figure 2.1. It is a 3D-recording of the absorption spectrum of the released fraction (from 200nm to 600nm) as a function of time: the x-axis unit is time (in min), the y-axis measures the absorption wavelength (in nm), and the z-axis is an absorption unit (A.U.) proportional to the optical density of the fraction being analyzed. For the sake of readability, we print a contour-plot of the chromatogram, whereby the AU is color coded in a 2D-plot of time and absorption wavelength. The color-coding is intuitive: blue between 0 and 0.1AU, cyan between 0.1AU and 0.2AU, green between 0.2AU and 0.3AU, yellow between 0.3AU and 0.4AU, and red above 0.4AU.

The analysis of the chromatogram (2.1) shows that four different fractions were separated during the purification. At $t=17.5\text{min}$, fraction (a) displays an absorption peak at 260nm (characteristic of DNA bases) and 490nm (characteristic of DABCYL): it consists of DNA oligonucleotides with DABCYL (these are the DNAs' which failed the bioconjugation with Rh6G). At $t=19.5\text{min}$, fraction (b) shows two absorption peaks at 260nm, and 525nm (characteristic of Rh6G): this fraction corresponds to the DABCYL-ated oligonucleotides which reacted successfully with Rh6G, and should thus be collected. At $t=21.0\text{min}$, fraction (c) displays one absorption peak at 525nm: this is bare-Rh6G which failed to react. At $t=26.7\text{min}$, fraction (d) shows the same spectrum as (c): it is unreacted Rh6G released by the HPLC column during the flushing phase.

The shapes of these absorption spectra are crucial to the accurate identification of each fraction. For example, the dye alone does possess some absorption peak in the UV range, and could be confused with the 260nm peak characteristic of DNA bases. However, the amplitude of the UV peaks for Rh6G is incompatible with a 1:1 ratio of DNA:dye. The ratio of absorption, measured as the ratio of the extinction coefficients ϵ , can be predicted to be 2.0, for a 25mer coupled to Rh6G:

$$\frac{\epsilon_{260\text{nm}}(\text{DNA})}{\epsilon_{525\text{nm}}(\text{Rh6G})} = \frac{300 * 25}{108.10^3 * 35.10^{-3}} \sim 2.0,$$

whereas it is only 0.5 for the ratio $\epsilon_{260\text{nm}}/\epsilon_{525\text{nm}}$ for the dye alone.

Upon closer examination of fraction (b) (Figure 2.1), we notice that the shoulder at 480nm on the absorption peak, attributed to the dye Rh6G, is in fact characteristic of DABCYL –the spectra of the dye alone can be seen on the fraction (d),. This fraction is thus the complete molecule to be

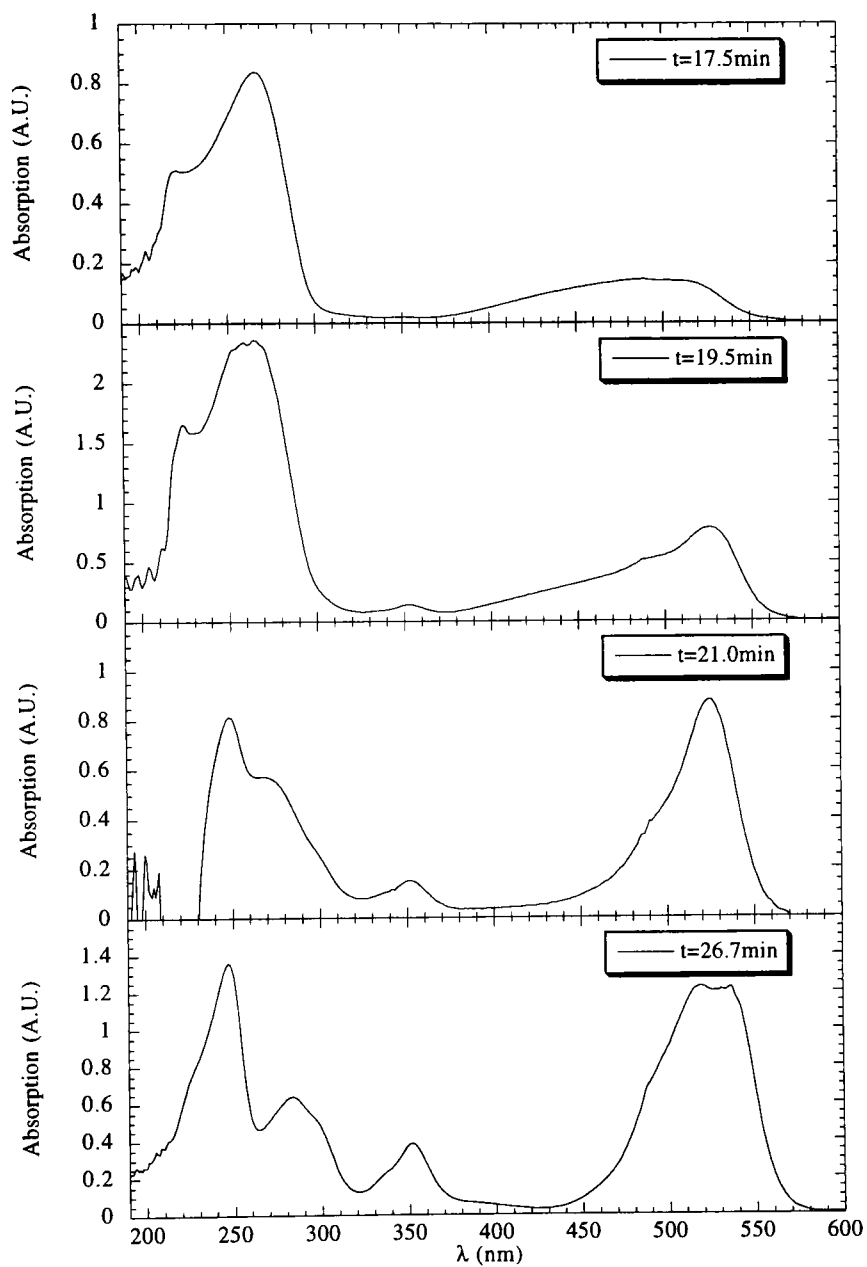


Figure 2.1: Chromatogram for the bioconjugation of Rh6G onto a DABCYLated 60-mer oligonucleotide.

collected: it is a full molecular beacon, with DABCYL and Rh6G at its two extremities (see § 2.1.5 for more details on the bioconjugation reactions).

The understanding of this HPLC chromatogram will be useful in the following paragraphs where we will be presenting the different bioconjugation reactions and subsequent purifications we carry out to prepare our DNA materials. We will then present a table of all the materials used in our experiments.

2.1.3 Coupling of a succinidimyl ester onto a primary amine.

Three bioconjugation reactions have been used to couple different dyes onto modified oligodeoxynucleotides (Figure 2.2). The most efficient bioconjugation consists of the reaction of the succinimidyl ester (SE) of the dye onto a primary amine inserted into the DNA. In our experiments, primary amines have been introduced during the oligo synthesis at the 5' end with a spacer consisting of 3 or 6 carbons. Primary amines could also be introduced via a modified thymidine base whereby the methyl group has been replaced by the primary amine desired: this results in the insertion of an internal dye grafted directly to the base. The bioconjugation is a one step reaction (Figure 2.2A), followed by a size-exclusion purification on SephadexG25 (to get rid of the bulk of the unreacted dyes), and a reversed-phase HPLC. The collected fractions are then ethanol-precipitated, dried and resuspended in the buffers required for future additional modifications or for storage.

The complete protocol is as follows:

1. Resuspend the newly-synthesized oligodeoxynucleotides in 0.1M Na₂CO₃, pH8.3, at a final concentration of 200μMol (*e.g.* 40nmol of material in 200μl)
2. Resuspend the dye in the succinidimyl ester form in anhydrous dimethylformamide at final concentration of 10mg/ml (here 1mg in 100μl)
3. Mix 200μl of the aqueous DNA solution with 100μl of the organic dye solution. Incubate with gentle vortexing for 2 hours at room temperature or overnight at 4°C.
4. Spin down the solution at 14000rpm for 2min to pellet down the excess of unreacted undissolved dyes, and collect the supernatant.
5. Pass the sample onto a Sephadex G25 column (Pharmacia,) with exchange to buffer A (0.1M triethyl amine acetate, pH6.5).
6. Run a reversed-phase HPLC onto a C18 column (Biorad), at 1ml/min with a gradient 0-50% of acetonitrile in buffer A (see § 2.1.2 for details). Collect the desired fractions 500μl by 500μl.
7. Ethanol-precipitate the fractions *i.e.* add 50μl of 3M Sodium Acetate solution (pH5.4), and 1ml of ice-cold 200-proof ethanol, store at -20°C for 1hr and spin down at 14000rpm at 4°C for 30min. Vacuum-aspirate and discard of the supernatant, and dry the DNA pellet with a Speed-Vacuum device.

The yield of the whole protocol is typically 80%, including 90 to 95% for the bioconjugation [103], and 90% for the purification-recovery.

- the 2:1 ratio of aqueous buffer to organic buffer in the coupling reaction must be preserved to achieve a good solubilization of the dye (mostly hydrophobic, hence the use of water-soluble anhydrous solvent), while maintaining the pH level (crucial for the reactivity of the primary amine).

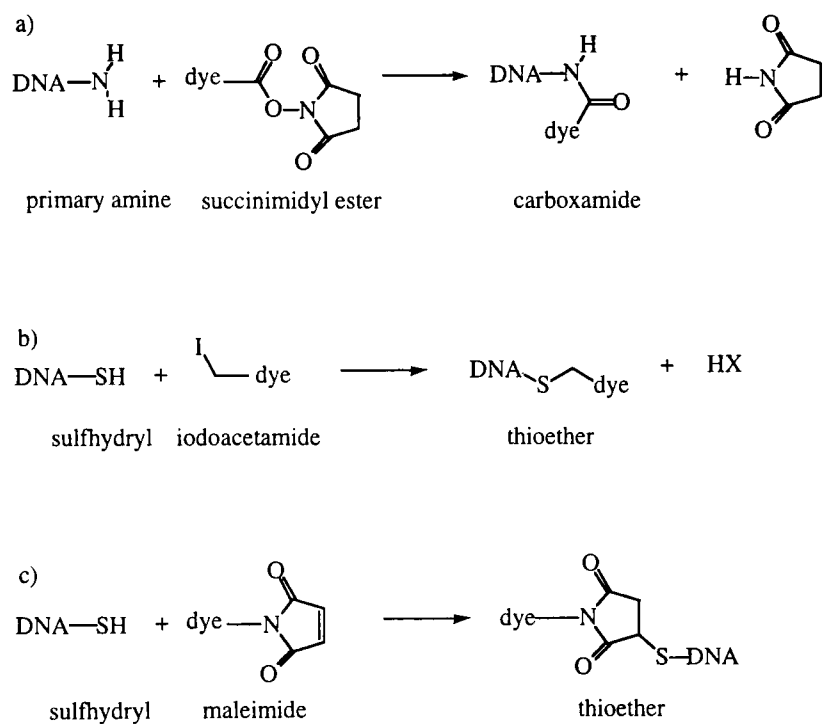


Figure 2.2: The three bioconjugation reactions used to couple dyes onto modified oligonucleotides: a) Succinimidyl ester onto primary amine. b) Iodoacetamide onto sulfhydryl. c) Maleimide onto sulfhydryl.

- Special care must be taken when coupling the quencher DABCYL onto primary amines. DABCYL is extremely hydrophobic and tends to precipitate in the protocol described above. The coupling yield is then decreased because the dye phase-segregates from the oligodeoxynucleotides. The solution of the succinyl ester of DABCYL must thus be added slowly at a pace of 10 μ l every 10min (*i.e.* in a total time of 1hour 20min), and the solution should be further incubated for 2 hours to achieve a decent yield (>75%).
- A typical spectrogram of an HPLC purification, after succinyl ester coupling onto the DNA, is presented in figure 2.3. The order of the incoming fractions follow their hydrophobicity: the bare oligodeoxynucleotides come first, followed by the dye-coupled DNA. The unreacted dyes do not come until hard washing of the HPLC column with 75% of acetonitrile is carried on. For short oligodeoxynucleotides, like in the chromatogram presented here, the different peaks are well resolved: a single-step purification is enough. For longer oligodeoxynucleotides (above 40 bases), the desired fraction comes at the same time as the uncoupled dye. The ethanol-purification is enough to get rid of the remaining dye (soluble in ethanol), but a second HPLC was usually carried on to assure the purity of the coupled material.

2.1.4 Coupling of an iodoacetamide or maleimide onto a sulfhydryl.

Sulfhydryls (commonly named thiols in the literature) can be inserted in synthetic oligodeoxynucleotides. This allows an alternative route for specific conjugation of dyes, used for the double-labeling of the synthetic oligodeoxynucleotides, or the labeling at different locations. Sulfhydryls can be introduced at the 3' or the 5' end of the oligonucleotides (available spacers are C3 and C7 [104]), or on the phosphate backbone [105, 106]. The second derivatization of the oligonucleotides does not involve new building blocks in the cyanoethylphosphoramidite methods, but simply milder deprotecting conditions (this modification is today widely used in antisense experiments, as these oligodeoxynucleotides are resistant against nuclease digestion). Note that the conversion of the backbone phosphate into a thiophosphate modifies the electric charge of the DNA backbone, and thus alters the structure and stability of the double-stranded DNA.

Sulfhydryls also offer the possibility of being introduced in a protected chemical form. For 5'end sulfhydryls, the trityl moiety routinely used in the cyanoethyl phosphoramidite method is the protecting group. For 3'end sulfhydryls, the protection is a disulfide bond. The thiols can be specifically deprotected with silver ions (when trityl is the protecting group) and reduced by DTT treatment, to become reactive. This is crucial for *ab initio* synthesis of the beacon, and for any other application where specific double labeling is required (see § 2.1.5 for further description).

The complete protocol for coupling of an iodoacetamide onto an unprotected sulfhydryl is as follows (see [107] for more details):

1. Resuspend the synthetic oligodeoxynucleotides in 0.1M Na₂CO₃ pH9.0 (prepared fresh) at a concentration of 100 μ Mol.
2. Resuspend the dye in the iodoacetamide form in anhydrous dimethylformamide at final concentration of 10mg/ml (here 1mg in 100 μ l *i.e.* 20mMol)
3. Mix 200 μ l of the DNA solution with 100 μ l of the organic dye solution. Shield the reaction tube from light with aluminum foil. Incubate with gentle vortexing for 2 hours at 50°C.
4. Spin down the solution at 14000rpm for 2min to pellet down the excess of unreacted undissolved dyes, and collect the supernatant.
5. Pass the sample onto a Sephadex G25 column (Pharmacia,) with reequilibration in buffer A (0.1M triethyl amine acetate, pH6.5).

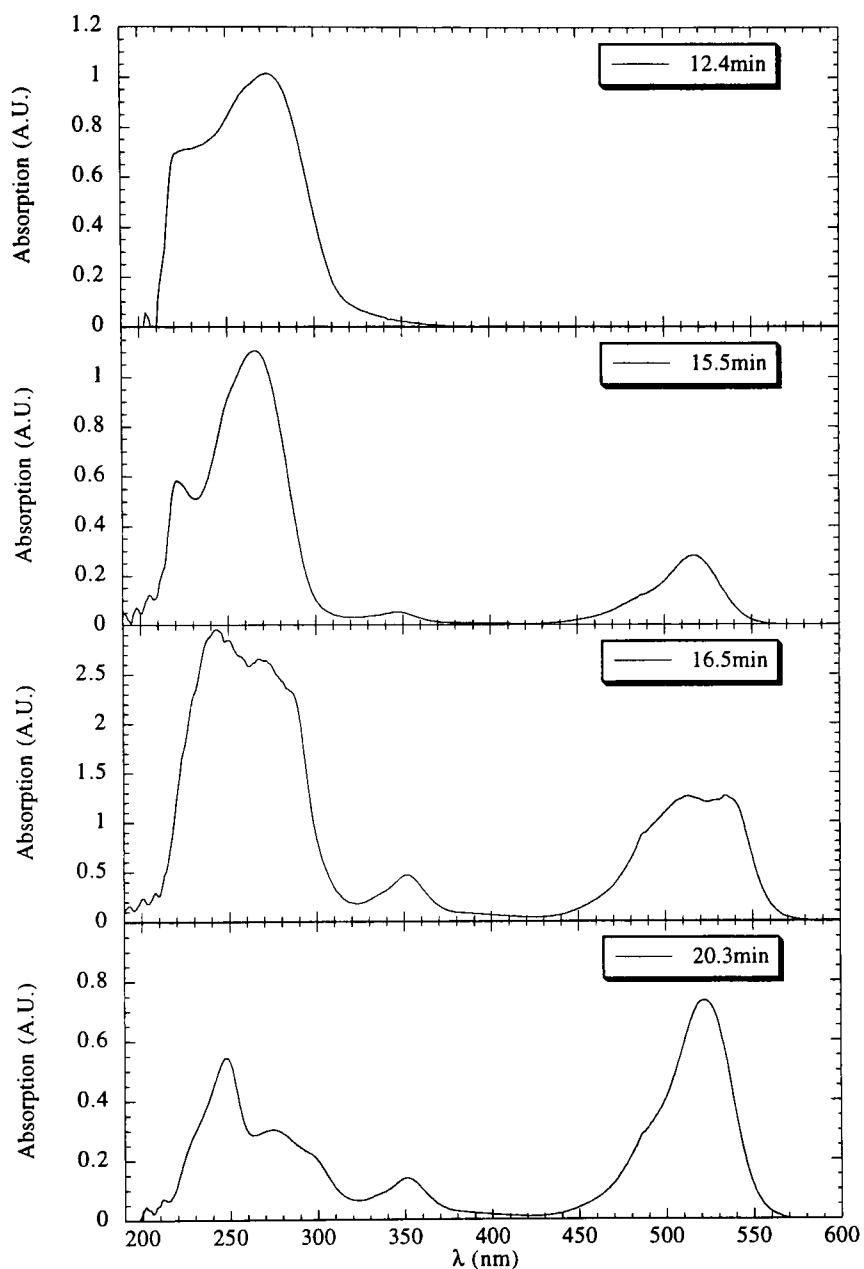


Figure 2.3: Chromatogram of the fractions after coupling a succinidimyl ester of 5-carboxyrhodamine 6G-SE onto a 5' amine of a 30-mer oligonucleotide. Three Fractions are separated: a) at $t=12.4\text{min}$, the uncoupled DNA; b) at 15.5min the DNA coupled to Rh6G; c) at 16.5min the DNA coupled to Rh6G (to be collected); d) at 20.3min , the unreacted dye. The fraction coming at $t = 15.5\text{min}$ corresponds to oligomers coupled to an alternative conformation of Rh6G.

6. Run a reversed-phase HPLC onto a C18 column (Biorad), at 1ml/min with a gradient 0-50% of acetonitrile in buffer A. Collect the desired fractions 500 μ l by 500 μ l.
7. Ethanol-precipitate the fractions *i.e.* add 50 μ l of 3M Sodium Acetate solution (pH5.4), and 1ml of ice-cold 200-proof ethanol. Store at -20°C for 1hr and spin down at 14000rpm at 4°C for 30min. Aspire and discard the supernatant, and dry the DNA pellet with a Speed-Vacuum device.

The chromatogram of a reversed-phase HPLC purification of the fractions resulting of this bio-conjugation reaction is similar to the one presented in § 2.1.2. The yield of the whole protocol is typically 90%, including practically 100% for the bioconjugation [107], and 90% for the purification-recovery. Note that a DTT reduction of the sulfhydryl is not needed for synthetic oligonucleotides.

2.1.5 Complete synthesis of a molecular beacon.

Molecular beacons have been synthesized from modified synthetic oligonucleotides with a primary amine at the 3' end and a protected (tritylated) sulfhydryl at the 5' end, according to Tyagi & Kramer's protocol [1]. This protocol merges the two coupling reactions presented in the two previous sections: A) first a succinimidyl ester of the quencher DABCYL is reacted onto the 3' primary amine; a first HPLC purification allows the selection of the material containing the DABCYL and the trityl group; B) this material's thiol is then deprotected and C) reduced into a sulfhydryl; D) an iodoacetamide of the fluorophore is then reacted onto the sulfhydryl, and a second HPLC purification allows the purification of fully-reacted material. Here is the full protocol of the complete synthesis of a molecular beacon.

1. Initial material :

Modified oligonucleotides with a sulfhydryl covalently linked to the 5' phosphate via a C6 spacer, and a primary amino group covalently linked to the 3' hydroxyl end via a C7 spacer.

2. Coupling of the quencher to the 3'-end.

- Suspend the oligos in 0.1M Na₂CO₃ pH9.0, to a final concentration of 0.6mM (this protocol is made for 300nmol of oligos). Use a light-tight glass vial.
- Prepare 500ml of a solution of DABCYL at 60mg/ml in high-quality anhydrous dimethyl formamide (DMF) and add the quencher solution to the oligos.
- Incubate for 3 hrs at room temperature.

3. Purification of the oligos efficiently coupled to the quencher.

- Prepare a Sephadex G-25 column (NAP10 for a 1ml-solution): reequilibrate the column with buffer A (Triethyl ammonium acetate -TEAA- 0.1M, pH6.5, filtered at 0.22 μ m) by adding three times the volume of the column.
- Add the reaction mix (1ml), let it equilibrate in the column.
- Add 1.5ml of buffer A, and collect the fraction.
- Run an HPLC purification on a C18 reversed phase column, with a linear gradient elution of 0 to 75% acetonitrile in buffer A in 30min at 1ml/min. Five fractions are detected (Figure 2.4): a) at t=16.5min, unreacted and unprotected DNA (fast arrival and no absorption peak in the visible range); b) at t=18.5min, DABCYL-coupled but unprotected (fast arrival and peak at 450nm); c) at t=20.2min, DABCYL-coupled and protected DNA; d) at t=22.2min, DABCYL-coupled and protected DNA: this fraction comes the latest (trityl is very hydrophobic) and displays the characteristic absorption peak of DABCYL at 460nm. The fifth fraction in this chromatogram comes at t=26min and is unreacted DABCYL.

- Ethanol precipitation of the material : add 1:10 Sodium Acetate (3M pH5.2), and twice the volume of the fraction of 100%-proof ethanol. Keep at -20°C in the freezer for 1 hour. Centrifuge at 4°C at 14000RPM for 30min. Discard the ethanol supernatant, resuspend with cooled (4°C) ethanol (70%). Give a quick spin at 14000RPM. Discard the supernatant again and dry the pellet under vacuum.
4. Deprotection and coupling of the fluorophore to the 5'-end.
 - Resuspend the pellet in 500ml of buffer A.
 - Add 10ml of freshly prepared solution of AgNO₃ (0.15M in buffer A). Incubate shaking for 30min.
 - Add 10ml of freshly prepared solution of DTT (0.15M in buffer A). Incubate shaking for 15min.
 - Centrifuge at room temperature at 14000RPM for 3min to remove the silver particles. KEEP the supernatant.
 - Prepare a fresh solution of iodoacetamide of fluorescein (I3-Molecular Probes) in 500ml of Na₂CO₃ 0.2M-pH9.0. Add to the collected supernatant and incubate shaking at room temperature for 1hr30min.
 5. Purification of the oligos efficiently coupled to the fluorophore.
 - Prepare a Sephadex column as described in step 3 and repeat the purification as described in that paragraph. In this HPLC purification (Figure 2.5), three main fractions are separated: a) at t=14.5 & 15.7min, the unreacted dye; b) at t=17.5min, the complete beacon (DNA coupled with DABCYL and OregonGreen)*i.e.* the fraction to be collected ; c) at t=20min, the DABCYL-DNA which failed to couple to dye. Note that the presence of OregonGreen reduces the hydrophobicity of the DABCYL-ated DNA. In this bioconjugation, the yield of IA-OregonGreen coupling onto the DNA is excellent (practically 100%).
 - After HPLC purification, ethanol precipitate the collected fraction (cf section A). Resuspend in 0.1M Trizma-1mM EDTA (TE) pH8.0. Check the signal/noise of the molecular beacon by measuring the fluorescence of a solution of 1nMol of beacon alone, *vs.* a solution of 1nMol of beacon with an excess of target (50nMol) (buffer: 1M NaCl, 10mM NaCacodylate, 1mM EDTA pH8.0). Freeze in a dry environment for long-term storage.

The final product (a full molecular beacon) is sketched in Figure 2.6. Note that the quencher DABCYL (Molecular Weight MW=366Da) as well as the fluorophore OregonGreen (MW=556Da) are of comparable size of a nucleotide (MW(Adenosine)=313Da, MW(Cytosine)=289Da, MW(Guanosine)=329Da, MW(Thymidine)=304Da).

2.1.6 Synthesis of DNA molecules to study DNA breathing.

To study the fluctuations of basepair association inside a double-strand of DNA, we needed to graft internal dyes onto DNA oligonucleotides. Modified thymidines bases, whereby a primary amine is substituted to the methyl group, are the only specific tagging sites (other modifications, such as phosphate derivatization, are very disrupting to a B-DNA structure). This primary amine can be conjugated with a succinimidyl ester of the dye (cf § 2.1.3). The resulting modified bases (for Rh6G and DABCYL) are presented in Figure 2.7). A standard construct would contain a AT^*/T^*A step (where T^* is the modified Thymidine base): DABCYL is coupled to one T^* , Rh6G is coupled to the other one, such as the two dyes are sticking out, only 10Å apart, in the B-DNA major groove.

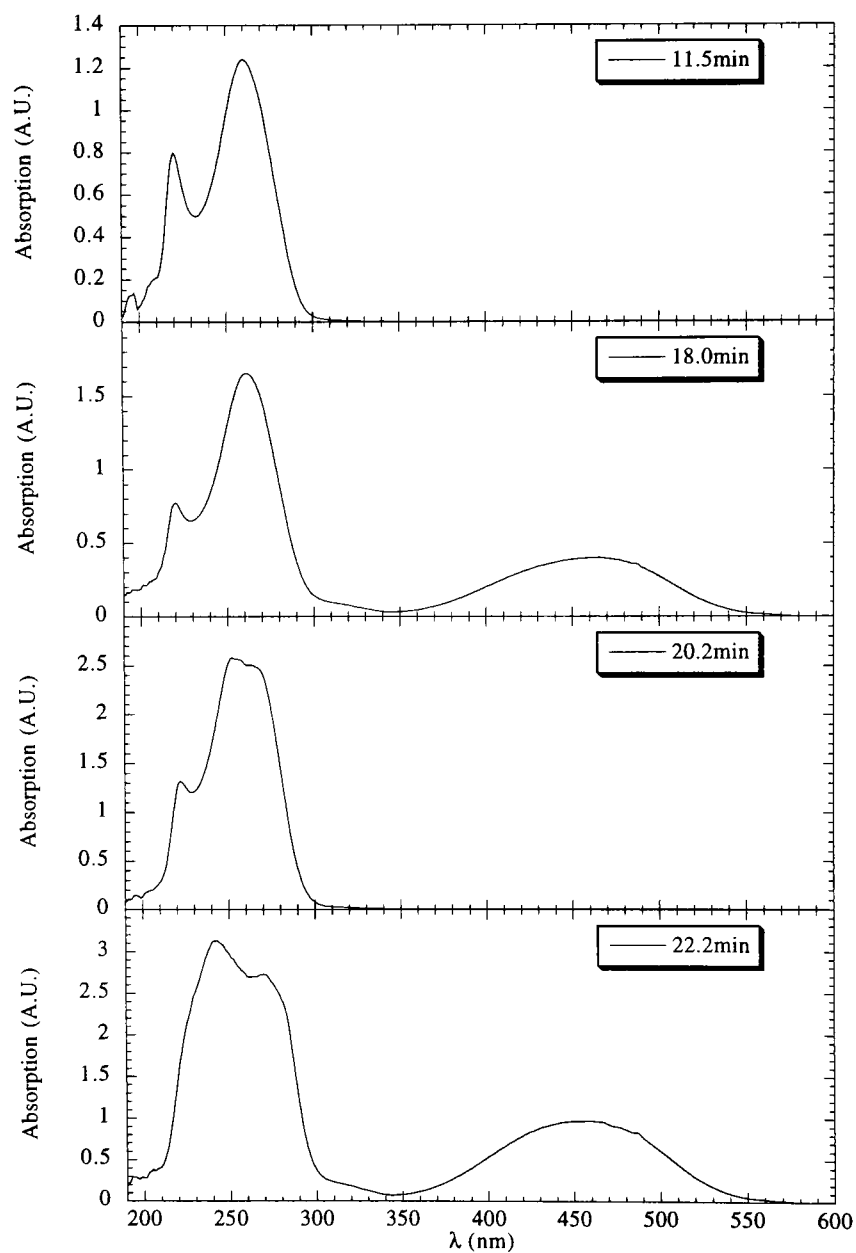


Figure 2.4: Chromatogram of the first HPLC purification (after DABCYL coupling onto the 5'-tritylated-DNA).

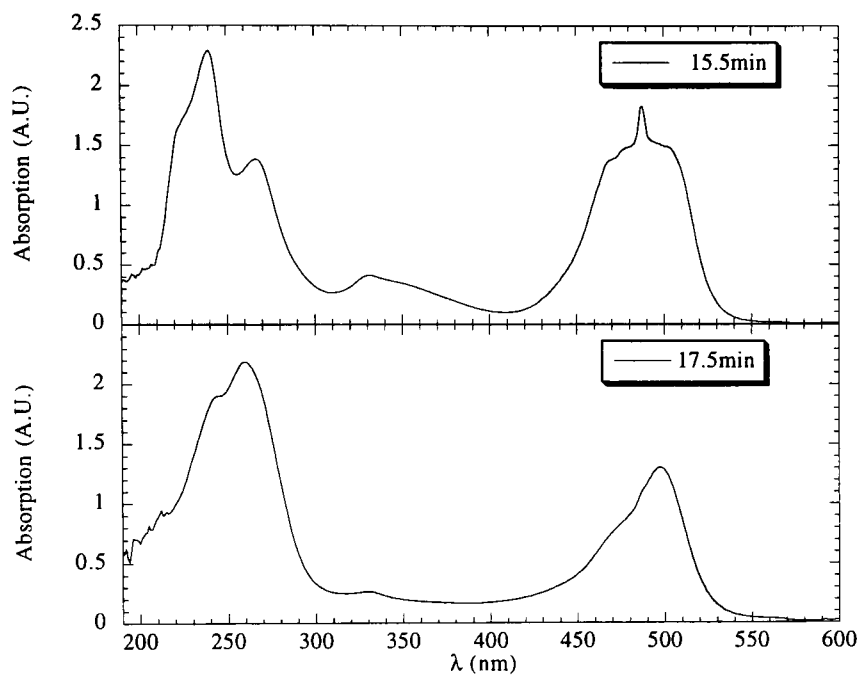


Figure 2.5: Chromatogram of the second HPLC purification (after OregonGreen coupling onto the Dabcyl-ated-DNA).

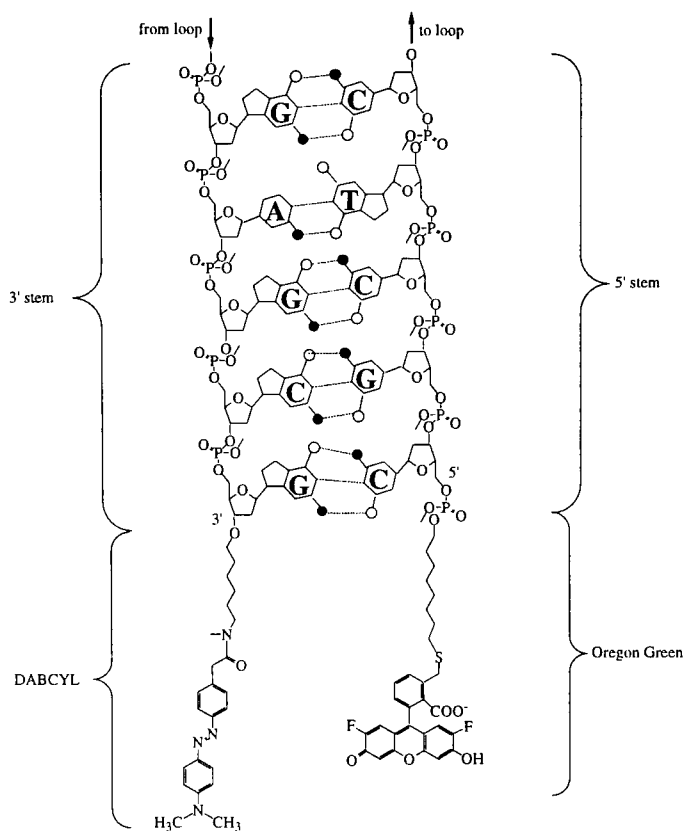


Figure 2.6: Molecular Sketch of the stem a molecular beacon.

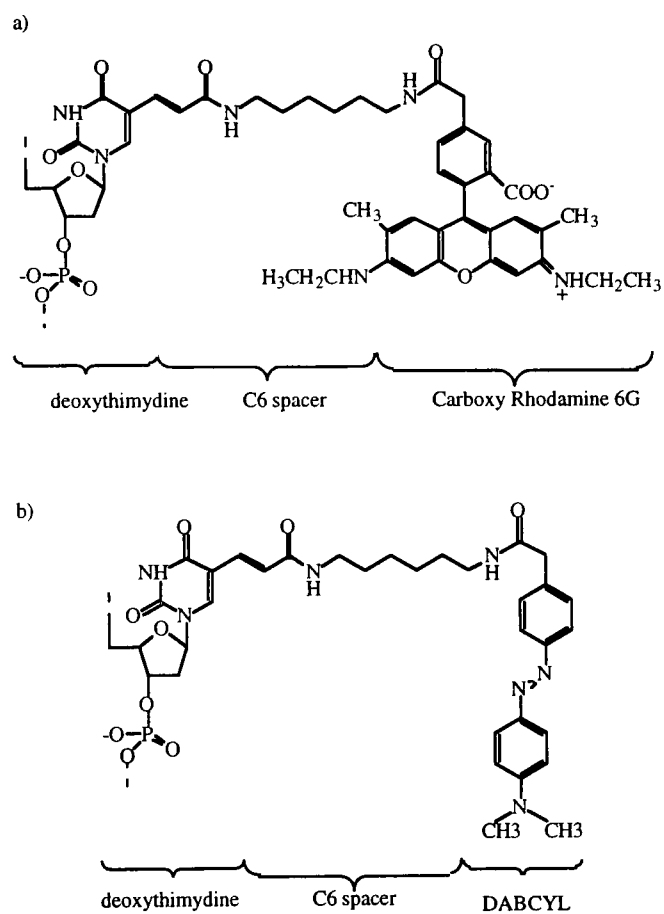


Figure 2.7: Internal tagging on modified thymidine bases: a) Rh6G with C6 spacer; b) DABCYL with C6 spacer

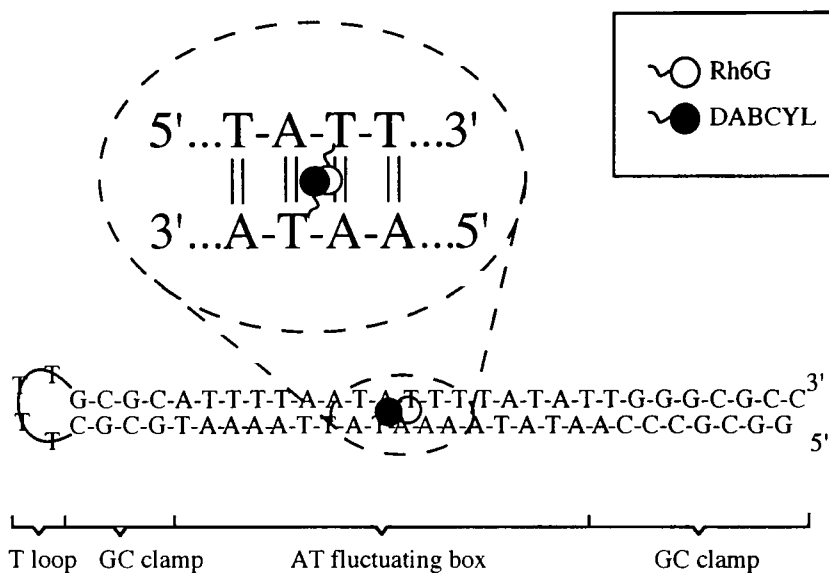


Figure 2.8: Construct for DNA breathing studies: the oligonucleotide is made of two complementary 30-base stands, linked together by a highly-flexible T_4 loop; at low temperature, in saline buffer, this construct folds into a hairpin-loop whereby Rh6G is quenched by nearby-DABCYL.

Rh6G and DABCYL (our fluorophore and quencher of choice for our FCS studies) are commercially available only as succinimidyl esters. We are lacking the diversity of bioconjugation chemistry to carry out the same protocol as for the molecular beacon synthesis (§ 2.1.5). Our solution consisted in conjugating two complementary strands of DNA, with either Rh6G or DABCYL, and then annealing and ligating them together. The purification of the ligation products was carried on by denaturing gel electrophoresis. This protocol is not very efficient and strongly damages the dye-DNA link: poor signal-to-noise (around 4) were achieved in term of fluorescence quenching-dequenching of Rh6G by DABCYL.

Fortunately, Glen Research Corporation [104] released internal modified thymidine phosphoramidites with DABCYL grafted onto the C5 carbon of the base (the same modification with Rh6G is too delicate for the harsh deprotection reactions of the oligonucleotide synthesis, and will never be released). The final synthesis of a double-strand of DNA with Rh6G and DABCYL in *vis-à-vis* is then a one-step reaction: we order the DNA with a $AT(NH_2)/T(DABCYL)A$ step, coupled Rh6G onto the primary amine (see § 2.1.3). The HPLC of this synthesis has been presented in § 2.1.2. The final construct is sketched in Figure 2.8.

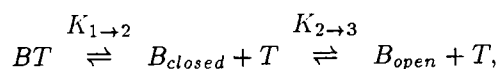
2.2 Thermodynamic measurement on nucleic acid structures.

2.2.1 Thermal denaturation profiles.

The fluorescence of solutions of a molecular beacon (*e.g.* fluorescein-5' – CGCTC – CCA₁₁CC – GAGCG – 3'-DABCYL *cf* Figure 1 on page 2), in the presence or absence of synthetic target oligonucleotides, was measured as a function of temperature. All measurements were made in 100 μ l solutions containing 50 nM molecular beacon, 100 mM KCl, 1mM MgCl₂, 10mM Tris-HCl (pH 8.0), and no target, 300 nM perfectly complementary target (5' – GGT₁₁GG – 3'), or 300 nM mismatched target (5' – GGT₅GT₅GG – 3'). The fluorescence of these reaction mixtures was measured by excitation with a 488-nm laser light source in a spectrofluorometric thermal cycler (Applied Biosystems Prism 7700)¹. The temperature was increased in steps of 1°C, from 15°C to 80°C, with each step lasting 5 min. To confirm that nonequilibrium hysteresis did not occur, the temperature was then decreased in 1°C steps from 80°C to 15°C, with each step lasting 5 min. Fluorescence was measured during the last 30 sec of each step. The intrinsic fluorescence of fluorescein varies with temperature. We therefore measured the fluorescence of a DABCYL-less analogue of the molecular beacon as a function of temperature, and these data were used to correct the thermal denaturation profiles so that the adjusted fluorescence intensities were independent of the intrinsic variation of fluorescein fluorescence. Furthermore, to compare different thermal denaturation profiles, the fluorescence intensities of each profile were divided by the fluorescence intensity measured at 80°C. Consequently, all profiles were normalized to a value of 1.00 at 80°C.

2.2.2 Equilibrium analysis for the denaturation profiles of the molecular beacons.

Molecular beacons in solution with their targets can exist in three states: bound to target (phase 1), free of target in the form of a hairpin (phase 2), and free of target in the form of a random coil (phase 3). At equilibrium:



where BT is the probetarget duplex, B_{closed} is the molecular beacon in the form of a hairpin, B_{open} is the molecular beacon in the form of a random coil, and T is the free target. The fluorescence of the solution at a given temperature, F , is the sum of the fluorescence of the molecular beacons in each of the three states:

$$F = \alpha \frac{[BT]}{B_0} + \beta \frac{[B_{closed}]}{B_0} + \gamma \frac{[B_{open}]}{B_0}, \quad (2.1)$$

where α , β , and γ are the characteristic fluorescence intensities the molecular beacons in each state, and

$$B_0 = [BT] + [B_{closed}] + [B_{open}].$$

The equilibrium constant for the dissociation of the hairpin stem is described by the equation

$$K_{2\rightarrow 3}^0 = \frac{[B_{open}]}{[B_{closed}]},$$

and the equilibrium constant for the dissociation of the probetarget duplex is described by the equation

¹ An alternative custom-made thermal cycler has also been used to work with 3ml-solution –see Appendix C for complete description.

$$K_{1 \rightarrow 2}^0 = \frac{[B_{closed}][T]}{[BT]}, \quad (2.2)$$

the total concentration of targets, T_0 , is much greater than the total concentration of molecular beacons, B_0 , as is the case in our experiments, T_0 can be substituted for $[T]$ in equation [2.2], and the fraction of molecular beacons in each state as a function of temperature can be expressed in terms of the two equilibrium constants as follows:

$$\begin{aligned} \frac{[BT]}{B_0} &= \frac{T_0}{T_0 + K_{1 \rightarrow 2}^0 + K_{1 \rightarrow 2}^0 K_{2 \rightarrow 3}^0} \\ \frac{[B_{closed}]}{B_0} &= \frac{K_{1 \rightarrow 2}^0}{T_0 + K_{1 \rightarrow 2}^0 + K_{1 \rightarrow 2}^0 K_{2 \rightarrow 3}^0} \\ \frac{[B_{open}]}{B_0} &= \frac{K_{1 \rightarrow 2}^0 K_{2 \rightarrow 3}^0}{T_0 + K_{1 \rightarrow 2}^0 + K_{1 \rightarrow 2}^0 K_{2 \rightarrow 3}^0}. \end{aligned}$$

Substituting these three expressions into Equation 2.1, the fluorescence of a solution of molecular beacons in equilibrium with its target is fully described by the equation

$$F = \frac{\alpha T_0 + \beta K_{1 \rightarrow 2}^0 + \gamma K_{1 \rightarrow 2}^0 K_{2 \rightarrow 3}^0}{T_0 + K_{1 \rightarrow 2}^0 + K_{1 \rightarrow 2}^0 K_{2 \rightarrow 3}^0}. \quad (2.3)$$

2.2.3 Thermodynamic measurements of the hairpin to random-coil transition.

The thermodynamic parameters that describe the transition from a stem-and-loop structure (phase 2) to a random coil (phase 3), enthalpy ($\Delta H_{2 \rightarrow 3}$) and entropy ($\Delta S_{2 \rightarrow 3}$), were determined by analyzing the fluorescence data obtained from the thermal denaturation profile of a 50 nM solution of molecular beacons incubated in the absence of targets. The equilibrium constant for this transition can be expressed in terms of the fluorescence measured in the experiments by taking $T_0 = 0$ in equation [2.3] and rearranging:

$$K_{2 \rightarrow 3} = \frac{F - \beta}{\gamma - F}, \quad (2.4)$$

where F is the fluorescence intensity at a given temperature, β is the characteristic fluorescence intensity of the molecular beacons in the form of a hairpin (obtained from the fluorescence measured at 15°C), and γ is the characteristic fluorescence intensity of the molecular beacons in the form of a random coil (obtained from the fluorescence measured at 80°C). Since $\Delta G^0 = -R\theta \ln K(\theta) = \Delta H^0 - \theta \Delta S^0$, where ΔG^0 is the free energy, R is the gas constant, and θ is the temperature in Kelvin, the fluorescence-temperature data could be fitted to a straight line having the equation

$$R \ln \left(\frac{F - \beta}{\gamma - F} \right) = -\Delta H_{2 \rightarrow 3}^0 \frac{1}{\theta} + \Delta S_{2 \rightarrow 3}^0, \quad (2.5)$$

where $\Delta H_{2 \rightarrow 3}^0$ is the slope and $\Delta S_{2 \rightarrow 3}^0$ is the intercept. This method assumes an all-or-none transition between the stem-of and-loop structure and the random coil, and it assumes that enthalpy and entropy do not vary with temperature.

2.2.4 Thermodynamic measurements of the dissociation of probe-target duplexes.

The thermodynamic parameters that describe the dissociation of a probetarget duplex (phase 1) into a free target and a molecular beacon in the form of a hairpin (phase 2), $\Delta H_{1 \rightarrow 2}^0$ and $\Delta S_{1 \rightarrow 2}^0$, were determined by analyzing the fluorescence data obtained from the thermal denaturation profiles of 50 nM solutions of molecular beacons incubated in the presence of six different concentrations of added target oligonucleotide (between 300 nM and 300 mM). The analysis was based on the determination of the melting temperature of the probe-target duplexes, θ_m , at each target concentration. θ_m was determined by fitting all of the data in each thermal denaturation profile to Equation 2.3, which fully describes fluorescence at every temperature in terms of the two equilibrium constants. At the melting temperature of the duplex, $[B_{closed}] = [BT]$, and from equation [2.2], $K_{1 \rightarrow 2}^0$ equals $[T]$, whose value at θ_m is $T_0 - 0.5B_0$. Since $\Delta G^0 = -R\theta \ln K(\theta) = \Delta H^0 - \theta\Delta S^0$, the melting temperature data can be fitted to a straight line having the equation

$$R \ln (T_0 - 0.5B_0) = -\Delta H_{1 \rightarrow 2}^0 \frac{1}{\theta_m} + \Delta S_{1 \rightarrow 2}^0,$$

where $\Delta H_{1 \rightarrow 2}^0$ is the slope and $\Delta S_{1 \rightarrow 2}^0$ is the intercept. The following rearrangement of equation [2.3] illustrates why θ_m could be obtained from the fluorescence-temperature data:

$$K_{1 \rightarrow 2}^0 = \frac{(\alpha - F)T_0}{(F - \beta) + (F - \gamma)K_{2 \rightarrow 3}^0}$$

The value of β , the value of $\Delta H_{2 \rightarrow 3}^0$ and the value of $\Delta S_{2 \rightarrow 3}^0$, which are needed to calculate $K_{2 \rightarrow 3}^0$ at each temperature, were obtained from the measurements made with molecular beacons incubated in the absence of targets. For each target concentration, the value of α was obtained from the fluorescence measured at 15°C and the value of γ was obtained from the fluorescence measured at 80°C. Therefore, the fluorescence-temperature data could be used to calculate the value of $K_{1 \rightarrow 2}^0$ at every temperature, and the melting temperature θ_m of the probe-target duplex is the temperature at which $K_{1 \rightarrow 2}^0$ equals $T_0 - 0.5B_0$.

2.3 Fluorescence Correlation Spectroscopy

Recently, FCS regained attention as a very practical technique to analyze molecular dynamics (diffusion and conformational fluctuation). The revival came at the cross-roads of many technical improvements in light detection and optics.

The introduction of a confocal geometry enabled the reduction of the background fluorescence and a better definition of the sampling volume: diffusion coefficients can be readily measured *in vitro* and *in vivo* without altering the samples (a great advantage for long term stability of the system compared to the fluorescence after recovery photobleaching (FRAP) technique) [108]. Furthermore, the reduction of the sampling volume to the diffraction limit $(0.5\mu m)^3$ increases the sensitivity of the technique: practically, FCS can be carried on single molecules[79, 109], which makes it amenable for applications in analytical chemistry and biophysics.

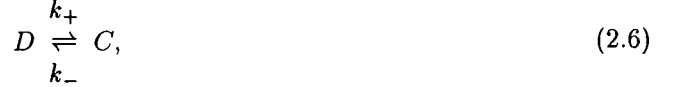
The development of single-photon counting avalanche photodiodes (APD) for the visible wavelengths greatly improves the quantum efficiency of photon detection. Classical photomultiplier tubes (PMT) have a typical quantum efficiency of 7% at a wavelength of 550nm whereas the avalanche photodiode can detect up to 70% of the incoming photons. This enhancement of light detection efficiency by a factor of 10 reduces the acquisition time by 10 as will be discussed later 2.3.4.

In this section, we will derive the analytical expression used to fit the autocorrelation function computed from the collected fluorescence of the molecular beacons. We will also present our experimental setup and its implementation for Fluorescence Correlation Spectroscopy.

2.3.1 Chemical-diffusion equation for the molecular beacon.

In this section, we solve the equations coupling chemical fluctuations and diffusion of the molecular beacon. The solution of these equations describes the relaxation dynamics of the fluctuations, to be plugged in the final computation of the fluorescence correlation function.

The system consists in the two conformations of the molecular beacon (Figure 2.9): closed (C) or open and denatured (D). The exchange between these two states is assumed to occur in an all-or-none fashion *i.e.* only two chemical rates are necessary to describe the fluctuations:



The equations coupling diffusion and chemical fluctuations are then:

$$\begin{cases} \frac{\partial [C](\vec{r}, t)}{\partial t} = D\Delta[C](\vec{r}, t) + k_+[D](\vec{r}, t) - k_-[C](\vec{r}, t) \\ \frac{\partial [D](\vec{r}, t)}{\partial t} = D\Delta[D](\vec{r}, t) - k_+[D](\vec{r}, t) + k_-[C](\vec{r}, t) \end{cases} \quad (2.7)$$

The equilibrium concentrations are such that:

$$\frac{[D]_{\text{equilibrium}}}{[C]_{\text{equilibrium}}} = K = \frac{k_-}{k_+}$$

The conservation of matter ($[D]_{\text{equilibrium}} + [C]_{\text{equilibrium}} = C_{\text{total}}$) gives:

$$\begin{cases} [D]_{\text{equilibrium}} = C_{\text{total}} \frac{K}{1+K} \\ [C]_{\text{equilibrium}} = C_{\text{total}} \frac{1}{1+K} \end{cases}$$

We can solve the equations in terms of fluctuations around equilibrium:

$$\begin{cases} [D](\vec{r}, t) = [D]_{\text{equilibrium}} + d(\vec{r}, t) \\ [C](\vec{r}, t) = [C]_{\text{equilibrium}} + c(\vec{r}, t) \end{cases}$$

such that the conservation of matter is now simply: $d(\vec{r}, t) + c(\vec{r}, t) = 0$. The diffusion-chemical equations (2.7) can be rewritten with fluctuations in the Fourier Space:

$$\begin{cases} \frac{\partial c(\vec{q}, t)}{\partial t} = -(Dq^2 + k_-)c(\vec{q}, t) + k_+d(\vec{q}, t) \\ \frac{\partial d(\vec{q}, t)}{\partial t} = -(Dq^2 + k_+)d(\vec{q}, t) + k_-c(\vec{q}, t) \end{cases}$$

The solution to this linear system of equations becomes the superposition of two s eigenmodes:

$$c(\vec{q}, t) = -d(\vec{q}, t) = Ae^{-Dq^2t} + Be^{-(Dq^2+k_-+k_+)t}$$

The amplitudes of these modes are evaluated in direct space, as standard fluctuations in a Poissonian distribution.

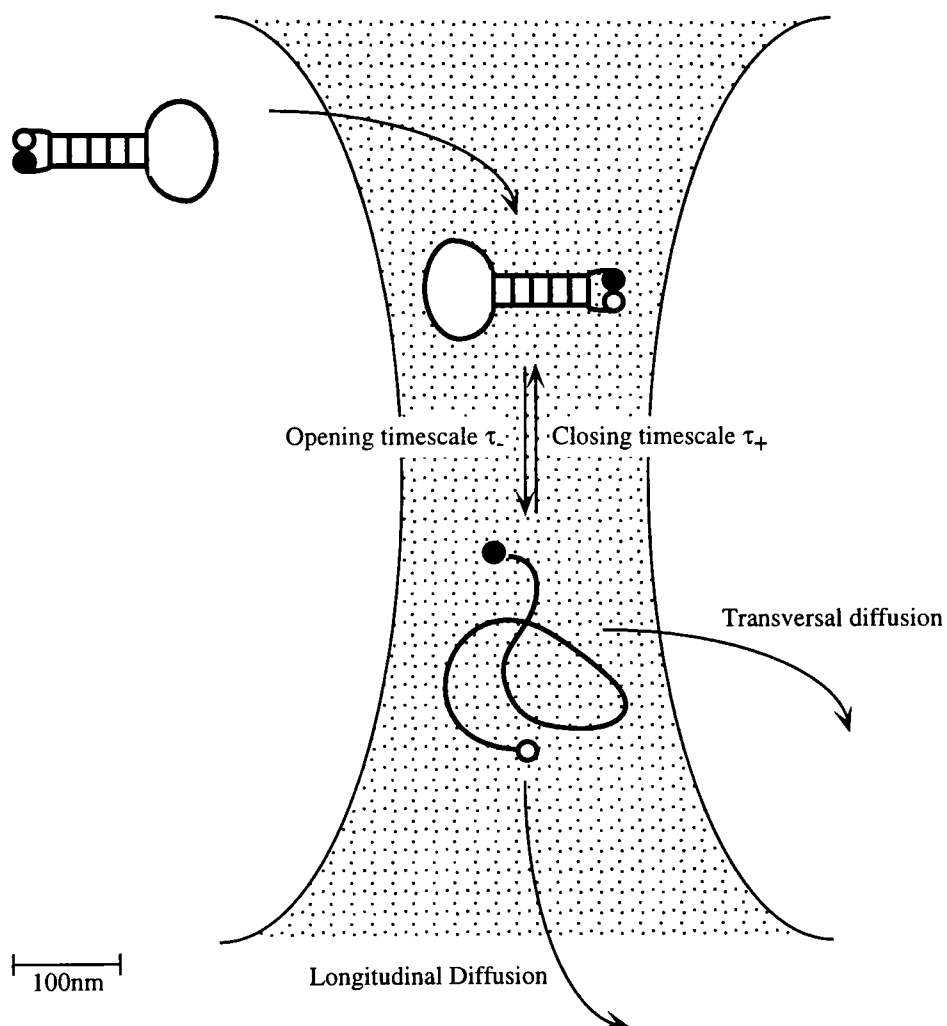


Figure 2.9: Sources of fluctuation in an FCS experiment with the molecular beacons. Four modes (*i.e.* four timescales) introduce fluctuations in the collected fluorescence from molecular beacons under confocal microscope: τ_{\perp} is the transversal-diffusion timescale to cross the beam waist by diffusion (typically $150\mu\text{s}$ for a 30-mer oligonucleotide), τ_{\parallel} is the longitudinal-diffusion timescale to cross the beam waist by diffusion (typically 15ms for a 30-mer oligonucleotide), τ_+ is the lifetime of the denatured state (typically, from 1 to $100\mu\text{s}$) τ_- is the lifetime of the closed state (typically, from $10\mu\text{s}$ to 10ms).

2.3.2 Analytical expression for the correlation function.

The correlation function of the fluorescence intensity is by definition normalized by the intensity $I(t)$ of the collected fluorescence ² :

$$G(\tau) = \frac{\langle I(t)I(t+\tau) \rangle_t - \langle I^2(t) \rangle_t}{\langle I(t)^2 \rangle_t}.$$

The correlation function is rewritten in terms of fluctuations around the equilibrium:

$$I(t) = \langle I(t) \rangle_t + \delta I(t) \Rightarrow G(\tau) = \frac{\langle \delta I(t) \delta I(t+\tau) \rangle_t}{\langle I(t)^2 \rangle_t}$$

To compute these averages, one has to integrate the fluorescence contribution from all the points in space. Three fields are convoluted to yield the collected fluorescence intensity: the local concentration of fluorescent molecules ($c(\vec{r}, t)$ for the closed molecules or $d(\vec{r}, t)$ for the denatured molecules), the intensity of the excitation light $I_{excitation}(\vec{r}, t)$, the efficiency of fluorescence collection $E(\vec{r})$. The fluctuation of collected fluorescence from a point \vec{r} in the sampling volume is:

$$\delta I(\vec{r}, t) = (Q_c c(\vec{r}, t) + Q_d d(\vec{r}, t)) I_{excitation}(\vec{r}) E(\vec{r}),$$

where Q_d and Q_c are the quantum efficiencies of the denatured and closed beacons.

To simplify the computation at first, we will assume that the closed beacon is perfectly quenched ($Q_c = 0$) such as only the open beacons contribute to the correlation function. We will assume also that the fluorescence collection efficiency is constant throughout the sample: $E(\vec{r}) = 1$.

In our experiment, the excitation light consists in a laser beam focused onto the sample through a microscope objective. The beam profile is then Gaussian:

$$I_{excitation}(\vec{r}) = I_0 e^{-\frac{x^2+y^2}{2w^2}} e^{-\frac{z^2}{2Z^2}},$$

where w is the transversal beam waist and Z is the longitudinal beam waist (see figure 2.10 for coordinate definition).

The autocorrelation function is easily manageable in the Fourier space [86]:

$$\begin{aligned} G(\tau) &\propto \int d\vec{q} e^{-\frac{1}{2}q_{\perp}^2 w^2} e^{-\frac{1}{2}q_{\parallel}^2 w'^2} (\alpha e^{-D(q_{\perp}^2 + q_{\parallel}^2)\tau} + \beta e^{-(D(q_{\perp}^2 + q_{\parallel}^2) + k_- + k_+)\tau}) \\ &\propto \frac{1}{1 + \frac{\tau}{\tau_{\perp}}} \sqrt{\frac{1}{1 + \frac{\tau}{\tau_{\parallel}}}} (\alpha + \beta e^{-(k_- + k_+)\tau}), \end{aligned}$$

where $\tau_{\perp} = \frac{w^2}{4D}$ is the diffusion timescale to cross the beam waist, and $\tau_{\parallel} = \frac{Z^2}{4D}$ is the diffusion timescale to cross the depth of field; α and β are amplitudes that we will determine in the next section (§ 2.3.3) In our setup, Z is 10 times larger than w , thus $\tau_{\parallel} = 100\tau_{\perp}$: the fluctuations of number is mostly due to in-plane diffusion and:

$$G(\tau) \propto \frac{1}{1 + \frac{\tau}{\tau_{\perp}}} (\alpha + \beta e^{-(k_- + k_+)\tau}).$$

²The symbol $\langle \rangle_t$ means that we take an average over time. For example,

$$\langle I(t) \rangle_t = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T I(t) dt.$$

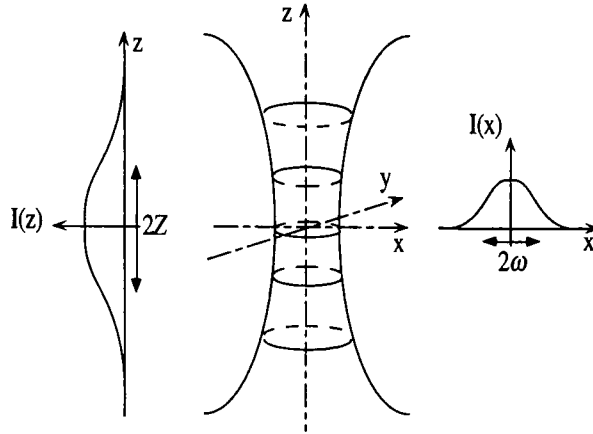


Figure 2.10: Geometry of the excitation beam in an epifluorescence microscope. The beam is Gaussian in every ordinates with a typical transversal waist of $w \sim 200nm$, and a longitudinal waist of $Z \sim 3\mu m$.

A more accurate computation would take into account the fluorescence collection efficiency $E(\vec{r})$, as defined by the confocal geometry. This correction would only rescale the shape of the gaussian illumination, and does not change the overall analytical expression for the correlation function.

Note that for freely diffusing molecules of constant fluorescence (*e.g.* dyes), the autocorrelation function is reduced to the freely-diffusing part:

$$G(\tau) \propto \frac{1}{1 + \frac{\tau}{\tau_d}} \quad (2.8)$$

2.3.3 Data analysis for a molecular beacon.

Amplitude of the correlation function. The amplitudes of the autocorrelation function as evaluated in section 2.3.2 are for an ideal case where no fluorescence comes from the background and the closed beacon is perfectly quenched. This section presents the autocorrelation function we expect from two non-ideal cases.

Diffusion case. For a solution of freely-diffusing fluorescent dyes, the only sources of fluctuation of the collected fluorescence are the fluorescence shot-noise and the fluctuation of the average number of particles in the field of view. The shot-noise is a statistical noise well compressed by the autocorrelation function (see section 2.3.4 for further discussion). The average number of particles in the field of view determines the amplitude of the autocorrelation function:

$$G(0) = \frac{\langle \delta I^2(0) \rangle}{\langle I(0)^2 \rangle} = \beta.$$

Our system is an open sample (the confocal volume) defined within a much larger total volume (1fl in $200\mu l$), thus β is the normalized variance of the gaussian distribution of number of particles:

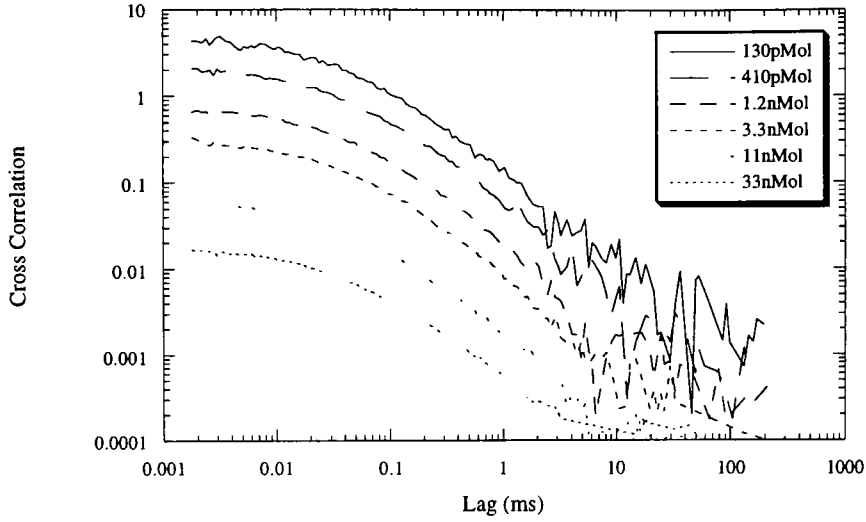


Figure 2.11: Cross-correlation function for a serial dilution of carboxyrhodamine 6G in 0.1M TrisCl pH8.0. The excitation laser intensity is 8 kW/cm^2 , the collection pinhole has a $25\mu\text{m}$ diameter.

$$\beta = \frac{1}{N},$$

where N is the average number of particles in the field of view. This results describes an ideal case where no stray fluorescence is collected. However, for low concentrations of fluorescent objects, one cannot neglect the background fluorescence. We can characterize the intensity of the constant background fluorescence as an equivalent constant number of fluorescent particles N_B . The corrected expression for the amplitude of the correlation function [77] is then:

$$\beta = \frac{N}{(N + N_B)^2} = \frac{c}{(c + c_B)^2} \frac{1}{V N_A}, \quad (2.9)$$

where c is the concentration of fluorescent particles, c_B is the equivalent concentration of background fluorescence, N_A is the Avogadro number and V is the sampling volume.

In figure 2.11, we present the cross-correlation function of collected light for a series of samples of different concentrations of carboxyrhodamine 6G in 0.1M TrisCl pH8.0.

As predicted, the intercept β of the correlation function increases as the concentration of objects decreases, *i.e.* as the amplitude of the number fluctuation increases. On figure 2.12, we present the intercept beta obtained from the diffusion fit (equation 2.8) of the cross-correlation function of figure 2.11. A tentative fit to the equation 2.9 shows a good agreement between the predicted and the measured intercepts, and enables us to measure the equivalent concentration of particles in the background:

$$c_B = 20\text{pMol}.$$

This study also enables us to better characterize the confocal geometry of our setup, as we can experimentally measure the dimensions of the sampling volume. From the fit of the data in figure 2.11, we evaluate the transversal diffusion timescale and thus the diameter of the sampling volume (see equation 2.8) :

$$\tau_{\perp} = 32 \pm 0.3\mu\text{s} \Rightarrow w = \sqrt{4D\tau_{\perp}} = 190\text{nm},$$

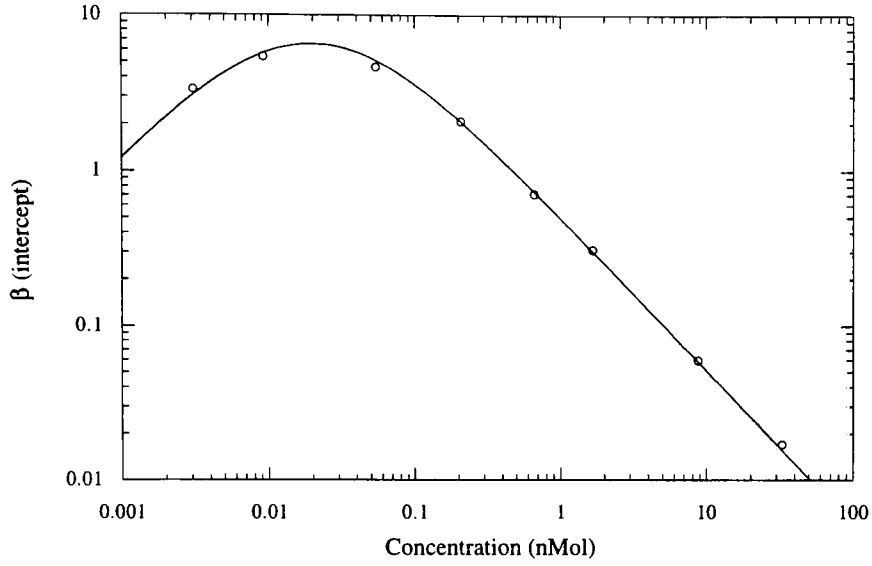


Figure 2.12: Intercept β of the cross-correlation function for a serial dilution of carboxy rhodamine 6G in 0.1MTrisCl pH8.0. These intercepts are measured from the cross-correlation presented in figure 2.11. The fit presented here corresponds to the equation (2.9). The fit gives $c_B = 20pMol$ and $V = 0.8fl$.

the diffusion coefficient of Rh6G being $D = 2.8 \times 10^{-6} cm^2 s^{-1}$). From the fit of the data in figure 2.12, we evaluate the sampling volume:

$$V_{sample} = 0.8fl = \pi w^2 Z \Rightarrow Z = 6.3 \mu m.$$

Thus, the sampling volume, from which fluorescent light is collected, is 380nm wide, and 6.3 μm deep

Molecular beacon case. The theoretical expression derived in section 2.3.2 is simplified into:

$$G(\tau) \propto \alpha + \beta e^{-\tau R},$$

for non-diffusing molecular beacons. The purpose of this paragraph is to determine α and β . To compute theses amplitudes of the correlation function of fluorescence emitted by a fluctuating molecular beacon, one has to take into account the residual fluorescence intensity emitted by the closed-quenched state (I_C) as much as the fluorescence intensity emitted by the denatured-unquenched state (I_D). The ratio of these two fluorescences is the signal/noise of the molecular beacon: $\delta = I_D/I_C$.

The fluorescence emitted by a single molecular beacon is $\langle i \rangle = pI_D + (1-p)I_C$, where p is the probability of the molecular beacon to be in an open state. Then,

$$\begin{aligned} \langle i^2 \rangle &= pI_D^2 + (1-p)I_C^2, \\ \text{and } \langle (\delta i)^2 \rangle &= \langle i^2 \rangle - \langle i \rangle^2 = p(1-p)(I_D - I_C). \end{aligned}$$

For N molecular beacons, the central limit theorem gives the variance of the total collected fluorescence I :

$$\langle I^2 \rangle = N \langle i^2 \rangle = Np(1-p)(I_D - I_C).$$

Figure 2.13: Amplitude of the cross-correlation function as the function of the open probability p for the molecular beacon.

Thus, the correlation function at 0 lag time is:

$$G(0) = \frac{\langle (\delta I)^2 \rangle}{\langle I^2 \rangle} = \alpha + \beta = \frac{1-p}{p} \left(\frac{(\delta-1)p}{(\delta-1)p+1} \right)^2 \quad (2.10)$$

The amplitude of the autocorrelation function is thus non-monotonic as the probability of the fluorophore to be quenched increases, as the temperature decreases (Figure 2.13).

For a large signal/noise of the molecular beacon (large compared to the probability of being closed *i.e.* $\delta p \gg 1$), we recover the equilibrium constant K as the amplitude of the correlation function:

$$G(0) = \frac{1-p}{p} = K \quad \text{for } \delta p \gg 1.$$

In the other limit where denatured state and closed state are indistinguishable (*i.e.* $I_D = I_C$ or $\delta = 1$), or when very few molecular beacons are open and fluorescent ($p \ll 1$), there is no fluorescence fluctuation, the amplitude of the autocorrelation function should be null. The equation (2.10) is consistent with:

$$G(0) = 0 \quad \text{for } \delta \sim 1 \text{ or } p \sim 0.$$

In conclusion, the amplitude of the autocorrelation function in an FCS experiment is never a reliable measure of the equilibrium constant K for the fluctuations.

2.3.4 Statistical accuracy in fluorescence correlation spectroscopy.

To assess the signal/noise ratio of an FCS experiment, one must take into account the inner stochasticity of the detection (shot-noise of the fluorescence emission by a dye, and shot-noise of the light collection efficiency), as well as systematic errors (background fluorescence, photobleaching, partial quenching...).

To evaluate the statistical accuracy of the acquired autocorrelation function, we must acquire a distribution of many acquisition $\hat{G}(\tau)$. We define the following signal/noise ratio $S/N(\tau)$ for each lag time τ :

$$S/N(\tau) = \frac{G(\tau)}{\sqrt{\text{Var}[\hat{G}(\tau)]}},$$

where $G(\tau)$ is the median of all the experimental acquisition of $\hat{G}(\tau)$, and $\text{Var}[\hat{G}(\tau)]$ its variance. Note that the signal/noise must be low at infinite lagtime (very few objects stay in the field of view long enough to contribute to the autocorrelation function at this lag time) and at 0 lagtime (the emission and collection of fluorescence at fast timescale are not efficient enough to give enough photon statistics).

This signal/noise ratio depends on the average fluorescence f collected per object as well as the concentration of objects, and the total acquisition time T .

D. Koppel [77] derived the theoretical expression for the signal/noise ratio when the solution of fluorophores is dilute and the background fluorescence negligible. The main result is that the signal/noise ratio does not depend on the total number of collected photons *per se*, but rather on the number of collected photons per fluorescent molecule. Moreover, in the high-counting rate limit, the experimental noise comes solely from the shot-noise of the fluorescence intensity. Koppel's asymptotic solution is:

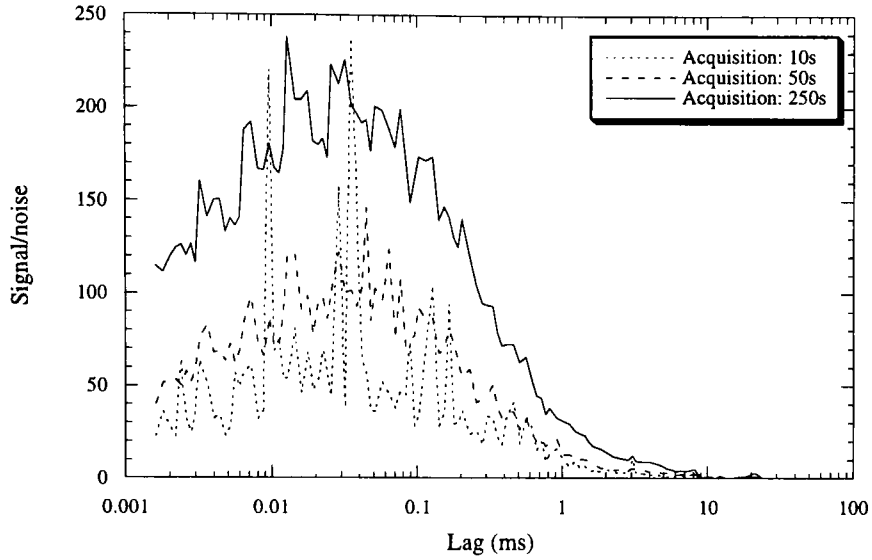


Figure 2.14: Signal/noise of the cross-correlation function for freely Rh6G at different concentration ($P_{\text{laser}} = 0.8 \text{ kW/cm}^2$, pinhole diameter = $25 \mu\text{m}$)

$$S/N(0) \sim \sqrt{fT}.$$

i.e. the signal/noise ratio increases as the square-root of the time of acquisition and the square-root of the fluorescence per object. We confirm this prediction for freely diffusing dyes (Figure 2.15). This equation implies that the improvement of light collection by a factor of 10 enables us to decrease by a factor of 10 the acquisition time. More generally, Koppel's definition of the signal/noise of the fluorescence autocorrelation provided a quantitative tool to assess its statistical accuracy, and optimize the experimental setup (choice of optical filters, choice of microscope objective). However, Koppel's formalism holds in a low-tech version of an FCS setup where the collection efficiency is low and the background fluorescence is high: the theoretical expression he predicted for S/N is not generally valid in our experiment. We will use its quantitative tools (definition of the signal to noise), but optimize phenomenologically this signal to noise for our own setup.

2.3.5 Experimental setup for fluorescence correlation spectroscopy.

Our experimental setup is fully homemade, following the design implemented by Rigler's group [79]. We used the definition of the signal to noise ratio from § 2.3.4 to optimize the different components.

Description of the optical setup. Figure 2.16 is a schematic of our setup. The excitation light comes from a continuous-wave water-cooled Argon Ion Laser (Lexel Model 9500) operated in a single-line mode at 514.5nm, in light control (for laser stability). The typical output power used is 100mW, and a neutral density filter(optical densities 2.0) cuts down the excitation intensity: the light control was found to be unstable for laser powers below 30mW. An additional removable neutral density filter of optical density 2.0 is used to further diminish the incoming light during sample handling.

The laser beam is expanded to a 5mm-diameter parallel beam with a beam expander (lens of focal distance 54mm conjugated to the focal plane of a 60X objective). This beam expander is aligned with the optical axis of the setup by aligning its axis by rotation (align a back reflection

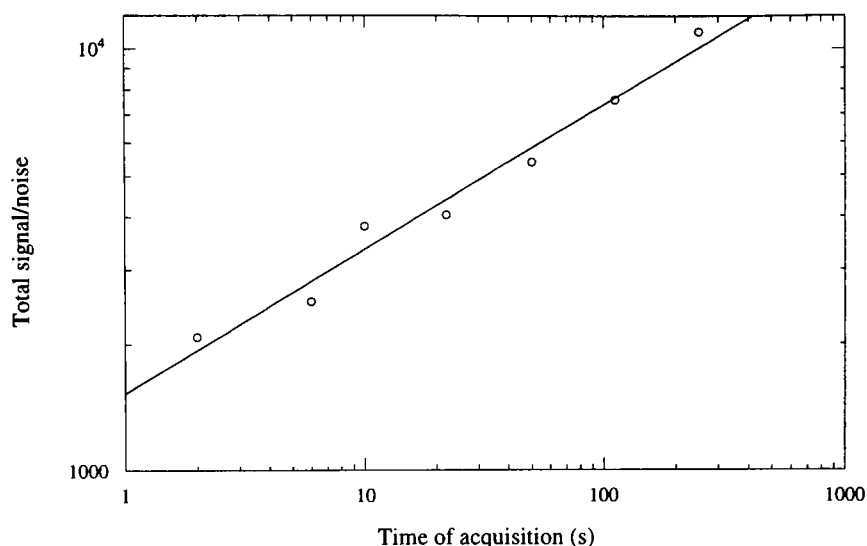


Figure 2.15: Square of the total signal/noise of the cross-correlation function for freely Rh6G for different acquisition time ($P_{\text{laser}} = 0.8 \text{ kW/cm}^2$, pinhole diameter = $25 \mu\text{m}$). The total signal/noise is defined as the integral of the curves presented in figure 2.14, between $2 \mu\text{s}$.

with two rotation degrees of freedom) and translation (center the beam with two translation degrees of freedom).

The light is reflected at 90° by a dichroic mirror (Chroma Optics DRLP525), and fed into a high-numerical aperture objective (Olympus UApoPlanAchromat, Magnification 60X, Numerical Aperture 1.2, water immersion), and focused onto the sample.

The sample is a chamber made of a wax-coated O-Ring (diameter 1cm, width 2mm) glued onto a coverslip n°1 (dimension $50 \times 24 \text{ mm}$, thickness $\sim 150 \mu\text{m}$). This chamber can be closed by an additional coverslip to avoid sample evaporation. The sample is typically made of 10nMol of fluorescent objects in a saline buffer.

The fluorescence of the objects excited by the focused laser beam is collected by the same objective, filtered by the dichroic mirror (acting as a long-pass filter), further filtered by a notch holographic filter, centered at 514.5nm with a notch width of 9nm and a notch optical density of 6.0 (Kaiser Optics), and finally focused onto a $25 \mu\text{m}$ -diameter pinhole aperture by a lens of focal length 175mm. The lens is mounted onto a lens tube (Thor Lab) to allow z-translation with the pinhole. The pinhole itself is mounted onto a differentially-adjustable XY stage (Thor Lab) for further alignment.

This geometry (combination of epifluorescence and spatial filtering in the back focal plane) is the classical design for a confocal microscope [110].

Collection of fluorescence. Two detection channels are implemented to collect the incoming fluorescence photons: PMT (photomultiplier) or APD (avalanche photodiode). A removable mirror after the pinhole could redirect the flux of photons towards a photomultiplier tube (PMT Thorn EMI). The advantage of the PMT is its large collection photocathode (1cm diameter), which allows the alignment of the $25 \mu\text{m}$ -diameter pinhole without further movement of the detector. The output of the PMT is a TTL signal, fed into a standard counter (Hewlett Packard) for frequency measurement (sampling rate: 1Hz). We align the pinhole by simply optimizing the measured countrate. The use of this low quantum efficiency measurement by a PMT has also the advantage to control the flux of photons before feeding it onto the extremely sensitive (and thus damage-prone) APDs'.

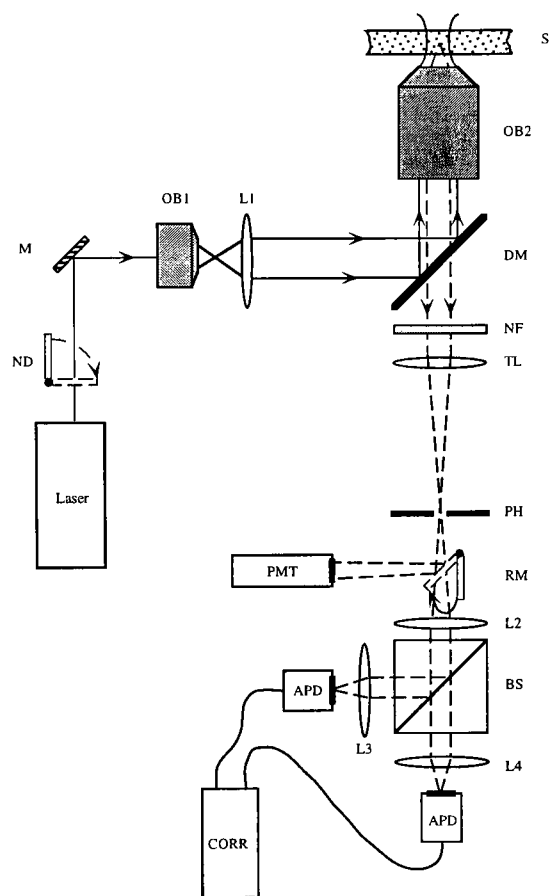


Figure 2.16: Optical setup for Fluorescence Correlation Spectroscopy. Optical components: ND: flippable neutral density filter of optical density 2.0; M: mirror mounted on rotational stage; OB1: 60X objective; L1: conjugated lens to OB1 $f=54\text{mm}$ (OB1 and L1 are mounted together onto a XY stage); OB2: UApoplan Achromat 60X N.A1.2 objective lens; S: sample (mounted onto XYZ stage); DM: dichroic mirror; NF: notch filter; TL: tube lens of focal length 175mm (mounted onto a Z stage); PH: pinhole of diameter $25\mu\text{m}$ (mounted onto an XY stage); BS: cube beamsplitter; RM: removable mirror to redirect the light onto the photomultiplier tube (PMT), for pinhole alignment; L2 & L3 & L4: collection coated lens $f=75\text{mm}$ (mounted onto XYZ stage); APD: avalanche photodiode; CORR: correlator.

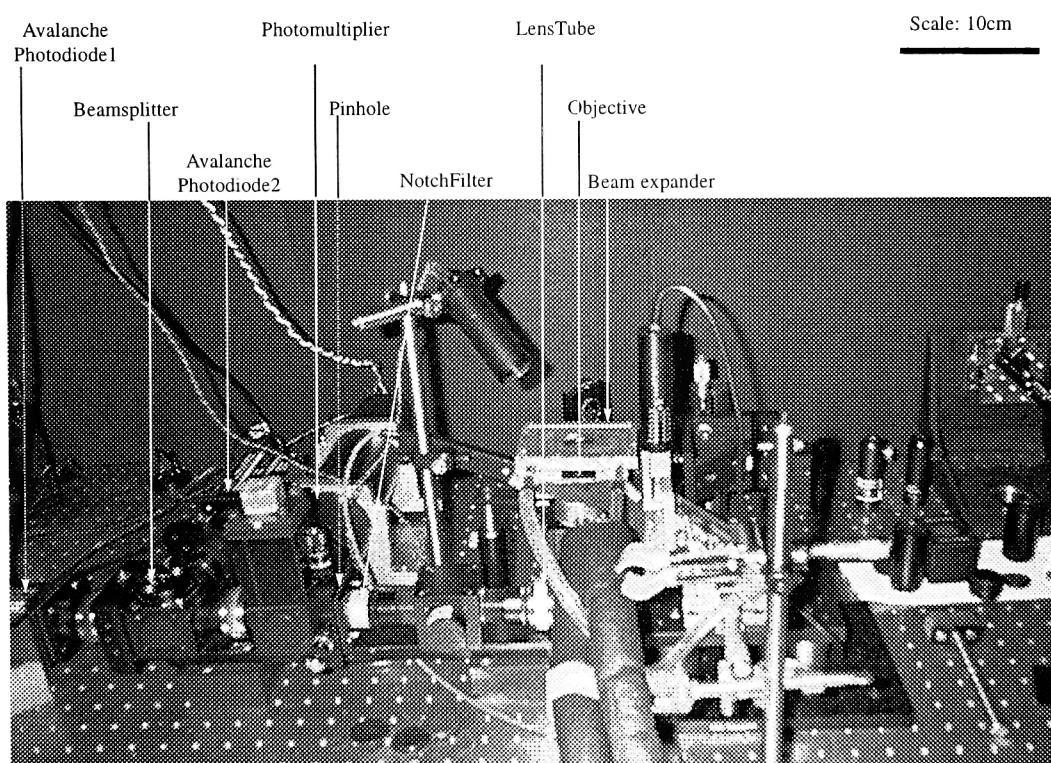


Figure 2.17: Photography of our FCS setup.

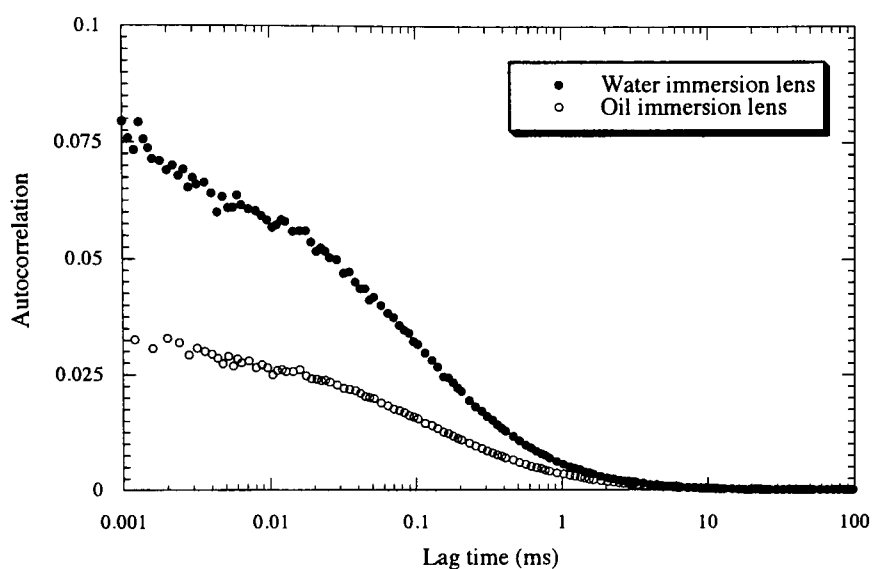


Figure 2.18: Autocorrelation function of the fluorescence intensity collected from a solution of 10nMol of Rh6G with a water immersion objective (•), and with an oil-immersion objective (◊).

Once the position of the pinhole is optimized, the mirror is removed to redirect the flux of photons towards the APDs' (SPCM-AQR14 - EG&G Optoelectronics, now Perkin Elmer). Because the active area of these diodes is only $150\mu\text{m}$ wide, the incoming light has to be refocused by a lens (focal length: 54mm), mounted onto an XYZ stage. The fluorescence signal is in fact split by a beamsplitter cube (Newport) onto two APDs', to reduce the highly correlated noise generated by the detector afterpulsing. The output of the APD is also a TTL signal. The TTL signals were fed into a Correlator board (ALV 5000E) to build the crosscorrelation function of the two signals.

Optimization of the optical components.

Objective lens The choice of objective lens implies the optimization of the collection efficiency, combined with a preserved definition of the confocal volume [111, 112]. Two possible objective lenses were tested: UApoplan Achromat 60X Numerical Aperture $\text{NA}=1.2$, water immersion (Olympus), 100X $\text{NA}=1.4$ oil immersion (Olympus). The figure 2.18 shows the autocorrelation functions, collected with the two possible lenses: the intercept of the autocorrelation function at 0-lag time is twice bigger for the water-immersion objective than the oil-immersion objective, i.e. the confocal volume is twice smaller. In fact, the water immersion lens achieves a signal/noise ratio twice superior to the oil immersion lens.

The fluorescence collection efficiency is proportional to the square of the numerical aperture, such as we should have picked the oil-immersion lens ($\text{NA}=1.4$) against the water-immersion lens ($\text{NA}=1.2$). However, the definition of the confocal volume is more reliable with a water immersion lens. Indeed, the UApoPlan lens from Olympus is designed for inverted microscopes, implying the presence of a coverslip between the objective and the sample. The lens contains an iris to correct for the thickness of the coverslip, in such a way that the working distance can be anything between $130\mu\text{m}$ (the minimal thickness of a coverslip) and $250\mu\text{m}$. In the case of an oil-immersion lens, the sample has to be placed in such a way that the lens focuses the excitation light right above the coverslip (interface glass-sample): as soon as the lens is set to focus "deep" inside the sample, the optimal excitation-collection geometry is lost, and the confocal volume is then poorly defined. The figure 2.19 shows how the sample height can be adjusted without affecting the geometry of

the excitation-collection in the case of a water-immersion lens, while the focusing right above the coverslip is crucial for an oil-immersion lens.

Special filtering and depth of field. To achieve an efficient collection of fluorescence for correlation spectroscopy, one must spatially filter the emitted fluorescence. This is classically done with a confocal microscopy geometry: the emitted fluorescence collected from the objective is focused onto a pinhole in the image focal plane. Any fluorescent object outside of an effective volume to be defined will see its collected fluorescence cut by the pinhole aperture.

In this section, we compare two pinholes (diameter $25\mu\text{m}$ and $100\mu\text{m}$) to optimize the signal/noise ratio of the autocorrelation function of fluorescence intensity collected for diffusing particles (see 2.3.4 for the definition of this signal/noise). Our objective lens (UApo 60XNA1.2, Olympus) has a working distance of $250\mu\text{m}$; the expanded excitation beam having a diameter of 5mm , the angle of the excitation cone is:

$$\delta = 2\text{Arctan}\left(\frac{2.5}{0.25}\right) \sim 170^\circ,$$

to be compare with the numerical aperture angle:

$$\alpha = \text{Arcsin}\left(\frac{1.2}{1.33}\right) \sim 64^\circ.$$

A precise Gaussian-optics estimation shows that the relevant parameter for focusing the beam is the smaller of the two [79]. The radius of the diffraction limited spot in the field of view, for a wavelength of $\lambda = 514.5\text{nm}$, is theoretically [113, 114]:

$$\omega = \frac{\lambda}{n_{\text{water}} \pi \tan(\alpha)} \sim 60\mu\text{m}. \quad (2.11)$$

With a magnification of $M=60\text{X}$, the radius of the first Airy disk in the object focal plane should thus be of $\omega * M \sim 23\mu\text{m}$. This is the basis for picking a pinhole of diameter $\emptyset = 25\mu\text{m}$ (any smaller pinhole would block some of the emitted fluorescence without shrinking further the illumination volume, any larger pinhole would uselessly enlarge the sampling volume). Once the tube lens and the pinhole are set, one can estimate the depth of field of the excitation-collection of the fluorescence. Fig.2.20 shows the tracing of a ray through the optical system. The definition of the depth of field is somewhat arbitrary: a fluorophore will be said to be in field of view if 13% ($1/e^2$) of its fluorescence can pass through the pinhole. With this threshold set, one can move a fluorophore along the optical axis, compute the collection efficiency, and define the depth of field [115]. The beam waist $w(z)$ is dependent on the longitudinal coordinate:

$$w(z) = \omega \sqrt{1 + \left(\frac{\lambda z}{\pi \omega^2}\right)^2}$$

A fluorophore at the coordinate z' along the optical axis will be in the field of view if $w(z') = e * \emptyset / M / 2 = 0.56\mu\text{m}$, which defines z' :

$$\begin{aligned} z' &= \frac{\pi \omega^2}{\lambda} \sqrt{\left(\frac{w(z')}{\omega}\right)^2 - 1}, \\ Z &= 2z' = \frac{2 * 3.14 * 0.30^2}{0.51} \sqrt{\left(\frac{0.56}{0.30}\right)^2 - 1} \sim 1.8\mu\text{m}. \end{aligned}$$

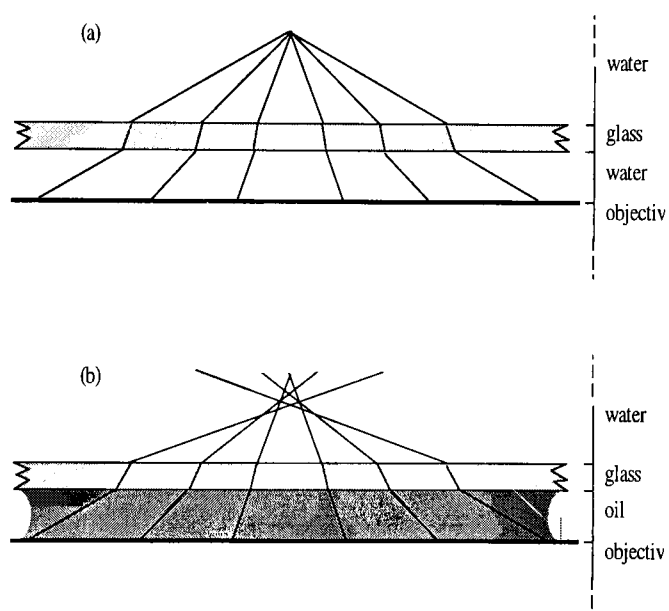


Figure 2.19: Schematic ray tracing in the sample: (a) for a well-corrected objective (water immersion with coverslip correction) (b) for an ill-compensated objective (oil immersion). In the first case, the light crosses two water-glass interfaces, the refraction at the interfaces is well compensated and a paraxial parallel beam of light is focused into a well-defined spot: the water immersion objective allows a focusing deep inside the sample. In the second case, the light crosses one oil-glass interface and one glass-water interface, and the refraction angles are not compensated: the oil-immersion objective easily introduces aberrations

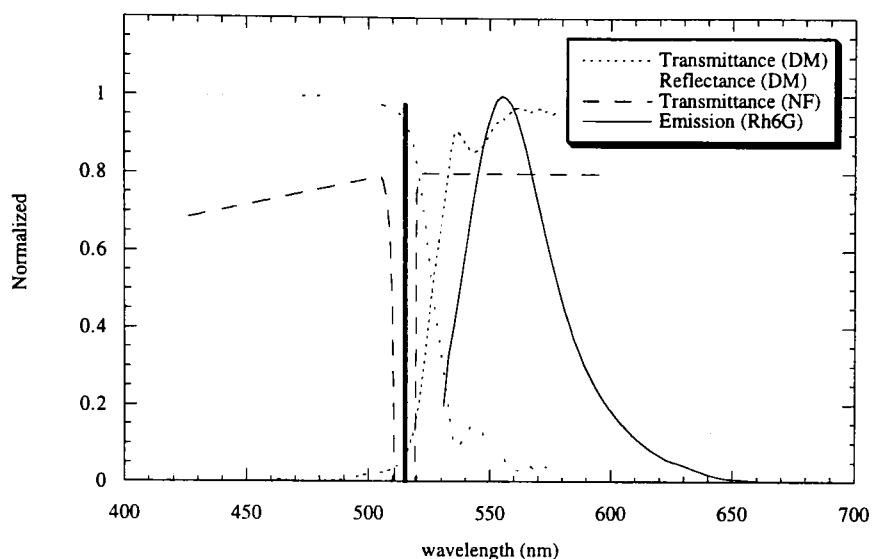


Figure 2.21: Normalized transmission and reflection coefficients, as a function of the light wavelength, for the optical filters used in our FCS setup: a) Reflection coefficient of the dichroic mirror (Chroma Optics^{inc} DRLP525); b) transmission coefficient of the dichroic mirror (Chroma Optics^{inc} XF104); c) transmission coefficient of the notch filter (Kaiser Optics^{inc} NotchPlus); d) emission spectra of Rh6G. Note: the maximum optical density of the Notch filter is 6.0.

Wavelength filtering Filtering the collected fluorescence via optical filters enables us to optimize the signal/noise ratio of our correlation setup. At stake is the rejection of the Rayleigh scattering (quasi-elastic scattering from the buffer) as well as the Raman scattering (inelastic scattering with the vibration modes of the buffer), and the collection of only the fluorescence photons.

Standard optical filters used for imaging applications are bandpass filters. Their window of transmission is centered to the right of the fluorescence peak to reject most of the scattered excitation light, while transmitting half of the peak of fluorescence (see 2.21). Thus these filters are quite “gourmand” and do absorb a lot of the fluorescence photons.

For FCS application, we found the optimal filter was simply a notch filter on the excitation wavelength. Indeed, the quasi-elastic scattering of the excitation light is of high amplitude but absolutely uncorrelated. Collecting the fluorescence in the correlation mode naturally minimizes the background contribution. Moreover, this notch filter transmits the entire fluorescence peak, whereas a bandpass emission filter (classically used in fluorescence microscopy applications) would absorb half of the peak of fluorescence to minimize the background scattered light.

Photodetector afterpulsing. In our experimental setup (§ 2.3.5), you must have noticed that the fluorescence emission was not collected and processed as an autocorrelation function, but was rather splitted between two detectors and processed as a cross-correlation function of the half-signal. This technical detail has been implemented to reduce the noise due to the photodetector afterpulsing.

Afterpulsing is a intrinsic problem of any photon counting detector, the detector being either an APD or a PMT. Whenever a photon hits the diode or the photocathode and generates a cascade of electron captured to generate a current to be amplified, discriminated and digitized into a TTL signal, there is a non-negligible probability that the ions or positive holes left in the detector generate an “echo”, more commonly named afterpulsing when recombining. This afterpulsing leads to a highly correlated noise in the detector, and typically generates a peak at 100ns in the auto-correlation

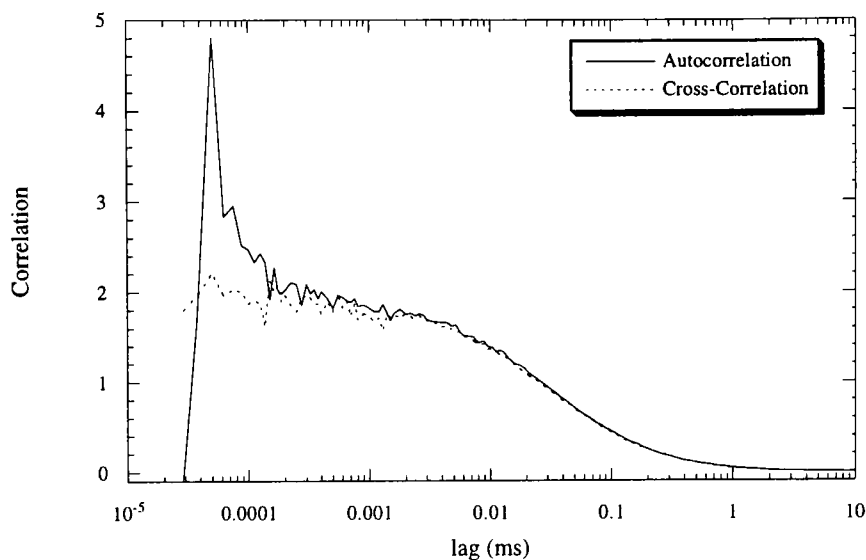


Figure 2.22: Comparison of the autocorrelation and the cross-correlation of the collected fluorescence of freely-diffusing Rh6G. Note the peak at 100ns in the autocorrelation function, due to the afterpulsing of the detector.

function of any signal fed to the detector (see figure 2.22).

Photochemistry of the dyes Fluorescence Correlation Spectroscopy is a very convenient and straightforward technique to study the photochemistry of fluorescent dyes [116]. By measuring the fluctuation relaxation, one can gain insights into the relaxation rates of a fluorescent molecules. Most relevant to our experiment is the measurement of the triplet lifetimes of carboxyrhodamine-6G and carboxy fluorescein. Triplet lifetimes are classically measured by phosphorescence relaxation, fluorescence saturation, or relaxation by absorption of triplet state quenchers.

The triplet lifetime is a critical parameter for our FRET-FCS experiment on DNA as it introduces a spurious quenched state in the analysis of the conformational fluctuations³. Indeed, if the dye reaches a long-lifetime triplet state (typically 1 μ s), it cannot emit any fluorescence photon for that period of time: any DNA conformational fluctuation occurring in that time window will not be recorded.

The measurement of the triplet lifetime relies on the analysis of a simplified Jabłoński diagram of the dye (figure 2.23). Three states are sufficient to describe the dye relaxation as recorded by the Fluorescence Correlation Spectroscopy technique. S_0 is the ground singulet state, S_1 is the excited singulet state, T is the excited triplet state, k_{ij} is the relaxation rate from the state i to the state j . Note that the lifetime of the excited triplet state is long compared to all the other photochemical relaxation as it involves a parity breaking. However, once spin-orbit and other electron-spin interactions result in a perturbation mixing singlet and triplet.

The linear equations of fluctuations are diagonalized to give the two timescales of relaxation.

$$\frac{\partial}{\partial t} \begin{pmatrix} S_0(t) \\ S_1(t) \\ T(t) \end{pmatrix} = \begin{bmatrix} -k_{01} & k_{10} & 0 \\ k_{01} & -k_{10} - k_{12} & 0 \\ 0 & k_{12} & -k_{20} \end{bmatrix} \begin{pmatrix} S_0(t) \\ S_1(t) \\ T(t) \end{pmatrix}. \quad (2.12)$$

The three eigenvalues are:

³All singlet to triplet transitions are strictly quantally forbidden in the zero-order approximation of the orbital configuration description.

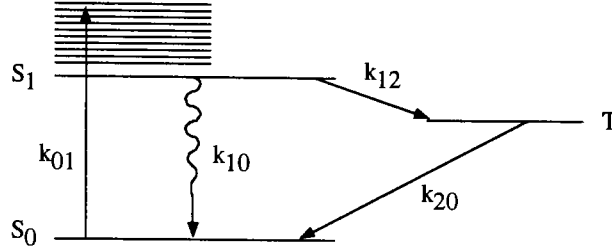


Figure 2.23: Jablonsky diagram [117] for carboxy rhodamine 6G (Rh6G). The energy levels and the associated lifetimes are shown for the most relevant photochemical states of carboxyrhodamine 6G, excited with $12\text{kW}/\text{cm}^2$ in a saline aqueous buffer (0.1M TrisCl pH8.0).

$$\begin{aligned}\lambda_1 &= 0 \\ \lambda_2 &= -(k_{10} + k_{01}) \\ \lambda_3 &= -(k_{20} + \frac{k_{01}k_{12}}{k_{01} + k_{10}}).\end{aligned}$$

The first timescale corresponds to the conservation of the matter ($S_0(t) + S_1(t) + T(t) = 1$), the second timescale is dominated by a fast timescale (lifetime of the singlet excited state at low laser power, or pumping time of the singlet ground state at high laser power), only the third timescale is accessible to a FCS measurement. The trick of the measurement consists in noticing that the pumping rate k_{01} from S_0 to S_1 is linear with the excitation intensity, $k_{01} = \sigma I$ (where σ is the cross-section of the ground singlet state) thus λ_3 is dependent on the laser intensity I as:

$$\lambda_3(I) = -(k_{20} + \frac{\sigma I k_{12}}{\sigma I + k_{10}}). \quad (2.13)$$

At low intensity, the relaxation of this third mode is dominated by the triplet to ground singlet transition (k_{20}). At high laser intensity, the relaxation is dominated by the excited singlet to triplet transition $-(k_{20} + k_{12})$.

The experiment is straightforward: we record the crosscorrelation of freely-diffusing Rh6G for different laser intensities. As we increase the laser intensity, the amplitude of the third mode of relaxation appears, and an exponential contribution is superposed onto the diffusion contribution (see § 2.3.1 for a complete analytical derivation):

$$G(\tau) = \frac{1}{N} \frac{1}{1 + \tau/\tau_D} (1 + \beta \exp(-\lambda_3 \tau)). \quad (2.14)$$

At low laser intensities, β is small and one can fit the cross-correlation function directly with a diffusion contribution

$$G(\tau) = \frac{1}{N} \frac{1}{1 + \tau/\tau_D}. \quad (2.15)$$

Thus, as we increase the laser intensity, one can fit the whole cross-correlation 2.14 with two parameters : β and λ_3 . Finally, we fit the different timescale λ_3 as a function of the laser intensity with the equation 2.13 to extract the triplet decay rate k_{20} and the intercrossing rate k_{12} .

For Rhodamine 6G, in 0.1M TrisCl pH8.0,

$$\begin{aligned} k_{12} &= 1.1\mu s^{-1} \\ k_{20} &= 0.5\mu s^{-1} \\ \sigma &= 2.0(\text{\AA})^2 \end{aligned}$$

We presented this analysis of the photochemistry of Rh6G to emphasize the importance of the triplet state in Fluorescence Correlation Spectroscopy experiments. Most of the timescales we are interested in in our study of the conformational fluctuations of nucleic acids range from 100ns to 100 μ s. The triplet state lifetime ranges typically from 100ns to 1 μ s and can superpose its quenched-fluorescence signature on top of the the conformational fluctuations dynamics. However, the amplitude of the contribution of the triplet state in the autocorrelation function is dependent on the excitation laser power. Practically, we minimized the artefacts related to the triplet state by working with Rh6G (the lifetime of its triplet state is much shorter than the triplet lifetime of fluorescein), and by checking the robustness of the dynamics measurement by respect to the laser excitation intensity.

Collection efficiency of the confocal microscope. Divergent estimates of the collection efficiency of a confocal setup have been reported. The acquisition in the autocorrelation mode allows us to estimate the absolute number of molecules in the field of view. We measured the average countrate collected *per* dye molecule to be 70000 counts/second (cps) for $P_{\text{laser}} = 30\mu W$ in the case of Rh6G.

We still have to estimate the total number of photons emitted by a single fluorophore to get an estimate of the collection efficiency of our confocal setup. FCS allows us to gain some insight on the photochemistry of the dyes we used (see § 2.3.5). In particular, we measure the absorption cross-section, the excited state lifetime, the triplet-to-singlet rate of dye relaxation (see § 2.3.5).

Knowing the extinction coefficient ϵ and the quantum yield of the dye, one can estimate the dye cross-section and thus, for a given flux J of excitation light, the total number of photons emitted per fluorophore.

$$\epsilon(Rh6G) = 108 \cdot 10^3 \text{ Mol}^{-1} \text{ cm}^{-1} = 108 \cdot 10^3 \mathcal{N}_A^{-1} \text{ cm}^{-1} l^3 = 1.8 \times 10^{-20} \text{ m}^2 = 1.8 \text{ \AA}^2,$$

thus, the cross section of the fluorophore is: $\sigma = 1.8 \text{ \AA}^2$

The flux of excitation photons is computed knowing the laser power $P_{\text{laser}} = 30\mu W$. The energy E per photon of wavelength $\lambda = 514.5 \text{ nm}$ is:

$$E = \frac{hc}{\lambda} = \frac{6.2 \cdot 10^{-34} \times 3 \cdot 10^8}{5.1 \cdot 10^{-7}} = 3.6 \cdot 10^{-19} \text{ J/photon}.$$

The beam waist being of the order of the laser wavelength λ , the flux J of excitation photons is estimated to:

$$J = \frac{P}{E \lambda^2} = \frac{30 \cdot 10^{-6}}{3.6 \cdot 10^{-19} (5100)^2} = 3.3 \cdot 10^6 \text{ photons/s/}(\text{\AA})^2.$$

Then, the total number of photons absorbed by one Rh6G *per second* is:

$$\mathcal{N} = J \times \sigma = 3.3 \cdot 10^6 \times 1.8 \text{ \AA}^2 = 5.9 \cdot 10^6 \text{ photons/s/fluorophore.}$$

The Quantum Yield Q of Rh6G in a saline buffer at pH8.0 has been can be evaluated from standards measured in pure ethanol. The conversion formula is [118]:

$$Q = Q_{EtOH} \frac{\epsilon_{EtOH}}{\epsilon_{water}} \left(\frac{n_{water}^2}{n_{EtOH}^2} \right) = 0.94 \times \frac{124 \cdot 10^3}{108 \cdot 10^3} \times \frac{1}{1.36^2} = 0.58. \quad (2.16)$$

Thus, the total number of photons emitted by one Rh6G is $\mathcal{N} \times Q = 5.9 \cdot 10^6 \times 0.58 = 3.4 \cdot 10^6$ cps.

The collection efficiency is the ratio of the total number of collected photons and the the total number of emitted photons *per dye per second*. For our setup, we showed that, for an excitation power $P_{laser} = 30 \mu W$, we collect 70000 photons/s/fluorophore. Hence, our collection efficiency is $7 \cdot 10^4 / 3.4 \cdot 10^6 \sim 3\%$.

This number may seem low, but a rough estimate of the optical loss in our system shows that our setup is close to optimal: the objective lens has a numerical aperture of 1.2, thus we collected the fluorescence from a solid angle of 3.6 steradian ⁴*i.e.* we are collecting only $\Omega/4\pi = 44\%$ of the emitted photons. The objective lens absorbs typically 20% of the transmitted light. The collected light passes a dichroic mirror (with a transmission T of 90%), a notch filter ($T=80\%$) and four optical components (tube lens, collimating lens, APD collection lens, and APD windows): each optical component has an antireflection coating, so the losses are minimized to 10% per optical surface. Finally the detector (avalanche photodiode) has a quantum efficiency of 80% in the range of wavelength. The theoretical collection efficiency would thus be $0.44 \times 0.8 \times 0.9 \times 0.8 \times 0.9^4 \times 0.8 \sim 10\%$.

Our optical setup is thus near to optimal (we can account for almost all the lost photons). The only possible improvement would consist in building a 4π microscope, as proposed by [119]. This geometry circumvents the loss of photons inherent to the inverted microscope, by placing another collecting objective in *vis-à-vis* of the excitation objective: the light is the collected from a solid angle of 4π instead of the 2π in the standard inverted microscope...

⁴By definition of the Numerical Aperture, $NA = n_{water} \sin \alpha$, with α is the angle of collection and n_{water} is the index of refraction of water. In our case, $\alpha = \text{Arcsin}(\frac{1.2}{1.33}) \sim 64.4^\circ$, thus the collection solid angle is $\Omega = 2\pi(1 - \cos \alpha) \sim 3.6$ steradian

蘿蔔煮牛肉

材 料：白蘿蔔 10 兩（約 400 克），牛肉 4 兩（約 160 克），薑 2 片，葱 2 條，生粉 2 茶匙。

醃 料：鹽 $\frac{1}{4}$ 茶匙，生粉 1 茶匙，水 2 湯匙，油 1 湯匙。

芡味料：上湯 1 杯，糖、鹽各 $\frac{1}{4}$ 茶匙，麻油、胡椒粉各少許。

做 法：①蘿蔔去皮，切成角形；牛肉切薄片，用醃料醃二十分鐘。

②燒熱一湯匙油，爆香薑片後放入蘿蔔略炒，將芡味料放入煮滾，加蓋，用慢火將蘿蔔煮約二十分鐘。

③牛肉片泡油瀝乾，放入蘿蔔內煮約一分鐘，放入葱段攪勻。

④生粉用一湯匙水調勻，慢慢注入牛肉內煮稠即成。

心 得：蘿蔔能利五臟、消痰及化積滯，與牛肉同煮，具有補脾健胃、強壯筋骨、利水散血等功能。

Beef and Turnip Stew

Ingredients: 400g of turnip, 160g of beef, 2 slices of ginger, 2 pieces of spring onion (cut into sections), 2 tablespoons of cornflour.

Marinade: 1 teaspoon of salt, 1 tablespoon of cornflour, 2 tablespoons of water, 1 tablespoon of oil.

Sauce: 1 cup of stock, 1 teaspoon of salt, 1 teaspoon of sugar, sesame oil and pepper.

Method:

1. Peel the turnip and cut it into triangular pieces. Cut the beef into thin slices and add to the marinade.
2. Heat one tablespoon of oil, and add the ginger slices then the turnips. Mix well. Pour in the sauce and bring it to the boil. Cover and simmer until the turnips are tender (about 20 minutes).
3. Slightly deep fry the beef slices until the colour changes, drain and add them to the turnip. Cook for about one minute. Add the spring onion.
4. Dissolve the cornflour with one tablespoon of water. Gradually add it to the beef. When the sauce thickens, serve hot.

Practical tips:

Turnip is good for the internal organs, it clears the throat and relieves indigestion over eating. Turnips with beef can strengthen the stomach, spleen and bones, clears the blood and rids the body of excess water.

Nourishing dishes for all seasons

Chapter 3

Results.

In this section, we discuss the thermodynamics and optimization of the molecular recognition of ssDNA sequences by molecular beacons. By definition, molecular recognition is the specific formation of stable complexes between two biomolecules (a ligand and a receptor for example). In the particular case of the genotyping field, DNA probes are used to read DNA sequence in targets, by forming stable duplexes with ssDNA target. When probe and target are only partially complementary, the formed duplex will contain mismatches whose energy cost destabilizes the duplex. This energy penalty is the basis for the specific molecular recognition of single-stranded target oligonucleotide by DNA probes.

We will show that molecular beacons, thanks to their hairpin loop structural constraints, have a tunable and optimizable specificity for their target. The applications lie in the design of better DNA probes of genetic information (for *in vivo* RNA expression monitoring, DNA microarrays...) [120]. The second part of our thermodynamic analysis of the molecular beacon will lead us to propose a theoretical algorithm to optimize the sensitivity of the molecular beacon [121]: a simple tuning of the strength of the beacon's stem will make the hybridization of the DNA probe onto its target marginal in thermodynamic stability such as the probe becomes exquisitely sensitive to the presence of a mismatch.

In a second part, we will extend the concept of the molecular beacon to the design of more complex DNA probes. Two logical gates (OR and AND) have been constructed and tested to process at the molecular level two separate oligonucleotide targets. We will then discuss the possibility of building large logical networks to process DNA information with DNA molecules. Thermodynamic and Kinetic limitations of these logical gates will also be presented.

In the third part, we study the conformational fluctuations of the most elementary DNA secondary structures: the hairpin loop. By combining Fluorescence Correlation Spectroscopy measurement and bulk fluorescence measurement, we measure opening and closing timescale of hairpin structures. A systematic study of the closing timescale of the hairpin leads us to analyze the length, salt and sequence dependence, in the flexibility of ssDNA [122]. The stacking in polyadenosine is shown to rigidify a DNA loop to the point where the hairpin loop can not form, when a similar loop with polythymidine would be flexible enough to close [123].

Finally, in the fourth part, we study the dynamics of breathing fluctuation in dsDNA by fluorescence correlation spectroscopy. We show that, for large AT domain, the distribution of lifetimes for the open state is well described with standard nearest-neighbor model. For shorter AT domain, unusual long-range effects are unraveled and comprehended as alternative destabilizing structures of dsDNA. The relevance of this study to understand the specificity of the initiation of transcription in prokaryotes will be introduced to be further discussed in Chapter 4.

3.1 Thermodynamics and Optimization of molecular recognition.

DNA Oligonucleotide probes are commonly used to identify complementary strands in the presence of unrelated nucleic acids in genotyping application. However, their specificity is limited when they are used to discriminate targets that differ from one another by only a single nucleotide. As the length of the probe is increased, the free energy penalty resulting from a mismatched base pair in the probetarget duplex represents a smaller and smaller fraction of the total free energy of binding [124]. Consequently, for oligonucleotides that are sufficiently long to distinguish a particular sequence in the presence of unrelated DNA the size of a genome, a mismatched base pair will have only a marginal effect on the stability of the duplex. This limitation often necessitates the use of proteins or chemical agents to assist in the identification of mismatched base pairs [125, 126, 127].

In this section, we show that molecular beacons readily distinguish targets that differ by only a single nucleotide, and that they are significantly more specific than conventional oligonucleotide probes of equivalent length. We demonstrate that their enhanced specificity is due to their ability to form a stem-and-loop structure. Utilizing a thermodynamic approach, we construct a phase diagram that illustrates how molecular beacons behave at different temperatures in the presence of perfectly complementary targets and in the presence of targets that contain a mismatched nucleotide. This analysis reveals that enhanced specificity is a general feature of structurally constrained probes.

A simple thermodynamics analysis is then carried out (§ 3.1.2) to define the specificity of molecular beacons, and propose an "algorithm" to optimize its sequence discrimination by tuning the strength of the stem.

3.1.1 Thermodynamics of molecular beacons as DNA probes.

Characteristics of molecular beacons. The principal feature that distinguishes molecular beacons from conventional nucleic acid probes is the presence of a hairpin stem. The complementary arms are so close to one another in the stem and loop configuration that direct coupling takes place between the fluorophore and the quencher [54]. Consequently, nonhybridized molecular beacons are quenched, while hybridized molecular beacons fluoresce, eliminating the need to isolate probe–target hybrids from nonhybridized probes to determine the number of probes that are bound to targets. The most intriguing feature of molecular beacons is that they are considerably more specific than the corresponding linear probes. Targets that differ from one another by as little as a single nucleotide substitution can readily be distinguished [54, 128].

To understand the basis of this enhanced specificity, we carried out a series of thermodynamic measurements that compared the behavior of molecular beacons in the presence of perfectly complementary target oligonucleotides to their behavior in the presence of targets whose sequence created a single mismatched base pair in the probe–target duplex. The experiments were designed to take advantage of the change in fluorescence intensity that occurs upon thermal dissociation of the probe–target duplexes [43, 44, 45].

Design of the probes. A molecular beacon was synthesized that possessed a 15–nucleotide loop and 5–nucleotide arms. The nucleotide sequence of the loop consisted of an 11–adenosine homooligomer flanked by two cytidines at each end (Figure 1 on page 2):



This homogeneous probe sequence was chosen to reduce the probability of secondary structures forming within the loop and to avoid any context dependence in studies on the effect of mismatch

position on the stability of probe–target duplexes. The cytidines at the ends of the loop sequence served to minimize the sliding of the hybridized target oligonucleotide along the probe sequence. The molecular beacons were labeled with fluorescein and dabcyI. When an excess of targets was added to a solution of the molecular beacons, fluorescence increased 50-fold.

Thermal denaturation profiles of molecular beacons in equilibrium with their targets.

We first monitored the fluorescence of a solution of molecular beacons incubated in the absence of targets at different temperatures. The results (Figure 3.1A, trace a) demonstrate that at lower temperatures the molecular beacons are in a closed state, the fluorophore and the quencher are held in close proximity to each other by the hairpin stem, and the molecular beacons fluoresce only weakly. However, at high temperatures the helical order of the stem gives way to a random-coil configuration, separating the fluorophore from the quencher, restoring fluorescence. This transition occurred at 54°C.

This experiment was repeated in the presence of a 6-fold excess of single-stranded DNA targets that were perfectly complementary to the loop sequence in the molecular beacon.

The results (Figure 3.1A, trace b) show that the molecular beacons fluoresce brightly at low temperatures, but as the temperature is slowly raised, fluorescence diminishes significantly, followed by an increase in fluorescence at the highest temperatures. Figure 3.1B summarizes the phase transitions that occur. At low temperatures, probe–target duplexes form spontaneously. Since the hairpin stem is less stable than the probe–target helix, the stem unwinds to allow the probe to bind to its target (Figure 3.1B, phase 1). Thus, molecular beacons in duplexes are open and fluorescent, whereas nonhybridized molecular beacons are closed and nonfluorescent. As the temperature is raised, the probe–target duplexes become destabilized, the molecular beacons are released, returning to their closed conformation, and fluorescence decreases (Figure 3.1B, phase 2). For the molecular beacons used in our experiments, the transition from phase 1_P (duplexes with perfectly complementary targets) to phase 2 occurred at 42°C. As the temperature is raised further, the closed molecular beacons melt into fluorescent random coils (Figure 3.1B, phase 3). This transition occurred at the same temperature (54°C) as the transition seen when the molecular beacons were incubated in the absence of targets. The temperature–fluorescence profile of the probe–target mixture was fitted to Equation 2.3 (page 58), which was derived from this two-phase-transition model. The resulting curve, plotted in Figure 3.1A, closely correlates with the experimental data, validating the model.

This experiment was also repeated in the presence of a 6-fold excess of single-stranded DNA targets that were complementary to the probe sequence except for one nucleotide (mismatched target). Similar changes in fluorescence were observed as the temperature was raised (Figure 3.1A, trace c). However, the transition from phase 1_M (duplexes with mismatched targets) to phase 2 occurred at a significantly lower temperature, 28°C, than the transition from phase 1_P to phase 2, which occurred at 42°C. This result reflects the lower stability of probe–target duplexes that contain a mismatched base pair. The difference between the melting temperature of perfectly complementary probe–target duplexes and the melting temperature of mismatched probe–target duplexes was 14°C. Within this temperature range, perfectly complementary targets form duplexes that lead to fluorescence, while mismatched targets do not form duplexes and no fluorescence occurs. The range of temperatures within which perfectly complementary targets form hybrids but mismatched targets do not form hybrids is significantly wider than the corresponding range observed for conventional probes [129, 130, 131].

Determination of thermodynamic parameters. To understand why molecular beacons are able to distinguish closely related target sequences with higher specificity than linear probes can, we measured the thermodynamic parameters that describe the phase transitions that occur when the temperature of a mixture of molecular beacons and oligonucleotide targets is slowly increased. The increase in fluorescence that accompanies dissociation of the hairpin stems (phase 2 to phase 3) in a 50nM solution of molecular beacons was fitted with Equation 2.4 (page 58) to determine the enthalpy

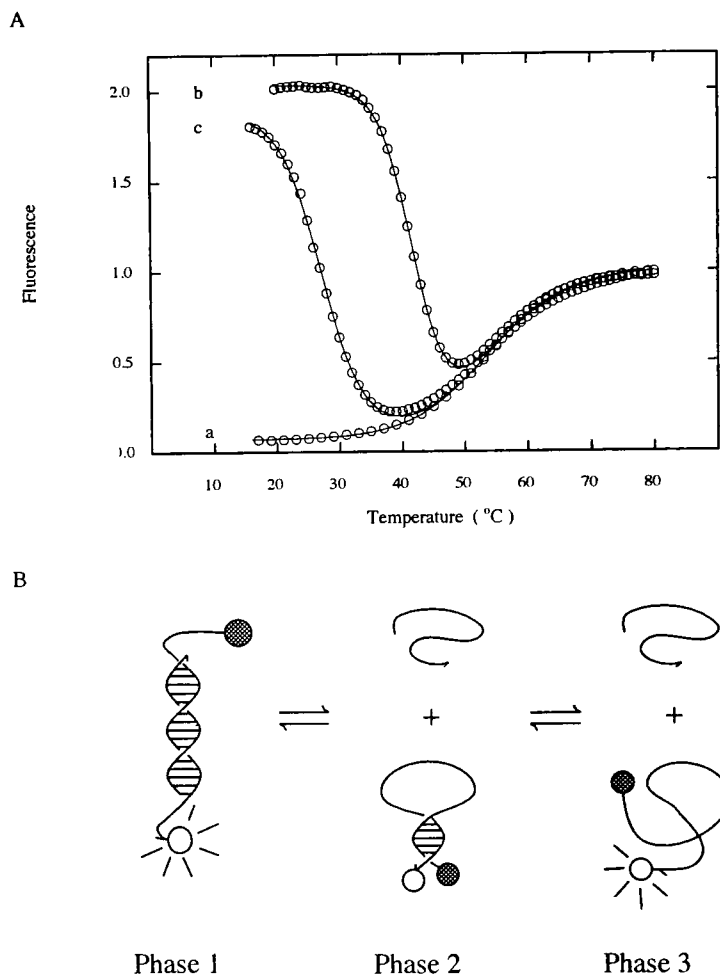


Figure 3.1: Phase transitions in solutions containing molecular beacons. (A) Thermal denaturation profiles of solutions containing molecular beacons: trace a, in the absence of targets; trace b, in the presence of a 6-fold excess of perfectly complementary targets; and trace c, in the presence of a 6-fold excess of mismatched targets. The data were fitted with equation 2.3 (page 58) and plotted as continuous lines. (B) Schematic representation of the phases. As the temperature is raised, the fluorescent probe–target duplex (phase 1) dissociates into a nonfluorescent molecular beacon in a closed conformation and a randomly coiled target oligonucleotide (phase 2). As the temperature is raised even higher, the hairpin stem of the molecular beacon unravels into a fluorescent randomly coiled oligonucleotide (phase 3).

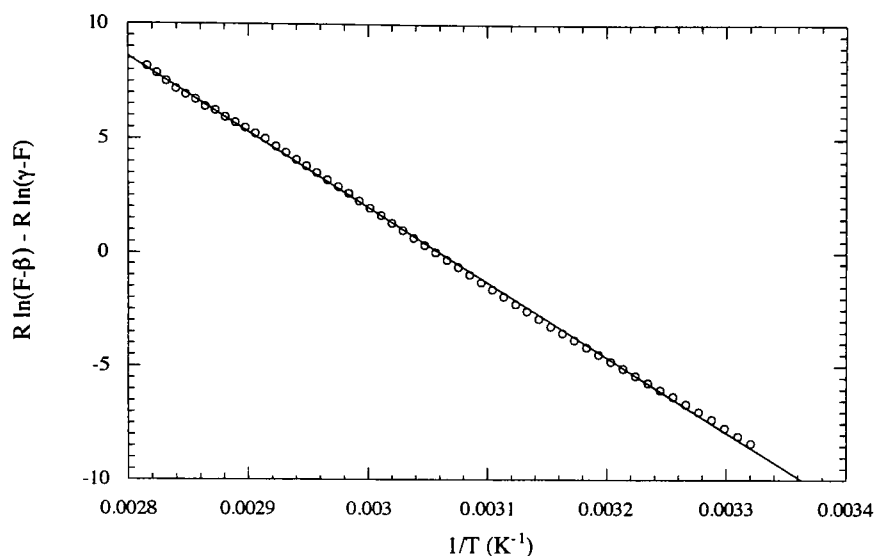


Figure 3.2: Determination of thermodynamic parameters for the melting of a molecular beacon alone. The increase in fluorescence that accompanies the melting of the molecular beacons hairpin stem was used to determine the thermodynamic parameters that describe this transition (see § 2.2 and equation 2.5 for further description). The slope of the fitted line is equal to the negative value of the enthalpy ($-\Delta H_{2 \rightarrow 3}^0$) and the intercept is equal to the entropy ($\Delta S_{2 \rightarrow 3}^0$).

and entropy describing this transition (Figure 3.2). We found that $\Delta H_{2 \rightarrow 3}^0 = 34 \pm 1 \text{ kcal.mol}^{-1}$ and $\Delta S_{2 \rightarrow 3}^0 = 104 \pm 3 \text{ cal.mol}^{-1}.\text{K}^{-1}$ (1 cal = 4.18 J). The experiment was repeated with concentrations of molecular beacons ranging from 5 nM up to 100 nM, and the resulting thermodynamics parameters were shown to be unaffected by the molecular beacon concentration. The transition (phase 2 to phase 3) is indeed intramolecular.

To determine the enthalpy and entropy describing the dissociation of probe–target duplexes (phase 1 to phase 2), we measured the melting temperature of the duplex as a function of the concentration of target oligonucleotides. We obtained temperature–fluorescence profiles from a series of solutions containing 50 nM molecular beacons and different excess concentrations of target oligonucleotides (between 300 nM and 300 μM – cf Figure 3.3). The higher the concentration of the target oligonucleotides, the greater the melting temperature of the probe–target duplexes. Since at higher target concentrations it is difficult to distinguish the decrease in fluorescence due to the dissociation of the probe–target duplexes from the increase in fluorescence due to the unwinding of the molecular beacon stems, we utilized Equation 2.3 (page 58), which describes the entire temperature–fluorescence profile, to accurately determine the melting temperature of the duplexes. The resulting melting temperatures were then used to calculate the thermodynamic parameters that describe the transition from phase 1 to phase 2 (Figure 3.3). For the dissociation of probe–target duplexes formed from perfectly complementary target oligonucleotides, we found that $\Delta H_{1P \rightarrow 2}^0 = 84 \pm 3 \text{ kcal.mol}^{-1}$ and $\Delta S_{1P \rightarrow 2}^0 = 237 \pm 9 \text{ cal.mol}^{-1}.\text{K}^{-1}$; and for the dissociation of probe–target duplexes formed from mismatched target oligonucleotides, we found that $\Delta H_{1M \rightarrow 2}^0 = 65 \pm 3 \text{ kcal.mol}^{-1}$ and $\Delta S_{1M \rightarrow 2}^0 = 185 \pm 10 \text{ cal.mol}^{-1}.\text{K}^{-1}$;

Phase diagram for solutions containing molecular beacons and targets. We constructed a diagram showing free energy for the three phases of molecular beacons in equilibrium with their targets (Figure 3.4). The free energy of each phase was plotted as a function of temperature ($\Delta G(\theta) = \Delta H - \theta \Delta S$). Phase 3 was chosen as the reference state ($\Delta G = 0$), because the

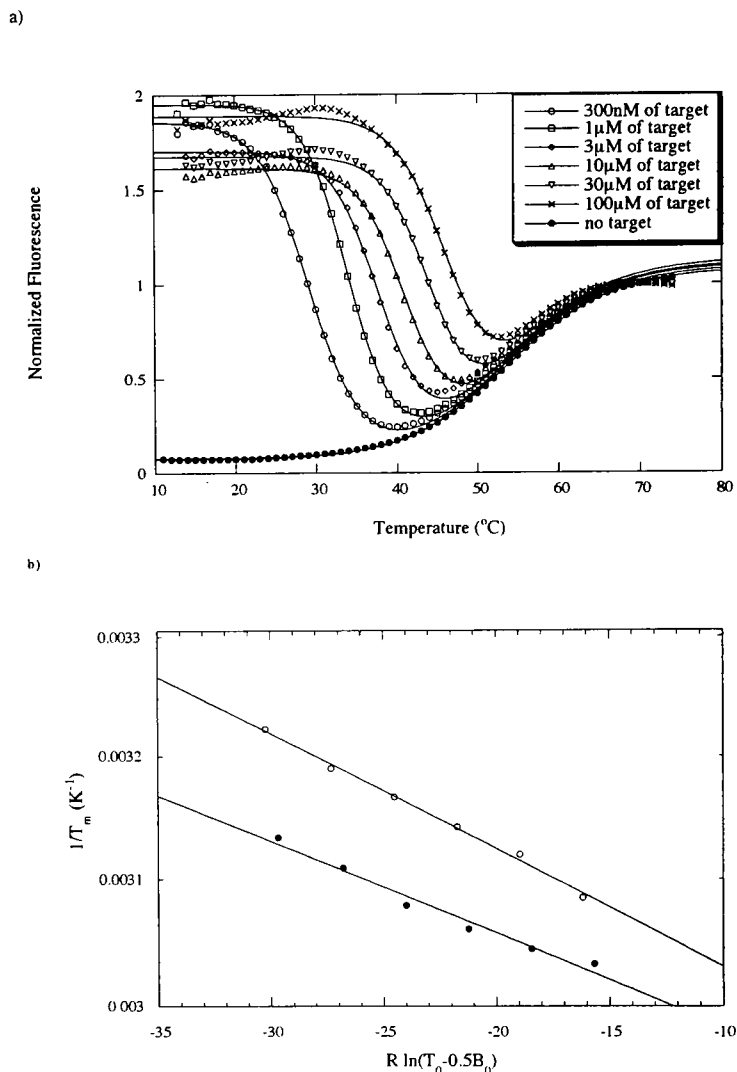


Figure 3.3: Determination of thermodynamic parameters for the melting of a duplex beacon-target by Van' Hoff analysis. A) Melting profiles for a solution of 50nM of molecular beacon in equilibrium with different concentrations of mismatched targets. The fits from Equation 2.3 (page 58) are used to determine the melting temperature of the probe-target duplex for each concentration of target. B) The increase in the melting temperature of the probe-target duplex that results from increasing the concentration of target oligonucleotides was used to determine the thermodynamic parameters that describe the dissociation of probe-target duplexes. Separate determinations were carried out with perfectly complementary duplexes (•) and with mismatched duplexes (○). The slope of each fitted line is equal to the negative value of the enthalpy ($-\Delta H_{1 \rightarrow 2}^0$) and the intercept is equal to the entropy ($-\Delta S_{1 \rightarrow 2}^0$). In this graph, the independent variable, $R \ln(T_0 - 0.5B_0)$ (where T_0 and B_0 are the total concentrations of molecular beacon and of target), is plotted on the abscissa and the dependent variable, $1/T_m$, is plotted on the ordinate, to illustrate the similarities in the manner in which the enthalpies and entropies were determined for the dissociation of probe-target duplexes and for the dissociation of hairpin stems (Figure 3.2).

molecular beacons and the target oligonucleotides are present as disorganized random coils. Since $\Delta G_2 = \Delta G_3 - \Delta G_{2 \rightarrow 3}$, the free energy of phase 2 is calculated to be $\Delta G_2 = -\Delta H_{2 \rightarrow 3}^0 + \theta \Delta S_{2 \rightarrow 3}^0$; and since $\Delta G_1 = \Delta G_3 - \Delta G_{2 \rightarrow 3} - \Delta G_{1 \rightarrow 2}$, the free energy of phase 1 is calculated to be $\Delta G_1 = -(\Delta H_{2 \rightarrow 3}^0 + \Delta H_{1 \rightarrow 2}^0) + \theta(\Delta S_{2 \rightarrow 3}^0 + \Delta S_{1 \rightarrow 2}^0 - R \ln T_0)$.

The predominant phase at each temperature is the phase with the lowest free energy. From the diagram it can be seen that in a system of molecular beacons and perfectly complementary targets at temperatures below 42°C, phase 1_P predominates; at temperatures between 42°C and 54°C, phase 2 predominates; and at temperatures above 54°C, phase 3 predominates (this progression is indicated in the figure by a line containing long dashes). By contrast, in a solution containing mismatched targets, the transition from phase 1_M to phase 2 (indicated by a line containing short dashes) occurs at a significantly lower temperature ($\theta_M = 28^\circ\text{C}$) than the corresponding transition from phase 1_P to phase 2 ($\theta_P = 42^\circ\text{C}$). Consequently, in the range between these temperatures ($\Delta\theta = \theta_P - \theta_M = 14^\circ\text{C}$), perfectly complementary targets can be distinguished from mismatched targets.

The phase diagram illustrates why molecular beacons are able to distinguish closely related target sequences with higher specificity than the corresponding linear probes can. If we had used a probe with the same loop sequence, but with arm sequences that cannot form a hairpin structure (unstructured probe), then phase 2 could not occur, and the only phase change would have been a direct transition from phase 1 to phase 3. If unstructured probes had been used to form perfectly complementary probe–target duplexes, the phase diagram predicts that the transition from phase 1_P to phase 3 would have occurred at 45°C (at the intersection of the line describing the free energy of phase 1_P with the line describing the free energy of phase 3); and if unstructured probes had been used to form mismatched probe–target duplexes, the transition from phase 1_M to phase 3 would have occurred at 37°C. The difference between these two transitions is relatively small ($\Delta\theta' = 8^\circ\text{C}$). A comparison of $\Delta\theta$ with $\Delta\theta'$ shows that probes possessing a stem-and-loop structure distinguish mismatches over a wider range of temperatures than do unstructured probes.

Although the phase diagram was constructed from thermodynamic parameters determined for a particular molecular beacon, the relative placement of the lines that describe the free energy of each phase is common to all molecular beacons, and more generally, to all probes that can form a structured intermediate when they dissociate from their target. Since a phase's enthalpy is dependent on the number of base pairs formed, the enthalpies of each phase will always occur in the following order: $0 = \Delta H_3^0 > \Delta H_2^0 > \Delta H_{1_M}^0 > \Delta H_{1_P}^0$. Similarly, entropy is dependent on the number of configurations that can be assumed by the probe in each phase: $0 = \Delta S_3^0 > \Delta S_2^0 > \Delta S_{1_M}^0 > \Delta S_{1_P}^0$. In particular, the number of configurations that can be assumed by an unstructured probe (ΔS_3) is substantially higher than the number of configurations that can be assumed by a structured probe (ΔS_2). Consequently, the phase diagram will look similar for all structured probes. It is thus generally true that if a probe can form a structure after dissociating from its target, phase 2 will occur, and the line describing phase 2 will intersect the lines describing phase 1_M and 1_P over a wider range of temperatures than the corresponding intersections with the line describing phase 3.

Since molecular beacons are constrained polymers, upon dissociation (or formation) of probe–target duplexes they undergo a greater reorganization than do unstructured probes, leading to a higher sensitivity to the presence of a mismatch. Similarly, all conformationally constrained probes should display higher specificity in molecular recognition than unconstrained probes.

Effect of mismatch position and mismatch identity. We assumed an all-or-none mechanism for the formation of probe–target hybrids. However, when a molecular beacon hybridizes to its target its stem must unwind, so it is not clear whether its hybridization is a one-step or a multistep process. The all-or-none mechanism predicts that the position of a mismatch should have little effect on the stability of the probe–target duplex, as long as the mismatch is flanked on either side by at least one base pair [19].

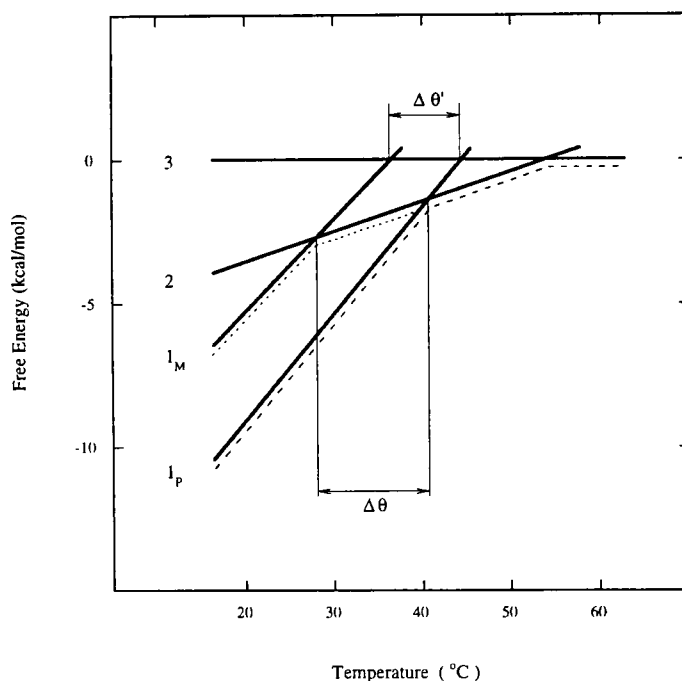


Figure 3.4: Free energy of the three phases of a solution of molecular beacons in equilibrium with target oligonucleotides. Although this phase diagram was calculated for a solution containing a particular molecular beacon at 50 nM and its target at 300 nM, the relative positions of the phase lines are generally the same for all molecular beacons under all conditions. The equation of each line is $\Delta G = \Delta H - TS$, where $\Delta G_3 = 0$; $\Delta G_2 = 0.104T - 34 \text{ kcal.mol}^{-1}$; $\Delta G_{1_P} = 0.371T - 118 \text{ kcal.mol}^{-1}$; and $\Delta G_{1_M} = 0.320T - 99 \text{ kcal.mol}^{-1}$. The difference between the melting temperatures of perfectly matched duplexes (phase 1_P) and mismatched duplexes (phase 1_M) is greater if the probe can form a structure after dissociation ($\Delta\theta = 14 \text{ }^\circ\text{C}$) than it is if the probe cannot form a structure ($\Delta\theta' = 8 \text{ }^\circ\text{C}$).

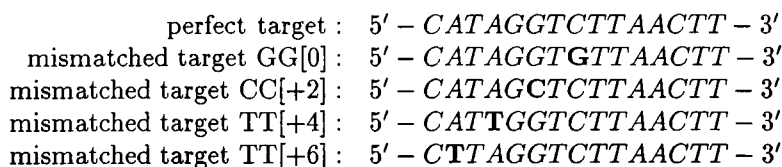
To check whether the all-or-none mechanism of hybridization applies to molecular beacons, we prepared a series of target oligonucleotides whose sequences ($5' - GGT_{5+n}GT_{5-n}GG - 3'$, where $-4 \leq n \leq 4$) created a G-A mismatch (instead of a complementary T•A base pair) at different positions within the duplex formed by the molecular beacon shown in Figure 1 (page 2). Sequence-dependent effects on hybrid stability were minimized because, irrespective of location, the mismatches were always embedded within neighboring T•A base pairs. For each target, we obtained a thermal denaturation profile at six different target concentrations (between 300 nM and 300 μ M) and used the results to determine the enthalpy and entropy for the dissociation of its probe-target duplex. We then calculated the melting temperature of the duplex in the presence of 300 nM target. The results show that the position of the mismatched base pair has a negligible effect on melting temperature (Table 3.1.1), implying that molecular beacons form hybrids through an all-or-none mechanism.

To determine whether the identity of the mismatched base pair influences the stability of the hybrid, we also prepared targets containing different nucleotides at the position corresponding to the central nucleotide of the probe sequence ($5' - GGT_5NT_5GG - 3'$, where $N = A, C, \text{ or } G$), and we studied the dissociation of their probe-target hybrids in the same manner. The results show that the identity of the mismatched basepair does not have an appreciable effect on the melting temperature of the hybrid (Table 3.1.1). The enthalpy of the probe-target duplex with a A-A mismatch is 12% bigger than the enthalpy if the duplexes with G-A and C-A mismatch (8kcal/mol over 65kcal/mol). The entropy of the duplex with a A-A mismatch is also 14% bigger than the enthalpy if the duplexes with G-A and C-A mismatch ($26\text{cal.mol}^{-1}.\text{K}^{-1}$ over $187\text{cal.mol}^{-1}.\text{K}^{-1}$). Consequently, the melting temperatures of the duplex varies only by 1.5% (5K over 300K). Thus, there is a surprising enthalpy-entropy compensation in the thermodynamics of the mismatch duplex. Since the melting temperatures of the mismatched duplexes are similar, and since they are all significantly lower than the melting temperature of the perfectly complementary duplex, molecular beacons can easily detect the presence of a mismatch, yet they are insensitive to variations in identity and position. These features greatly simplify the detection of point mutations.

We tested the effect of the position and identity of a mismatch with a molecular beacon with a non-homogeneous probe sequence. The sequence of the molecular beacon is:



The sequences of the targets we probed with this inhomogeneous molecular beacon are (we count the position of the mismatch from the center of the target):



We recorded the bulk melting curve of a solution of 5nMol of this inhomogeneous molecular beacon in equilibrium with its target in the following buffer: 100mM TrisCl, 1mM MgCl₂ pH8.0 (Figure 3.5). We show that the melting temperature of the probe-target duplex is very much dependent on the position and identity of the mismatch basepair.

$$\begin{aligned} T_m &= 51.3^\circ\text{C} \text{ (perfect target),} \\ T_m &= 32.9^\circ\text{C} \text{ (mismatched target GG[0]),} \\ T_m &= 21.9^\circ\text{C} \text{ (mismatched target CC[+2]),} \\ T_m &= 38.9^\circ\text{C} \text{ (mismatched target TT[+4]),} \\ T_m &= 43.3^\circ\text{C} \text{ (mismatched target TT[+6]).} \end{aligned}$$

Table 3.1: Enthalpies, entropies, and melting temperatures for the dissociation of a perfectly complementary beacon–target duplex (first entry), beacon–target duplexes containing different mismatched base pairs at the same position (next three entries), and beacon–target duplexes containing the same mismatched base pair at different positions (last nine entries)

Mismatch	Position (n)	ΔH^0 , kcal.mol ⁻¹	ΔS^0 , cal.mol ⁻¹ .K ⁻¹	θ_m , °C
T•A	0	84	237	42
A-A	0	69	201	27
C-A	0	61	175	23
G-A	0	65	185	28
G-A	-4	75	218	30
G-A	-3	68	194	29
G-A	-2	78	228	30
G-A	-1	72	208	29
G-A	0	65	185	28
G-A	+1	74	213	29
G-A	+2	74	213	29
G-A	+3	74	212	29
G-A	+4	77	221	31

Standard enthalpies and standard entropies are shown for solutions containing 50 nM molecular beacons and 1 M target oligonucleotides in the presence of 100 mM KCl and 1 mM MgCl₂. Melting temperatures are shown for solutions containing 50 nM molecular beacons and 300 nM target oligonucleotides. The errors associated with ΔH^0 and ΔS^0 averaged $\pm 6\%$.

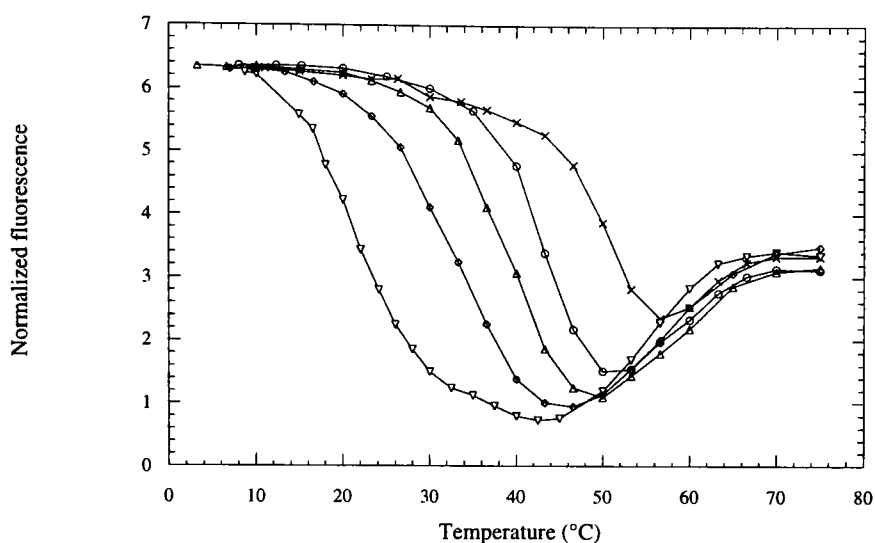


Figure 3.5: Melting curves of a solution of 10nMol of inhomogeneous molecular beacon with 10 μ Mol of different targets: (x) perfect target; (o) mismatched target TT(+6); (Δ) mismatched target TT(+4); (∇) mismatched target CC(+2); (\diamond) mismatched target GG(0) (buffer: 100mM TrisCl, 1mM MgCl₂ pH8.0).

One could have predicted that the closer to the edge of the probe–target duplex the mismatch is positioned, the lower the structural perturbation on the B-DNA structure, the higher the melting temperature. In fact, the melting temperature is not dependent on the position of the mismatch, as there is no systematic trend for the duplex melting temperature: $T_m(\text{TT}[+6]) > T_m(\text{CC}[+2])$ but $T_m(\text{GG}[0]) > T_m(\text{CC}[+2])$. Thus, if one changes the flanking bases of a mismatch, the entropy–enthalpy compensation unraveled with the homogeneous is lost, and the melting temperature of the probe–target duplex becomes very much sequence–dependent, and also, through this sequence–dependence, position–dependent.

Fourth phase of the molecular beacon. At low temperatures (below 20°C), for high salt concentration of salt ($[Na^+] = 1M$), a fourth phase of the molecular beacon in equilibrium with its target has been unraveled: we will call it Phase 4. Its signature is a decrease of the collected fluorescence (Figure 3.6A&B) at low temperature. Thus, this phase has often been considered as a potential source of artefact in DNA hybridization experiments, as it decreases unexpectedly the fluorescence of the formed duplex target+beacon at low temperature.

Two hypothetical structures can explain the decrease of fluorescence at low temperature:

- #1 aggregation of the duplex beacon–target (*via* hybridization of the complementary stems of two separate molecular beacons),
- #2 reclosing of the molecular beacon while still hybridized to its target.

To decipher the correct hypothesis, understand the predominance of Phase 4, we investigated its thermodynamics. Melting curves for a solution of molecular beacon and complementary targets were recorded as described previously (§ 3.1.1), but with extended temperature range (3 to 80°C), and special care to avoid hysteresis (in general, reaction kinetics slow down as the temperature of the solution is decreased: it takes up to 1 hr to reach thermal equilibrium –as monitored by the convergence of the fluorescence to a stationary value– at 5°C). The melting curves were recorded for a molecular beacon of sequence 5′ – GCGAG CCA₁₁CC CTCGC – 3′ (Figure on page 2), and

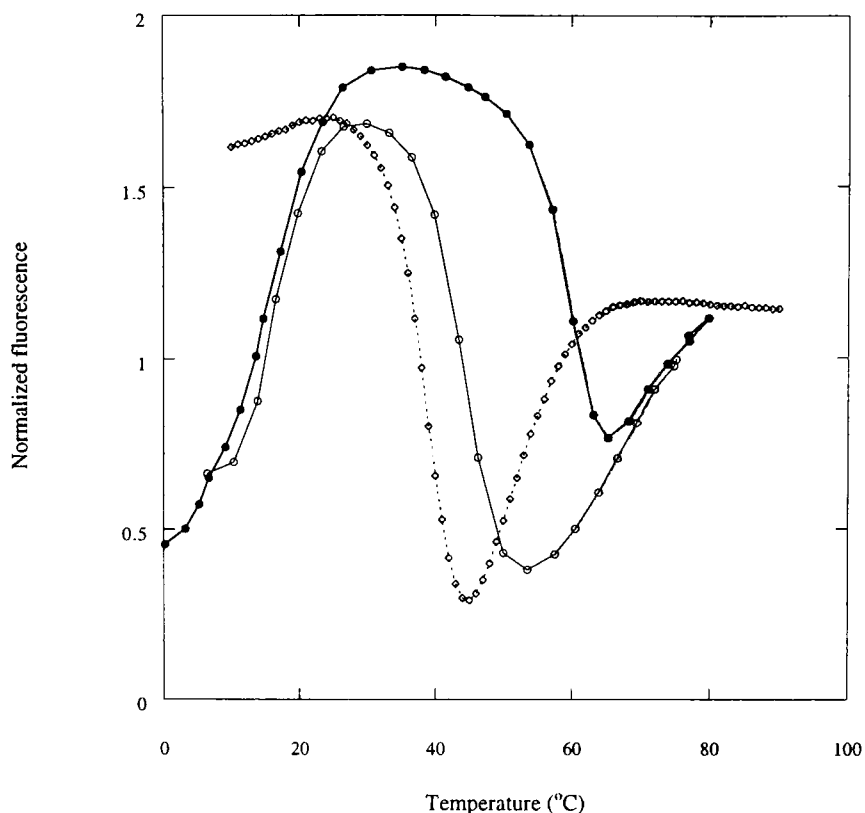


Figure 3.6: Low-temperature transition in the melting curve of a solution of molecular beacon and target in high salt: (●) perfect target (buffer: 1M NaCl, 10mM Cacodylic acid, 1mM EDTA, pH8.0); (○) mismatched target (buffer: 1M NaCl, 10mM Cacodylic acid, 1mM EDTA, pH8.0); (◊) mismatched target with flanking bases (buffer: 100mM KCl, 10mM TrisCl, 1mM Mg^{2+} pH8.0)

targets of sequence 5' – GGT₁₁GG – 3' (perfect match) or 5' – GGT₅GT₅GG – 3' (mismatched target). The concentration of beacon is 5nMol, the concentration of target is 100nMol in a 1M NaCl buffer (Figure 3.6A&B).

We find that the low-temperature phase melts at 14°C, whatever the concentration of targets, and whatever the sequence of target oligonucleotide (match or mismatch). Thus the transition from Phase 4 to Phase 1 must be an intramolecular rearrangement in the molecular beacon. The hypothesis #1 (aggregation of the beacon) in the interpretation of the Phase 4 formation must then be rejected. The best explanation to the decrease of fluorescence at low temperature must then be the reclosing of the beacon stem, while still duplexed with the ssDNA target.

S. Tyagi carried out the same kind of measurement for molecular beacons with longer stems and/or longer duplex sequence (*unpublished results*): the longer beacon stems induce a higher melting temperature for the transition Phase 4 → Phase 1, *i.e.* the Phase 4 must be stabilized by basepairing in the stem. Furthermore, the longer loops (*i.e.* the longer duplex dsDNA) induce a higher melting temperature for the transition Phase 4 → Phase 1, *i.e.* the Phase 4 is more stable with longer dsDNA duplex on the loop. These findings prove that the stability of the stem basepairs is a crucial parameter but not the only thermodynamic determinant for the formation of Phase 4.

Let us present a semi-quantitative analysis of the Phase 4 → Phase 1 transition as a surprising re-closing of the beacon stem. We can evaluate the free energy cost of bending a duplex dsDNA of $L=15$ bases to close it into a loop:

$$\Delta G_{bending} = k_B T L_p \int_0^L \frac{1}{R^2} ds = 4\pi^2 k_B T \frac{L_p}{L}. \quad (3.1)$$

where $L_p = 150$ bases is the persistence length of double stranded DNA, R the radius of curvature of the loop, and s the curvilinear coordinate along the loop. Although not very appropriate to this small-scale molecular problem, this continuous approximation of the elasticity of dsDNA is the best estimate we can provide for the elastic cost of closing the dsDNA loop. Here,

$$\Delta G_{bending} \sim +400 k_B T. \quad (3.2)$$

To compensate the energetic loss of bending the dsDNA, the reformation of the basepairs in the stem is considered. We measured (from the melting curve of molecular beacons in 1M NaCl, 10mM Cacodylate, 1mM EDTA pH8.0):

$$\begin{aligned} \Delta H_{3 \rightarrow 2}^\circ &= -38.8 \pm 7 \text{ kcal/mol} \\ \Delta S_{3 \rightarrow 2}^\circ &= -114 \pm 13 \text{ cal/K/mol.} \end{aligned}$$

These values correspond to the free energy of melting the complementary strands of the beacon stem, under the flexible constraints of the loop. We will see in the following sections (§ 3.3) that the free-energy barrier associated with the closing of a molecular beacon is quite negligible: most of the free energy associated with the melting transition of a molecular beacon corresponds to the melting of the stem itself. Thus, the free energy gained in closing the five basepairs of the stem at 5°C can be evaluated:

$$\Delta G_{closing} \sim -38800 + 114 * 278 = -7100 \text{ kcal/mol} \sim -13 k_B T$$

This free energy gain does not make up for the elastic loss associated with the bending of the 15-mer loop-target duplex (Equation 3.2). However, some of the basepairs in the duplex can have unzipped to soften the rigidity of the double-strand of DNA and accommodate the formation of the stem basepairs (this flexibility associated with the unzipping of basepairs in the loop-target is hard to evaluate). Thus, the fourth phase must be a delicate balance between the stabilizing energy of the stem basepairs and the elastic cost of looping the probe-target duplex.

As pointed out earlier, this fourth phase of a solution of the molecular beacon in equilibrium with its target is a potential source of artefacts, “unexpectedly” decreasing the fluorescence. Three practical tricks have been implemented to destabilize this fourth phase, and avoid the decrease of fluorescence at low temperature:

- Carry out the experiment in low salt buffers. High salt concentrations are needed to screen the charges of the target-beacon duplex, and reduce the bending energy of dsDNA (*via* reduction of the persistence length in equation 3.1). At lower salt concentration, the duplex domain is too rigid to accommodate the reformation of the stem basepairs. A good hybridization buffer contains 100mM KCl, 10mM $MgCl_2$ and 10mM TrisCl pH8.0: the Mg^{2+} ions stabilize the duplex DNA without screening efficiently the phosphate charges, and the dsDNA remains rigid. On Figure 3.6C, we present the melting curve of the same molecular beacon in solution with a mismatched target in this hybridization buffer: note that, down to 10°C, the fluorescence of the probe-target duplex stays constant.
- Introduce 2 (or more) additional bases flanking the target sequence: the spatial constraints on stem basepairing associated with these unpaired bases in the duplex seem to destabilize the fourth phase (observations by S. Tyagi, *unpublished*).

- Design the molecular beacon such as half of the stem sequence is complementary to the target itself: when the beacon–target duplex forms, half of the bases in the stem are involved in dsDNA pairing, such as there is no energetic gain in reclosing the stem.

Those 3 adjustments in the design of the molecular beacon annihilate the formation of the fourth phase and reduce the possible fluorescence artefacts.

Conclusion on the thermodynamics of the molecular beacon as a DNA probe.

In this section, we experimentally measured the thermodynamics parameters associated with the melting transitions in a solution of molecular beacon in equilibrium with its target. The effect of one mismatch in the target sequence has been shown to destabilize the duplex target–beacon. A thermodynamic diagram has been drawn to unravel the higher sequence–specificity of molecular beacons compared to classical denatured probe.

A rule of thumb can summarize this result. The molecular beacon is conformationally constrained by the five basepairs in its stem. When it hybridizes to its target to form a 15–basepair complex, it disrupts the five basepairs in the stem, hence a net energy gain of 10 basepairs. It is thus easier for a molecular beacon to discriminate one mismatch among the target sequence as it will read it among 10 basepairs, compared to a classical denatured probe which would read the same mismatch among 15 basepairs.

The position of a mismatch along the target sequence was shown not to influence the thermodynamics of the duplex formation, as predicted by standard Nearest–Neighbor model (§ 1.2 on page 12). More surprising is that, for given flanking bases, the identity of the mismatch induces a perfectly compensated enthalpic and entropic destabilization of the duplex, such as the melting temperature of the duplex beacon–target is only weakly dependent on the identity of the mismatch. Thus, the molecular beacon as a DNA probe can not read the position or identity of a mismatch in a probed target.

Finally, we discussed the possibility of the reclosing of the molecular beacon hybridized to its target (in high salt buffer, below 10°C), and proposed a buffer solution to avoid this troublesome decrease of fluorescence of a molecular beacon hybridized to its target.

3.1.2 Optimal sensitivity of a DNA probe.

Molecular beacons introduce new rules in the field of DNA hybridization assays. The loop sequence is complementary to the target DNA sequence. The stem is designed to act as a “counterweight”, in a “double–weight” like method [132]. Upon hybridization to its complementary target, the beacon stem unzips (the rigidity of double–stranded DNA is such that it forbids the closing of the beacon). The optimum mismatch sensitivity is achieved in the limit that target hybridization energy approaches stem self–hybridization energy. With the increase of mismatch sensitivity, there is a corresponding decrease in the speed of recognition.

This discussion of the optimal sensitivity of a DNA probe is organized as follows. First, we will define a sequence discrimination sensitivity to quantify the mismatch sensitivity and compare a molecular beacon to a classical DNA probe without stem–loop structure. Optimization of this sensitivity can be achieved by tuning the stem strength for the beacon, in an experiment where the temperature is fixed (in vivo case). When temperature is a

free parameter (in vitro experiment), the stem can be optimized under the constraint of an operational time to reach equilibrium.

Second we will address the target concentration problem: for a given concentration of perfect target, there is always a concentration of mismatched target leading to the same hybridization level of the probe. This concentration ratio will be estimated in the context of molecular beacons.

Third, we will compute the concentration sensitivity to assess how quantitative a molecular beacon hybridization assay is. This discussion is particularly relevant to DNA chips applications.

The hybridization problem. The standard probing of long DNA strands consists in choosing a small length of the gene sequence (typically 20 bases), synthesizing the complementary strand, and monitoring its level of hybridization to the gene. The buffer, the temperature and the DNA strand concentration in the sample are imposed by the experiment. In such a case, optimization is achieved by choosing the sequence (length and content) such that the designed probe hybridizes only to the targeted gene.

In Figure 3.7A, we represent the energy levels of the different objects (hybridized probe to a specific target, hybridized probe to a mismatched target and unhybridized probe).

The introduction of the molecular beacon as another kind of DNA probe enables an optimization the signal/noise ratio of DNA hybridization [1]. The molecular beacon has a stem-loop structure (Figure 3.7B). The loop is (by design) complementary to the target sequence. The stem length and base content is tuned to optimize the sensitivity of DNA probing.

In Figure 3.7B, the energy levels of a molecular beacon hybridizing to a perfectly matched target or a one-base mismatched target are represented. The energies of the duplex are the same as in Figure 1A, but the free energy of the closed unhybridized probe is decreased by the stem basepairing, when compared to the unhybridized classical probe. **Note that this free energy has been positioned between the two duplex energies, by tuning the stem.**

It is important to remark that, whatever the probe, the ratio of the concentration of matched duplex to mismatched is fully determined by the difference in their free energy. The key point is that, for the sake of measurement, the absolute probability of hybridization is essential (it is always difficult to measure minute differences in an experiment). Let us give an example (see § 3.1.2 for further): in the case of Figure 3.7, with a classical probe, the probability of hybridization is $(1 \text{ minus } 10^{-7})$ for the matched target, and $(1 \text{ minus } 2 \cdot 10^{-5})$ for the mismatched one. With the beacon, at the same temperature, those values are 0.7 and 0.01. Obviously, it is easy to discriminate matched and mismatched target with a beacon, whereas, for the classical probe, the noise level would have to be zero.

The signal-to-noise ratio is the main problem for the experimentalist. It is defined as the ratio between the signal level of the hybridized to unhybridized probe. The signal/noise ratio depends on the type of detection. Any hybridization measurement relies on the change of state of the DNA probe upon hybridization. Traditionally, UV absorption, gel mobility assay, or fluorescence labeling are used to distinguish the hybridized probes from the free probes. In the case of fluorescence probes where fluorescence is switched on upon hybridization, the background is determined by stray fluorescence (in the case of the molecular beacon, the fluorophore might be imperfectly switched off by the quencher).

Let us now derive the sensitivity of a DNA probe with a finite noise level.

Computation of the sensitivity of a DNA probe for sequence discrimination. We define the sequence discrimination sensitivity S of a DNA probe as the ratio of the hybridization levels, monitored by fluorescence, of the probe onto a matched target and a mismatched one. Buffer conditions, target and probe concentrations are fixed by the experiment.

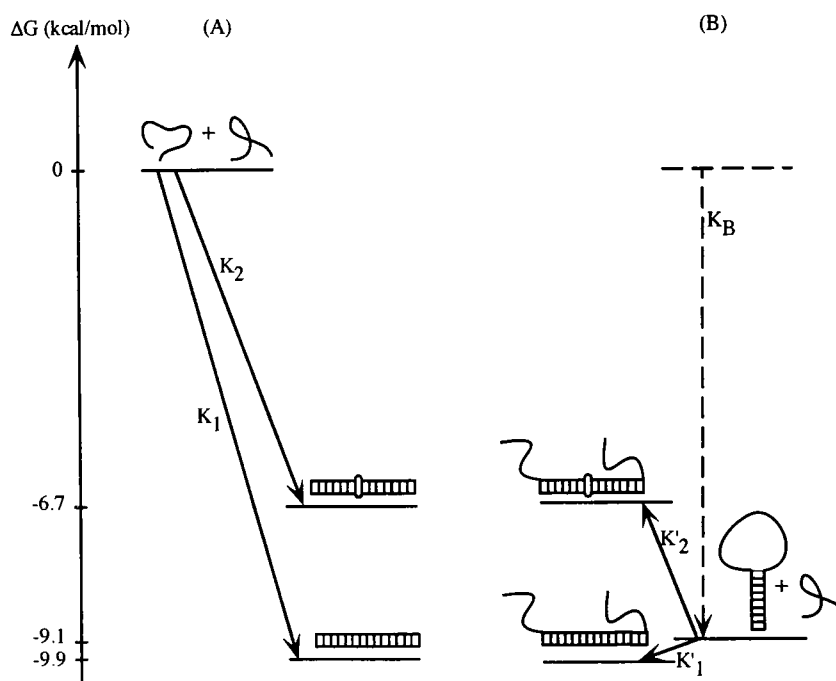
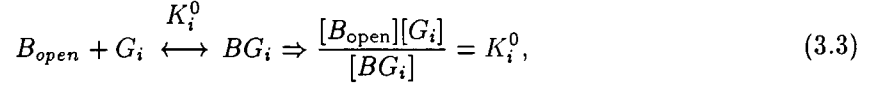


Figure 3.7: Free energy diagram of hybridization: A) for a classical denatured probe of 15 bases; the two arrows correspond to the hybridization with a perfect matched target or a one-base-mismatch target. In both cases, there is complete hybridization as the free energies are negative; B) for an hypothetical molecular beacon with a loop of 15 bases, and a stem of 13 basepairs; in that case, the beacon hybridizes to its perfect matched target ($K'_1 \sim 1.5$) but not to the one-base-mismatched target ($K'_2 \sim 0.007$). The free energy parameters used to draw this diagram were measured in 1M NaCl, 10mM Cacodylic, 1mM EDTA pH8.0, with the method presented in § 3.1

Given all the tabulated thermodynamics parameters of DNA hybridization [27, 19, 25], we can completely derive the expression for the sequence discrimination sensitivity for a molecular beacon.

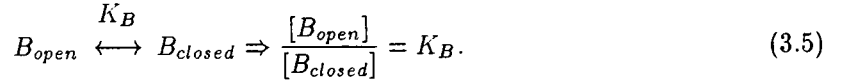
The melting of the duplex of the beacon (B) and the target gene (G) into two coiled DNA ($B_{open} + G$) corresponds to the equilibrium:



where K_i^0 is the equilibrium constant for the hybridization reaction ($i = 1$ for a perfectly match, $i = 2$ for a mismatch). We assume that the targets are in excess such as $[G_i]$ is constant (for most applications, this is a realistic assumption, and it simplifies the calculus). Thus,

$$\frac{[B_{open}]}{[BG_i]} = \frac{K_i^0}{[G_i]} = K_i. \quad (3.4)$$

One must take also into account the possible melting of the stem-loop structure:



Combining the two equilibria 3.4 and 3.5, with the conservation of probe ($[BG_i] + [B_{closed}] + [B_{open}] = B_0 = \text{constant}$), we get the three concentrations:

$$\begin{aligned} [BG_i] &\propto (1 + K_i + K_i/K_B)^{-1}, \\ [B_{open}] &\propto K_i(1 + K_i + K_i/K_B)^{-1}, \\ [B_{closed}] &\propto K_i/K_B(1 + K_i + K_i/K_B)^{-1}. \end{aligned}$$

The signal-to-noise ratio of the molecular beacon is defined by δ , the ratio of fluorescence of the open molecular beacon with the fluorescence of the closed one. The fluorescence f_i of the solution of a molecular beacon in equilibrium with its target G_i is then proportional to:

$$f_i \propto \delta([BG_i] + [B_{open}]) + [B_{closed}]. \quad (3.6)$$

as the beacon can open upon hybridization or by thermal melting of its stem-loop structure.

The sequence discrimination sensitivity (S) of the beacon is then:

$$S = \frac{f_1}{f_2} = \frac{\delta(1 + K_1) + K_1/K_B}{\delta(1 + K_2) + K_2/K_B} * \frac{1 + K_2 + K_2/K_B}{1 + K_1 + K_1/K_B}. \quad (3.7)$$

Where K_1 , K_2 , and K_B are the equilibrium constant for the melting of a perfect matched duplex (1), a mismatched duplex (2) and the stem-loop of the beacon (B).

Understanding the sequence discrimination sensitivity. Let us show first what the sequence discrimination sensitivity represents, and why a classical denatured probe has a low sensitivity. The target is the 15-base-long polymer of sequence 5'–GGTTTTTTTTTTTGG–3'. The mismatched target is 5'–GGTTTTTATTTTGG–3'. The probe sequence is 5'–CCAAAAAAAAAAACC–3'. The free energies (as a function of the temperature T) for the binding of the probe (at 1nmol/l) onto a target (at 50nmol/l) in a standard buffer (100mM KCl, 10mM cacodylate pH8.0, 0.5 mM ethylenediaminetetraacetate) were measured with the same method as in [120] (data not shown):

$$\Delta G_1(T) = 134 - 0.400T \text{ kcal/mol for the matched target,}$$

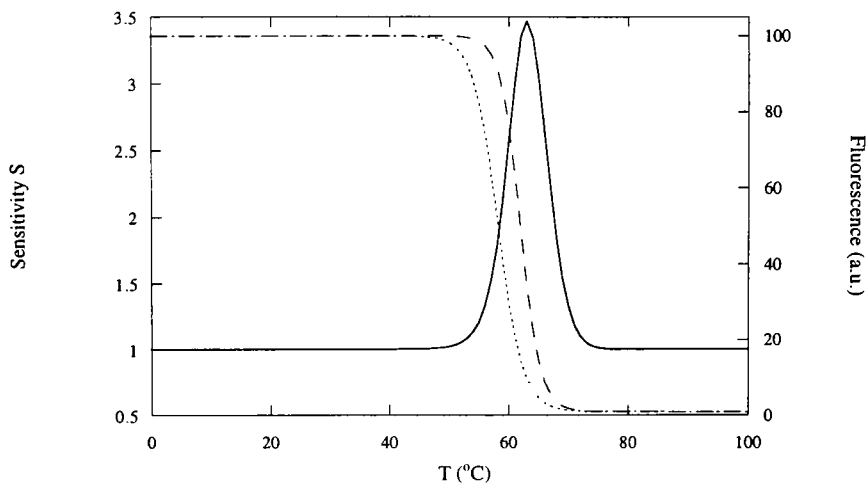


Figure 3.8: Melting curves for matched (---) and mismatched duplexes (- - -) (simulated fluorescence) in the case of a denatured probe. The sequence discrimination sensitivity (—) is the ratio of the two melting curves. A maximal sensitivity of 3.5 is obtained at 63°C.

$$\Delta G_2(T) = 105 - 0.317T \text{ kcal/mol for the mismatched target.}$$

Assuming that the hybridized classical probe has a fluorescence $\delta = 100$ times bigger than the fluorescence of the free one, one can recalculate the melting curves (Figure 3.8) of the duplexes. At low temperatures, the probe is completely hybridized onto its target (matched or mismatched), thus the fluorescence is equal to 100. At high temperature, the probe unbinds from its target, the fluorescence is given by the free probe, and is equal to 1.

On the same figure, we plot the sequence discrimination sensitivity \mathcal{S} , as a function of the temperature, using a similar derivation as for equation (3.7); it is the ratio of the two melting curves:

$$\mathcal{S} = \frac{f_1}{f_2} = \frac{\delta + K_1}{\delta + K_2} * \frac{1 + K_2}{1 + K_1} \quad (3.8)$$

At low and high temperatures, the fluorescence is the same for matched and mismatched targets. Thus, the sensitivity \mathcal{S} is equal to 1. Obviously, the best sensitivity is obtained between the two melting temperatures. The discrimination of the two targets cannot be achieved with a sensitivity better than 3.5 (maximal sensitivity at 63 °C). For an *in vivo* study of discrimination at 37°C, this classical probe does not work at all ($\mathcal{S} = 1$).

Optimization of the sequence discrimination sensitivity. At a given temperature T_0 , K_1 and K_2 are completely determined. One way to maximize \mathcal{S} is to introduce a molecular beacon. A proper design of its stem-loop structure is achieved by tuning K_B , such as (from equation(3.7):

$$\left(\frac{\partial \mathcal{S}}{\partial K_B} \right)_{T_0} = 0, \quad (3.9)$$

$$\text{thus } K_B(T_0) = \sqrt{\frac{1}{\delta} \frac{K_1(T_0)}{1 + K_1(T_0)} \frac{K_2(T_0)}{1 + K_2(T_0)}}, \quad (3.10)$$

$$\max\{\mathcal{S}(T_0)\} = \frac{2K_1K_2 + (\delta K_2(1 + K_1) + K_1(1 + K_2))\sqrt{\frac{1}{\delta} \frac{K_1}{1 + K_1} \frac{K_2}{1 + K_2}}}{2K_1K_2 + (\delta K_1(1 + K_2) + K_2(1 + K_1))\sqrt{\frac{1}{\delta} \frac{K_1}{1 + K_1} \frac{K_2}{1 + K_2}}}. \quad (3.11)$$

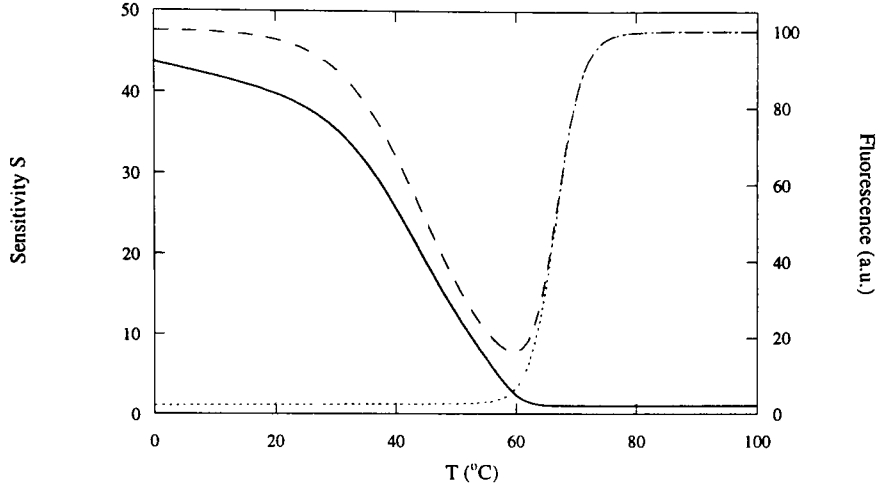


Figure 3.9: Melting curves for matched (—) and mismatched (---) duplexes (simulated fluorescence) in the case of a molecular beacon. At low temperature, the beacon hybridizes with its perfect target: the fluorescence is high. The duplex melts at 45 °C into a closed beacon and a coiled target: the fluorescence goes down. Ultimately, above 67 °C, the molecular beacon melts and opens its stem-loop structure: the fluorescence goes up again. In the case of a mismatched target, the duplex never forms between 0 and 60 °C; the fluorescence goes up above 67 °C as the beacon melts. The sequence discrimination sensitivity (—) is the ratio of the two melting curves.

Let us make a simple comment in the case of the targets given in section 3.1.2. As $K_1(T_0) \ll 1$ and $K_2(T_0) \ll 1$, eq. 3.10 becomes:

$$K_B(T_0) = \exp\left(-\frac{\Delta G_B(T_0)}{RT_0}\right) = \sqrt{\frac{K_1(T_0)K_2(T_0)}{\delta}} \quad (3.12)$$

In terms of free energy,

$$\Delta G_B(T_0) = \frac{1}{2} (\Delta G_1(T_0) + \Delta G_2(T_0) + RT \ln(\delta)) : \quad (3.13)$$

this explains the position of the free energy of the beacon stem-loop structure in Figure 3.7; it is the average of the free energies of the duplexes shifted by the signal-to-noise ratio of the beacon.

The design of the beacon implies the tuning of ΔH_B and ΔS_B (melting enthalpy and entropy of the stem-loop structure) such as:

$$K_B(T_0) = \exp\left(-\frac{\Delta H_B - T_0 \Delta S_B}{RT_0}\right) = \sqrt{\frac{1}{\delta} \frac{K_1(T_0)}{1 + K_1(T_0)} \frac{K_2(T_0)}{1 + K_2(T_0)}}. \quad (3.14)$$

To fully determine ΔH_B and ΔS_B , one must impose another constraint, like the melting temperature T_m of the beacon alone:

$$\Delta H_B = T_m \Delta S_B. \quad (3.15)$$

Tuning ΔH_B and ΔS_B can be achieved by changing the length and the base composition of the stem.

Figure 3.9 shows the hybridization of a molecular beacon optimized to work at 37°C. The optimized free energy for the melting of its stem-loop structure is:

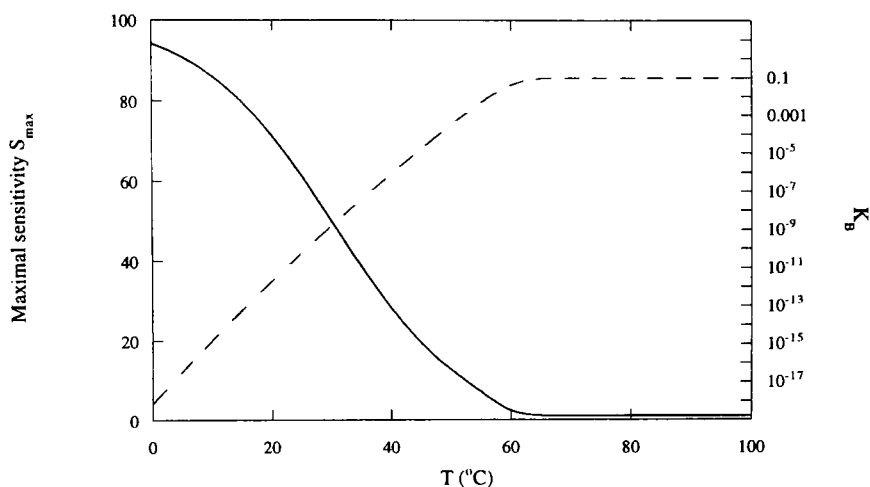


Figure 3.10: Maximal sequence discrimination sensitivity \mathcal{S} (—), and corresponding equilibrium constant K_B (---) for the melting of the beacon stem-loop structure.

$$\Delta G_B = 106 - 0.312T \text{ kcal/mol} = -RT \ln(K_B) \quad (3.16)$$

A possible sequence for the molecular beacon is



the stem is made out of 13 basepairs with a 46%GC content (this sequence was designed with the help of Zuker's folding algorithm [29, 26]). With this optimized beacon, it is striking that the mismatched duplex does not form at all, and the beacon remains closed as the melting temperature of its stem-loop structure (67°C) is much higher. The sensitivity is equal to 29 at 37°C, to be compared to 1.0 for the classical probe (Figure 3.9).

Let us finally show how $\max\{\mathcal{S}(T)\}$ and $K_B(T)$ vary with temperature T (Figure 3.10). Obviously, above 60°C (the melting temperature of the classical duplex into two coiled DNA's in Figure 3.8), no DNA probe can hybridize, thus the sensitivity is one. As the temperature is decreased, the molecular beacon sensitivity increases, and saturates at the signal-to-noise ratio δ of the probe itself at low temperatures. At the same time, the equilibrium constant of the beacon structure, achieving the optimal sensitivity, increases as the temperature goes down.

Nevertheless, the penalty for better sequence discrimination sensitivity lies in longer hybridization time. In the kinetic timescales involved for DNA hybridization, the timescale for the opening of the beacon is the limiting factor [122]. The timescale of closing is more or less constant, therefore a decrease in K_B implies an increase in the opening time for the beacon. In the example given above (at 37°C) the opening timescale would be of the order of 100 s. The interplay between sensitivity and measurement timescale must be chosen according to the needs of the application.

Competition for coexisting targets. A more realistic experiment consists in probing the genetic information inside a pool of diverse sequences. Discriminating two closely related sequences is critical in the study of single-nucleotide polymorphism: a lethal point mutation might appear in a sample, hardly detectable in a pool of unmutated sequences. Molecular beacons have been proven successful to monitor such critical mutation in medical applications (TB drug resistance, HIV resistance [133]).

We discuss here the limit of resolution of a point-mutated sequence in a pool of unmutated sequence, by beacon hybridization. The strategy in this case is to target the mutated sequence:

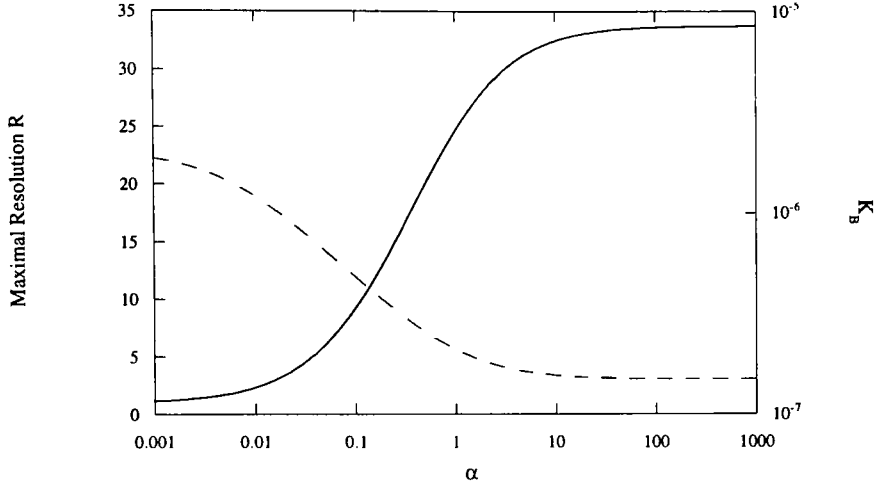


Figure 3.11: Resolution (—) of a match and a mismatch target, competing for hybridization to an optimized beacon and corresponding K_B (---)(equilibrium constant for the melting of the stem-loop structure). α is the population ratio of match to mismatched targets.

the DNA probe is designed to be complementary to the mutated sequence (target sequence). The total quantity of target DNA is fixed ($50 \mu\text{mol/l}$), as well as the ratio of target DNA's to non target DNA's (denoted α). The resolution \mathcal{R} of the DNA probe is the ratio of $f(\alpha)$ (fluorescence of a mixed solution of matched and mismatched sequences in a ratio α), with $f(0)$ (fluorescence of the solution containing only mismatch): $\mathcal{R} = f(\alpha)/f(0)$. Experimentally, one can define a threshold of resolution: the perfect target is said to be detected when $\mathcal{R} > 2.5$.

A calculation (similar to the one presented in section 3.1.2) is carried out for matched G_1 competing with a mismatched G_2 for hybridization onto a beacon (B).

The fluorescence of the solution is:

$$f \propto \delta([BG_1] + [BG_2] + [B_{open}]) + [B_{closed}] \quad (3.17)$$

$$f \propto \frac{\delta(K_1^{-1} + K_2^{-1} + 1) + K_B^{-1}}{1 + K_1^{-1} + K_2^{-1} + K_B^{-1}} \quad (3.18)$$

The total concentration of target DNA's is: $[G] = [G_1] + [G_2]$. The ratio of target DNA to mismatched DNA is $\alpha = [G_1]/[G_2]$. We note $K'_1 = [G]/K_1^0$, $K'_2 = [G]/K_2^0$ and $K'_B = 1/K_B$, then

$$f(\alpha) \propto \frac{\delta(\frac{\alpha}{1+\alpha}K'_1 + \frac{1}{1+\alpha}K'_2 + 1) + K'_B}{1 + \frac{\alpha}{1+\alpha}K'_1 + \frac{1}{1+\alpha}K'_2 + K'_B} \quad (3.19)$$

$$\mathcal{R} = \frac{\delta(\frac{\alpha}{1+\alpha}K'_1 + \frac{1}{1+\alpha}K'_2 + 1) + K'_B}{1 + \frac{\alpha}{1+\alpha}K'_1 + \frac{1}{1+\alpha}K'_2 + K'_B} * \frac{1 + K'_2 + K'_B}{\delta(1 + K'_2) + K'_B} \quad (3.20)$$

As in section 3.1.2, \mathcal{R} can be optimized by tuning K_B (i.e. tuning the strength of the stem-loop structure).

$$\left(\frac{\partial \mathcal{R}}{\partial K'_B} \right)_{T_0} = 0 \Rightarrow K_B^{-1}(T_0) = \sqrt{\delta(1 + K'_2) \left(1 + \frac{\alpha}{1+\alpha}K'_1 + \frac{1}{1+\alpha}K'_2 \right)} \quad (3.21)$$

This implies that, for each value of α , one must tune the strength of the stem-loop structure to reach this optimal $K_B(T_0)$. Once $K_B(T_0)$ is fixed, \mathcal{R} is determined (cf eq. 3.20). In Figure 3.1.2, we plot the optimal \mathcal{R} as a function of α . For $\alpha > 0.01$, $\mathcal{R} > 2.5$. Thus, if 1 out of 100 strands of DNA has a mutated sequence, it can be detected.

In conclusion, we have introduced and computed a sensitivity parameter to assess the discrimination power of a DNA probe between two closely related target oligonucleotides. We designed molecular beacons to optimize this sensitivity parameter. At a given temperature (*in vivo*), the optimal beacon is such that its stem-loop structure has a comparable energy to the duplex to be formed: any mismatch shifts the stability towards non-hybridization. When the temperature is a free parameter (*in vitro*), the sensitivity of the molecular beacon increases as the temperature decreases, at the expense of slower kinetics of hybridization: the optimal beacon is then a compromise of better signal-to-noise with an acceptable timescale. In the more realistic case of competing hybridization, the beacon is again more sensitive by three orders of magnitude than a classical probe, and allows the detection of a sparse single-point mutation (a point mutation on a 15-mer DNA is detectable even if only 1 out of 100 DNA's' has been mutated).

This discussion on the optimization of DNA probe can be generalized towards any molecular recognition: we argue that, contrary to classical arguments in molecular recognition, the free energy difference between match and mismatched complexes is not a sufficient definition of the specificity of the probe. Rather, one must take into account the signal/noise ratio of the recognition scheme to evaluate accurately the specificity of the probe. The problem of specificity in molecular interactions is thus a convolution of structural information (thermodynamics) as well as signal detection (signal/noise of the detection).

3.2 DNA logic gates.

We just presented the classical application of molecular beacons for genotyping application, whereby the conformational transition of the molecular beacon upon hybridization to its complementary target switches on a fluorescence signal. This constitutes the most elementary logical operation one can envision:

IF "target is present" THEN fluorescence is switched ON.
IF "target is not present" THEN fluorescence is switched OFF.

In fact, our original interest in the molecular beacons is related to their function as molecular logic operators. The introduction of DNA computation by Adleman [134] suggested applications to DNA hybridization into solving complex algorithm (see § A). The nature of DNA self-assembly is appealing as a way to solve algorithms in a massively-parallel computation. In the context of DNA computation, molecular beacons can be considered as elementary DNA molecular switches, and good building blocks for complex logical operations.

In this section, we describe the design of DNA logic gates with the blueprint of the molecular beacon. Two logic gates have been implemented (OR and AND). A concatenation scheme is also proposed to carry the result of one logic operation for further molecular computation.

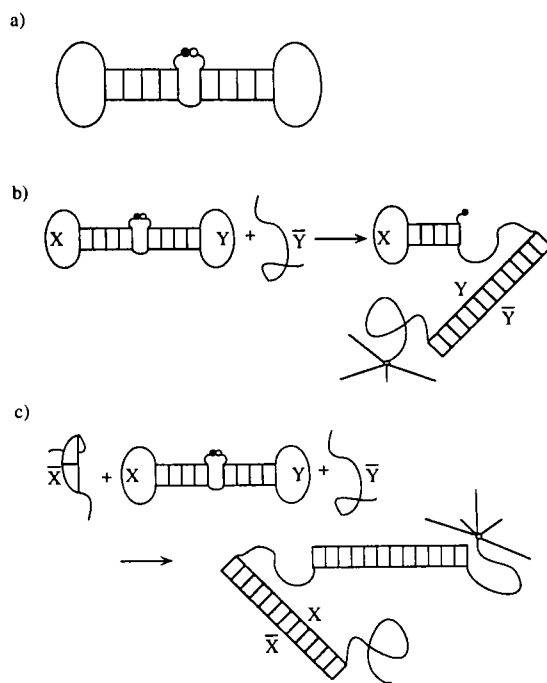


Figure 3.12: Sketch of the structure and function of the OR gate.

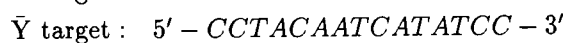
3.2.1 OR gate.

The structure of the OR gate is presented in Figure 3.12. Its sequence is:



with a fluorescein coupled at the 5' end, and a DABCYL at the 3' end.

The synthesis of this gate is straightforward, as it involves the same chemistry as the molecular beacon. We tested the functioning of the Or gate with the two possible targets of sequence:



The experiment is carried on in the hybridization buffer (1M NaCl, 10mM NaCacodylate, 1mM EDTA pH8.0), at room temperature. The background scattering and fluorescence from this buffer

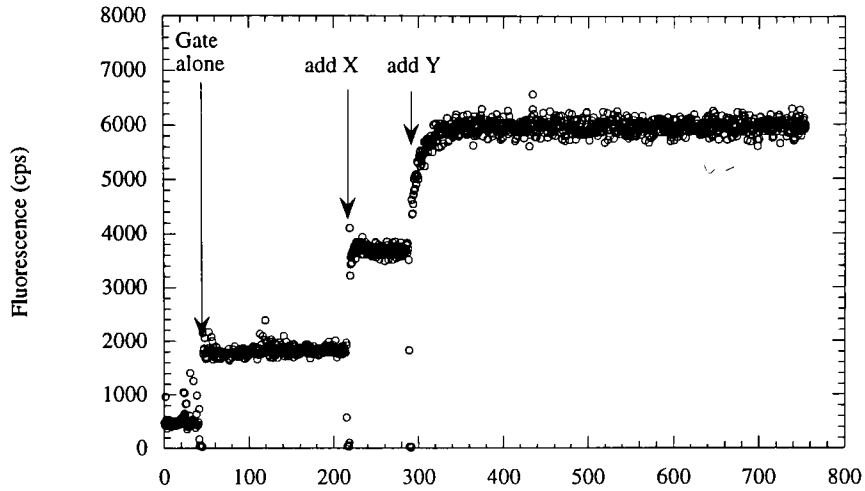


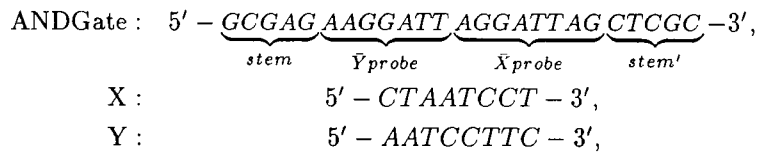
Figure 3.13: Fluorescence signals from the OR gate: for $40 < t < 200$ s, buffer alone; for $40 < t < 220$ s, OR gate alone; at $t=220$ s, add the target \bar{X} ; at $t=300$ s, add the target \bar{Y}

is $f_B=400$ cps. The residual fluorescence of the closed OR gate in these conditions is $f_C=1800$ cps. For a concentration of 6nM of target \bar{X} and a concentration of 1nM of OR gate, we observe the full opening of the beacon ($f_{\bar{X}}=3.8 \cdot 10^3$ cps). Adding 6nM of target \bar{Y} increases a little but not substantially the fluorescence of the gate ($f_{\bar{X}+\bar{Y}}=6.0 \cdot 10^3$ cps) (Figure 3.13). Thus, the presence of one of the two targets is enough to open the OR gate.

The signal/noise ratio of the OR gate is as good (~ 60) as the signal/noise ratio of the molecular beacon from which it is derived. The OR gate is in fact two concatenated molecular beacons: the optimization of its design implies the same thermodynamic arguments that we presented in § 3.1.2. Special care must be taken in the design of the OR Gate to avoid spurious folding of the gate itself: the sequence proposed here was chosen after trial and error, tested with Zuker's folding algorithm [29].

3.2.2 AND gate.

The structure of the AND gate is presented in Figure 3.14. It is simply a molecular beacon whose loop consists of two independent probes, (\bar{X} and \bar{Y}) complementary to the targets (X and Y).



with a 5' fluorescein and a 3' DABCYL on the AND gate, and a 5' phosphorylation on the Y target. The thermodynamics of the hybridization is such that the duplex $\{X+\text{Gate}\}$, or $\{Y+\text{Gate}\}$ is not stable. When the X target and the Y target are presented together, they can form a triplex with one molecular beacon and destabilizes the stem structure: we then get an opening of the beacon conditioned by the simultaneous presence of both targets.

The experiment is carried on in a special buffer (30mM TrisCl, $\text{pH}7.8$, 10mM MgCl_2 , 1mM DTT, $500\mu\text{M}$ ATP), classically used for DNA ligation reaction. The background scattering and fluorescence from this buffer is 5000cps . At 10°C , for a concentration of $3\mu\text{M}$ of target X and a concentration

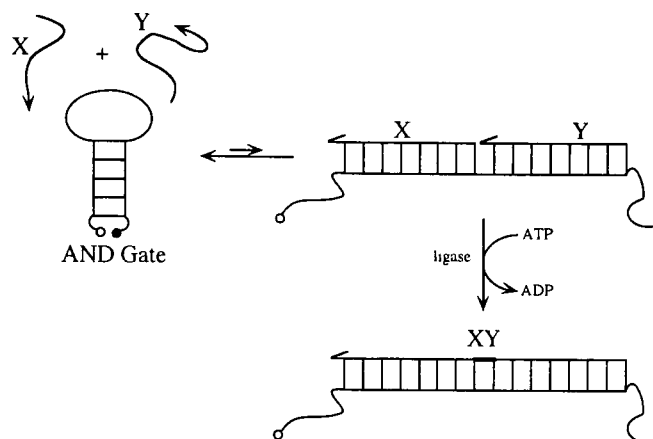


Figure 3.14: Sketch of the structure and function of the AND gate.

of 2nM of AND gate, no noticeable change of fluorescence is measured compared to a baseline level defined by the residual fluorescence of the AND gate in this buffer ($f=8000\text{cps}$). The same baseline level of fluorescence has been measured for a solution of $3\mu\text{M}$ of target Y and a concentration of 2nM of AND gate (Figure 3.15, trace B). Thus, the presence of one of the two targets is not enough to open the AND gate.

For a solution containing $3\mu\text{M}$ of target X AND $3\mu\text{M}$ of target Y, the fluorescence goes up to $f=10500\text{cps}$ (a rise of 183%, taking into account the background fluorescence of 5000cps of the buffer). The presence of both targets induce the opening of the molecular beacon which then acts as an AND gate.

An additional step increases the signal/noise of this gate: we use T4-ligase to catalytically stabilize the formed triplex, by joining the 5'-phosphate and 3'-hydroxyl groups of adjacent nucleotides of the two targets. Hence, it converts the triplex of poor thermodynamic stability into a duplex of higher stability. In Figure 3.15, we show that the addition of 8 units of T4 ligase increases the fluorescence of the solution of AND gate and two targets to reach 15000cps (a rise of 330%, taking into account the background fluorescence of 5000cps of the buffer).

Note that the signal/noise ratio of this AND gate is not optimal (we only recover 10% of the full signal/noise of the initial molecular beacon): we were compelled to perform the reaction at 10°C , to stabilize enough the triplex $\{\text{Gate}+\text{X}+\text{Y}\}$ to get a decent ligation yield.

One nice feature of the AND gate action however is its cooperativity: three molecules must interact together to yield a stable product, thus we could design a AND gate to open up when two similar targets hybridize. This high stoichiometry (higher than the classical bimolecular reaction) has been shown to be crucial to the successful construction of large kinetic networks with computational purpose [135].

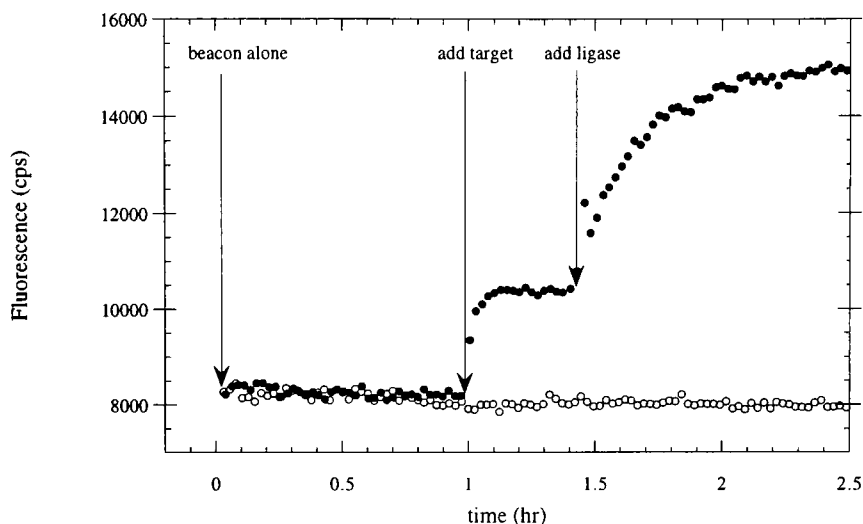


Figure 3.15: Fluorescence signals from an AND gate: (●) AND gate with the two targets X and Y, (○) AND gate with the target X only.

3.2.3 Concatenation of DNA logic gates.

The OR and AND gates are proof that the molecular beacon concept can be applied to perform simple logical operations in bimolecular recognition. Could one perform more complex logical operations by making the output of one molecular operation the input of another molecular operation?

The molecular output of our DNA gates is the unraveling of the stems' sequence (the switching ON of the fluorescence is merely considered as a monitoring device here). Indeed, when the molecular beacon is closed, the stem sequence is not accessible for trans-hybridization as it is involved in a more stable cis-hybridization. If the molecular beacon opens up by thermal fluctuation, the kinetics to reclose on itself is intramolecular thus very fast (around $10\mu\text{s}$ [122]), not leaving enough time for a complementary oligonucleotide to hybridize (one would need a concentration around 10mM of complementary oligonucleotide to efficiently compete with the complementary stem). The induced opening of molecular beacons, AND gates, and OR gates, unravels new single-stranded sequences which are now available for hybridization. By this mean, one can envision to concatenate the positive output of the logical gates to further logical operation.

The major limitation to this scheme is imposed by thermodynamics of DNA hybridization. We have designed the logical gates to process at thermal equilibrium (ligation aside). Thus the computation with these logical gates must be driven by free energy gains. In the case of DNA secondary structure, free energy gains imply that the output structure must be stabilized by more basepairs than the input structure. This imposes tight constraints on the design of the gates: the stem duplex must be shorter than the loop-probing sequence, such as the duplex gate-target is stable enough to unzip the stem. For example, for our OR gate, if the probing sequence is 12 base-long, the stem is four-base long, and the next logical operation will have to perform on only these four bases. This sets hard constraints on the number of logical operation one will be able to perform.

The main escape consists in coupling the logical operation with energy consumption, such as the computation is driven with an energy reservoir. In a way, this is exactly the logic of most chemical biological networks, whereby catalytic reactions are hydrolysing ATP or GTP to advance the chemical operation.

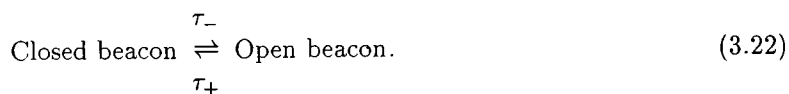
Two elementary logical gates (OR and AND) have been constructed. Their use in complex diagnosis application has been proposed to detect and pro-

cess many sequences at the molecular level and generate a simple YES/NO fluorescence signal. In fact, the idea of designing complex chemical networks out of molecular beacon is very appealing, as, once one finds an interesting kinetic scheme, one finds many by simply changing the encoding DNA sequence. We will discuss more thoroughly the design of possible complex chemical networks out of DNA in the Discussion Chapter.

3.3 Conformational fluctuations of DNA hairpin-loops.

The most elementary secondary structure that ssDNA can form are hairpin-loops: two complementary sequences on the same molecule hybridize to define a stem, the single-stranded section in between is constrained into a closed loop by the last basepair of the stem. Hairpin-loops constitute the “building block” of most secondary structures formed in large nucleic acids, such as tRNA, MDV-1 amplicon, ribosome 16S. Alternative secondary structures (such as bulges, or pseudo-knots) are formed with the same physics as the one driving the formation of hairpin-loops *i.e.* a competition between the stabilization of the structure by basepairs, and its disruption by entropic single-stranded domains. Thermal fluctuations induce conformational fluctuations of such hairpin-loop structures. Their dynamics are fast as they involve intramolecular rearrangements (on the scale of few nanometers).

The molecular beacons are ideal probes to address the issue of the dynamics of hairpin-loop conformational fluctuation for two reasons. First, their excellent fluorescence signal/noise enables very sensitive monitoring of the fluctuations over a wide range of temperature (*e.g.* from 10°C to 50°C). Second, their conformational fluctuations follow *a priori* simple kinetics with only two states:



where τ_- is the opening timescale and τ_+ is a closing timescale (Figure 3.16)

In this section, we will first present a new technique combining Fluorescence Correlation Spectroscopy and Fluorescence quenching of molecular beacons to measure the kinetic parameters of conformational fluctuation of DNA hairpin-loops. We will then present our measurements on opening and closing timescales of a molecular beacon and a thorough understanding of the energy landscape describing the secondary structure of a DNA hairpin-loop (we will use the molecular beacon presented in Figure 3.16 as a benchmark for this study). Finally, we will compare hairpin-loops, in different salt condition, with different loop length, different loop sequence.

3.3.1 Measurement of the opening and closing timescales of a DNA hairpin-loop.

Outline of the Experimental Procedure.

In a fluorescence correlation spectroscopy measurements, the autocorrelation function reflects any process leading to the fluctuations of the emitted fluorescence. For the molecular beacons, the main sources of the fluctuations in fluorescence are the diffusion of molecules in and out of the sampling volume and the opening and closing of its secondary structure. To extract the kinetic timescales of the opening and the closing, three independent measurements needed to be performed:

1. Measurements of the autocorrelation function of the molecular beacons G_{beacon} , which has both the diffusion and the chemical kinetics contributions.

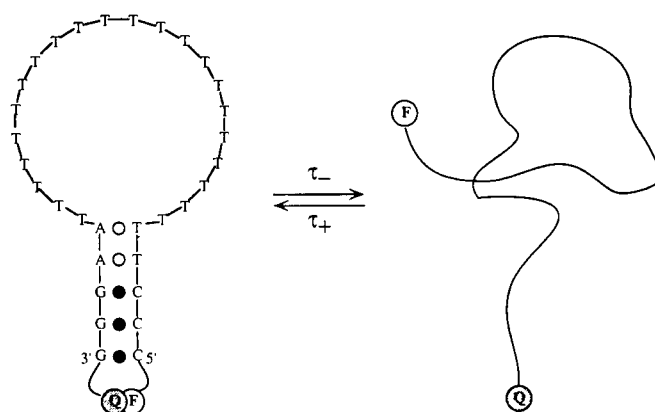


Figure 3.16: Sketch of the conformational thermal fluctuations of a molecular beacon.

2. Measurements of the autocorrelation function G_{control} from a specially designed control sample, for which the correlation function consists of the diffusion contribution only. The ratio $G_{\text{beacon}}(t)/G_{\text{control}}(t)$ (see text below) isolates the chemical kinetics contribution and yields a measure of the sum $\tau_+^{-1} + \tau_-^{-1}$
3. Measurements of the bulk fluorescence give the equilibrium constant $K = \tau_+/\tau_-$ (see § 3.3.1). Thus, τ_+ and τ_- can be determined separately.

Data Treatment of the FCS curves.

The theoretically predicted $G_{\text{beacon}}(t)$ is a product of a diffusion term and a chemical reaction term [109, 86]:

$$G_{\text{beacon}}(t) = \frac{1}{N} \frac{1}{1 + \frac{t}{\tau_D}} (\alpha + \beta e^{-\frac{t}{\tau_{\text{fluc}}}}), \quad (3.23)$$

where N is the average number of molecular beacons in the sampling volume, τ_D is a diffusion time across the illuminated region, α and β are amplitude factors related to the equilibrium constant and the signal to noise of the beacon (see § 2.3), and τ_{fluc} is the chemical reaction timescale:

$$\frac{1}{\tau_{\text{fluc}}} = \frac{1}{\tau_+} + \frac{1}{\tau_-} \quad (3.24)$$

For our experimental situation, τ_D ($\sim 150 \mu\text{s}$) falls in the range of τ_{fluc} (between $5 \mu\text{s}$ and 1ms), which makes a direct fit of experimental data with equation (3.23) unreliable. To extract τ_{fluc} , we designed control samples similar to the beacons with a fluorophore but no quencher. For the control, the autocorrelation function consists of the diffusion contribution only:

$$G_{\text{control}}(t) = \frac{1}{N} \frac{1}{1 + \frac{t}{\tau_D}} \quad (3.25)$$

The ratio $G^*(t) = G_{\text{beacon}}(t)/G_{\text{control}}(t)$ isolates the chemical reaction kinetics (Figure 3.17). Practically, we fit this ratio by a three-parameter exponential:

$$G^*(t) = \alpha + \beta \exp(-t/\tau_{\text{fluc}}). \quad (3.26)$$

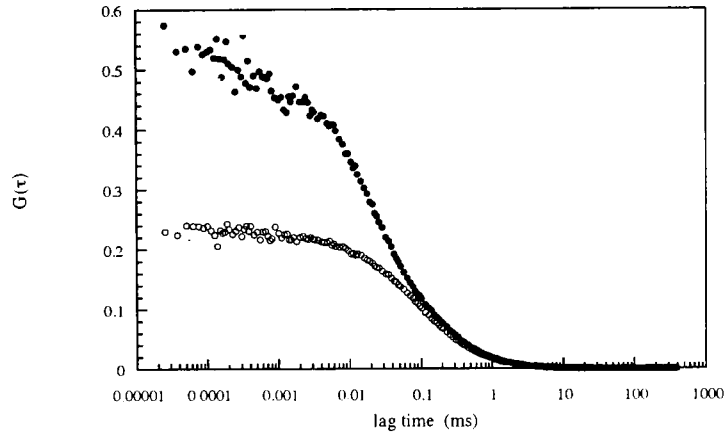
All of our measured $G^*(t)$ can be fitted very well with this single exponential over a wide interval of lag times, typically from 100ns to 10 ms (Figure 3.17B and Figure 3.18). This result validates the model of an all-or-none transition (reference [27, 19], and references therein) between the open and the closed states.

The acquisition of the autocorrelation function consists in taking the median of 60 10s-runs (for a total acquisition time of 10min), and determining the error bar by measuring the standard deviation of the 60 runs at each lag time (this procedure required a macro programmed by Oleg Krichevsky). The median was preferred to the simple average of the 60 runs, as we noticed that any perturbation during acquisition (sample impurity, laser fluctuation, electronic noise...) resulted in a non-zero baseline at long lagtimes: the distribution of correlation point at a given lagtime is always skewed to the right. The macro implements a classical reduction of noise by eliminating the runs out of 3 standard deviations, and computes the median and the standard deviation of the remaining runs.

Ultimately, the least-square fit of the autocorrelation function with equation 3.26 took into account the experimentally-determined error bars for the molecular beacon sample and for the control sample at each lag time:

$$\Delta G^*(t) = G^*(t) \left(\frac{\Delta G_{\text{beacon}}(t)}{G_{\text{beacon}}(t)} + \frac{\Delta G_{\text{control}}(t)}{G_{\text{control}}(t)} \right)$$

A)



B)

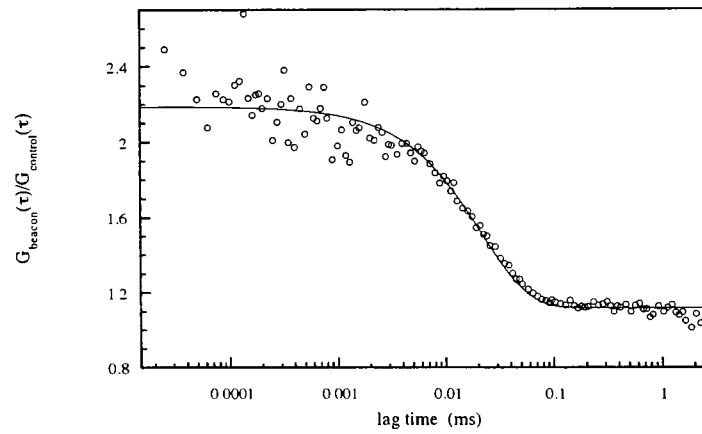


Figure 3.17: Data treatment for FCS of a molecular beacon. (a) Autocorrelation curves for beacon (●) and control (○) at $T = 45^\circ\text{C}$, in 0.1 M NaCl. Both beacon and control have loops of 21-T residues. (b) Ratio of the two curves shown in a. The line is a three-parameter exponential fit to the data giving $\tau_{\text{fluc}} = 24.2 \pm 0.6 \mu\text{s}$.

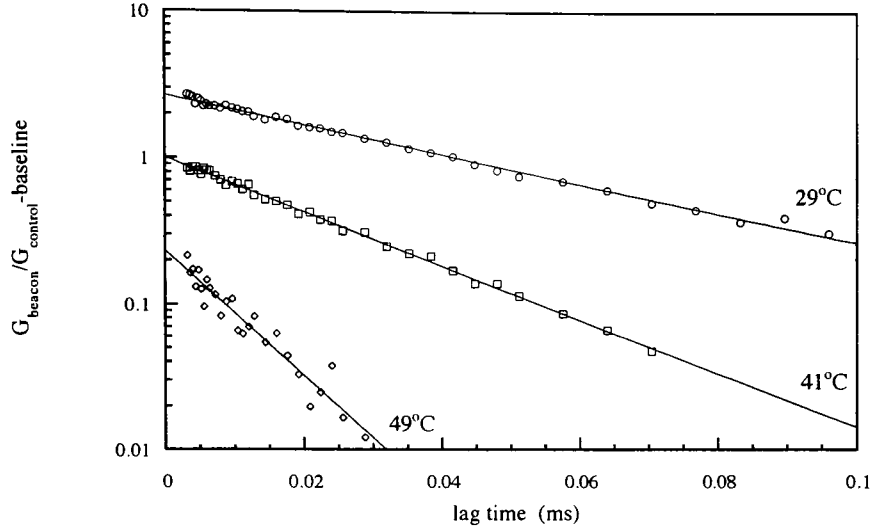


Figure 3.18: Autocorrelation function $G^*(t)$ for the chemical fluctuation of the molecular beacon at different temperatures. This plot, for a timescale range of 100ns to 10ms, is scaled in log-normal to show the validity of the single-exponential fit (continuous lines). The molecular beacon has a loop of 21 Ts', and the buffer contains 0.1M NaCl, 10mM NaCacodylate, 1mM EDTA pH8.0.

The error bar on τ_{fluc} is thus the “true” experimental error bar directly determined from the distribution of the autocorrelation runs. For example, for the curve on Figure 3.18, the fits yields $\tau_{fluc} =$, with the a Pearson coefficient of 0.998. The fit confidence Q [136] can be estimated with incomplete Γ functions of the χ^2 and the number of points N :

$$Q = \frac{\Gamma_c\left(\frac{N-2}{2}, \frac{\chi^2}{2}\right)}{\Gamma\left(\frac{N-2}{2}\right)},$$

where $\Gamma(a) = \int_0^\infty e^{-t} t^{a-1} dt$,

and $\Gamma_c(a, x) = \int_x^\infty e^{-t} t^{a-1} dt$.

In our experiment, fits were considered reasonable when $Q > 0.01$ (the error bars are usually underestimated). For example, for the fit in Figure 3.18, $\chi^2 = 42$ for $N = 90$ points, thus $Q = 10^{-4}$.

Fluorescence bulk melting curve of the molecular beacon.

To measure τ_+ and τ_- , we must complement the FCS measurement of the fluctuation timescale with an independent measurement. Monitoring the fluorescence of a solution of molecular beacon as a function of the temperature enables us to measure the equilibrium constant for the equilibrium:

$$\text{Closed beacon} \xrightleftharpoons{K(T)} \text{Open beacon}, \quad \text{where } K(T) = \frac{\tau_+}{\tau_-} \quad (3.27)$$

We fitted the melting curve with a single-equilibrium mass action law, taking into account the finite signal/noise ratio of the beacon:

$$F(T) = \frac{\alpha + \beta K(T)}{1 + K(T)} \quad \text{with } K(T) = e^{-\frac{\Delta H - T\Delta S}{RT}} \quad (3.28)$$

(α is the residual fluorescence of the closed beacon -measured as the baseline of the melting curve at low temperature-, and β is the fluorescence of the open beacon -measured as the plateau of the melting curve at high temperature). Experimentally, the error bar on the baseline of the melting curve (α) is much higher than the error bar on the plateau (β). Thus a more controlled fit was carried out at low temperature to measure α :

$$F(T) = \alpha + \gamma e^{-\frac{\Delta H}{RT}} \quad (3.29)$$

The two fits (equations 3.28 and 3.29) yielded baseline α and plateau β (with experimental error bars $\Delta\alpha$ and $\Delta\beta$), which were used to normalize the fluorescence and estimate the equilibrium constant:

$$K(T) = \frac{F - \alpha}{\beta - F}.$$

The error bars on the equilibrium constant are then:

$$\Delta K(T) = K(T) \left(\frac{\Delta\alpha}{F - \alpha} + \frac{\Delta\beta}{\beta - F} + \frac{\sqrt{F}}{\beta - F} + \frac{\sqrt{F}}{F - \alpha} \right)$$

The validity of the All-or-None model can be tested by fitting $K(T)$ with the mass-action (linearity of the plot $\ln(K(T))$ as a function of $1/T$). In Figure 3.19(A), we present the melting curve of the molecular beacon with a T_{21} loop, in 0.1M NaCl. In these condition, we measure a melting temperature of 37.9°C, a melting enthalpy of 28.7kcal/mol, a melting entropy of 92.1cal/mol/K. Our measurements span 3.5 decades of $K(T)$, and thus validate the fit with a single equilibrium constant (Figure 3.19B).

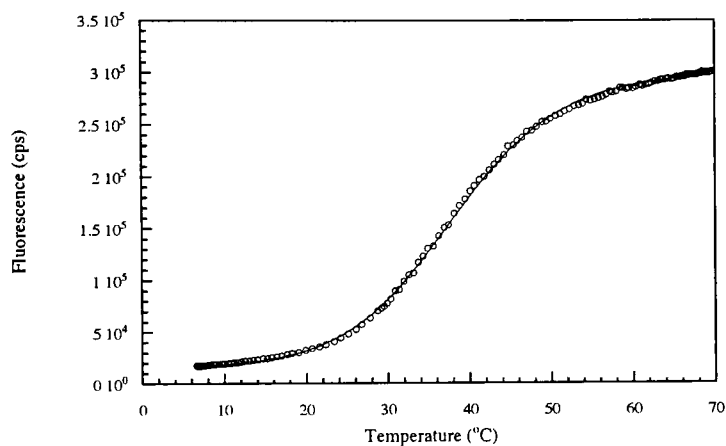
Results: kinetic timescales for opening and closing a DNA hairpin-loop.

Typical Arrhenius plot for τ_- and τ_+ . To extract τ_+ and τ_- from τ_{fluc} , we supplement the FCS experiments with a fluorescence bulk measurement, which gives $K(T)$. From equations (3.23) and (3.24), we get:

$$\tau_- = \tau_{fluc} \frac{1 + K}{K}, \quad \tau_+ = \tau_{fluc} (1 + K). \quad (3.30)$$

The experiments were carried out in the range of temperatures T from 10 to 50°C. We present data as Arrhenius plots for the opening and closing timescales as a function of $1/T$. We measure the activation enthalpy ΔH_a by fitting the experimental points with an exponential $\tau(T) = \tau_\infty \exp(+\Delta H_a/RT)$ (the + sign is used to be consistent with the Kramers-Eyring theory of the chemical activation). Typically, for the beacon of 21 T residues in the loop, τ_{fluc} passes from 134 μ s at 10°C to 28 μ s at 45°C, while the equilibrium constant $K(T)$ of open-closed transition is changing from 0.016 to 1.5. This leads to τ_- spanning the range from 44 μ s to 8.3ms and τ_+ from 71 μ s to 135 μ s (Figure 3.20). Note that for T_{21} most of the measurements are performed below the melting temperature of the secondary structure $K(T) < 1$. In this temperature range τ_+ is mostly determined by τ_{fluc} (see equation 3.30) such that it varies by a factor of 2 only. On the other hand, the variation of τ_- is dominated by the temperature dependence of $K(T)$, τ_- varying by two orders of magnitude. For the closing of this beacon the activation energy is +3.4kcal/mol, whereas the activation energy for the opening is +30.6kcal/mol (1 kcal = 4.18 kJ).

A)



B)

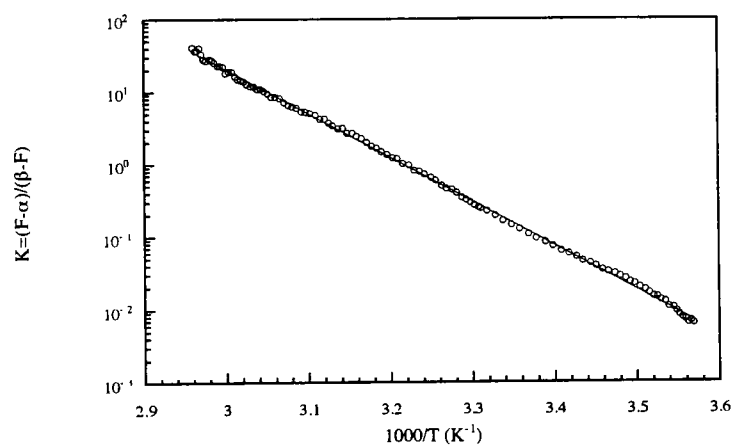


Figure 3.19: A) Bulk melting curve of the molecular beacon from Figure 3.16 B) Derived equilibrium constant and mass-action fit with $\alpha = 1.6 \cdot 10^4$ cps, and $\beta = 3.0 \cdot 10^5$ cps (buffer: 0.1M NaCl, 10mM Sodium Cacodylate, 1mM EDTA pH8.0).

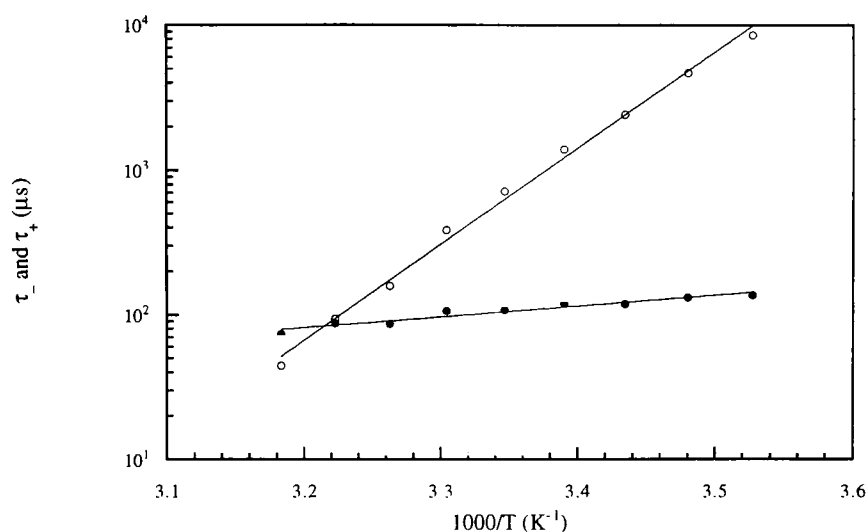


Figure 3.20: Arrhenius plot for τ_- and τ_+ for a molecular beacon with a 21-T loop (buffer: 0.1M NaCl, 10mM Na cacodylate, 1mM EDTA pH8.0).

Both activation enthalpies are positive: the limiting step is thus unfavorable, yet the transition open-closed does occur. Classically, one must argue that the limiting step must generate an entropy gain to compensate the enthalpic loss to form the activation molecule.

Note that one can read the melting temperature of the molecular beacon on the Arrhenius plot of the kinetic timescales. By definition, at T_m , the molecule is half of the time open, half of the time closed, thus $\tau_-(T_m) = \tau_+(T_m)$, and the melting temperature can be visualized as the intersection point of the two timescale fits: on Figure 3.20, the intersection point corresponds to $1000/T = 3.215$, thus $T_m = 37.9^\circ\text{C}$ (it is of course the same melting point as in the bulk melting curve of Figure 3.19).

Interpretation of the experimental results: limiting steps in hairpin-loop fluctuations.

Measuring the opening and closing timescales for the molecular beacons provides quantitative parameters for understanding the dynamics of fluctuation of DNA hairpin loops. In particular, we would like to interpret the order of magnitude of opening and closing timescales as well as the enthalpic barriers presented in the previous paragraph (for the sake of simplicity, we present our model only for the 21-T loop).

The opening timescale for the molecular beacon with the stem $5' - GCGAG - 3'$ varies drastically with temperature, with an activation enthalpy of +32kcal/mol. This enthalpy is comparable to the opposite of the enthalpic gain associated with the formation of dsDNA, as predicted by the Nearest-Neighbor model (see Chapter 1):

$$\begin{aligned}
 & \Delta H(5'GCGAG3'/5'CTCGC3') \\
 = & \Delta H(GC) + \Delta H(CG) + \Delta H(GA) + \Delta H(AG) + \Delta H(\text{initiation}) \\
 = & -9.8 - 10.6 - 8.2 - 7.8 + 2 * 0.1 \sim -36.2\text{kcal/mol}.
 \end{aligned}$$

We can attribute the enthalpic barrier associated with the opening of the molecular beacon to the disruption of the five basepairs at once.

The prefactor τ_-^∞ in the Arrhenius plot of the opening timescales is of the order of 10^{-18}s . It is difficult to predict accurately this prefactor as it corresponds to an extrapolation of the Arrhenius fit to infinite temperature, but we can present an estimation of its order of magnitude. The opening of

the stem corresponds to the unzipping of the five basepairs in the stem. Only vibrational fluctuations of the double-helix can produce this large-scale disruption. A typical quantum frequency is $\omega = k_B T/h$, where h is the Planck constant (Eyring's classical theory of activated chemical complex predicts this frequency of conversion of vibrational modes into translational modes [24]). At room temperature, $\omega = 6 \cdot 10^{12} \text{ Hz}$. The limiting timescale for opening the hairpin-loop structure is then:

$$\begin{aligned}\tau_- &= \frac{1}{\omega} e^{-\frac{\Delta G_-}{k_B T}} \\ &= \frac{1}{\omega} e^{+\frac{\Delta S_-}{k_B}} e^{-\frac{\Delta H_-}{k_B T}} \\ &= \tau_-^\infty e^{-\frac{\Delta H_-}{k_B T}}\end{aligned}$$

Thus, the entropic barrier ΔS_- can be estimated to be:

$$\begin{aligned}\Delta S_- &= k_B \ln(\tau_-^\infty \omega) \\ &\sim 2.0 * \ln(1.7 \cdot 10^{18} * 6 \cdot 10^{12}) \sim 143 \text{ cal.mol}^{-1} \cdot \text{K}^{-1}\end{aligned}$$

Interestingly, this entropic barrier is comparable to the entropy loss associated with the formation of the five basepairs of the stem according to the nearest-neighbor model¹ (see Chapter 1):

$$\begin{aligned}&\Delta S(5'GCGAG3'/5'CTCGC3') \\ &= \Delta S(GC) + \Delta S(CG) + \Delta S(GA) + \Delta S(AG) + \Delta S(\text{initiation}) + R \ln(c) \\ &= -24.4 - 27.2 - 22.2 - 21.1 - 2 * 2.8 + 2 \ln(0.05) \\ &\sim -100 \text{ cal.mol}^{-1} \cdot \text{K}^{-1}\end{aligned}$$

This confirms the modelization for the unzipping of the stem as a the nearest-neighbor-model disruption of the five basepairs in the stem.

In 0.1M NaCl, $\Delta H_+ \sim +3.4 \text{ kcal/mol}$. This activation barrier is a little less than half the enthalpic gain associated with the formation of a doublet of basepair in the nearest-neighbor model (from 7.2 to 10.6 kcal/mol). In fact, it is comparable with the enthalpic gain of forming a single basepair, without additional stacking basepair [137]. Thus the limiting step for closing a molecular beacon with a T-loop is the formation of the first basepair in the stem (the zipping of the rest of the stem is so fast that it is not a limiting step at all).

The prefactor τ_+^∞ in the Arrhenius plot of the closing timescales is of the order of $1 \mu\text{s}$ for a molecular beacon with a loop made out of 21 Ts'. Eyring's theory of thermal activation of chemical reaction is not applicable in the case of the closing timescale as the limiting step is mostly an hydrodynamic timescale [28]. The physical parameter relevant to the closing timescale are: $k_B T$, η (viscosity of the buffer), a (activation distance), R_g (radius of gyration). A dimensional analysis yields a fluctuation timescales for the polymer coil of ssDNA :

$$\tau_0 \propto \frac{\eta R_g^3}{k_B T} \sim 4 \text{ ns}.$$

(R_g is estimated to be $5\text{\AA}\sqrt{31} \sim 28\text{\AA}$ for a 31 mer of ssDNA). Among all the configurations adopted by the random coil, few are compatible with the formation of the first basepair in the stem. The

¹This entropy is estimated for a concentration of complementary oligonucleotides of 0.05M: we have to correct for the concentration in the duplex where the stem's complementary strands are bound within 20\AA , i.e. $c \sim 1/(4\pi/3 * (20\text{\AA})^3 * 6 \cdot 10^{23}) \sim 0.05\text{M}$.

probability to reach that special fluctuation is estimated to be the ratio of the activation volume a^3 compared to the whole volume of the coil R_g^3 . Thus, we estimate the prefactor of the Arrhenius plot to be:

$$\begin{aligned}\tau_+^\infty &= \tau_0 \left(\frac{R_g}{a} \right)^3 \\ &\sim 4 \left(\frac{28}{5} \right)^3 \sim 0.7 \mu\text{s}.\end{aligned}\tag{3.31}$$

This estimate is quite comparable with the experimental value of $1 \mu\text{s}$ we found by fitting the Arrhenius plot of Figure 3.20.

On Figure 3.21, we present a sketch of the free enthalpy landscape describing the dynamics of fluctuation of the hairpin-loop. The abscissa is an arbitrary reaction coordinate. The enthalpic barriers are represented for the case of a molecular beacon with 21-T loop, in 0.1M NaCl buffer: they show that both limiting steps are positive (a generic case of thermally activated folding, unlike the case of β -hairpin peptides' folding [138])

In conclusion, for the molecular beacon with a 21-T loop, we used a combination of FCS and bulk measurements to measure the two timescales relevant to the dynamics of conformational fluctuations. Arrhenius plots of the

We understand quantitatively the energy landscape associated with the fluctuations of the hairpin-loop secondary structure. The opening timescale can be modeled with standard nearest-neighbor model and thus does not teach us anything new about nucleic acid fluctuations. The closing timescale comprehends two contributions: the preequilibration of the coil to nucleate a basepair in the stem, and the thermally-activated formation of the first basepair. The preequilibration contribution is very interesting as it provides information on the intrinsic dynamics of ssDNA: in the next section, we will be using the measure of the closing timescales for different hairpin loops to better understand the dynamics of ssDNA.

3.3.2 Systematic study of the hairpin-loop fluctuation dynamics.

Dependence of the timescales on loop length and sequence.

In Figure 3.22, we compare the Arrhenius plots of the extracted timescales for the beacons with different sizes of T-loop in 0.1M NaCl buffer: the timescales for opening τ_- are similar for different N_{loop} , whereas the timescales for closing τ_+ decrease significantly with the loop size. The activation energy for opening these beacons is always 30kcal/mol whatever the loop; the activation energy for closing these beacons is also practically constant at 3.4kcal/mol. The only difference between different T-loops is the activation entropy for the closing of the molecular beacon.

Scaling laws for the closing timescale of the molecular beacon with different lengths.

On Figure 3.23, we plot the closing timescales at a given temperature as a function of the loop length for molecular beacons with T-loop. This closing timescale is related to the probability of end-to-end collision for the ssDNA strand. It is expected to scale with the loop size N_{loop} [139, 140, 141]. Assuming that the collision of any of the five basepairs in the hairpin stem can lead to the hairpin closure, we plot τ_+ as a function of $N_{loop} + 5$ (the number of bases between the midpoints of the stem). The linearity of this log-log plot is fair, and we can fit our data with:

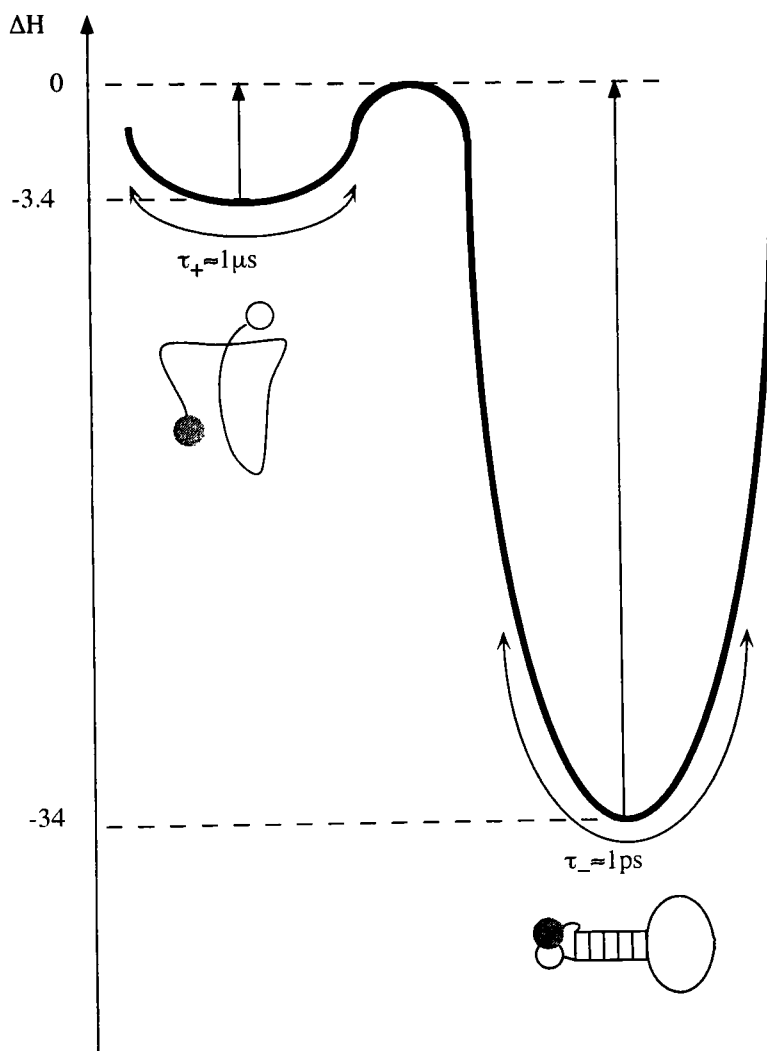


Figure 3.21: Energy landscape for the fluctuation dynamics of the hairpin-loop structure.

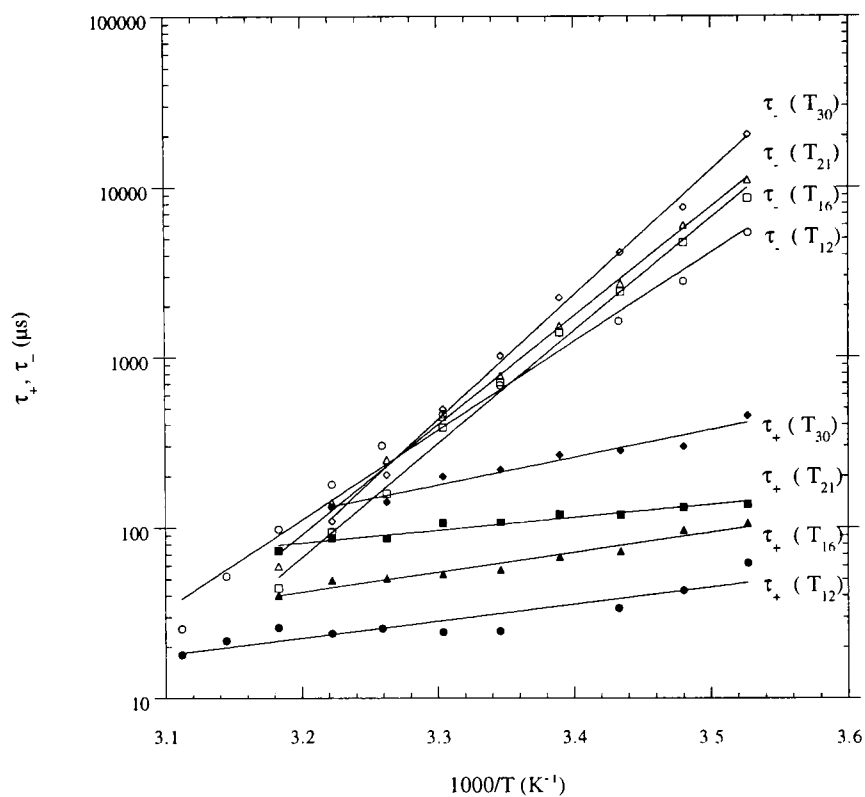


Figure 3.22: Arrhenius plots of the opening rates (open symbols) and the closing rates (filled symbols) of beacons with different loop lengths: T_{12} (\circ), T_{16} (\square), T_{21} (\diamond), and T_{30} (\triangle). The lines are exponential fits to the data. The lines corresponding to opening and closing intersect at the melting temperature. The buffer contained 0.1 M NaCl for all the data.

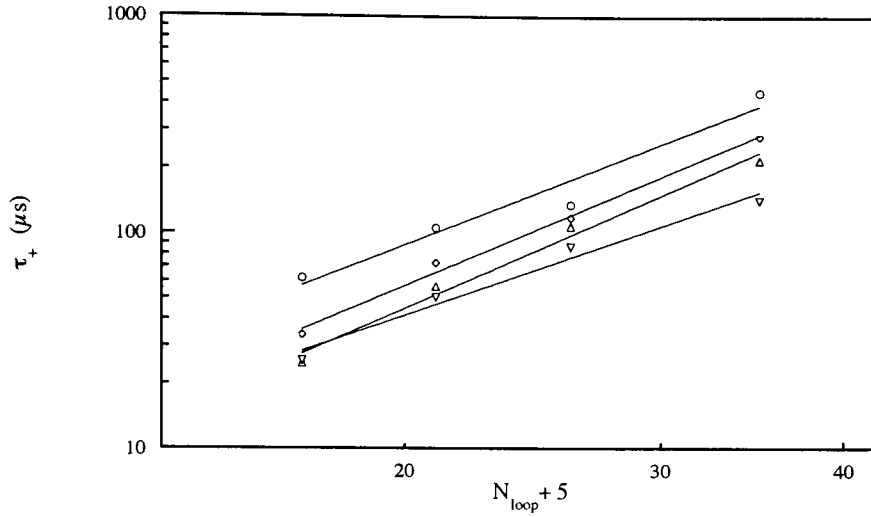


Figure 3.23: Scaling laws for the closing timescale of T-loop molecular beacons as a function of the loop length N_{loop} : (\circ) 10°C; (\diamond) 18°C; (\triangle) 26°C; (∇) 33°C (buffer: 0.1M NaCl, 10mM NaCacodylate, 1mM EDTA pH8.0).

$$\tau_+(N_{loop}, T) \propto (N_{loop} + 5)^{\alpha(T)}. \quad (3.32)$$

For T-loops in 0.1M, the exponent $\alpha(T)$ is between 2.6 and 2.9 (± 0.3). These scaling exponents are mostly temperature-independent even though lower temperatures are slightly larger than those above 30°C. These values are to be compared with the theoretical predictions made for a number of similar problems (DNA cyclization, intrachain reactions, etc.), and assessed from equation 3.31: all length dependence can be attributed to the dependence of the radius of gyration of ssDNA with the loop length. If the ssDNA chain could be modeled as random walk, then $R_g \propto (N_{loop} + 5)^{0.5}$ and $\tau_+ \propto R_g^3 \propto (N_{loop} + 5)^{1.5}$. A more realistic description of a polymer takes into account the swelling of the coil by volume exclusion [142]: $R_g \propto (N_{loop} + 5)^{0.6}$ and $\tau_+ \propto (N_{loop} + 5)^{1.8}$. These theoretical values are obviously much smaller than the ones we find in T-loops in 0.1M NaCl buffer.

In fact, these theoretical scaling laws are valid only in the asymptotic regime of very long polymers. In our case, we are dealing with short ssDNA chains, 15 to 40 nucleotide-long. Podtelezhnikov *et al.* [141] showed by molecular dynamics simulation that large excluded-volume effects are crucial to short polymers' dynamics, inducing further swelling of the coil: they predict a scaling exponent of 2.6 for the cyclization probability of short polymers. The scaling exponent that we find for T-loops are consistent with these simulations. They also further confirm the model for hairpin closure presented in § 3.3.1: the limiting step for hairpin closure is the formation of the first basepair, it is thus dependent on the probability of end-to-end contacts in the ssDNA coil.

Dependence of the opening and closing timescales on the salt concentration. We compare the kinetics of hairpin fluctuations for different concentrations of monovalent ions (Figure 3.24). Closing and opening timescales are measured with a molecular beacon with a T₂₁-loop. We observe that the overall trend for these timescales as a function of salt concentration follows intuitive ideas: the opening time increases with the salt concentration as higher salt concentration implies better screening of the phosphate negative charges, and thus higher stability of the stem basepairs; the closing time decreases with the salt concentration as the electrostatic screening decreases the rigidity of the ssDNA, and thus increases the probability of forming an end-to-end contact.

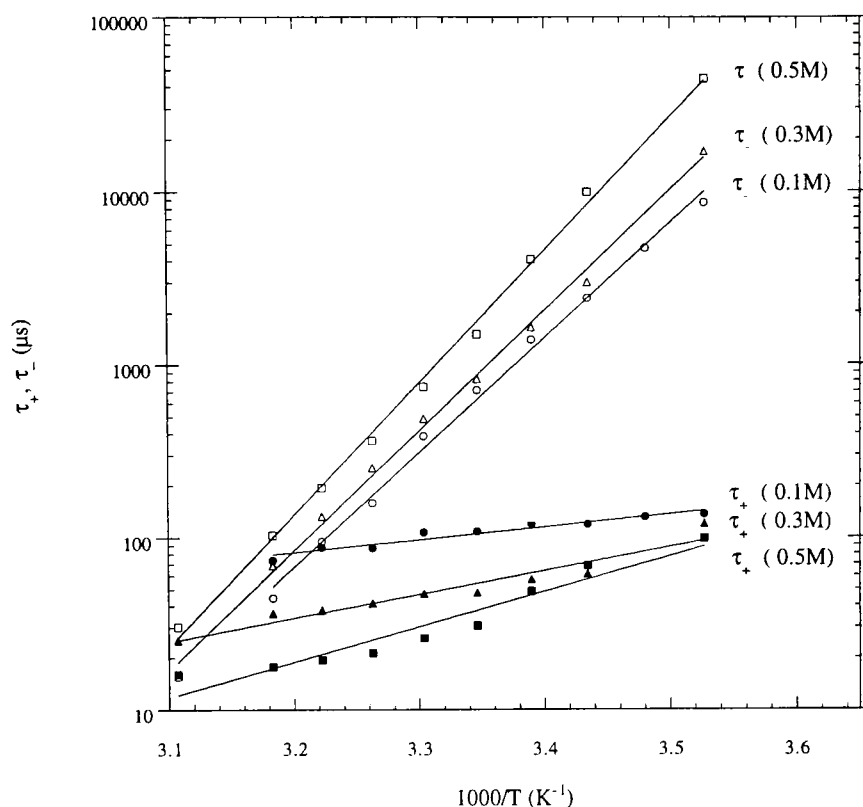


Figure 3.24: Closing and Opening timescales for a molecular beacon of loop T₂₁ for different salt concentrations.

The activation enthalpy for the closing of the beacon increases strongly with the salt concentration: 13.4 kcal/mol for $[\text{Na}^+] = 0.1$ M, 16.6 kcal/mol for $[\text{Na}^+] = 0.2$ M, and 19.0 kcal/mol for $[\text{Na}^+] = 0.5$ M. These results are quite surprising as one would have expected that the limiting step associated with the formation of the first basepair in the stem would be favored by the higher salt concentration. However, one must argue that the limiting step for closing must include the exclusion of monovalent ions from the nucleation volume, this exclusion being harder at higher salt concentration.

The enthalpic barrier for the opening of the beacon stays approximately the same for all the salt concentrations used: the effect of higher salt concentration is thus mostly entropic as one must evaluate the condensation of screening monovalent anion onto the phosphates (in the range of concentration we use, $[\text{Na}^+] > 0.1$ M, the electrostatic repulsion between opposite strands is fully-screened and irrelevant).

Sequence-dependent rigidity of ssDNA.

Classical experiments on ssDNA fluctuations measure a persistence length of 2 nucleotides in 0.1M NaCl (cf § 1.1.5). These measurements were made with very large polymers (from 200 nucleotides up to 48000 nucleotides [11]), where the detail of the DNA sequence is irrelevant. Using our technique combining Fluorescence Correlation Spectroscopy and Fluorescence quenching to study the dynamics of fluctuations of DNA hairpin-loops, we can address the issue of the sequence-dependent rigidity of ssDNA by measuring the closing time τ_+ of different short homopolymers.

In this section, we report our measurements on the closing timescales of molecular beacons of

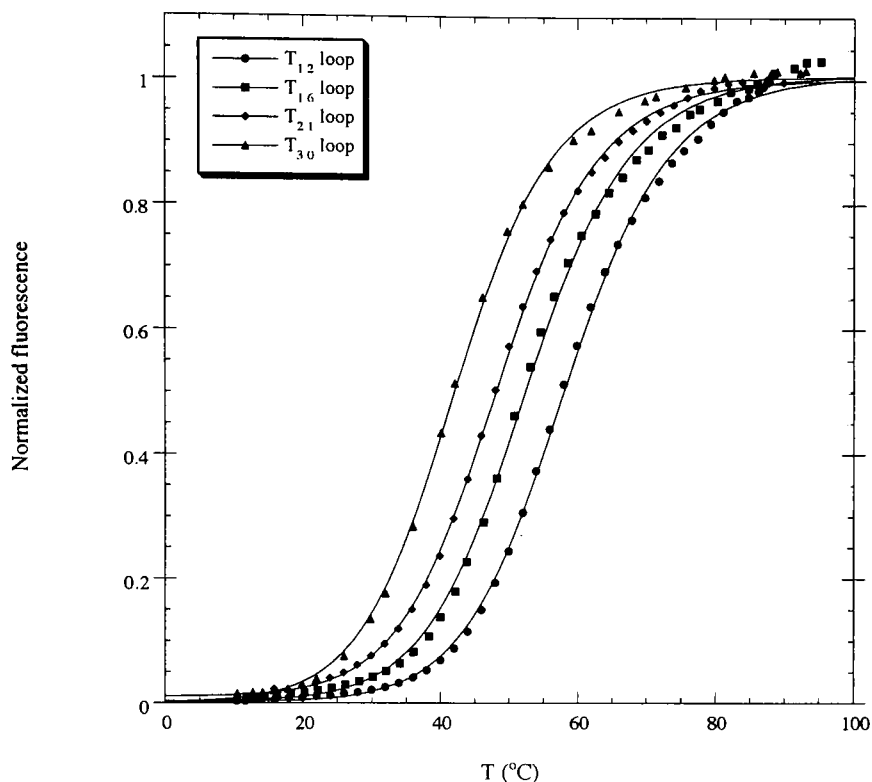


Figure 3.25: Melting curves of T-loop beacons in 0.25M NaCl, 10mM Cacodylic acid, 1mM EDTA pH8.0.

sequence $5' - CCCAA(X)TTGGG - 3'$, where X is an homopolymer of Ts' or As'. The buffer we used for this study contains 0.25M NaCl, 10mM NaCacodylate, 1mM EDTA pH8.0.

In Figure 3.25, we present the melting curves for beacons with Thymidines (T) loops of different lengths (T_{12} , T_{16} , T_{21} , T_{30}). In Figure 3.26, we present the melting curves for beacons with Adenosines (A) loops of different lengths (A_8 , A_{10} , A_{12} , A_{16} , A_{21} , A_{30}). In Figure 3.27, we plot the melting temperatures of the T-loop and A-loop beacons in 0.25M NaCl: the melting temperatures are plotted as $1000/T_m$ vs. the logarithm of the loop length. This representation is proposed to emphasize the length dependence of the entropic term in the melting temperature.

These bulk measurements already show that the sequence-independent rigidity of ssDNA is not valid: a loop made out of As' decreases drastically the melting temperature of a hairpin loop compared to the same loop length but made out of Ts', A-loops are obviously more rigid than T-loops. Two possible scenarii can explain the equilibrium constant of the hairpin-loop structure.

#1 The closed hairpin is destabilized by the presence of As' in the loop, *i.e.* $\tau_-(A \text{ loop}) < \tau_-(T \text{ loop})$.

#2 It is harder to close a denatured hairpin with a A-tract rather than a T-tract, *i.e.* $\tau_+(A \text{ loop}) > \tau_+(T \text{ loop})$.

To decipher between the two models, we carried out FCS measurements on all these hairpin-loop structures and measured systematically τ_- , and τ_+ as a function of the temperature.

In Figure 3.28 and Figure 3.29, we compare the Arrhenius plots of the extracted timescales for the beacons with different sizes of T-loop and A-loop: we find that the rates for opening τ_- are

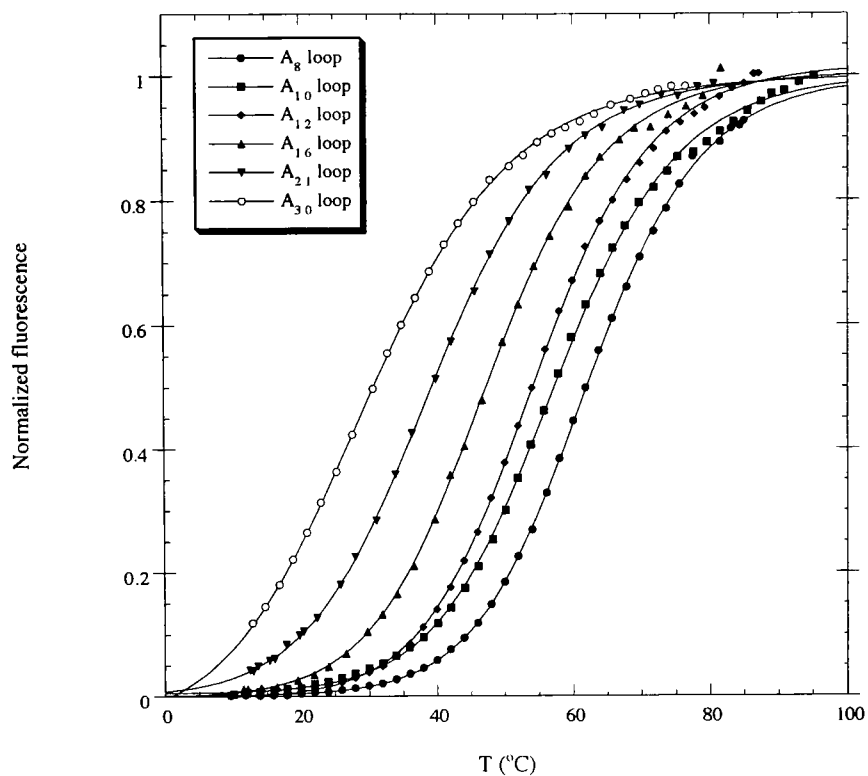


Figure 3.26: Melting curves of A-loop beacons in 0.25M NaCl, 10mM Cacodylic acid, 1mM EDTA pH8.0.

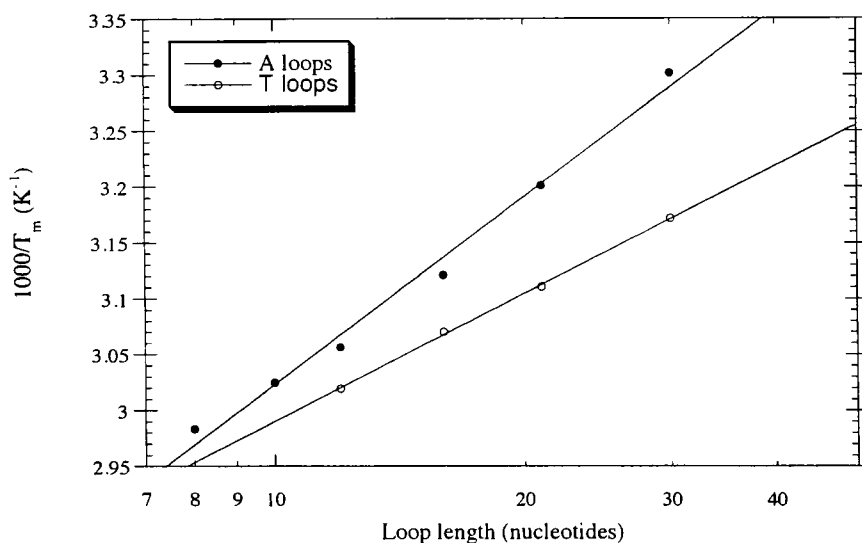


Figure 3.27: Melting Temperatures T_m for T-loop and A-loop beacons as a function of the loop length (Buffer: 0.25M NaCl, 10mM Cacodylic acid, 1mM EDTA pH8.0.)

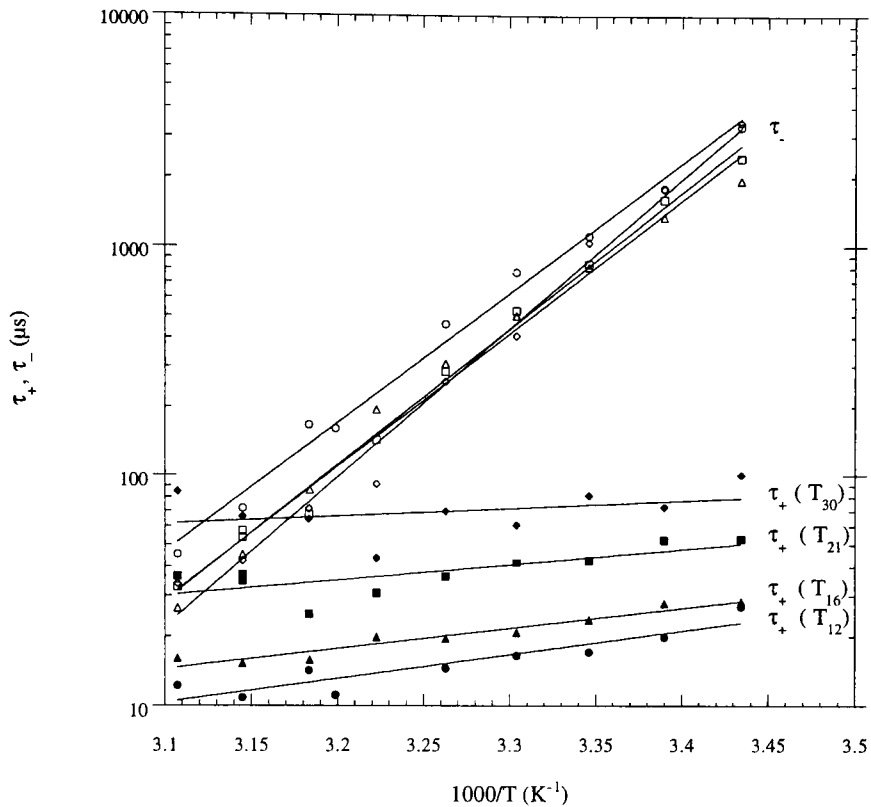


Figure 3.28: Closing and opening timescales for beacons of different lengths of T-loops (data acquired in 0.25M NaCl, 10mM NaCacodylate, 1mM EDTA pH8.0)

similar whatever the size of the loop, and whatever the sequence in the loop. On the other hand, the timescale τ_+ for hairpin loop closing decreases significantly with the loop size, and is much longer for an A-loop than for a T-loop of same length.

For all cases in Figure 3.28 the closing timescale grows with increased loop length as we might expect that a larger molecule has a longer fluctuation timescale. By concentrating on just one series, $(T)_{12} - (T)_{30}$, we observe two key points: (1) the slopes ($\Delta H_c/R$) are practically identical and (2) τ_{fluc} increases with increasing loop length. A different process is seen in poly(A): (1) $\Delta H_c/R$ increases with increasing loop length and (2) ($\tau_{closing}$ for $T \rightarrow \infty$), decreases with increased size. This sequence dependence is illustrated by directly comparing the closing enthalpic barriers for equal loop lengths of poly(A) to poly(T). For example we find that $\Delta H_c/R$ is much greater for As than Ts: ΔH_c of $(T)_{30}$ is 3.0kcal/mol, whereas for $(A)_{30}$ 17 kcal/mol. These enthalpic barriers unravel an unexpected sequence dependence in ssDNA dynamics. Figure 3.30 displays the enthalpic barriers extracted from Figure 3.29 as a function of loop length for poly(T) and poly(A) hairpin. This dependence is weak for poly(T)s ($-0.1\text{kcal}\cdot\text{mol}^{-1}\cdot\text{base}^{-1}$) whereas it is strongly linear for poly(A)s ($+0.5\text{kcal}\cdot\text{mol}^{-1}\cdot\text{base}^{-1}$).

As a conclusion, the comparison of the closing timescales of T-loop beacons and A-loop beacons unravel a sequence-dependent rigidity of A-loops. T-loops behave mostly are purely entropic coils, whereas A-loops contain an enthalpic rigidity most probably due to the base-stacking in ssDNA. This conflicts with the standard model of ssDNA [11] as a fully flexible polymer. We propose that molecular rigidity could explain the additional enthalpy contribution of poly(A).

Numerous experimental studies (reviewed in [13]) suggest that poly(A) has a significant helical

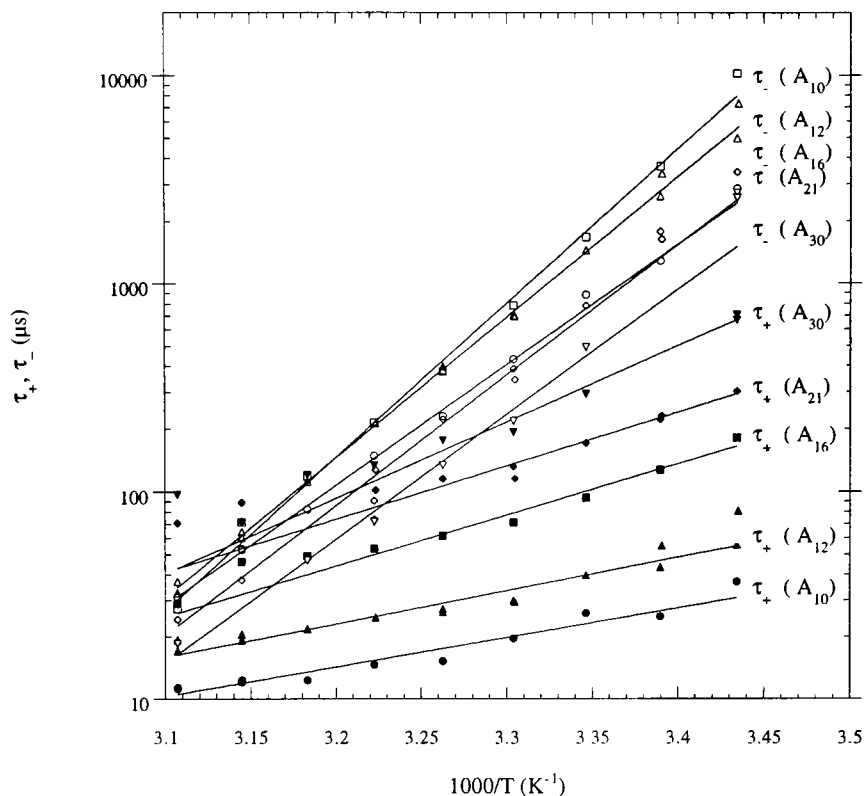


Figure 3.29: Closing and opening timescales for beacons of different lengths of A-loops (data acquired in 0.25M NaCl, 10mM NaCacodylate, 1mM EDTA pH8.0)

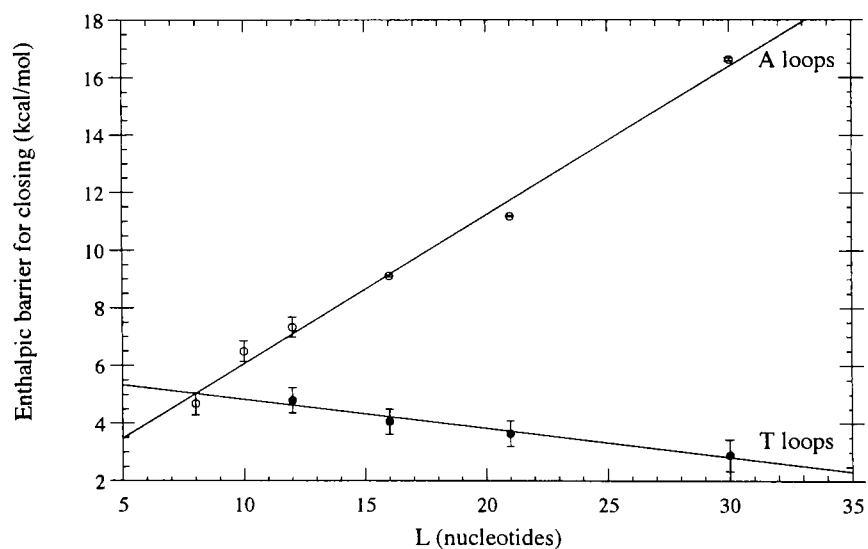


Figure 3.30: Barrier enthalpy for the closing of the molecular beacons as a function of the loop length: for A-loops (\circ), and for T-loops (\bullet) (buffer: 0.25M NaCl, 10mM NaCacodylate, 1mM EDTA pH8.0).

structure stabilized through base stacking. In contrast poly(T) has been observed to have little long range order [143]. Although a formal description of base stacking is still under debate, a recent work by Friedman and Honig [15] unifies the various contributions of electrostatic interactions, hydrophobic effects, surface tension, and dispersive interactions. Could base stacking lead to sequence dependent rigidity inferred from the hairpin closing timescale?

The energetic barrier for closing is complex. The ssDNA has to distort whichever conformation it is in to accommodate the formation of the first base pair in the stem. The neighboring base pairs in the stem are then zipped very fast ($<100\text{ns}$). Thus the barrier accounts for this distortion and the nucleation of the first base pair. Since the stem is identical in all hairpins, the loop dependence for the closing barrier reflects the distortion energy of ssDNA. Our simple analysis of the thermal fluctuations provides insight into the mechanical rigidity of poly(A)s compared to poly(T)s. We then understand the linearity of the enthalpic barrier for hairpin closing with respect to loop length. It reflects and measures the disruption of ssDNA base stacking. We get an estimate of ($+0.5\text{kcal mol}^{-1}\text{ base}^{-1}$) comparable to predicted values [16] for base stacking.

This sequence-dependent rigidity of ssDNA is of crucial importance to understand and predict the integrity of DNA secondary structures (Fig. 3.1). Let us emphasize that this effect has been overlooked in standard Nearest-neighbor models for analytical tractability (ssDNA rigidity introduces long-range correlation in the partition function of a ssDNA secondary structure). In fact, this sequence-dependent rigidity of ssDNA is irrelevant to most DNA hybridization study: the buffer of choice for these studies contains 1M Na^+ , which means that all the melting transitions are shifted towards very high temperatures (above 70°C). Moreover, classical spectroscopic studies (UV hyperchromicity and circular dichroism) have a poor signal/noise, requiring the use of high concentration of nucleic acid material, which constitutes another reason why all the melting transitions are shifted towards higher temperature (base stacking in ssDNA melts around 40°C [16]). Consequently, base-stacking in ssDNA was made irrelevant in these studies. Our technique combining FCS and Fluorescence quenching to study the dynamics of intramolecular secondary structure unraveled the sequence-dependent rigidity of ssDNA of crucial importance in more biologically relevant systems (see discussion in § 4.1.1).

Defect in ssDNA stacking.

Let us illustrate the strength of this sequence-dependent rigidity by perturbing the poly(A) stacking with a single “defect” (a different base) in the loop sequence. $(A)_{21}$ loops containing a single C (cytosine) defect in three varying positions along the sequence were studied. Melting curves and FCS measurements were taken for the series in 0.1M NaCl (lower salt concentration enhances the rigidity of ssDNA). The closing timescales for the series are shown in Fig. 3.31. We find that the closing timescales for these hairpins are shorter than for hairpins of $(A)_{21}$ with no defect. The cytosine defect induces a disruption in the base stacking, analogous to a mechanical weak point in a rigid structure.

Moreover, the enthalpic barrier for closing decreases when the defect is placed closer to the center of the loop ($\Delta H_c=17.4\text{ kcal/mol}$ for $(A)_{21}$, 16.7 kcal/mol for A_5CA_{15} , 14.2 kcal/mol for A_8CA_{12} , 13.4 kcal/mol for $A_{10}CA_{10}$).

This is consistent with a hinge description of the mechanical fluctuations which facilitates the alignment and consecutive base pairing of the stem. Think of a macroscopic rigid rod: if we introduce a mechanical weak point in its middle, then it is easy to bend and fold the bar such as its two ends match. If the weak point is away from the center (say at one third of the rod length), then bending will not bring the two bar ends in proximity. Ultimately, our experiment with A-tracts of ssDNA, with a C defect within, shows that ssDNA partially behaves as a rigid rod.

We studied the closing of different hairpin loops as a probe of ssDNA fluctuation dynamics. For loops made out of Ts', scaling laws, for the closing timescale as a function of the loop length, show that these ssDNA behave as

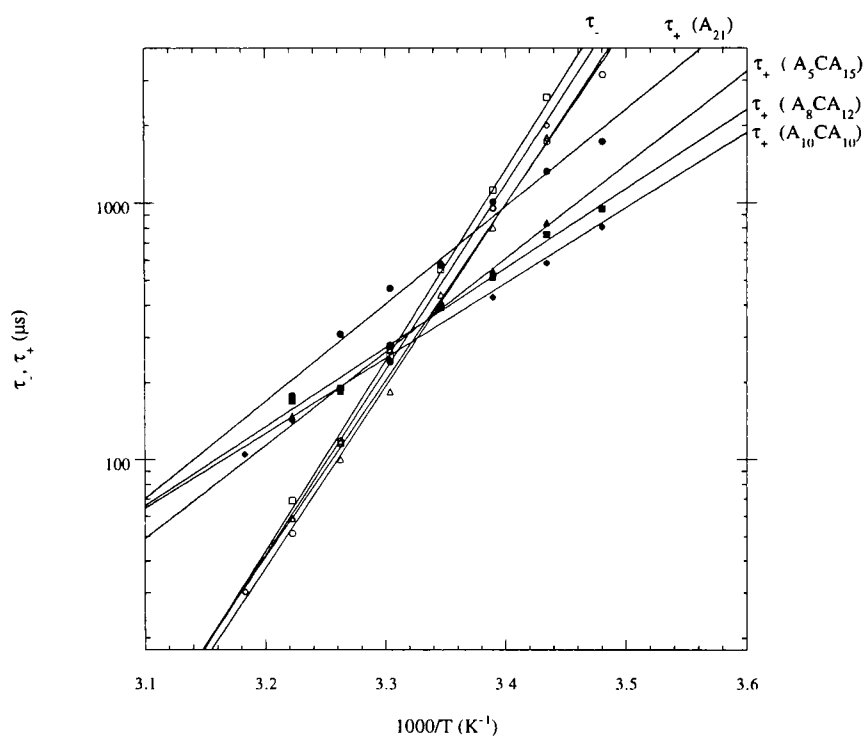


Figure 3.31: Closing timescales of four Poly(A) series with a single cytosine defect compared to (A)₂₁ (buffer: 0.1M NaCl, 10mM Cacodylic acid, 1mM EDTA pH8.0). Loop sequences are represented by the symbols \circ =A₂₁, \square = A₁₀CA₁₀, \triangle = A₈CA₁₂, $+$ = A₅CA₁₅

entropic coils, with a substantial swelling by excluded-volume interaction. For a loop made out of As', the hairpin closure limiting step requires the disruption of the base stacking with an enthalpic cost linear to the size of the loop. We conclude that ssDNA displays very sequence-dependent rigidity, which is determinant to the stability of secondary structures such as hairpin loops.

3.4 Breathing modes of double-stranded DNA.

Breathing of double-stranded DNA is a prerequisite step to DNA transcription initiation, or intercalation of drugs. More generally, breathing of dsDNA unravels ssDNA domains, whose sequence readability is enhanced compared to paired domains (the hydrogen bond pattern, involved in the formation of Watson-Crick basepairs, becomes accessible upon breathing of the double helix). Breathing modes are expected to be excited by thermal fluctuations of the helical structure, even much below the melting transition associated with the full separation of the two complementary strands. The only dynamic measurements on dsDNA breathing were achieved by titration of the imino-proton exchange time with proton acceptor in the buffer (see Chapter 1): the lifetime of an AT basepair among GC basepairs has been estimated around 1-10ms at 15°C, in 0.1M NaCl [59].

No measurement on the lifetime of an internal open bubble has been reported so far, no matter how relevant these breathing timescales are. Moreover, no dynamic information on the shape of the fluctuations (distribution of DNA bubbles' sizes and their relaxation times) is available. We use the combination of Fluorescence Correlation Spectroscopy and Fluorescence Quenching, developed to study the molecular beacons), to study the dynamics of dsDNA breathing.

Breathing in AT-rich regions of dsDNA were investigated for different lengths of AT domain (6,8,18 basepairs), and for different sequences. We will present the constructs we designed to monitor dsDNA breathing modes. For the sake of discussion, we will focus first on the construct with a random 18-AT domain, present the bulk measurement assessing the control on the fluctuation modes, present the analytical derivation of the autocorrelation function expected in the FCS measurements, and the FCS measurements themselves. The dynamics of breathing of the random 18-AT domain was found to comply with standard nearest-neighbor model.

We will then present results for shorter AT domain. Bulk measurements will show us that the compliance of the GC clamp is not robust at medium temperature, and large-scale premelting of GC basepairs occurs. These results do not comply with the nearest-neighbor models. We will investigate this long-range disruption of basepairs for different AT domains, and present a possible alternative model for dsDNA breathing.

3.4.1 Design of the probes to study dsDNA breathing modes.

To study dsDNA breathing modes, we designed specific fluorescent DNA probes to monitor the local fluctuation of one specific basepair, among a large double-stranded template. All the constructs are hairpin-loops, with a small 4 – T loop, and a large double-stranded stem (typically 30 basepairs). The loop is used only to concatenate the two complementary strands. The stem's double-stranded domain contains one AT-rich domain (the “breathing domain” in the center), flanked by two GC-rich domains (“GC clamps”), whose high stability constrains the breathing modes to the central AT-domain (Figure 3.32).

Two fluorescence-quencher tags are used:

1. Internal tagging: Rh6G and DABCYL are coupled to internally-modified thymidine bases (see § 2.1 on page 43). The two thymine bases must be in *vis-à-vis* to bring DABCYL and

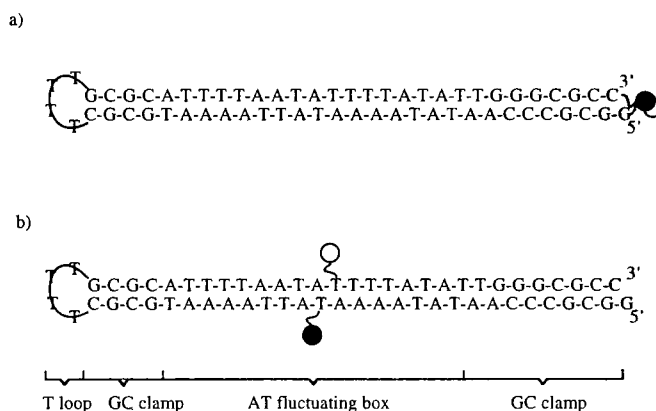


Figure 3.32: DNA construct for dsDNA breathing study: a) Construct with internal tagging with the fluorophore-quencher pair; b) Construct with end-tagging.

Rh6G within 10\AA , when the AT basepair is formed. The internal tagging is used to monitor the breathing of the dsDNA.

2. End tagging: Rh6G is coupled onto the 5' end of the stem, DABCYL being at the 3' end by synthesis. This end-tagging is used as a control to measure the integrity of the end-GC-clamp.

We will focus in the next section on the end-tagged and internally-tagged breathing constructs with a fluctuation box of sequence $5' - ATT\ TTA\ ATT\ ATT\ TAT\ ATT - 3'$ (18 basepairs).

3.4.2 Breathing fluctuations for 18AT fluctuation box.

Bulk fluorescence measurement.

We measure the bulk fluorescence of a solution of 10nMol of the two breathing constructs (internally-tagged and end-tagged). The melting curve is fully reversible with respect of temperature sweep, and independent of the concentration of DNA constructs: this confirms that all the fluorescence change is related to intramolecular rearrangements in the construct. The signal-to-noise (ratio of the fluorescences at high temperature and low temperature) is of 60 for the end-tagged construct, and 15 for the internally-tagged object. The difference of signal-to-noise can be explained by the difference of the structural geometry of the tagging: end-tagging positions Rh6G and DABCYL at 20\AA at the end of a double helix, whereas internal tagging fixes Rh6G's and DABCYL's anchorage points within 10\AA in the major groove of the DNA double helix. The quenching is less efficient for the internal tagging because the motion of Rh6G and DABCYL must be more constrained than for the end tagging.

In Figure 3.33, we present the two melting curves, in a buffer of 0.1M NaCl, 10mM Cacodylic acid, 1mM EDTA pH8.0. The melting curve of the end-tagged construct follows essentially an all-or-none transition, at 80°C , and the transition is very sharp (within 5°C). These results are consistent

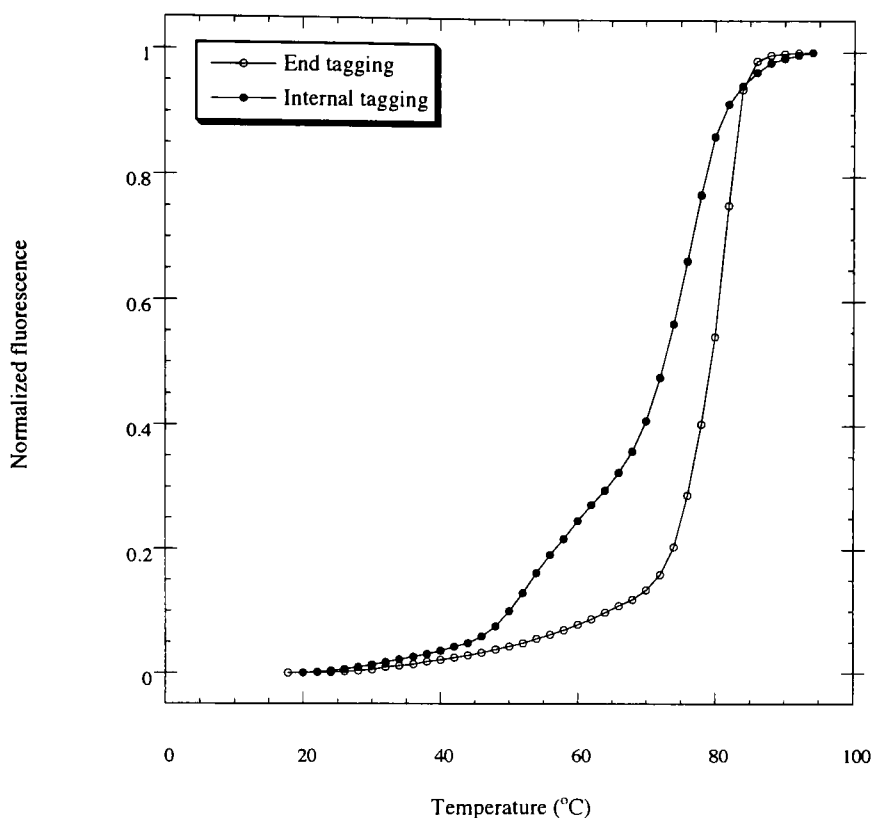


Figure 3.33: Melting curve of the end-tagged (\circ) and internally-tagged (\bullet) DNA constructs with a random 18AT domain (buffer: 0.1M NaCl, 10mM cacodylic acid, 1mM EDTA pH8.0)

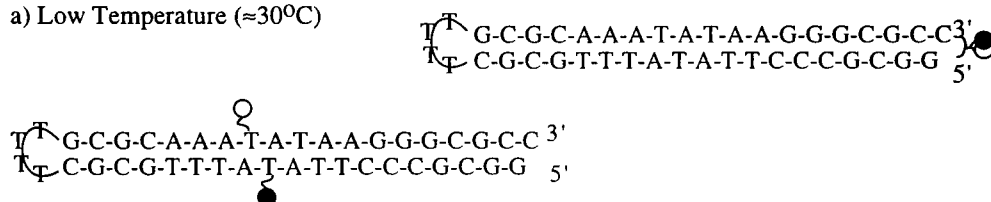
with the all-or-none melting of the terminal GC-clamp. The melting curve of the internally-tagged construct consists of two steps: the first melting occurs at 55°C and plateaus at 20% of the total fluorescence, the second melting occurs 75°C and restores the full fluorescence. The first melting is rather wide, which is a sign of a not very cooperative melting transition, typical of AT-rich domains. The second melting is much sharper and almost overlaps with the opening of the GC clamp as measured in the end-tagged constructs: this melting must corresponds to the complete melting of the GC-clamps, monitored internally by the separation of the complementary strands (in the first melting, the two strands are held together in a single-stranded bubble).

All and all, these bulk measurements validate the design of our construct: from 45°C to 65°C, the AT-rich domain opens up (high fluorescence for the internally-tagged) while the GC clamp is preserved (low fluorescence for the end-tagged). Figure 3.34 presents a sketch illustrating the conformational rearrangement and fluorescence change associated with the melting transitions. We then carried out FCS measurements on these DNA constructs to monitor the breathing dynamics of the AT-rich domain.

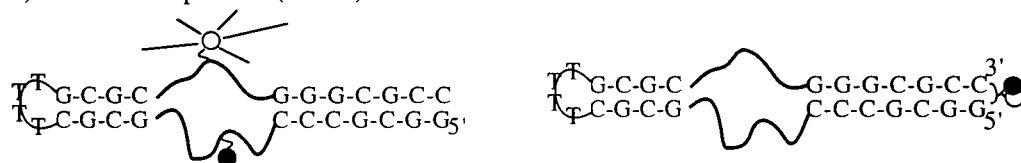
Stretched exponentials in breathing modes.

Before presenting the FCS measurements on our DNA breathing constructs, let us present a model deriving the autocorrelation function we predict for a FCS measurement on internally breathing modes. In the case of dsDNA breathing fluctuations monitored with the internal tagging, we expect the correlation function measured by FCS to be a stretched exponential:

a) Low Temperature ($\approx 30^{\circ}\text{C}$)



b) Medium Temperature ($\approx 50^{\circ}\text{C}$)



c) High temperature ($\approx 90^{\circ}\text{C}$)

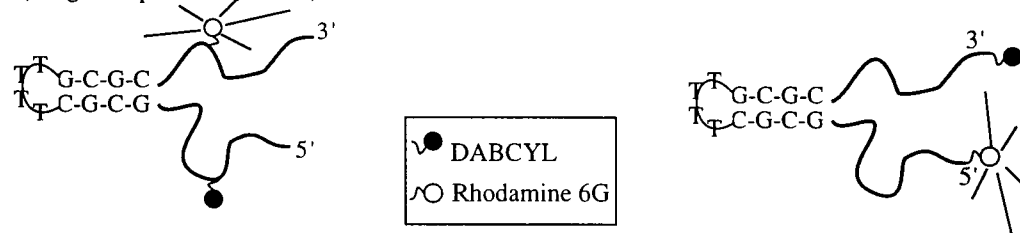


Figure 3.34: Conformational transitions associated with the melting transition of the breathing constructs. Left column: internally-tagged construct. Right column: end-tagged construct. Note that the fluorescence is partially restored at 50°C for the internally-tagged construct, whereas the fluorescence is dequenched only at 80°C for the end-tagged construct.

$$G(\tau) \propto e^{-\left(\frac{\tau}{\tau_0}\right)^\beta}.$$

where β is comprised between 0 and 1. This anomalous stretched relaxation is very generic in complex physical system: it has been observed in viscoelastic systems, spin glass magnets or dielectric relaxation. Palmer *et al.* [144] introduced a model of hierarchically constrained dynamics to show how universal (*i.e.* independent of the underlying microscopic dynamics) stretched exponentials are. In the classical theory of relaxation dynamics (Debye relaxation), one gets the conventional single exponential for the relaxation of a single mode ($\beta = 1$):

$$G(\tau) \propto e^{-\frac{\tau}{\tau_0}},$$

thus G can be characterized by a single relevant relaxation timescale τ_0 . This single mode of relaxation was observed for the fluctuations of the molecular beacon, and constitutes the simplest relaxation process. To reach a higher degree of complexity, one has to take into account a distribution of relaxation timescale, associated with different degrees of freedom in the system.

In the case of the fluctuations of a single basepair inside a large piece of double-stranded DNA, one can assume that the basepair can unzip with $(L-1)$ neighboring basepairs. Assuming that these modes of relaxation are independent, and contribute additively in the correlation function, we get:

$$G(\tau) = \sum_L w(L) e^{-\frac{\tau}{\tau(L)}}, \quad (3.33)$$

where, $w(L)$ is the probability to get a mode of size L , and $\tau(L)$ is the relaxation time of this L -mode.

To estimate this sum, we need a model to estimate the weight of a mode of size L , as well as its relaxation time $\tau(L)$. The simplest consists in assuming that $w(L)$ is the Boltzmann weight for thermally unzipping L basepairs:

$$w(L) \propto e^{-\frac{\epsilon L}{k_B T}}, \quad (3.34)$$

where ϵ is the free energy to unzip one basepair.

As far as the timescale $\tau(L)$ is concerned, the simple case of the molecular beacon hints towards estimating it as the closing timescale of L basepairs. Indeed, a fluorescence correlation measurement always picks up the fastest timescale of a relaxation process (see § 2.3 on page 59). Bulk measurements show that, between 10°C and 50°C, the fluorophore is more than 80% of the time quenched, thus the basepairs must be closing faster than they are opening, and the timescale that FCS will pick up is a closing timescale. The timescale $\tau(L)$ should then scale with the size of the mode (*cf* § 1.1):

$$\tau(L) = \tau_0 \left(\frac{L}{L_0} \right)^\alpha.$$

This estimate for the weight $w(L)$ and for the relaxation timescale $\tau(L)$ contains all the physics to compute the sum (3.33). A classical approximation by the saddle-point method enables us to estimate the sum and yield to a stretched exponential for the correlation function $G(\tau)$:

$$G(\tau) = \int w(L) e^{-\frac{\tau}{\tau(L)}} dL = \int e^{-g(L, \tau)} dL \quad \text{where} \quad g(L) = \frac{\epsilon L}{k_B T} + \frac{\tau}{\tau(L)} \quad (3.35)$$

The sum (3.33) is approximated by the biggest contribution of the smallest $g(L)$, given at \mathcal{L} :

$$\begin{aligned} \frac{\partial g}{\partial L}(\mathcal{L}, \tau) &= 0 \Rightarrow \frac{\epsilon}{k_B T} - \alpha \frac{\tau}{\tau_0} \mathcal{L}^{-\alpha-1} L_0^\alpha = 0, \\ \mathcal{L} &= \left[\alpha \frac{\tau}{\tau_0} \frac{k_B T}{\epsilon} L_0^\alpha \right]^{\frac{1}{\alpha+1}} \end{aligned}$$

A Taylor development around the maximum gives:

$$\begin{aligned}
 g(L, \tau) &\sim g(\mathcal{L}, \tau) + \frac{1}{2} \frac{\partial^2 g}{\partial L^2}(\mathcal{L}, \tau) (L - \mathcal{L})^2 \\
 \Rightarrow G(\tau) &\sim e^{-g(\mathcal{L})} \int_0^{+\infty} e^{-\frac{1}{2} \frac{\partial^2 g}{\partial L^2}(\mathcal{L}) (L - \mathcal{L})^2} dL \\
 &\sim \frac{1}{\sqrt{\frac{\partial^2 g}{\partial L^2}(\mathcal{L}, \tau)}} \exp \left(- \left(\alpha^{\frac{1}{\alpha+1}} + \alpha^{-\frac{\alpha}{\alpha+1}} \right) \left(\frac{\tau}{\tau_0 \left(\frac{k_B T}{\epsilon L_0} \right)^\alpha} \right)^{\frac{1}{\alpha+1}} \right)
 \end{aligned}$$

This expression can be further simplified noticing that the prefactor contributes only an algebraic decay to the correlation function, then

$$G(\tau) \propto \exp \left(- \left(\frac{\tau}{\tau_1} \right)^\beta \right) \quad (3.36)$$

$$\text{with } \beta = \frac{1}{\alpha+1} \text{ and } \tau_1 = \tau_0 \left(\frac{k_B T}{\epsilon L_0} \right)^\alpha$$

This model of DNA breathing modes presents a simple microscopic description of the origin of the DNA fluctuations and allow us to estimate the expected autocorrelation function as a stretched exponential. Its virtue lies in the physical interpretation of the stretching parameter β and of the timescale τ_1 .

FCS measurement on 18AT-breathing constructs.

We carried on FCS measurement on the internally-tagged random 18AT construct. The method is very much similar to the method used for the molecular beacons (§ 3.3.1 on page 107), *i.e.* we used a control construct (with Rh6G but no DABCYL) to measure and divide out the diffusion contribution in the autocorrelation:

$$\begin{aligned}
 G_{\text{construct}}(t) &= G_{\text{diffusion}}(t) \times G_{\text{chemical fluctuation}}(t) \\
 G_{\text{control}}(t) &= G_{\text{diffusion}}(t) \\
 \Rightarrow G^*(t) &= G_{\text{construct}}(t)/G_{\text{control}}(t) = G_{\text{chemical fluctuation}}(t) \quad (3.37)
 \end{aligned}$$

$$= B + P e^{-(\frac{t}{\tau_{fluc}})^\beta} \quad (3.38)$$

One resulting autocorrelation function is presented in Figure 3.35. The typical relaxation timescale is around 50 μs , and the autocorrelation curve is obviously not a mono-exponential: it must be a superposition of many fluctuation modes relaxing with timescales from 1 μs up to 1ms.

Using the model of § 3.4.2, we expect the the superposition of the breathing modes to be a stretched exponential. To measure the stretching parameter β , we measure the plateau P (below 100ns) and the baseline B (between 1 & 10ms), and plot: $g(t) = \ln(P) - \ln(G^*(t) - B)$ as a function of t . In a log-log representation, this plot unravels the expected power law ($\ln(g(t)) = \beta \ln(t/\tau_{fluc})$), yielding β and τ_{fluc} . In the case of the internally-tagged random 18AT construct, at 23°C, in 0.3M NaCl, we find $\beta = 0.50 \pm 0.01$, and $\tau_{fluc} = 76 \pm 4 \mu\text{s}$.

To interpret these values, we use the analytical expression presented in equation 3.36:

$$\begin{aligned}
 \beta = 0.5 &\Rightarrow \alpha = 1, \\
 \Rightarrow \tau(L) &= vL \text{ (where } v = \frac{L_0}{\tau_0} \text{).}
 \end{aligned}$$

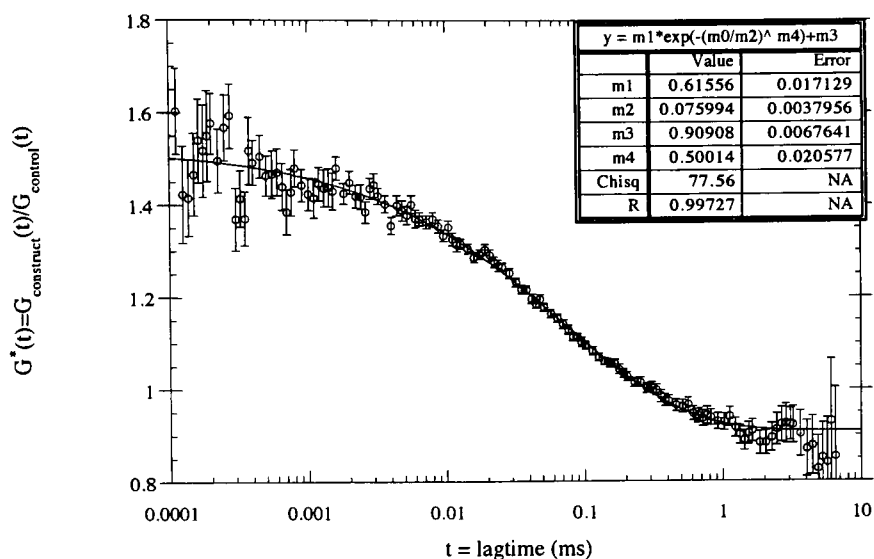


Figure 3.35: Autocorrelation function for the internally-tagged random 18AT construct at 23°C (buffer: 0.3M NaCl, 10mM cacodylic acid, 1mM EDTA pH8.0)

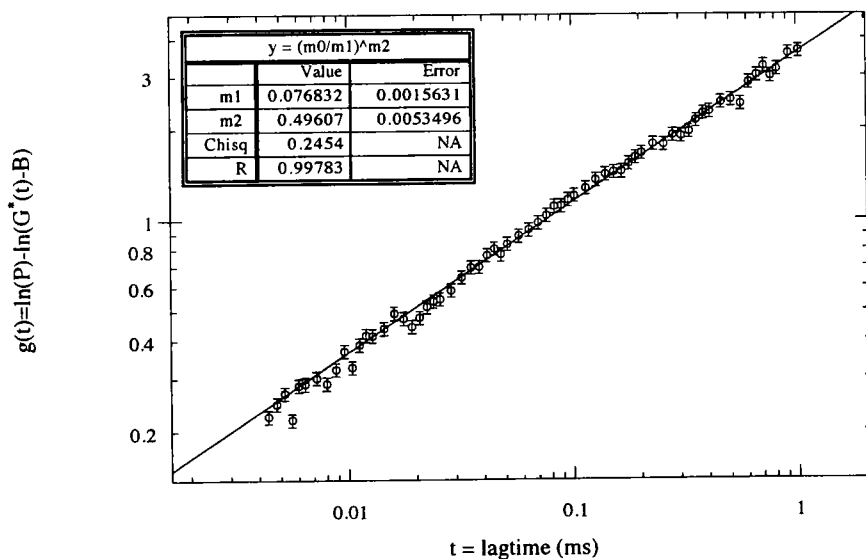


Figure 3.36: Autocorrelation function for the internally-tagged random 18AT construct, rescaled into $g(t)$ to unravel the stretching parameter ($T_{sample} = 23^\circ\text{C}$, buffer: 0.3M NaCl, 10mM cacodylic acid, 1mM EDTA pH8.0)

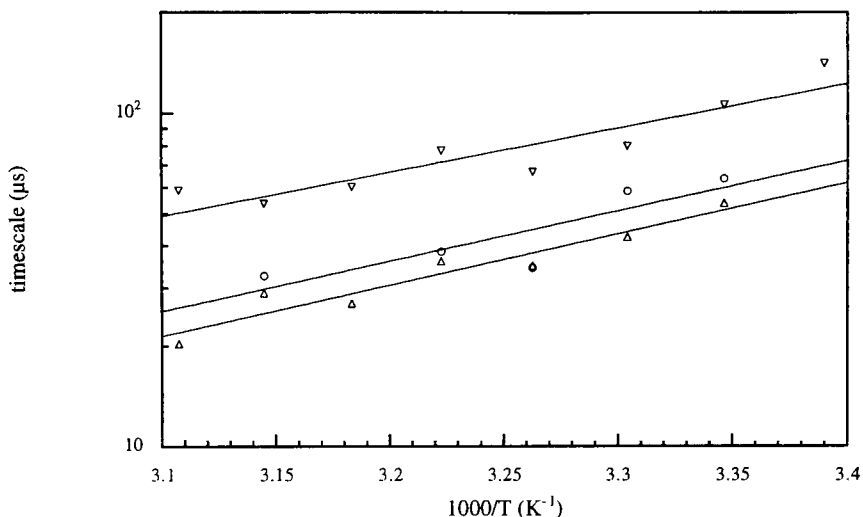


Figure 3.37: Arrhenius plot of breathing timescales for 18AT-fluctuation box constructs of sequence: (o) Random sequence *AATATAAATAATTAAAAT • ATTTAATTATTTATATT*; (Δ) *A₁₈ • T₁₈*; (∇) *(AT)₉ • (TA)₉* (buffer: 0.1M NaCl, 10mM cacodylic acid, 1mM EDTA pH8.0).

Thus, we find that the closing timescale for a bubble of size L scales linearly with L , *i.e.* a ssDNA bubble must simply zip from a prenucleated basepaired domain. To estimate the zipping speed v , we use an estimate of the free energy of one AT basepair: $\epsilon \sim 0.5k_B T$, such as $v \sim 1/2/(76\mu s) \sim 1$ base *per* $40\mu s$. This zipping speed seems rather small for an intramolecular basepairing, but we might be

A systematic study of the temperature dependence of the autocorrelation function collected from this construct showed that the stretching parameter is 0.5 at any temperature (our temperature study ranges from 22 °C to 48 °C), and the timescale of fluctuation τ_{fluc} weakly decreases as the temperature increases (from 80 μs at 22°C to 25 μs at 48°C). The temperature dependence of the fluctuation timescale could be fitted with an Arrhenius plot with an enthalpic barrier of 6.9kcal/mol (Figure 3.37). Most of the temperature dependence must come from the temperature-dependent rigidity of ssDNA (L_0 can not be considered as a constant when A-rich ssDNA stretches are under consideration – *cf* § 3.3).

As a conclusion for the study of the 18AT bubble construct, the dynamics of breathing seems to “well-behave” as predicted by a simple nearest-neighbor model. The distribution of modes simply follow a thermally-activated Boltzmann weights, and the relaxing timescale scales linearly with the size of the bubble.

3.4.3 Breathing fluctuations for shorter AT-tracts.

Premelting of the GC clamp for short AT-rich regions.

Surprising results were recorded with breathing DNA constructs with shorter AT-fluctuation box. On Figure 3.38, we present the bulk fluorescence melting curve of the 8AT-construct, with internal tagging in the middle of the AT-rich region, or with end-tagging at the end of the GC-clamp (the construct itself is presented in Figure 3.40).

The melting curve for the internally-tagged construct is qualitatively similar to that of the internally-tagged 18AT-construct (Figure 3.33A). A premelting occurs at 42°C, to reach a plateau of 8% of the total fluorescence. This premelting can be associated to the opening of the AT-rich region, the lower plateau corresponds to a higher probability of quenching in smaller bubbles

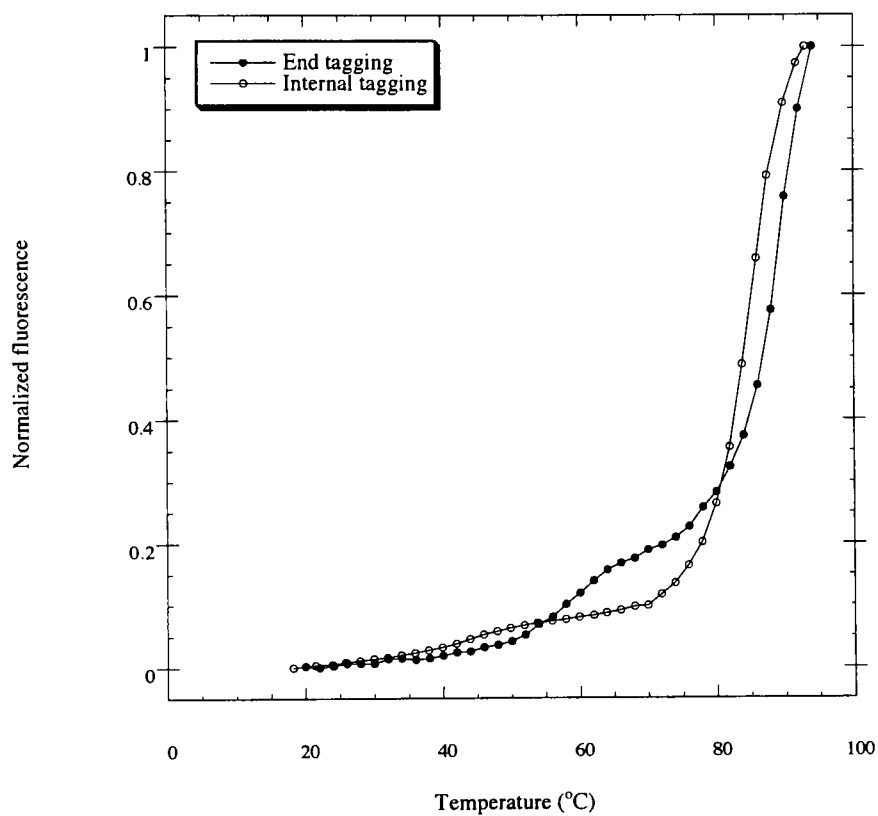


Figure 3.38: Melting curve of the end-tagged (●) and internally-tagged (○) DNA constructs with a random 8AT domain (buffer: 0.1M NaCl, 10mM cacodylic acid, 1mM EDTA pH8.0)

(fluorophore and quencher are less pulled apart). The second melting occurs at 85°C: this melting can be attributed to the total melting of the GC clamp, releasing the spatial constraint on the fluorophore–quencher pair, and thus restoring the full fluorescence of the construct.

The surprising result is the bulk fluorescence melting curve of the end-tagged 8AT-construct (Figure 3.38B). Contrary to the end-tagged 18AT-construct, the 8AT-construct melts in two steps. First, at 57°C, a first melting occurs to bring 20% of the full fluorescence; second, at 88°C, a second melting occurs to restore the full fluorescence. The occurrence of this premelting of the GC clamp in the end-tagged 8AT-construct comes as a surprise by comparison to the all-or-none melting of the GC clamp in the end-tagged 18AT-construct: whatever the sequence is, 7 basepairs upstream of the GC clamp (in the AT rich region) matters to the stability of the end of the GC clamp. This premelting seems to challenge the standard all-or-none description of DNA melting.

Is the premelting of the GC-clamp caused by the entropic constraint in the ssDNA domain ?

We design two DNA constructs to test the relevance of entropic fluctuations in the melting of the 7GC clamps. These constructs consist of the same flanking GC-clamps as in Figure 3.32, but the fluctuation box is replaced by unmatched bases: the ‘‘A-bubble’’ construct consists of 8 As’ on either side; the ‘‘T-bubble’’ construct consists of 8 Ts’ on either side. These two constructs are presenting a permanently melted domain next to the GC-clamp, whereas the end-tagged 8AT-construct presents a potentially double-stranded domain (Figure 3.39)

We measured the melting curves of these three constructs in 0.1M NaCl, 10mM Cacodylic acid, 1mM EDTA pH8.0. As expected, the ‘‘T-bubble’’ construct melts in an all-or-none fashion at 75°C. The ‘‘A-bubble’’ construct premelts weakly at ~35°C (the fluorescence plateau is 3%), and melts completely at 76°C, in a similar fashion to the ‘‘T-bubble’’ construct. The construct with 8AT matched basepairs displays a more surprising melting curve: as expected by the additional stability provided by the 8AT basepairs, this construct undergoes the large melting at higher temperature (88°C) than the ‘‘T-bubble’’ or ‘‘A-bubble’’ constructs. However, this construct displays a surprising premelting of large amplitude (with a fluorescence plateau at 20%) at 60°C: for example, at 64°C, the end-tagged 8AT-construct’s fluorescence is 15%-ON, when the ‘‘A-bubble’’ and ‘‘T-bubble’’ constructs’ fluorescence are only 7% and 5%-ON, respectively.

The premelting of the GC-clamp in the end-tagged 8AT-construct is thus not a consequence of an entropic constraint imposed by a ssDNA-domain in the AT-region: when an already-formed bubble (unmatched bases in the ‘‘T-bubble’’ construct or ‘‘A-bubble’’ construct) is present, the GC-clamp is not as destabilized. The sequence of the fluctuation box in the end-tagged 8AT-construct is 5’ – AAATAATT – 3’ is quite purine-rich (the complementary is thus pyrimidine-rich). The sequence-dependent rigidity of ssDNA (unraveled with the molecular beacon closing kinetic in § 3.3 on page 106) led us to conjecture that the GC-clamp could be destabilized by the elastic constraints imposed by the base stacking of purines in the ssDNA domain: the AT-rich domain being less stable than the GC-rich domain would premelt into a single-stranded DNA bubble, which could display some rigidity because of base-stacking, this rigidity distorting the flanking GC-clamps and inducing the melting of the basepairs in the GC-clamp to relax the elastic constraint.

The melting curve of the ‘‘A-bubble’’ construct proves this conjecture wrong: the premelting of the GC-clamp for this construct is not as strong as in the case of the end-tagged 8AT-construct. Thus, the premelting of the GC-clamp in the end-tagged 8AT-construct is a consequence of the AT-basepairs in the fluctuation box.

Sequence-dependence of the premelting of the GC clamp 6AT fluctuation boxes.

The premelting of the GC-clamp comes as a surprise as it does not fit the nearest-neighbor model for the formation of secondary structure in DNA. This model would have predicted that the melting of the 7GC basepairs at the end of our DNA constructs would melt in an all-or-none transition,

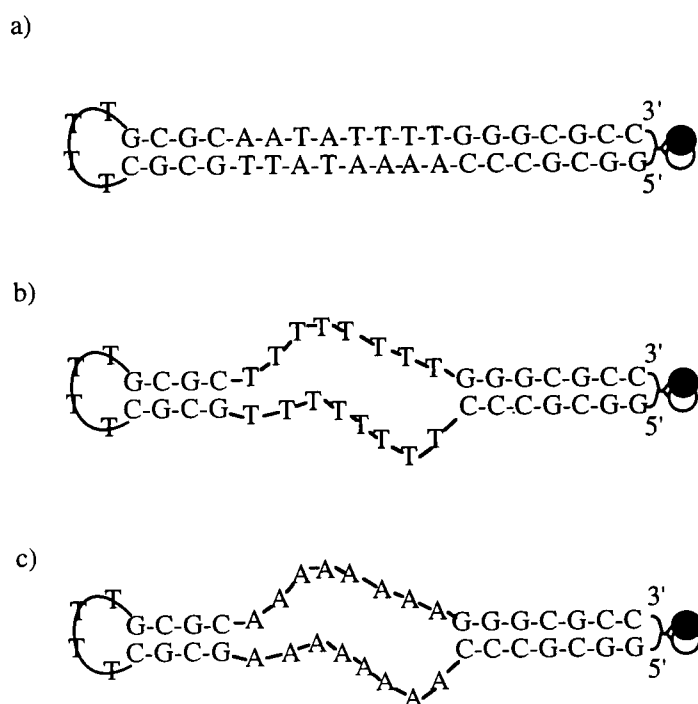


Figure 3.39: DNA construct for study of the entropic constraint on the GC-clamp: a) end-tagged 8AT-construct; b) "T-bubble" construct; c) "A-bubble" construct.

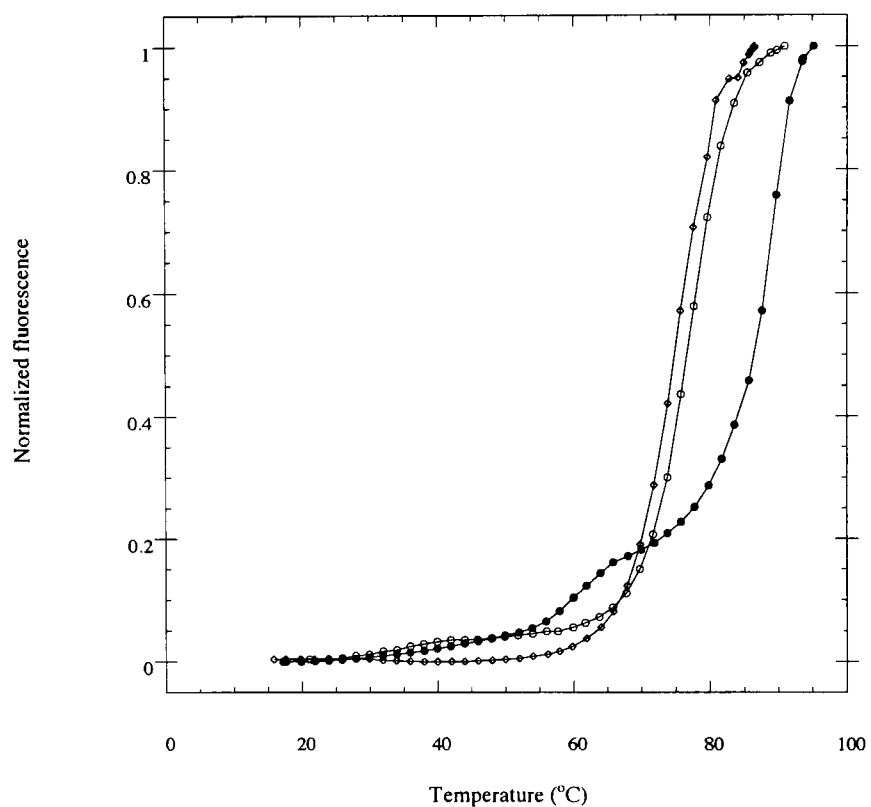


Figure 3.40: Influence of entropic bubble on the melting of the GC-clamp for different breathing DNA constructs: (○) construct with an entropic bubble made out of 8 unmatched As' on both strands; (◊) construct with an entropic bubble made out of 8 unmatched Ts' on both strands; (●) construct with 8-matched basepairs (buffer: 0.1M NaCl, 10mM cacodylic acid, 1mM EDTA pH8.0)

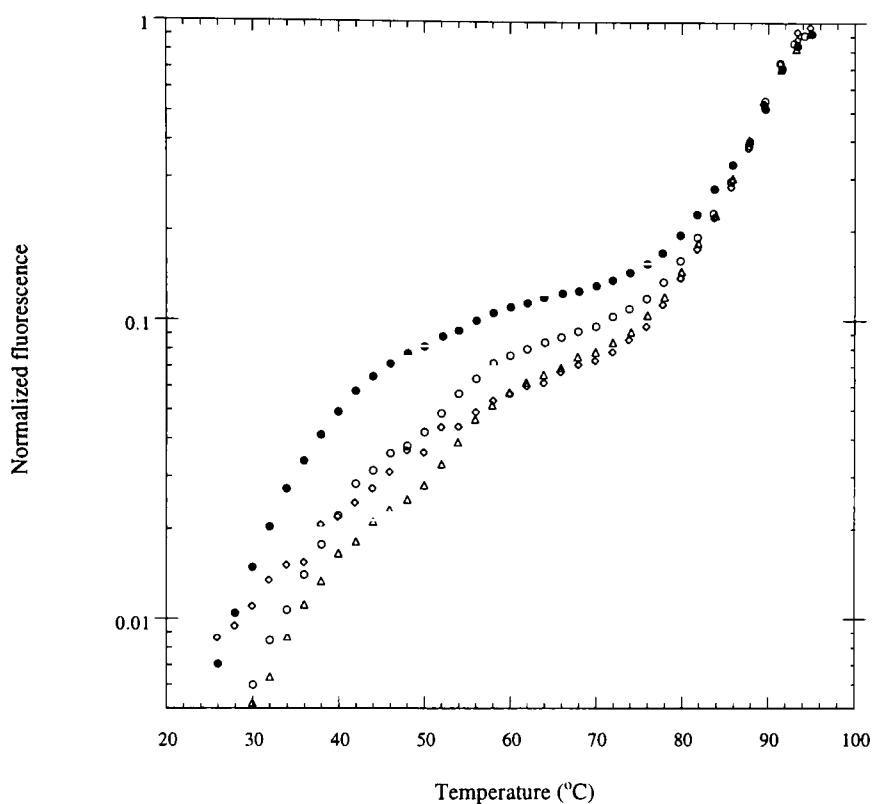


Figure 3.41: Melting curves of DNA breathing constructs with a 6AT-basepairs fluctuation box: (●) A Box construct; (○) Pribnow Box construct; (◊) Anti-Pribnow Box construct; (△) AT Box construct.

whatever the sequence upstream. We showed that this is not the case for short AT-rich fluctuation domains, and that we could not explain the premelting by the entropy provided by the ssDNA domain next to the 7GC-clamp. This observation is interesting as it unravels long range breathing modes in dsDNA which may be of biological relevance in the understanding of the specificity of the transcription initiation in prokaryotes (see § 1.3.3 on page 29).

In this paragraph, we present melting curves of 4 end-tagged DNA breathing constructs with a fluctuation box made out of different 6AT boxes (Figure 3.41). The sequences of the boxes (upstream of the 3' end of the construct) are:

5' – AAAAAA – 3'	(A Box construct)
5' – TATAAT – 3'	(Pribnow Box construct)
5' – ATTATA – 3'	(Anti – Pribnow Box construct)
5' – ATATAT – 3'	(AT Box construct)

The melting curves are monitored for a solution of 10nMol of DNA construct in 0.1M NaCl, 10mM Cacodylic acid, 1mM EDTA pH8.0. The fluorescence is normalized by taking the baseline fluorescence at low temperature and the plateau fluorescence at high temperature. The signal/noise ratio of these DNA constructs being typically 60, the dynamic range of the melting profile spans 2.5 decades (from 0.003 to 1 in normalized fluorescence). We present the melting curves in log-normal coordinates to emphasize the premelting difference between the different constructs.

These melting profiles for the end-tagged breathing constructs with 6 AT-basepairs fluctuation boxes with different sequences show that the premelting of the 7GC basepairs-clamp is very sequence

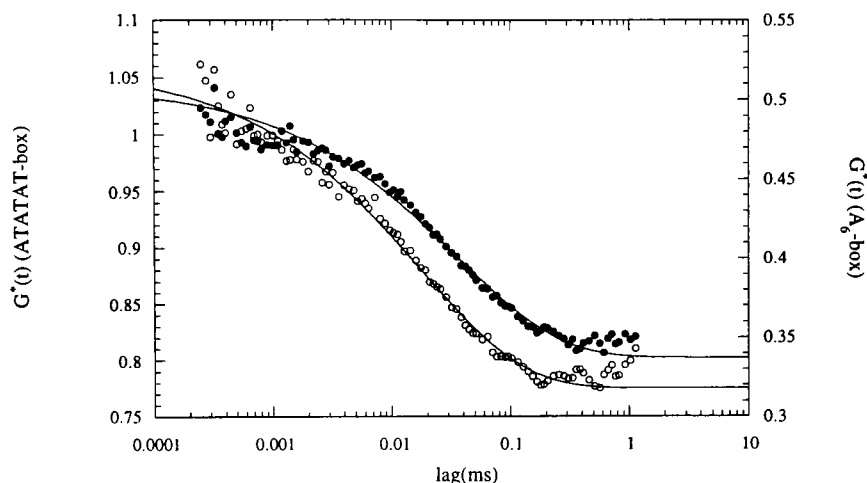


Figure 3.42: Autocorrelation functions for the chemical fluctuation of two 6AT breathing constructs with end tagging at 45°C: (●) A-Box, (○) AT-Box. The continuous lines are fits with $G^*(t) = BS + P \exp(-\sqrt{t}/\tau)$ (buffer: 0.1M NaCl, 10mM Na cacodylate, 1mM EDTA, pH8.0)

dependent. At 37°C, the A-Box construct opens up and restores 4% of the fluorescence while the AT-Box construct stays more closed and only restores 1% of the fluorescence. This premelting, though of low amplitude in this buffer, unravels sequence-specific long range deformations of dsDNA. Note that the Pribnow box construct does not induce any premelting of the GC clamp like the A-Box construct. As in the case of the sequence-specific rigidity of molecular beacons, there are two possible explanations for the sequence-specific premelting of the GC-clamp: either the AT-rich ssDNA bubble is stabilized (longer open time), either the AT-rich dsDNA is destabilized (shorter closed time). To decipher between these two kinetic possibilities, we carried out FCS measurements on these DNA constructs.

FCS measurement on breathing constructs, with short AT domains.

The FCS measurements on the end-tagged 6-AT breathing constructs yield autocorrelation functions of the chemical fluctuation which can be fitted with stretched exponential with a stretching parameter of 0.5 and a timescale τ_1 around 50 μ s at 45 °C (Figure 3.42).

We know, from the molecular beacon experiment, that the FCS measurement always unravels the fastest timescale in a fluctuation mode: as the DNA construct are statistically closed on average, we are below the melting temperature, and the timescale we measure is the lifetime of the open state (*i.e.* the closing timescale). Moreover, the temperature dependence of the stretched exponential is weak (from at 65 μ s at 23°C down to 35 μ s at 45°C): this is consistent with a very weak activation barrier (~ 9 kcal/mol), typical of a closing process (see the case of molecular beacons in § 3.3).

At low temperature (below 30°C), the autocorrelation functions are the same whatever the AT-rich sequence. However, at 45°C, where the average opening differs between the A-Box construct and the others, the timescales measured by FCS differ: we find $\tau_1 = 35 \pm 2 \mu$ s for the A-Box construct, and $\tau_1 = 17 \pm 1 \mu$ s for AT-Box construct. The longer lifetime of the A-Box construct (by a factor of 2) must be compared to the enhanced opening of the GC-clamp (by a factor of 4) at 45°C. Thus the lifetime of the open GC clamp in all these DNA breathing constructs accounts partially for the average opening of the clamp. The lifetime must also be shorter for the closed GC clamp in the A-Box construct to make up for the enhanced opening, but this can not be measured by Fluorescence Correlation Spectroscopy but should be investigated with the NMR imino-proton exchange technique (§ 1.3).

The enhanced premelting of the A-Box construct compared to the other constructs can be explained as a combination of a dynamic distortion of the double-stranded DNA structure (shorter lifetime of the paired complex) as well as stabilization of the single-stranded bubble (longer lifetime of the unpaired complex). The base stacking of poly(A) strands in single-stranded domains must present an alternative competing structure to the B-DNA at medium temperature (45°C) in 0.1M NaCl buffer. Hence the longer lifetime of the open bubble in the A-Box construct, and the enhanced premelting of the GC clamp.

3.4.4 Distortion of the GC clamp for different sizes of fluctuation box.

We constructed 5 different DNA breathing constructs, with identical GC clamps, and different sizes of fluctuation box. The fluorescence-quencher are tagging the 5' and 3' ends of the construct respectively. Thus, these constructs are designed to monitor the integrity of the GC clamp. The basepair sequences of the boxes are:

0 basepair :	5'—	—3'.
6 basepairs :	5'—	ATTATT —3'
8 basepairs :	5'—A	ATTATT T—3'.
12 basepairs :	5'—TTA	ATTATT TAT—3'.
18 basepairs :	5'—ATTTTA	ATTATT TATATT—3'

We measured the bulk fluorescence melting curves of each of these constructs in a 0.1M NaCl buffer (Figure 3.43&3.44). The melting temperature for the high-temperature transition (total opening of the GC clamp) decreases when the size of the fluctuation box (upstream of the clamp) increases. This is an entropic effect, due to the size of the ssDNA bubble, similar to the one observed in § 3.4.3.

However, the premelting transitions are very dependent on the length of the fluctuation box. At 65°C, the constructs with 0 AT basepairs and 18 AT basepairs have only melted to 7%, whereas the construct with 8 AT basepairs is 15% open. Thus, there is an optimal length of the fluctuation box (8 basepairs) to induce the premelting of the GC clamp.

Although we lack a quantitative modeling of the long-range distortion of dsDNA, we want to argue that, for large AT fluctuation boxes, distortions of dsDNA can be relaxed by unzipping further AT basepairs and thus does not perturb the neighboring GC basepair. When the AT fluctuation box is small (less than 6 basepairs), the tight structural constraint of the flanking GC clamp forbids any distortion of the AT-rich domain. When the size of the AT fluctuation box is around 6–8 basepairs, distortion modes can nucleate and perturb the neighboring GC clamp. Although the details of this model are missing, the phenomenological observation remains: large-scale distortions of dsDNA occur next to domains made out of 6AT basepairs. These breathing modes imply long-range interactions in the breathing modes of dsDNA which are incompatible with the classical nearest-neighbor description of DNA secondary structures.

In conclusion, we have unraveled long range breathing modes in dsDNA: a sequence of 6 or 8 AT basepairs can destabilize in a sequence-dependent manner the stability of 7 GC basepairs downstream. The effect is small (4 fold increase in breathing amplitude) but we will show in § 4.1.2 that it can be of crucial relevance for the understanding of the initiation of transcription in prokaryotic systems. Our FCS technique shows that a stabilization of the open structure occurs for constructs with enhanced premelting, but does not account completely for the premelting amplitude: there must be a sequence-specific destabilization of the closed complex.

For longer breathing constructs, there is no disruption of the GC clamp, which enables us to keep the breathing modes localized to the AT-rich re-

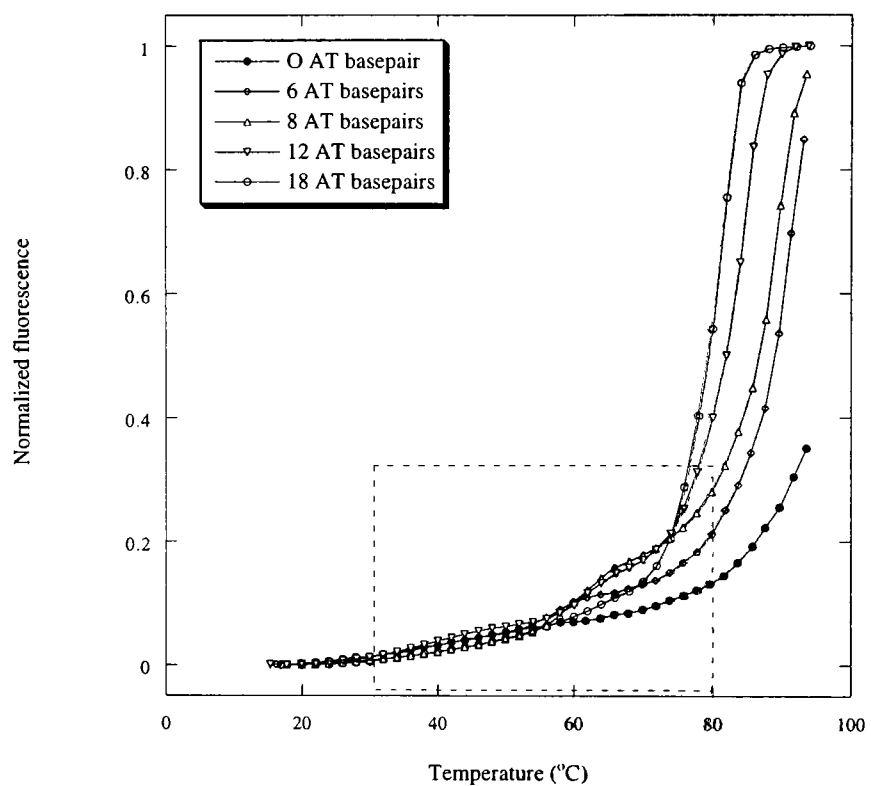


Figure 3.43: Melting curves of DNA breathing constructs with **different** lengths of *AT*-fluctuation box (buffer: 0.1M NaCl, 10mM Sodium Cacodylic 1mM EDTA **pH8.0**).

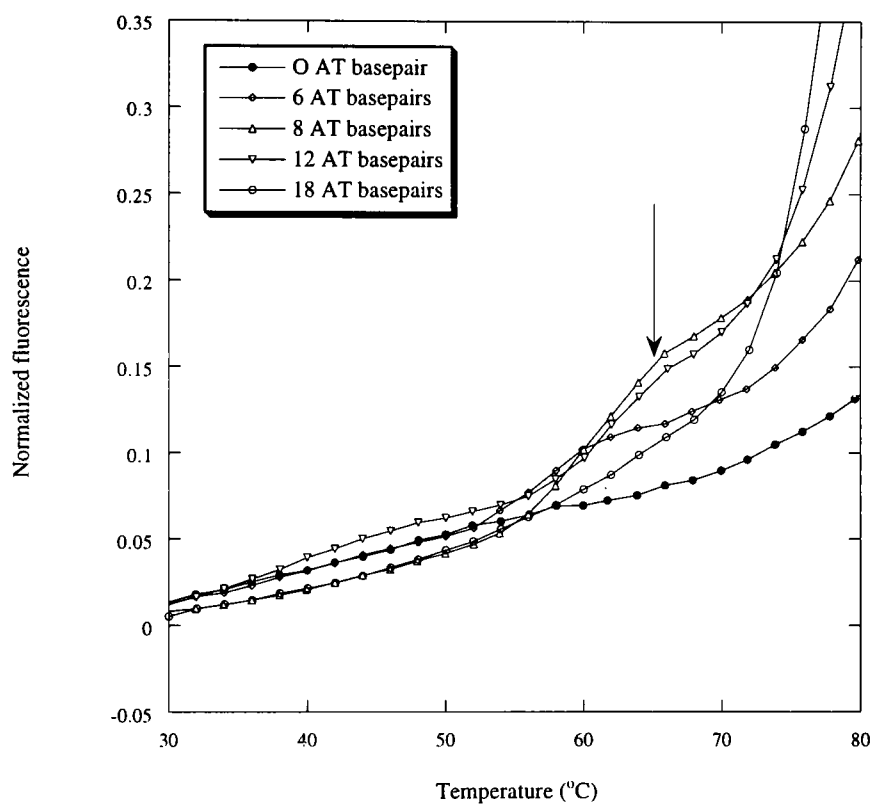


Figure 3.44: Closup on the melting curves of DNA breathing constructs with different lengths of AT-fluctuation box, from 30°C to 80°C (buffer: 0.1M NaCl, 10mM Sodium Cacodylic 1mM EDTA pH8.0).

gion. The lifetime of the open bubble (typically $50\mu s$) has been measured for the first time. The distribution of bubble-size/bubble-lifetime leads to an autocorrelation function fittable with a stretched exponential with a stretching parameter of 0.5. The case of localized breathing modes in large AT-rich domains is consistent with the nearest-neighbor model.

*Jesus Christ was a man who traveled through the land
Hard working man and brave
He said to the rich, "Give your goods to the poor.
So they laid Jesus Christ in his grave.*

*Jesus was a man, a carpenter by hand
His followers true and brave
One dirty little coward called Judas Iscariot
Has laid Jesus Christ in his grave*

*He went to the sick, he went to the poor,
And he went to the hungry and the lame;
Said that the poor would one day win this world.
And so they laid Jesus Christ in his grave.*

*He went to the preacher, he went to the sheriff,
Told them all the same;
Sell all of your jewelry and give it to the Poor,
But they laid Jesus Christ in his grave.*

*When Jesus came to town, the working folks around,
Believed what he did say;
The bankers and the preachers they nailed him on a cross,
And they laid Jesus Christ in his grave.*

*Poor working people, they followed him around,
Sung and shouted gay;
Cops and the soldiers, they nailed him in the air,
And they nailed Jesus Christ in his grave.*

*Well the people held their breath when they heard about his death,
And everybody wondered why;
It was the landlord and the soldiers that he hired.
That nailed Jesus Christ in the sky.*

*When the love of the poor shall one day turn to hate.
When the patience of the workers gives away
"Would be better for you rich if you never had been born"
So they laid Jesus Christ in his grave.*

*This song was written in New York City
Of rich men, preachers and slaves
Yes, if Jesus was to preach like he preached in Galilee,
They would lay Jesus Christ in his grave.*

Woody Guthrie
Jesus Christ

Chapter 4

Discussion.

4.1 New insights on the biophysics of DNA conformational fluctuations.

Our biophysical study of ssDNA hairpin-loops' conformational fluctuations unravels the mechanical consequences of base stacking of single-stranded DNA, as a sequence-dependent rigidity. Further study on dsDNA breathing modes shows that singular distortions induce large premelting of the B-DNA structure. These two features were unpredicted by the standard nearest-neighbor (NN) model of DNA. In this section, we discuss possible implications for the prediction of ssDNA folding, the localization of dsDNA breathing and regulation of transcription. We also propose necessary adjustments to the NN model of DNA hybridization.

4.1.1 Single-stranded DNA stacking and consequences.

Biological relevance of base-stacking in ssDNA.

Studying DNA hairpin-loop fluctuations, we showed that a loop of 21As' closes 20 times slower than a loop of 21Ts' in 0.1M NaCl, at 10°C. The enthalpic barrier associated with the closing limiting step (cyclization of the ssDNA) contains an enthalpic contribution proportional to the loop length, associated to the destacking of the ssDNA bases, for loops made out of As', whereas it is essentially null for loops made out of Ts' (§ 3.3).

For dsDNA breathing modes, we discovered long-range opening of a B-DNA helix, associated with tracts of homopurines. The end GC-clamp premelts at 60°C if the sequence AAAAAA is inserted upstream, whereas it melts only at 80°C if the sequence ATATAT is inserted upstream. We pointed out that these long-range sequence-dependent distortions of B-DNA correlate well with a mechanical rigidity of the ssDNA bubbles, predicted from the hairpin-loop measurement. However, a closer analysis of the fluctuation dynamics by FCS proved that the bubble open state does have the same lifetime, whatever the sequence of the fluctuation domain. We concluded that the enhanced breathing modes must be a consequence of the premelting of the B-DNA structure into an alternative double-helical structure (§ 3.4)

Base-stacking in ssDNA, although discovered in the 70s', has always been overlooked in the predictions for the formation of DNA secondary structure (cf § 1.2 on page 12). For example, Cantor & Shimmel's textbook [19] states that:

"In practice, we already know that single-strand properties change little with temperature, or salt concentration or length. Thus, it will be a safe assumption that values of

these properties derived from double-strands represent almost entirely the thermodynamic interactions of basepairing and double-strand stacking."

This approximation was in fact legitimate for the buffer's salt conditions used in most hybridization study: 1M NaCl. This buffer was chosen to enhance the stability of double-stranded DNA (through electrostatic screening of the negative charges of the phosphate backbone). Moreover, monitoring DNA hybridization by UV spectroscopy requires "large" quantities of DNA to produce a signal above the noise level (typically, in the μMol range). Consequently, all the DNA melting transitions occur at high temperature (above 60°C) where, indeed, base stacking in the ssDNA domains is made irrelevant.

The introduction of fluorescence quenching techniques to monitor DNA conformational changes enables us to work in biologically-relevant conditions. The robustness of the fluorescence signal gives a better detection signal/noise, we thus work with more dilute DNA solutions (typically InMol range), and also achieved a better resolution for low-temperature premelting transitions. Furthermore, our constructs are DNA hairpins where all the changes in secondary structures are intramolecular: we can then afford to work in low-salt conditions (typically 0.1M NaCl), where stacking interactions were shown to be enhanced (§ 3.3 on page 106).

ssDNA base-stacking can be relevant for many biological functions. For example, the transcription factor DAX1 has been shown to bind specifically to hairpin-loops with T-loops, folding out of the chromatine's dsDNA [145]. The puzzling fact is that, if a hairpin-loop secondary structure is encoded upstream of a gene to be transcribed on the (+) strand, the hairpin-loop's complementary sequence is present on the (-) strand. By complementarity, there are potentially two complementary hairpin-loops on the (+) strand and on the (-) strand, with the same stem, but complementary loops. Then, how does the transcription factor binding to a hairpin-loop distinguish (-) and (+) strand? Zazopoulos *et al.* [145] argued that DAX1 must match structurally the T-loop to bind specifically and set the specificity of the initiation of transcription on the (+) strand. Reviewing their data, we argue that the DNA-binding specificity of DAX1 may be regulated dynamically with the hairpin-loop secondary structure formation: the hairpin-loop's sequence on the (+) strand contains a flexible T-loop, and the hairpin-loop closes freely, whereas the hairpin-loop's sequence on the (-) strand contains a rigid A-loop, and the hairpin-loop is less likely to form. Further studies are required to distinguish the two molecular scenarii: structural specificity of DAX1 for T-loops *vs.* dynamic asymmetry of the hairpin-loop formation)...

Theoretically, the importance of the base-stacking in ssDNA lies in the introduction of a new phase in the computation of the partition function for the secondary structures in a DNA solution (*cf* § 1.2). The free energy of the coiled DNAs' is no longer a relevant reference. To form a B-DNA duplex, two complementary oligonucleotides need to overcome the structural and energetic barrier associated with the unstacking of the bases in the ssDNA domains (base-stacking in ssDNA and in B-DNA do not match structurally *a priori*). The partition function of a solution of DNA molecules with different secondary structures should take into account the competition between ssDNA structures and dsDNA structures.

One can envision corrections to the nearest-neighbor (NN) model. One would like to measure systematically the sequence-dependence of the stacking-destacking melting transition in ssDNA (this can be done by circular dichroism or UV hyperchromicity measurements). We expect the stacking interactions in ssDNA to be mostly local, and a NN model of the ssDNA melting should be valid. Using well-defined oligonucleotides (homopolymer or repeats without any other secondary structure than basestacking), a table of 10 nearest-neighbor free-energy parameters (similar to the Table presented on page 19) would result. Then, the first correction to the secondary structure's partition function is to take into account the stacked-ssDNA phase. The resulting model would still be nearest-neighbor like, and manageable analytically.

One of the first test would be to predict a paradoxical effect, uncovered by Burdet *et al.* [146, 147] and called telestability: two complementary oligonucleotides of sequence $d(C_{15}A_{15})$ and $d(T_{15}G_{15})$

were shown to hybridize at low temperature and melt at high temperature in a single transition. The paradox is that the NN-model would predict that the two basepaired domains (*GC* rich and *AT* rich) should melt independently and generate two separate melting transitions. Burd *et al.* pointed out that these oligonucleotides' sequences must introduce large-scale correlation in the basepairing, this correlation (telestability) being inconsistent with the NN-model. The corrections of the NN-model with ssDNA stabilization through basestacking might predict this telestability effect. A second (more speculative) correction would take into account the sequence-dependent and base-stacking-induced rigidity of the ssDNA domains: this would introduce hard-to-compute corrections of further long-range correlation in the NN-model.

Alternative models for DNA dynamics with possible inclusion of ssDNA base-stacking.

A nonlinear model (alternative to the NN-model) has been proposed to predict DNA denaturation curves [148, 149, 150, 151]. Such model extends the nearest-neighbor model (whereby basepairs are either fully closed or fully open) by introducing a free energy for the stretching (transversal separation) of the two complementary strands: all the possible conformations between closed and open basepair are now considered. The cooperativity of the melting transition comes from longitudinal coupling between bases, *via* stacking interaction. Noting y_n the strand separation at the n^{th} nucleotide, Dauxois *et al.* [148] write the Hamiltonian for the DNA denaturation:

$$\mathcal{H} = \sum_n \frac{1}{2} m \dot{y}_n^2 + \frac{K}{2} (y_n - y_{n-1})^2 + D(e^{-\alpha y_n} - 1)^2, \quad (4.1)$$

where m is the mass of one nucleotide ($m = 300\text{Da}$), K is the elastic modulus for the base-base longitudinal stacking (typically $K = 0.06\text{eV}/\text{\AA}^2$), D is an elastic energy for base-base transversal stretching (typically, $D = 0.04\text{eV}$), and α is the range of this interaction (typically, $\alpha = 4.45\text{\AA}^{-1}$). The last term (exponential with the base separation y_n is a Morse potential for the stretching energy of the basepair bond: this additional term is the main difference between this nonlinear model and the NN-model.

The dynamics of base fluctuations is derived from this Hamiltonian:

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial \dot{y}_n} &= -\frac{\partial \mathcal{H}}{\partial y_n} - \xi m \dot{y}_n \\ \Rightarrow m \ddot{y}_n &= K(y_{n+1} + y_{n-1} - 2y_n) + 2\alpha D(e^{-\alpha y_n} - 1)e^{-\alpha y_n} - \xi m \dot{y}_n, \end{aligned} \quad (4.2)$$

where the friction term ξ is added. This equation is a discrete version of the nonlinear continuous diffusion equation [148]:

$$m \frac{\partial^2 y(x, t)}{\partial t^2} = K \frac{\partial^2 y(x, t)}{\partial x^2} + 2\alpha D(e^{-\alpha y(x, t)} - 1)e^{-\alpha y(x, t)} - \xi m \frac{\partial y(x, t)}{\partial t},$$

which shows that the denaturation dynamics can be analyzed with a nonlinear Schrödinger equation. A classical analysis predicts the existence of soliton-like solutions for the fluctuation opening of dsDNA. This model might describe more accurately the dynamics of breathing observed in our FCS measurement, and should be specifically applied in this context. The discreteness of the system (modeled in equation 4.2) may also explain the clamping of fluctuation bubbles to specific domains of the dsDNA *i.e.* long-lived bubbles (the interface between dsDNA and ssDNA is much harder to move around in a discrete model than in a continuous model [152]).

More explicitly, the Morse potential proposed in equation 4.2 to describe the stretching of the hydrogen bonds can be modified to take into account the possibility of ssDNA base-stacking. A second minimum of free energy then appears when the two complementary strands are sufficiently pulled apart (Figure 4.1). The longitudinal coupling (associated with backbone stretching)

introduces long-range distortions of the double-helix: as soon as one basepair hops from a closed configuration to an open configuration, it can drag its neighbor into the open configuration (now stabilized by base-stacking). The second minimum brings us back to a system very much similar to Ising-like models for ferromagnetism, whereby two phases (paired and unpaired) can coexist, and all the physics come from the presence of an interface.

So far, we have emphasized the importance of single-stranded base stacking-destacking for the dynamics of formation of secondary structures. This was derived from experiments on relatively long loops of homopolymer (10-20 nucleotides), not very relevant to biological systems (it is in fact nearly impossible to find or even to design an inhomogeneous sequence of 20 nucleotides without any secondary structure). For shorter loops (below 6 nucleotides) ssDNA base-stacking is also most relevant to the stability of secondary structures once formed. Resolution of the structures of specific RNA hairpin-loops (tetraloops) demonstrated substantial structural arrangement of the bases in the loop, by base-stacking or hydrogen-bonding [153]. The enhanced stability of the RNA tetraloops might be related either to a stabilization of their closed state, or to an increased closing speed of their open state. Note that this effect would be opposite to the destabilization by ssDNA base stacking, that we observed with large DNA loops. The enhanced stability of RNA tetraloops is an open issue very much addressable with our technique combining FCS and molecular beacons.

Base stacking and rigidity of ssDNA are sequence-dependent. They might determine the limiting step to the formation of specific secondary structures involved in docking of transcription factors. More generally, they must be relevant to the stability of any DNA secondary structures, as it introduces elastic constraints, and long-range distortion of double-stranded domains.

The introduction of nonlinear dynamics to model DNA breathing modes is appealing as it introduces long-range correlations to the nearest-neighbor model. This model of DNA denaturation suggests that large-amplitude breathing modes may occur in a dsDNA ladder by energy localization, specially when base-stacking in ssDNA domains is taken into account.

4.1.2 Initiation of transcription and DNA breathing.

In the course of our study of the dynamics of DNA breathing modes, we unraveled singular long-range distortion of dsDNA. In the introduction, we presented data on the specificity of the initiation of transcription in prokaryotes. In this discussion, we use our results on the dynamics of dsDNA breathing, to address the issue of the specificity of the Pribnow box (the promoter region initiating the transcription bubble).

We have shown that specific AT-rich sequences premelts before the rest of dsDNA and can further induce the opening of flanking basepairs, 10 bases downstream of its location. The lifetime of the open bubble is short (typically 50 μ s for a 6 AT-basepair domain), but long enough to interact with RNAP, if the polymerase has been beforehand localized next to the promoter region through binding to the -35 region.

The most simple model of the initiation of transcription would predict that the Pribnow Box sequence is a good promoter because it imposes large scale breathing modes in the promoter region. However, the specificity of large-scale distortion of dsDNA does not correlate well with the specificity of transcription initiation. For example, the sequence 5'-AAAAAA-3' distorts our DNA breathing construct four times more than 5'-TATAAT-3' and much more than any other sequence, at 37°C in 0.1M NaCl (§ 3.4. On the other hand this sequence (5'-AAAAAA-3') is not an efficient promoter of transcription [68]. Where does the specificity of transcription initiation come from ?

Standard models of transcription initiation with the *E.coli* RNA Polymerase (RNAP) sums up the kinetics with two steps[154, 155]:

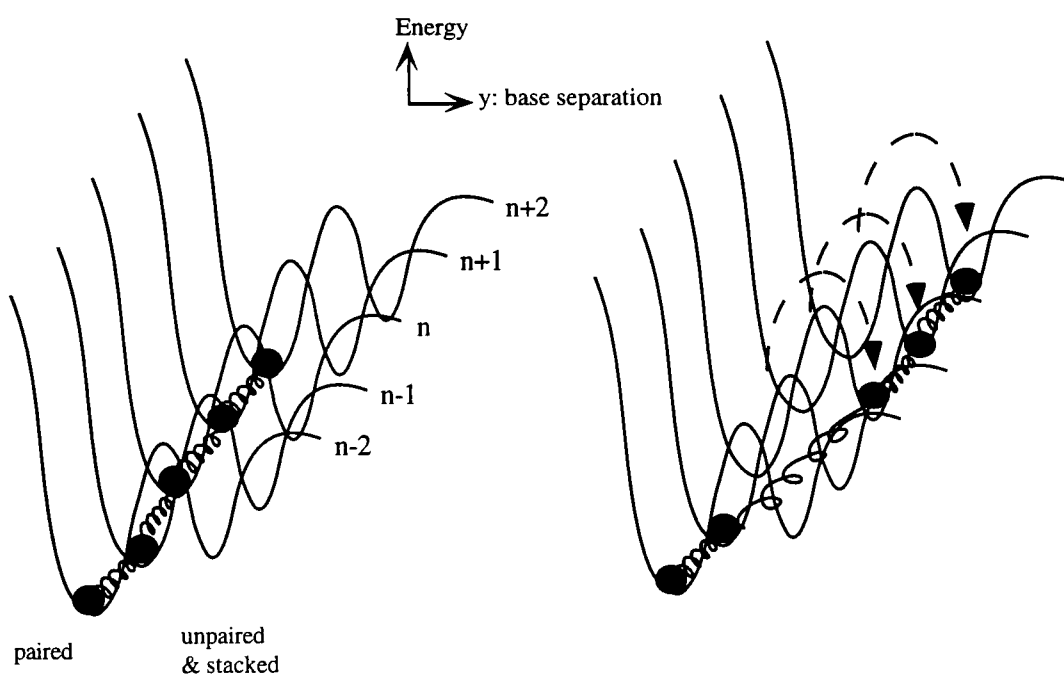
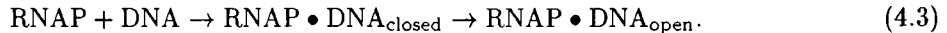


Figure 4.1: Double-well energy potential for DNA breathing and soliton-like solutions. The introduction of a second minimum of free energy for an open basepair –supposedly because of ssDNA base stacking– implies long range distortion of the DNA: as soon as one base hops to the open phase, it forces its neighbor to hop as well, hence the apparition of soliton-like deformations.



Many additional steps should be added, but we will focus on the second step of that scheme (isomerization), the subject of our work (the breathing modes of dsDNA). A crucial player in this kinetic scheme is the σ^{70} subunit of *E. coli* RNAP (RNAP is a holoenzyme consisting of 5 subunits). This subunit was shown to bind tightly with the -35 region, inducing a conformational change of a ssDNA-binding pocket in the σ^{70} subunit, which then binds to the nontemplate side of the ssDNA bubble [155].

Other prokaryotic systems have been studied (such as *Bacillus subtilis* RNAP), to extract the “substantifique moëlle” of the isomerization of the closed complex into an open complex [156]. The strong (flagellin) promoter sequence is different from the consensus Pribnow box of other organisms:

$$5' - \text{TCCGATAT} - 3',$$

but you will notice that it also contains *AT* steps. The σ subunit of the *B. subtilis* RNAP is σ^D in this experiment. The interest in the *B. subtilis* system is that it use a different σ subunit as well as a different promoter sequence, for the same function as in the *E. coli* system: the specific initiation of transcription. Chen *et al.* [156] systematically mapped the formation of the open complex with specific permanganate oxydation of thymines at different temperatures, and concluded that, in the case of this RNAP, the transcription bubble nucleates from the pseudo-Pribnow box, and propagates unidirectionally till the start site with a stabilization of the open complex procured by the σ subunit. For our discussion, the isomerization step from closed complex to open complex (Model 4.3) seems to be the same for the *Bacillus subtilis* RNAP and *E. coli* RNAP, with similar promoter sequence made out of *AT* steps, but different σ subunit. Thus, the comparison between these two transcription systems emphasizes the importance of the DNA promoter sequence in the isomerization step leading to the formation of the open complex.

Most of the recent papers concerning the formation of the open complex deal with the role of the σ subunit [157, 158, 159, 160, 155]. For example, a kinetic analysis of the formation of the σ^{70} subunit directed promoter recognition by RNAP has been carried out by time-resolved UV irradiation and controlled proteolysis [157]. This study establishes that the following initiation steps are sequential: RNAP binding to the -35 region, and strand separation at the -10 site with binding of the σ^{70} subunit onto the single-stranded non-template Pribnow box. Experiments by systematic mutation of the σ^{70} subunit [161], as well as *in vitro* selection of single-stranded oligonucleotide binding to the σ^{70} subunit [162] prove that the σ^{70} subunit contains a binding pocket interacting specifically with single-stranded oligonucleotide of sequence $5' - \text{TATAAT} - 3'$

These results limit the scope of relevance of our finding of the singular dsDNA breathing modes, as they emphasize the role of the RNAP (specifically of its σ^{70} subunit) to determine the site of transcription initiation. However, one can still assign some biological relevance to the large-scale singular distortion modes of B-DNA. As pointed out in the introduction (§ 1.3 on page 22), conformational fluctuations can be crucial in DNA-protein interaction. In the case of RNAP binding to its promoter, one can argue that, before evolving a specific binding pocket for the non-template strand of the Pribnow box, a primitive RNAP must have first been able to bind DNA bubbles. We want to propose that our singular breathing modes are evolutionary relevant to the definition of a transcription promoter. In other words, the σ subunit of RNAP evolved its bubble-stabilizing function on a sequence of dsDNA prone by itself of enhance breathing modes.

Two issues are then to be addressed: why is the Pribnow box so small (6 AT basepairs), and why did not the system evolve a larger promoter bubble as a docking site? We showed, with the large AT-domain construct, that it is crucial to keep the AT-domain small (6 or 8 basepairs) to distort efficiently the neighboring basepairs; when the AT-rich domain is larger (18 basepairs), the distortions of dsDNA seem to be confined to the AT-rich domain *i.e.* the neighboring AT basepairs may simply unzip to relax the distortion imposed by the Pribnow box. Our conclusion is that the size

of the Pribnow box might be a structural prerequisite to form a stable large scale bubble whatever the neighboring sequence.

This counter-intuitive argument emphasizes the necessity of cooperative breathing of dsDNA upon RNAP binding to open up the transcription bubble. When RNAP binds upstream of the Pribnow box, it may clamp the DNA strands into a closed structure, and thus enhance the distortion of the Pribnow box to be released downstream (§ 3.4). Although we can not provide direct evidence for this hypothesis, our FCS measurements show that GC-rich region opens up with the same timescale as the AT-rich region in the case of short fluctuation box. Thus, our study of bare-dsDNA breathing shows that the promoter region is “seeded” with potential transcription bubble.

One can argue that there is also an evolutionary pressure to keep the promoter region small: the binding pocket of the σ^{70} subunit must be limited in size to keep its specificity, and it must also be much harder to maintain large promoters. Thus, the size of the Pribnow box seems to be large enough to create large breathing bubbles, but not so large as to concentrate the disruption of B-DNA and be stable evolutionary.

The dynamics of DNA breathing modes is not sequence-dependent enough to explain the specificity of the initiation of transcription. However, the correlation between the isomerization step forming the $RNAP \bullet DNA_{open}$ complex and the enhanced breathing modes of specific dsDNA sequences with short AT regions gives an evolutionary relevance to these singular large-scale breathing modes.

4.1.3 Entropy-enthalpy compensation in DNA hybridization.

Our study of DNA hybridization with molecular beacons led us to uncover a puzzling fact concerning the thermodynamics of base-base recognition. In a nutshell, base-base interaction by Watson-Crick pairing inside a B-DNA structure involves large negative enthalpies (typically $-10\text{kcal.mol}^{-1}/\text{basepair}$) as well as large negative entropies (around $-25\text{cal.mol}^{-1}.\text{K}^{-1}/\text{basepair}$) such that the resulting free energies are small (around $-2\text{kcal.mol}^{-1}/\text{basepair}$, i.e. $1 k_B T$). In this section, we discuss this enthalpy-entropy compensation effect, and its implication in terms of DNA sequence recognition.

Many groups have worked on compiling Nearest-Neighbor thermodynamic parameters (enthalpy $\Delta H(NN)$, and entropy $\Delta S(NN)$) to predict the stability of short oligonucleotides complexes. The regularity of their results (whatever the salt condition, whatever the description of the DNA as a polymer or an oligonucleotide [26]) is striking: when reported on a $(\Delta H, \Delta S)$ graph, the nearest-neighbor parameters appear to follow a linear relationship, for example, in 1M NaCl, 10mM cacodylic acid, 1mM EDTA pH7.0 (Figure 4.2):

$$\frac{\Delta S(NN)}{1 \text{ cal.mol}^{-1}.\text{K}^{-1}} = -0.61 + 2.55 \cdot 10^{-3} \frac{\Delta H(NN)}{1 \text{ kcal.mol}^{-1}} \quad (\chi^2 = 3.9, N_{points} = 10).$$

Even more striking is the enthalpy-entropy compensation in the thermodynamics of DNA mismatch basepairs. We have found an example of such a compensation studying the hybridization of a molecular beacon onto its target: the identity of the mismatch did affect strongly the enthalpy and the entropy of the transition, but the melting temperature of the duplex was found to be unaffected by the nature of the mismatch (in our case, $G \diamond A$, $G \diamond T$, or $G \diamond G$ mismatch flanked by two $A \bullet T$ basepairs). That led us to conclude that we could not use the molecular beacon to read the nature of a mismatch for a given set of flanking basepairs.

In fact, this effect had been well documented by AbouEla *et al.* in 1985 [124], and we present their results in Figure 4.3. The enthalpic change due to the replacement of an $A \bullet T$ basepair, flanked by two $A \bullet T$ basepairs, with a mismatched basepair is strictly linear with the entropic change:

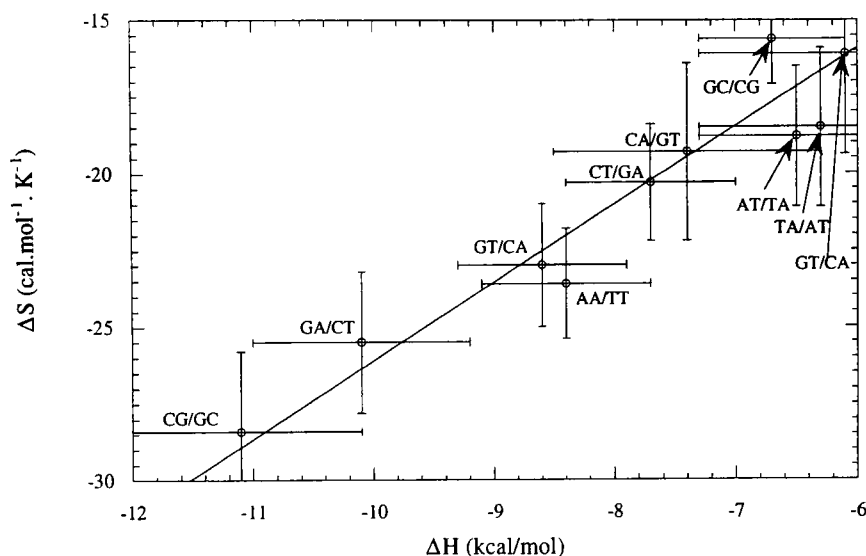


Figure 4.2: Nearest-neighbor thermodynamic parameters, in 1M NaCl, 10mM cacodylic acid, 1mM EDTA pH7.0, from [25]

$$\Delta S_{ij} = -1.9877 + 3.4426\Delta H_{ij} \quad (R = 0.98).$$

We want to argue that these realizations of enthalpy–entropy compensation in the association of complementary DNA strands is crucial to the biological function of DNA. *To be functional, a biomolecule must be able to undergo reversible association with other biomolecules.* That general statement can be justified along two lines of reasoning: first, if a functional biomolecule were to interact strongly with others, then we would introduce an irreversibility which is antinomial to any living matter; second, if a biomolecule interacts strongly with another, then it might interact strongly with many other biomolecules and thus bear no specificity. The example of the DNA molecules drove us to emphasize the importance of reversibility and specificity to any function of molecular recognition.

The main consequence of entropy–enthalpy compensation can be found in the value of the free energy of base–base interaction in DNA duplexes: whatever the sequence of the basepair, its free energy is a delicate balance of few $k_B T$, with a large enthalpy and a compensated large entropy. The stability of a DNA basepair (measured in terms of free energy) is thus marginal, and reversibly corrected by the thermal fluctuations of the bath.

It is puzzling to think that one bit (base) of genetic information is encoded with a molecular interaction weighting essentially one or two $k_B T$. This observation fits well with classical considerations on the Maxwell demon, sieving particles according to their speed to generate a temperature gradient [163]. The demon was introduced by Maxwell as a challenge to the second principle of thermodynamics. This challenge was resolved by evaluating the thermodynamic cost associated with the erasing of the memory of the demon necessary to its function. It was proven that the handling of one bit of information, necessitated at least $k_B T \ln(2)$ in free energy [164]. This entropy–enthalpy compensation makes the strength of DNA base–base interaction close to the fundamental thermodynamic limits for handling information bits.

We would like to present a speculative argument, related to the molecular structure of DNA, to justify this enthalpy–entropy compensation in DNA hybridization. As was shown in the introduction (page 5), the biggest contribution to the stability of B-DNA is the stacking interaction between adjacent bases. The hydrogen-bonding pattern of the bases provides only the specificity of the

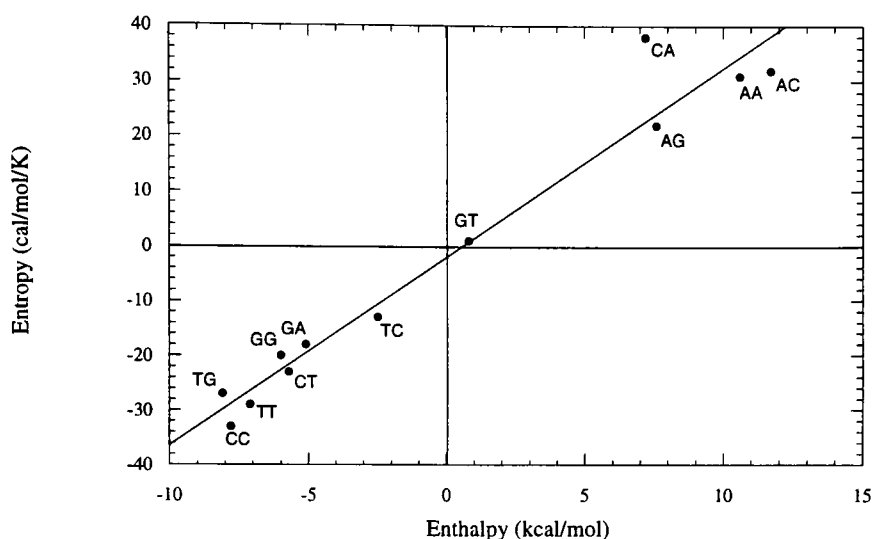


Figure 4.3: Nearest-neighbor thermodynamic parameters, in 1M NaCl, 10mM cacodylic acid, 1mM EDTA pH7.0, from [124]

Watson-Crick basepair. The stacking interaction between adjacent bases involves the electrostatic alignment of the π -electron cloud, as well as the hydrophobic effect of the water expulsion out of the base stacking. The first interaction is mostly enthalpic (it would be the same in vacuum), whereas the second is mostly entropic (it involves the number of water molecules trapped in the structure, as well as a constraint on the conformation of the phosphate backbone). DNA is such that the number of degrees of freedom frozen by the base stacking matches exponentially with the strength of the electrostatic interaction: the harder the bases stack (like in purine-purine interaction), the more constrained the bases are.

This enthalpy-entropy compensation in DNA hybridization is in fact a very general feature of ligand-receptor interactions. Qian&Hopfield have presented a general thermodynamics argument, extending the le Châtelier's principle, to show that any system will compensate enthalpy and entropy change under perturbation of its environmental constraints [165]. It is further argued that this entropy-enthalpy compensation is in fact a signature of the flexibility of the structural constraints of the thermodynamics system, flexibility which can be identified with an induced-fit conformational change upon interaction [166]. In the case of DNA molecules, one can see that this induced-fit interaction leading to the entropy-enthalpy compensation is associated with the transition from purely coiled to fully stacked conformations upon hybridization. This observation had been made by Searle & Williams [167] for ssDNA stacking, but our arguments presented here are more general to encompass dsDNA hybridization.

DNA molecules display an entropy-enthalpy compensation in their conformational transitions, which can be justified by the molecular details of base stacking. As far as biological function is concerned, one can see the advantage of compensating enthalpy and entropy: DNA basepairs are formed at the edge of stability (as measured by their free energy), thus easily and reversibly corrected by thermal fluctuations. One bit of genetic information is then encoded by one quantum of thermal energy $k_B T$. This entropy-enthalpy compensation is in fact a salient feature of any ligand binding interaction, making a molecular recognition biologically functional.

4.2 DNA molecular switches and computation.

4.2.1 DNA and molecular code.

What did we learn from the example of molecular recognition achieved by molecular beacons ? Two general features are discussed in this section: the importance of induced-fit molecular recognition (the enhanced specificity of molecular beacons as structurally-constrained DNA probes being an example) and the cascade of DNA hybridization reactions to design complex chemical networks with DNA molecules.

Molecular recognition and its specificity are ubiquitous issues to any molecular interaction inside a cell: localization of proteins, metabolites, and substrates is crucial to many metabolic pathways, yet the diversity of molecular templates must be tackled and deciphered, at the molecular level, by the enzymes to generate specific chemical reactions.

Most interactions involved in molecular recognition are weak interactions: hydrogen bonds, hydrophobic forces, electrostatic attractions and Van der Waals bonds are the interactions used in ligand-receptor binding events. The resulting equilibrium constant falls in range between 10^{-2} and 10^{-8} Mol, *i.e.* 0.5 to 20 $k_B T$ (the notable exception to these limits is the case of biotin-streptavidin binding with an equilibrium constant of 10^{-15} *i.e.* 40 $k_B T$). These binding constants are one order of magnitude smaller than the ones for covalent bonding (typically 100 $k_B T$). Thus, molecular recognition seems to involve weak binding events. The crucial consequence of this design is the reversibility of the interaction and the possibility of thermal corrections: if molecular recognition involved too strong interactions (such as covalent), then one could not take advantage of thermal noise imposed by the "bath" to dismantle wrongly-formed complexes.

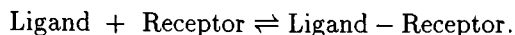
The second consequence of these weak binding constants is the stringency of molecular recognition. By definition, molecular recognition is stringent when only perfect complexes are stably-formed whereas defect complexes are not. In the case of the molecular beacons, we showed that the introduction of the self-hybridization of the beacon stem introduces a counter-weight to the trans-hybridization such that the stringency of the molecular recognition can be tuned. Weak binding constants to form perfect complexes are crucial for stringency: if the complexes were to be very stable, the energy difference between perfect complex and mismatched complex would still exist but be negligible in the overall binding energy. Then, complexes would quantitatively form, even in the presence of defects, and molecular recognition could not be stringent. The molecular beacons showed us how to optimize the stringency of molecular recognition: if the difference of binding energy between perfect and defect complexes is $\Delta\Delta G$, one must tune the structure of the probe such as the binding energy of the perfect complex $\Delta G_{perfect}$ itself is poised to almost 0 but negative, such that the binding energy of the defect complex is positive:

$$\Delta G_{defect} = \Delta G_{perfect} - \Delta\Delta G > 0, \quad (4.4)$$

and the defect complex does not form. Note that $\Delta\Delta G$ ought to be few $k_B T$ as it involves minute molecular mismatches between ligand and receptor. Thus, $\Delta G_{perfect}$ has to be tuned to be of the order of $k_B T$ to satisfy the condition 4.4, and thermal fluctuations become again the spontaneous correcting tool of imperfect complexes.

Molecular recognition of ssDNA by molecular beacons must be poised close to the thermal limit to maximize its specificity. More generally, the reversibility of a marginal molecular recognition makes it correctable by thermal noise. Thus, in presenting an algorithm to optimize the sensitivity of molecular beacons, we showed how to tune the induced-fit of a DNA probe into its target, and make the association marginal.

This discussion on stringency is limited to the simple case of molecular recognition, which, in a more general description, is a bimolecular interaction:



The thermodynamic limit on the stringency is determined by the ratio of the equilibrium constants between perfect and defect complexes ($K_{\text{perfect}}/K_{\text{defect}}$) *i.e.* by the free energy difference between perfect and defect complexes ($\Delta\Delta G$). Many schemes have been proposed to go beyond the thermodynamic limit on the stringency of the molecular recognition, such as kinetic proofreading [168] and regulated localization [169, 64]. A whole description of all the schemes involved in improving the stringency of molecular recognition would bring us too far from our main topic. In the following, we will focus on the concept of regulated localization, to show how it is applicable with molecular beacons.

Ptashne and Gann introduced a general paradigm to understand the specificity of gene regulation. Their initial remark was that allosteric change as well as induced-fit interaction between ligand and receptor seem harder to evolve than simple lock-and-key interaction [169]. Whatever the thermodynamics, one is always limited in the specificity of the molecular recognition. However, cells take advantage of cooperative binding of transcription factors onto DNA template to generate high level of stringency in the initiation of transcription. The solution consists in co-localizing the gene regulators with three-party interaction: two gene regulators bind weakly onto the DNA and weakly between each other, thus the three-body complex sums up the free energies and builds a specific tight complex. Thermodynamically, that colocalization works because the mere increase of the local concentration of the two gene regulators is enough to generate a higher level of stringency. Ptashne & Gann point out that this strategy is highly evolvable as it simply bundles up independently-evolved "lousy" machines to generate cooperative interaction.

This approach to enhance cooperativity and specificity of molecular recognition is attractive: it is modular and scalable to machineries of higher and higher complexity. The simplicity of this design makes it also applicable to the design of more specific DNA probes. Let us propose a scheme based on molecular beacons to read specifically long sequence of DNA (this scheme was designed with Sanjay Tyagi following his proposition of a ribosome-free machinery as an step in the evolution of protein translation [170]).

The DNA template to be "read" is considered as a concatemer of ssDNA targets (T_1, T_2, T_3, \dots) with a hairpin-loop at its 5' end (Figure 4.4), leaving a short overhanging single strand (called S) at the 5' end. Several molecular beacons (B_1, B_2, B_3, \dots) are designed with a loop complementary to the sequence in the template (the first beacon is targeted towards T_1 , the second towards T_2 , etc.), and a stem made of the sequence S and its complementary. The hybridization of the first beacon onto the template uses the overhanging S as well as the first ten bases of the template T_1 and switches on the first fluorophore (Figure 4.4b). The newly-formed complex presents a structure very similar to the initial templates: a single-stranded template with an overhanging single-strand of sequence S at the 5' end. The next hybridization (B_2 on sequence T_2) switched off the fluorophore of the beacon B_1 , switches on the fluorophore of the beacon B_2 , and presents again an overhanging S, to prolong the chain reaction of hybridization. And so on and so forth...

One may wonder whether we can enforce a controlled step-by-step hybridization (why doesn't B_3 hybridize onto the template before beacon B_2 ?). The trick is to tune the sequence S such that the duplex $B_i - T_i$ is not stable, but the "three-way" junction $S - B_i - T_i$ is: the beacon B_i must wait that beacons B_1, \dots, B_{i-1} have hybridized to be able to interact with its target (the thermodynamics of the three-way junction are well documented [171, 172]). This scheme is tunable as well as sequence specific: each step can be tuned by design of the stem $S - \bar{S}$ strength to be exquisitely stringent; as soon as one hybridization fails because of mismatches, the whole chain reaction is stopped. On Figure 4.5, we present the results of a small numerical simulation of the fluorescence trace as a function of time (simulation presented in the appendix § B). The fluorescence of each fluorophore is turned on/ off one by one at regular interval of times (beacon B_n at time n); diffusion and absence

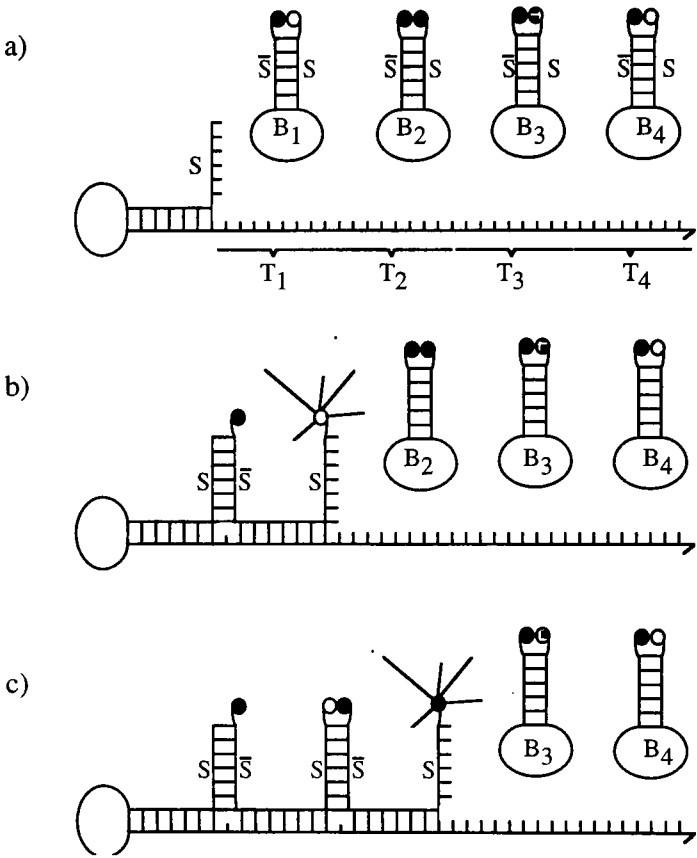


Figure 4.4: Schematic of the hybridization cascade

of synchronization spreads out the peak of the fluorescence (the width of the fluorescence peak of beacon B_n at time $t(n) = n$ is $\Delta t(n) = n^{0.7}$; as n increases, the fluorescence peaks are less and less separable because $\Delta t(n)/n = n^{-0.3}$ decreases); the presence of one mismatch is enough to stop the chain reaction and we would then be able to identify the location of a single mismatch along a very long ssDNA template.

This chain-hybridization scheme is a proto-translation system where the molecular beacons act as pseudo-tRNA: their loop sequence is complementary to the template, and their stem brings together different moieties. In fact, Sanjay Tyagi speculates that the colocalization of the two residues at the complementary ends of the stem of beacons B_i and B_{i+1} would be sufficient to produce a covalent bond (this scheme would then be a ribosome-free translation system: of course, its efficiency and processivity would be questionable but it would still constitute a proof of principle). The absence of energy consumption makes the stringency and processivity of the reaction questionable: tRNA anti-codons are 3-base long and thus highly stringent, the ribosome structure can also enhance the stringency, and the consumption of energy enables the synchronization and kinetic-proofreading of the assembly of the peptides...

4.2.2 DNA self-assembly and nanotechnology.

Molecular self-assembly is an appealing approach to design well-controlled nanometer-scale objects. It is classically described as a bottom-up approach, by opposition to the top-down approach used in most mesoscopic design and fabrication (CMOS, DNA chips). DNA molecules constitute a good building block for supramolecular assembly and bottom-up design of complex molecular objects.

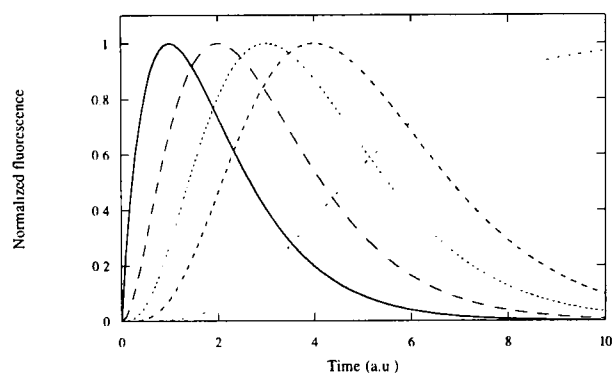
In this section, we review examples of self-assembly of DNA supramolecular structures and their applications. Most experiments have focused on the construction of complex DNA crystals (static structure), whose function would be mostly to impose a well defined crystallographic order to other biomolecules. Recent developments have introduced the design of DNA versatile assemblies whose function is to act as a DNA biosensor.

The tiling of a 2D surface by rectangular tiles (Wang tiling) with sticky edges has been shown to constitute an example of Turing computation. Winfree *et al.* [173] developed DNA building blocks to achieve a Wang tiling. This experimental *tour de force* constitutes an example of molecular self-assembly driven by a predetermined grammar: the hybridization between sticky ends of their DNA tiles self-assemble DNA sheets of mesoscopic scale.

Let us discuss the crucial experimental issue in this experiment of DNA self-assembly. Most DNA structures are too flexible to accommodate large-scale self-assembly. For example, if one designs a stretch of dsDNA with complementary sticky ends, these DNA molecules should self-assemble into an infinitely long 1D neck-lace of DNA. Unfortunately, the rigidity of dsDNA is not infinite, and a strand of dsDNA longer than 50nm (the persistence length) can accommodate some bending. Thus, any assembled necklace tends to close on itself (the sticky ends being complementary), and shortcuts any further hybridization of dsDNA, any self-assembly is then limited to finite sizes. In the Wang tiling with DNA molecules, Winfree *et al.* used a double-crossover molecule, whose rigidity forbids any impromptu termination of the self-assembly. The beautiful DNA tilings (visualized by atomic force microscopy) proved that the bottom-up approach for molecular nanotechnology could be carried on with DNA molecules: atomic structures could be designed to self-assemble into a mesoscopic order.

This self-assembly approach is very error-prone as it is driven by the huge free energy gain associated with the hybridization of the DNA tiles: like in a crystallization, it is impossible to grow a huge molecular assembly without defects. In case of the DNA tiling, there is no mechanism available to correct the defects and mistaken assembly. Thus the system will use very "inefficiently" the building blocks. That consideration is a reminder of the importance of error-correction and

a)



b)

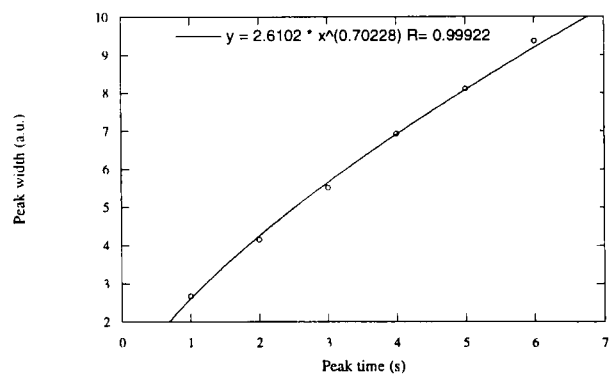


Figure 4.5: Simulation of the fluorescence curves collected from an hybridization cascade (simulation presented in the appendix § B).

recycling machinery in biological systems. Of course, biological machines (such as the the replication or the ribosome systems) do create a lot of junk products (mismatched dsDNA or misfolded proteins); however, there are often correcting machineries (RecA recombinaison, chaperones), to destroy and/or correct the mistakes.

A DNA self-assembly can be considered as the crystallization of versatile and designable elementary units made out of DNA. This approach led to many applications in the field of DNA biosensors, specifically as an hybridization probe. Mirkin *et al.* [174], and Alivisatos *et al.* [175] have developed a cute DNA hybridization sensor. In their experiments, gold nanoclusters of diameter 40nm, are functionalized with single-strands of DNA of two different sequences. The sequences are chosen to be complementary to the first and second half of a targeted oligonucleotide. Upon mixing with this particular target, the DNA strands anneal onto each other, entraining the condensation of the gold nanocluster with them, and the formation of a crystalline structure of colloidal gold. The nice feature of this experiment is that the hybridization can be monitored with the naked eye: the gold colloids, when in solution, have an absorption peak around 525nm (red). Upon clustering, these gold particles couple their plasmon resonance spectra and the peak of absorption is shifted towards shorter wavelengths at 460nm (blue). In the case of these DNA-coated nanoclusters, the introduction of the complementary oligonucleotide as a condensation catalyst can be seen directly by monitoring the color of the solution. Thus, the self-assembly of these gold nanoparticles (induced by DNA hybridization) changes the macroscopic spectral properties of the solution.

Another example of the self-assembly driven with DNA information is the connection of microscopic circuits with conducting silver wires. Braun *et al.* [176] showed that the sticky ends of a lambda DNA are sufficient to bridge specifically two conducting pads tagged with oligonucleotides complementary to the sticky ends. This molecular link is made conductive by replacing the DNA counter-ions by silver ions and reducing these ions into metallic clusters : this chemical reaction renders the silver metallic, building a conductive silver wire which bridges the two pads. In this approach, nanoscopic electronic circuits would be connected by self-assembly of DNA molecules. The advantage in terms of miniaturization of the circuitry (the silver connecting wires are nm-thick, the DNA framework being only 20Å thick) are a plus. Applications for DNA self-assembly have also been envisioned in the field of photonic crystals, or biomolecular chips. One must notice though that these self-assembled materials can only be a first generation in the design of more and more complex materials.

Our design of DNA logic gates and the proposition of the hybridization cascade (page 156) are other example of DNA self-assembly. We showed that such molecular computation with molecular beacons has the fundamental drawback of not-separating the processed information (the DNA sequences) from the energy required to carry out the computation (the hybridization free energy). To design a viable molecular machine to process DNA sequences, one must envision a non-equilibrium machine which uses external sources of energy to read and process the DNA sequences. So far, all the molecular computations and DNA self-assembly we have described are carried on at equilibrium: the information processing is too tightly linked to a fall towards a minimum of free energy. Only the classical DNA amplifiers ($Q\beta$ replicase system, or Polymerase Chain reaction) can escape the belittling free energy requirements of any DNA processing. Examples of such schemes (DNA computation and DNA amplifiers) are presented in an appendix (§ A)

A molecular self-assembly driven by DNA-DNA interaction is simply a crystallization problem thus it is not a very sophisticated assembly. The biological information encoded in the DNA molecules is simply used to drive thermodynamically the assembly, and this approach relies in the design of a global free energy minimum. The biggest limitation lies in the absence of error correction: when DNA tiles are self-assembled, when an hybridization reaction proceeds, one relies completely in the specificity of the DNA hybridization. Any mistake in the pairing of the tiles will introduce irreversibly a defect

in the final structure. This example points out the superiority of biological systems with their error correction mechanisms: a biological computation can be seen as a delicate molecular self-assembly where the building blocks, if misused, can be recycled.

4.2.3 DNA Molecular switches.

The design of the two logical gates (OR and AND) was our main motivation to develop our project around the molecular beacons. Although very limited in applications, these simple molecular switches help us to envision different means of computation. In this section, we present design of complex molecular switches with DNA molecules as a building block. We emphasize the need of DNA-modifying enzymes to build a more efficient molecular computation with these DNA switches.

The 90s' have seen the parallel development of quantum computation [177] as well as DNA computation [178]. Both fields are attacking the limits set by classical computation with silicon-based chips. Processor chips have been shrinking in size following an exponential law for the last 50 years. Extrapolations predict that the chip industry will hit the atomic limit (whereby the size of a bit storing unit is shrunk to the Angstrom scale) around 2015. Thus, researchers have been stimulated in developing molecular switches of nanometer-scales for computation purposes.

In fact, molecular switches are also appealing in biomedical applications. One can envision complex molecular systems, based on a concatenation of molecular switches, used to regulate and cure *in vivo* medical diseases. During our PhD work, we have witnessed the introduction of many kind of molecular switch as detection system. For example, the *chameleon* device was introduced in 1995 to monitor intracellular Calcium level (a Calmodulin folds on itself in the presence of calcium to increase the level of FRET between two GFP proteins fused at its two ends) [179]. A genetically encoded, fluorescent indicator for cyclic AMP was introduced in 1999, based on the molecular conformational switch of a protein kinase A (PKA) in the presence of cAMP [180]. Similarly, a chemical derivative of the SrK kinase was synthesized to detect tyrosine phosphorylation, based on the hypothetical conformational change of SrK kinase upon the addition of the phosphate group on its tyrosine (work in T. Muir lab at the Rockefeller University).

More generally, the paradigm of molecular switch seems to be useful in many biological system, such as bacterial phase variation [181], two-component systems in prokaryotes... It is also actively applied in the development of chemically assembled electronic nanocomputer (CAEN), as illustrated with the release of electronically configurable molecular-based logic gates in 1999 [182].

How far could we go with our own DNA molecular switch ? The limitations of our crude logical gates are related to thermodynamics limits: the OR gate simply concatenates two hybridization reactions, the AND gate takes advantage of a DNA ligase to drive an hybridization reaction. In both cases, the signal/noise ratio of these molecular gates was not very good (essentially the signal/noise of the beacon itself). Although the problem of signal/noise can be considered as a mere detection problem, it is in fact the fundamental limitation faced with any molecular switch.

For DNA molecular switches, the signal/noise is fully determined by the kinetic barriers associated with the conformational transition. The key to a successful design of a molecular switch lies in the lifetime of each of the conformational state. Think of a molecularly-encoded memory device: one wants to switch the conformation of the molecule to "write", and the new conformation should not switch back by fluctuation towards another conformation to avoid memory erasure. Thus, the reduction of the conformational fluctuations (or slowing down of the transition time) is crucial.

For example, for a simple molecular beacon, only two states are relevant: fully open, or fully closed. The lifetime of the closed state is fully determined by the stability of the stem basepairs, and can be large at low temperature; the lifetime of the open state is always short whatever the temperature. Thus, molecular beacons alone are not very good molecular switches as they are always

fluctuating between the two states. However, the molecular beacon in equilibrium with its target is a good molecular switch as the lifetimes of the two possible states (closed-unhybridized and open-hybridized) are both long. The kinetic barrier to leap from one configuration to another corresponds to the unzipping of typically more than 5 basepairs and thus, once all the targets are hybridized to their beacons, the solution contains two kinetically separated species (the unhybridized beacon and the hybridized beacons).

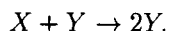
As pointed out earlier in the motivation to this thesis work, the predictability of nucleic acid folding makes the design of DNA switches very amenable. Two groups have already presented successful design of molecular switch out of nucleic acid: Seeman's group used the double-crossover motif (presented in the previous section) to build a molecule which switches ON its fluorescence when switching from a B-DNA to a Z-DNA configuration [48]. Breaker's group concatenated two RNA ribozymes (RNA domains with enzymatic activity): their two-module construct couples a ligand-specific allosteric switch in the aptamer (first module) to the activation of an RNA-cleaving ribozyme (second module) [183, 184].

It is easy to envision higher and higher kinetic separation between conformational states (and thus better molecular switch): one just has to increase the number of basepairs to be unzipped in the transition. In Figure 4.6, we present an example of a molecular-beacon-like construct which exists in essentially only two states: the two states almost exactly the same free energy (*modulo* a small entropic correction for loops vs. bubbles) but the kinetic barrier between the two states is arbitrarily large. At low temperature (25°C), in good saline buffers (1M NaCl), the lifetime of each state can be expected to exceed the experiment time.

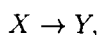
Once one has secured the kinetic separation between the conformational states of a molecular switch, one has to be able to address them, and to control the switching from one to another. In the case of the molecular beacon, the introduction of the target oligonucleotide was the experimental handle on the switch. However, as pointed out in the section (3.1.2) on page 94, the optimization of the molecular beacon requires a lengthening of the stem which introduces longer and longer kinetics (the operational time for the experimenter). This transforms a kinetically-robust molecular beacon into a slow finicky molecular switch.

This shortcoming emphasizes the importance of non-equilibrium transition for the molecular switch, as exemplified with the AND gate. One must envision the use of enzymatic active process to switch the DNA constructs from one conformation to the other (and thus bypassing the kinetic barriers). For example, in the case of long double-stranded DNA domains, strand-exchange can be actively driven with ATP-consumption and the use of RecA. Preliminary results show that RecA is indeed efficient to promote strand-exchange between a long hairpin-stem (40 basepairs with 5 mismatches) and perfectly complementary single-stranded oligonucleotide (40 bases): the final complex is more stable but would not have been accessible through spontaneous displacement of the stem by the ssDNA [185]. In conclusion, allosteric changes of the molecular switch must be driven catalytically, if we want to bypass the kinetic barriers necessary for a good kinetic separation and a good signal/noise.

In addition, one could take advantage of autocatalytic and amplifying schemes to enhance the signal/noise of the molecular switch. Autocatalysis is the ubiquitous condition for non-monotonous (oscillating) chemical kinetics, as was shown with the Belousov-Zhabotinsky reaction [186]: far-from-equilibrium chemical reaction generates oscillations of the reactants from the dissipation of the available chemical free energy. The autocatalysis is an extraordinary case of nonlinearity in chemical kinetics, summed up in the following reaction:



Classical consideration simplifies this reaction to the spontaneous conversion:



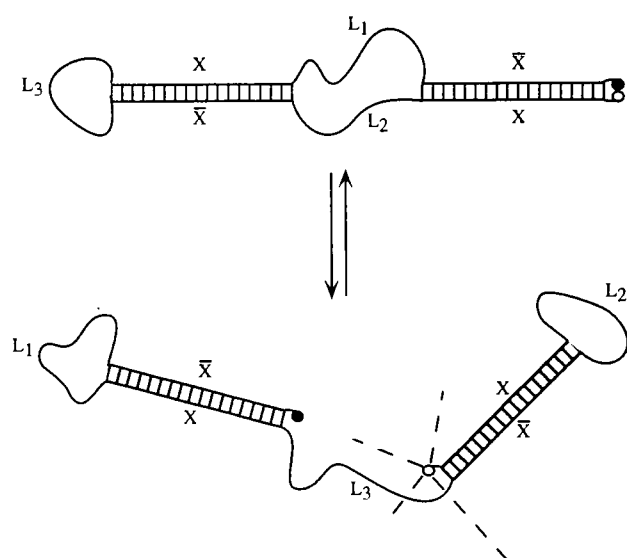


Figure 4.6: Example of DNA-based memory: this molecular-beacon like construct has two energetically-equivalent secondary structures, separated by huge kinetic barriers.

catalyzed by Y. The amazing feature is the autocatalysis whereby the reaction generates a product which is a catalyzer to its own production: it is a very "explosive" reaction by nature. We have spent a lot of time trying to design (at least on paper) such autocatalytic reaction with DNA molecules to no avail. The RecA-induced opening of large-stem molecular beacons is one trail we will be following. Two recent papers have presented an *in vivo* implementation of autocatalytic reactions, based on nucleic acid transcription-translation and transcriptional regulators [187, 188]. They show the fundamental role of autocatalytic reactions to build molecular oscillations.

In conclusion, it is possible to design complex molecular switches out of DNA. Our knowledge of DNA folding dynamics makes the design easier to implement with DNA molecules than with peptides. The need of enzymatic catalysis to drive the allosteric changes of the DNA was recognized to be crucial to the successful design of more complex molecular switches. Thus, once again, DNA, with its library of DNA-modifying enzyme, is a good building block to design molecular switches of increasing complexity.

Afterglow & Acknowledgments.

Now it is time to close this thesis work and thank the many people I interacted with during these five years at the Rockefeller University.

I would like to thank my advisor, Albert Libchaber, for brain-draining me to America, and enabling me to pursue *une éducation expérimentale*. There are many things one can learn while working with Albert: dimensional analysis, chaos theory, fuzzy logic, information theory... But for me, the best was the enthralling passion for good experiments which I wish to withhold. Amazing how much the Socratic pedagogy did teach me !

I would like to thank Steven K. Burley, for accepting to chair my thesis committee, and Magda Konarska, Stan Leibler and Marcelo Magnasco for accepting to ponder on my writing.

Two very dedicated experimentalists gave crucial pushes to this PhD work. Sanjay Tyagi, with his visionary craftsmanship, allowed me to play with his molecular beacons, and shared many brainstorming coffee discussions on the molecular switches. Didier Chatenay, through discrete puffing and humming, and a generous camaraderie, kicked off the tinkering on the FCS experiment. Also, coffee never tasted better than when the room filled up with cigarette smoke.

I want to thank David Thaler for acting as a dare-devil introducing me to the illuminating mysteries of living matter. Mark Goulian was the genius surfer of sublime ideas, sharing unselfishly all the tricks of the trade. John Marko and Éric Siggia taught me enough polymer statistical physics to fantasize on DNA contorsions. Fred R Kramer inoculated me the virus of typographic obsession.

Effie Asvestas, our dedicated secretary, has been the real beacon of this PhD, monitoring all the life of the lab, and entertaining the lab with sweets and cookies.

The gentlemen in the Instrument Shop of Rockefeller University, John Doherty, Heinz and Shao-Hui Li, taught me the beauty of aluminium dust and introduced me to the poetry of craftsmanship.

The Dean's Office at Rockefeller University was crucial to set up the free and stimulating environment for this PhD work. I would like to thank Marta Delgado, Kristen Cullen, Marguerite Mangin and Georges Cross for their incredible availability and efficiency.

I want also to name and thank my High School and Undergraduate teachers, who stimulated an unbounded fascination for science, with appropriate passion and skepticism: the late Marie-Louise Lenoble, Monique Dufour, Monique Rosier, Claude Mesnil, Rémy Briançon, Paul Manneville, Alain Arnéodo, Stéfan Fauve and Bernard Castaing.

Finally, here is my Academy-awards-like acceptance speech.

I wish to thank Monique and Roger, my Mom and Dad, for introducing me very early with Lego designing, library haunting, piano tickling, world remaking... This work is dedicated to them.

The biggest partner in crime, for most of the work and life during this PhD, is Oleg Krichevsky. Oleg, you are the big brother, the know-how chum with unlimited oversight on our endless scientific discussion. I wished you had stayed a little longer to perfect the powerplay. Remember: the best

preachers always came from the Neguev.

The lab was, very much, the cheery surrogate family Elian wished he had: Bahram Houchmanzadeh, Benoît Dubertret, Deborah Fygenson, Elisha Moses, Frank Vollmer, Jun Zhang, Mario Feingold, Masaki Sano, Michael Schnall-Levin, Michel Calame, Noel Goddard, Roi BarZiv, Shivashankar, Shumo Liu. I tried to render the cackling cacophony of the lab by quoting your enigmatic languages, and prove to myself that, after five years, *Ich bin ein New Yorker*.

The drunken raft of Rockefeller was also a great harvester of friends: Hervé, Sylvia and Guillermo (and Lucas), Philippe, Sylvie and Pascal, , Dennis, Kambiz, Joe, Diana, Luc, Pablo, Radha, Rassika and Philippe, Sidharta, Salvatore. And *les visiteurs du soir* (in chronological order): Hakan and Jeanette, Bérengère and Xavier, Nurca and Erkan, Claire, Charlotte and Cyrille, Gilles, Raphaële and Benoît, Alain, Hatem, Véronique and Hugues, Elcin, Sandrine and Jean-Baptiste, Valérie and Franck, Boris, Damien and Christophe, David, Joyce and David, Olivier, Isabelle and Éric. . .

There are some friends I did not meet enough while cramming this thesis, but were there during some exhilarating moments: Ludwig, Franz, Thelonious, Fats, LJ, Latrell & Marcus, Professor Longhair, Lou, Franck, Tony, Raymond O', John Z, Orhan, Naguib, Russell, the bald eagles, William Jefferson, Rudi, no, not Rudi, Woodie and Woodie, Jean-Luc, Steven (not Spielberg, the other Steven) and Spike, Jennifer, Ashley and John M., Léo, and all the others. . . I wish also to thank the M66, the subway n°6, Greyhound'71 and Air France for helping the great escapes.

The most monumental event of this PhD might not be directly related to research, but it is definitely the most enthralling. How would have I had expected, starting my PhD, that I would encounter a swinging lokum to share my life with. Nihal, the sunshine off-the-beaten-path, the biology fairy, the melting-pot cooking, the New York wanderer and discoverer, I love you, and I am ready for a new step in life.

New York, Monday May 1st 2000.

यदा संहरते चायं कूर्मोऽङ्गानीव सर्वशः ।
इन्द्रियाणीन्द्रियार्थेभ्यस्तस्य प्रज्ञा प्रतिष्ठिता ॥ ५८

यदा सम्-हरते च भयम् कूर्मः अङ्गानि इव सर्वशः ।
इन्द्रियाणि इन्द्रिय-अर्थेभ्यः तस्य प्र-ज्ञा प्रति-स्थिता ॥

The Bhagavad Gita (LVIII)

*When also, like a tortoise its limbs,
he can withdraw the sense from sense-objects
his wisdom is then set firm.*

The Bhagavad Gita
Translation by Srimath Swami Chidbhavananda

Appendix A

DNA Computation and amplifiers.

In this section, we review the introduction of complex chemical protocols to process DNA molecules. We will focus on the scheme of DNA computation proposed by Adleman to solve NP-complex algorithm problems, as well as examples of DNA amplifiers (PCR, Q_β replicase amplifier...). These examples show how clever DNA-based schemes can be designed to achieve complex processing tasks in solution. These schemes were inspirational in the development of the molecular beacons as molecular switches.

A.1 DNA computation.

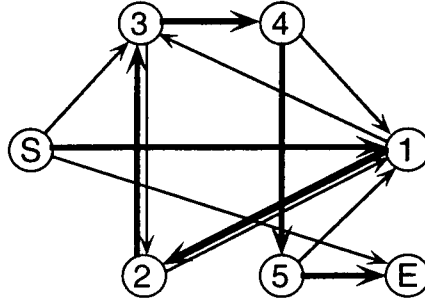
The idea that DNA molecules could be used to solve complex algorithms has been introduced by Adleman in 1994 [134]. In his groundbreaking experiment, Adleman encoded a NP-complete problem with DNA oligomers, and solved the problem (existence of and identification of eventual solutions) in a linear number of steps : this constituted a proof of principle that the parallelism of molecular assembly could be driven in order to solve complex algorithms.

The problem solved experimentally by Adleman is the classical "traveling salesman" problem. A salesman is supposed to visit a set of connected cities starting from a given one, passing once and only once through each city, finishing by another given one, A set of connecting paths between cities is given (Figure A.1A). The algorithmic complexity of this problem is non-polynomial (it would take steps to enunciate all the possible graphs achievable with the given set of cities and linking paths. Adleman used the combinatorial approach provided by DNA self-assembly to generate all the possible graphs compatible with the given set of paths in one single step: the algorithm was then limited to $N+1$ steps to check that the salesman visited all of the N cities, and all of them only once.

Each city ($O_1, O_2, O_3...$) is encoded by an oligomer. The path $O_i \rightarrow_j$ linking O_i to O_j is, by construction, complementary to the second half of O_i concatenated with the first half of O_j , such as it can self-assemble as shown in Figure A.1B. All the oligomers consisting of the N cities and all the oligomers representing the paths in the initial problem are mixed together and let to self-assemble. A ligation step enables the fixation of all the generated graphs self-assembled by DNA hybridization. Adlemans trick consist of fishing out all the concatenated oligomers of cities, selecting out the solutions starting from the city S , finishing at the city E , and passing them through N filters one by one selecting the ones containing each of the N cities. At the end, only the self-assembled graphs, solutions to the problem, will remain in solution (*i.e.* the oligomer consisting of the N cities, starting from S , finishing at E).

The severe practical limitation to Adlemans approach is that the NP-complexity of the problem requires an exponential amount of DNA to encode the problem: one must make sure that all

a)



b)

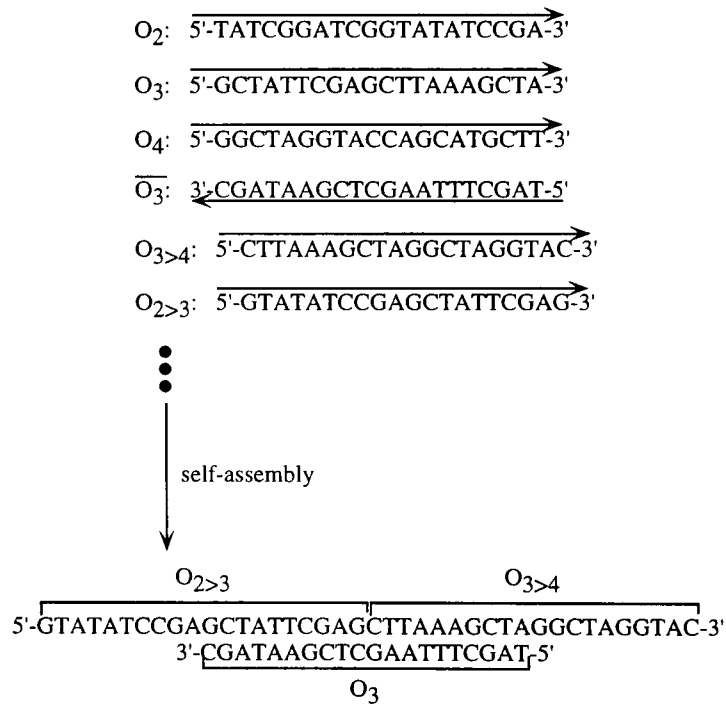


Figure A.1: a) The traveling salesman problem: a salesman is supposed to visit the N cities of this graph, traveling along the drawn paths connecting different cities, starting from city Start (S), finishing in city End (E), visiting all the cities once (1,2,...,5) and only once. A possible solution of this problem is highlighted with bold connecting paths. b) Encoding of the graph with DNA oligonucleotides: O_i is the 20-mer oligonucleotide coding for the city i , \bar{O}_i is the complementary oligonucleotide to O_i , and $O_{i>j}$ is the oligonucleotide encoding for the path between i and j .

the possible graphs are generated, such as the correct solution is generated with a non-negligible probability. For example, to solve a 7-city problem, Adleman used of oligonucleotides.

The success of Adleman's DNA computation relied heavily on the use of the polymerase chain reaction (PCR) as a DNA amplifier. It is the main "non-trivial" chemical reaction, which enables the specific and selective fishing of a correct graph out of all the non-correct graphs. In the next section, we discuss the polymerase chain reaction as one particular DNA amplifier used routinely in molecular biology. Other DNA amplifiers have been proposed for diagnostics purpose, such as the $Q\beta$ replicase system introduced by S. Tyagi *et al.* [189], or the ligase chain reaction (LCR, introduced by F. Barany [190]).

A.2 DNA amplifiers.

DNA or RNA processing enzymes enable the design of chemical reactions working strongly out of equilibrium. These enzymes couple a consumption of energy (in the form of ATP hydrolysis) to some chemical modification of DNA: they introduce a fundamental irreversibility in the chemical processing of the information encoded in the DNA. The practical applications of these enzymes in Molecular Biology are obvious, as exemplified by the development of Recombinant DNA technologies. In this section, we are discussing the use of these DNA processing enzymes to design DNA amplifiers or other chemical networks.

Polymerase Chain Reaction (PCR). The Polymerase Chain Reaction (PCR) is a clever scheme proposed by Mullis *et al.* [191] to amplify exponentially specific double-stranded templates. It relies on kinetic separation of oligonucleotide hybridization and template-mediated DNA polymerization to achieve an amplification by cycle. The initial solution contains the amplicon to be amplified (at low concentration around $1pMol$), two different primer oligonucleotides (complementary to the two 3' end of the amplicons) (at moderate concentration $1\mu Mol$), and nucleotides (at high concentration $1mMol$). The PCR consists in temperature cycles of the solution between three temperatures.

Let us present one typical temperature cycle. First, the DNA material is denatured by raising the solution temperature to $95^{\circ}C$. Second, the solution temperature is cooled to $50^{\circ}C$, and DNA materials anneal and hybridize onto each other. The primers are highly concentrated: their hybridization onto the amplicon is much faster than the reannealing of complementary amplicons ($\tau_{primers} = (k_{diffusion} * concentration)^{-1} = (10^6 * 10^{-6})^{-1} = 1s$) \ll $\tau_{amplicon} = (10^6 * 10^{-12})^{-1} = 10^6s$). Thus the primer-amplicon complexes are formed as initiation complex for the DNA polymerase. The third step of the PCR cycle is carried out at $72^{\circ}C$, which is the optimal temperature for the Taq Polymerase to polymerize the complementary amplicon from the seed of the primer.

Thus, from one amplicon, one PCR cycle synthesizes two amplicons. This exponential amplification of the DNA is sequence-specific as only amplicons accurately flanked by two primers at its 3' end will be amplified (an amplicon flanked by only one primer will be amplified only linearly).

PCR relies heavily on DNA hybridization between short primers and long amplicons. All the sequence specificity relies on the stringency of the primer hybridization. PCR is the best tool to fish out one gene with two flanking primers. However, it is insensitive to any single-nucleotide polymorphism between the two primer sequences.

$Q\beta$ replicase amplification. Tyagi *et al.* developed an alternative scheme of nucleic acid amplification to detect pathogenic agents in clinical samples [189]. This scheme was introduced out of practical reasons (reduced cost compared to PCR, and fast speed of amplification), but fundamental aspects of the $Q\beta$ replicase assay makes it quite complementary to the PCR: mainly, the main noise in PCR detection of a pathogenic DNA comes from the false-positive amplification of unrelated

pieces of DNA. In other words, a PCR scheme does not “proofreads” the primer hybridization and allows false amplification.

The DNA amplifier in this experimental scheme is based on the specificity and efficiency of the bacteriophage Q β replicase, an RNA polymerase. Its amplification template is the naturally-occurring MDV-1. The assay consists of two steps: first, two recombinant RNAs’ hybridize to their template and are ligated together into a MDV-1 particle; second, the reconstituted MDV-1 is amplified isothermally by the Q β replicase to generate a detectable signal. Thus this Q β replicase assay consists in one single step of hybridization-ligation onto the target DNAs’, followed by a robust amplification by the Q β .

The advantage of this detection-by-amplification method is its signal-to-background. Tuning the reaction to eliminate all the false-negative diagnosis¹ is easily achieved by diluting the template and asking the threshold below which the amplification scheme does not work. In the case of the Q β replicase system, Tyagi *et al.* showed that a single MDV-1 particle is enough to generate a detectable amplification. Thus, the Q β replication conditions have been tuned to avoid a false-negative diagnosis. The interesting feature is that, for these conditions, no false-positive detection is generated [189]. Indeed, the Q β replicase system relies on one single step of hybridization-ligation (the main source of noise in any amplification scheme). By comparison, the Polymerase Chain Reaction is understandably more prone to generate false-positive diagnosis as each amplification cycle involves one hybridization step: as soon as one wrong amplicon is assembled, it is flanked by the right priming regions and amplified exponentially as any right template.

If one detects only the final amount of amplified DNA, one can not rule out a false-positive diagnosis in a PCR reaction, whereas the Q β replicase system is very much less prone to generate such a false-positive.

A.2.1 Quantitative PCR.

The false-positive diagnosis problem has been solved in the case of a PCR detection scheme by being able to monitor the rate of DNA amplification. This technique has been named quantitative PCR.

Monitoring the advancement of a PCR reaction is achieved by tagging fluorescently the amplified DNA. Many alternative fluorescent tags have been introduced, from the least sequence-specific (ethidium bromide), to the most sequence-specific molecular beacons). Note that the development of the quantitative PCR has been driven by the development of PCR machines with fluorometer features (*light cycler*) by the Perkin Elmer Corporation.

Ethidium bromide is a DNA-intercalating dye whose fluorescence is magnified when the dye is stacked between DNA basepairs. Thus, if ethidium bromide has been introduced in a PCR reaction, the monitored fluorescence is low during the first cycles (the available nucleic acids -primers and nucleotides- are mostly single-stranded), and raises as the double-stranded DNA are assembled by the polymerase during the amplification cycles. The signal generated by ethidium bromide is the least sequence-specific as it only quantitates the amount of amplified double-stranded DNA.

DNA hybridization probes give the best signal for quantitative PCR. First introduced by Perkin Elmer, TaqMan probes are single-stranded oligonucleotides whose sequence is complementary to a specific sequence of the amplicon. A fluorescein and a rhodamine are tagging respectively the 5’ and 3’ ends of the probe: when intact, the two dyes are in close proximity and coupled by FRET. If excited at 488nm, the probe fluorescein is excited and transfers its energy to the rhodamine which relaxes by emitting some fluorescence at 550nm. During the PCR reaction, the TaqMan probes can hybridize onto the already-amplified DNAs (the same way the primers do hybridize but

¹By definition, a false-negative diagnosis is an experiment where no nucleic acid has been detected while the template to be detected was present in the initial solution. To the contrary, a false-positive diagnosis is encountered when the amplification scheme generates some detectable nucleic acid while no template was present in the solution. For medical applications, false-negative diagnosis are of course more dangerous than false-positive ones -it is always possible to cross-check a false-positive diagnosis by a complementary method-. Yet, one must work to minimize all the noise and consecutive false diagnosis.

at different loci along the template). When the Taq polymerase starts from the primer and knits the complementary amplicon, it encounters the hybridized TaqMan probe. Its exonuclease activity enables the polymerase to digest this obstacle, and the TaqMan probe is chewed into individual nucleotides. Thus, separate fluorescein and rhodamine are released in solution, and, if excited at 488nm, the solution will now fluoresce at 514nm. The ratio of fluorescence emitted at 488nm and 514nm is thus a direct measure of the concentration of intact vs. digested DNA probes, thus a direct measure of amplicon concentration. The TaqMan technique is thus a quantitative PCR as the sequence-specific probes are digested at each run of amplification to generate a quantitative fluorescent signal.

Another example of quantitative PCR can be achieved with molecular beacons [192]. The idea is similar to the TaqMan system except that one does not need to invoke the exonuclease activity of the polymerase. Molecular beacons are introduced in the PCR reaction volume such as it can hybridize onto the amplicons during the low-temperature phase (always the temperature when primers are annealing onto the amplicons). Monitoring the fluorescence during that phase enables a precise measure of the beacon hybridization onto the amplicons and thus a precise quantitation of the already-amplified DNA. The resulting amplification curve can be fitted with a one-parameter fit, yielding the initial concentration of template DNA. The nice feature of the molecular beacons are the possibility of multiplexing (*i.e.* mixing different beacons complementary to different alleles with different fluorophore reporters), the exquisite sequence specificity and the high signal-noise.

Stolovitzky and Cecchi [193] proposed a branching-process model of the PCR reaction to predict the amplification probability for an amplicon as a function of the number of cycles. Their model enables them to predict the initial number of amplicons once the replication probability is measured. In fact, these two parameters have to be taken as unknown, but measurements of the distribution of replication probability and initial template concentration yields their variances, from which the initial template concentration can be reliably extracted. The application of quantitative PCR are enormous in all gene expression, diagnostics and genotyping fields.

To conclude, many complex chemical reactions have been designed to process DNA sequences. PCR, LCR, Q_{β} replicase system are wonderful examples of DNA amplifiers. They all rely heavily on DNA hybridization as the sequence reading step, hence our interest in being able to tune the sensitivity of DNA hybridization. The example of DNA computation was also an incentive to develop molecular beacons as molecular switches.

Appendix B

Kinetic model for the hybridization cascade.

In this paragraph, we present a simple kinetic model for the hybridization cascade we introduced in § 4.2.1. A series of beacons ($B_1, B_2, B_3 \dots$) is designed to hybridize successively onto a large ssDNA template (whose sequence is divided into concatenated targets $T_1, T_2, T_3 \dots$).

Our goal is to present the expected shape of the successive fluorescence peaks (arrival time and width). The hybridization cascade can be modeled with n kinetic equation (this model is oversimplified as it assumes that the $B_i - T_i$ duplex is unstable whereas the $S - B_i - T_i$ -three-way junction is stable):

$$\left\{ \begin{array}{l} T + B_1 \rightarrow TB_1 \\ TB_1 + B_2 \rightarrow TB_1 B_2 \\ TB_1 B_2 + B_3 \rightarrow TB_1 B_2 B_3 \\ TB_1 B_2 B_3 + B_4 \rightarrow TB_1 B_2 B_3 B_4 \end{array} \right. \Rightarrow \frac{\partial [TB_1 \dots B_i]}{\partial t} \propto [TB_1 \dots B_{i-1}][B_i] - [TB_1 \dots B_i][B_{i+1}].$$

To simplify, we choose an excess of beacons such that $[B_i](t) \sim \text{cste}$. Then,

$$\frac{\partial}{\partial t} \begin{pmatrix} T(t) \\ TB_1(t) \\ TB_1 B_2(t) \\ TB_1 B_2 B_3(t) \\ TB_1 B_2 B_3 B_4(t) \end{pmatrix} = \begin{pmatrix} -2 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} T(t) \\ TB_1(t) \\ TB_1 B_2(t) \\ TB_1 B_2 B_3(t) \\ TB_1 B_2 B_3 B_4(t) \end{pmatrix} + \begin{pmatrix} T(0) \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Classical solutions for this linear differential equations are:

$$\left\{ \begin{array}{l} TB_1(t) \propto te^{-t+1} \\ TB_1 B_2(t) \propto \frac{t^2}{4} e^{-t+2} \\ TB_1 B_2 B_3(t) \propto \frac{t^3}{9} e^{-t+3} \\ TB_1 B_2 B_3 B_4(t) \propto 1 - te^{-t+1} - \frac{t^2}{4} e^{-t+2} - \frac{t^3}{9} e^{-t+3} \end{array} \right.$$

The graphs of the fluorescence profiles as a function of time are plotted in Figure 4.5A on page 158 (we normalized the curve peaks to 1, as one does not expect the different fluorophore-DABCYL pair to have the same signal/noise). This solution could perfectly be generalized to much larger cascades involving several more beacons. The maximum of the concentration of $[TB_1 \dots B_i]$ is at time $t = i$.

From these profiles, we can also deduce numerically the width Δt_i of each peak ($\Delta t_i = t_i^{(+)} - t_i^{(-)}$) where $[TB_1 \dots B_i](t_i^{(+)}) = [TB_1 \dots B_i](t_i^{(-)}) = \frac{1}{2}$). The resulting graph is presented in Figure 4.5B. As in a diffusive process, the width of the peak increases with the square root of the peak time (arrival time). Thus, for a very long template, the sharpness of the fluorescence peak may get lost with the previous peak.

This simple model shows the obvious limitation of this purely-kinetic scheme (where no energy consumption is involved): Experimentally, one has to design the beacons such as the spectral overlap of their fluorophore is minimal (Tyagi *et al.* [54] showed that, up to six fluorophore can be resolved in a standard fluorometer). More difficult (but achievable) is the balance of the stringency of the DNA hybridization reactions: one must make the three-way junction stable without stabilizing too much the duplexes. A successful design of such an hybridization cascade would be a proof of principle for a more specific DNA probes as well as for a proto-translation scheme [170].

Appendix C

Temperature-control device for bulk measurement.

To measure a melting profile of our fluorescently-tagged DNAs', we needed to custom-build a temperature-control turret to mount on our fluorometer. In this paragraph, we just present the design of the temperature controller, and its tuning to show how reliable the temperature sweep in the fluorescence melting curves of § 2.2 is.

Our application required to be able to scan the fluorescence of the 3-ml solution for temperatures ranging from 5 °C to 98 °C at a reasonable rate (typically 1 °C per minute). The turret is made of a cuvette cavity, flanked by two Peltier elements (Melcor CP1.4-35-0045L-1), clamped onto a water-cooled block. The cuvette was topped with a Teflon stopper with a hole in its center to introduce a thermistor. The thermistor itself (Thermometrics AA6A8-GC11KA143L/37C) was mounted onto a 5/16" screw, which was used to clamp and tightly close the cuvette. Special care was taken to avoid the solution evaporation: the solution was topped with 200l of mineral oil, the cuvette was sealed with a piece of parafilm, and the seals were waxed with vacuum grease (no more than 50μl would evaporate after keeping the 3-ml solution at 98°C for 30min).

The temperature of the solution was measured with a thermistor calibrated in the range of temperature of interest. A single fit (Figure fig:Tcalibration) was used to calibrate the thermistor:

$$R(T) = R_0 e^{-\frac{T_0}{273.15+T}} \Leftrightarrow T = 273.15 - \frac{T_0}{\ln\left(\frac{R}{R_0}\right)} \quad (\text{C.1})$$

For the calibration of the thermistor presented in Figure C.1, $R_0 = 1.6971 \cdot 10^{-5} \text{ k}\Omega$, and $T_0 = 3956 \text{ K}$. We present this calibration to emphasize that the temperature resolution of the thermistor is 0.01°C.

The resistance of the thermistor was measured with a Keithley197 multimeter, and acquired by the computer through GPIB under Labview (National Instrument). The precision on the resistance measurement was of 0.1Ω over a resistance of 30kΩ (at 5°C) to 1.5kΩ (at 98°C). Thus the precision of the temperature measurement was of 0.01°C(1Hz acquisition).

Peltier elements were used as thermoelectric coolers as well as heaters. The interesting part of the design of this temperature-controlled turret was the tuning of the controller for the Peltier elements. We used a classical PID scheme (proportional-integrative-differential) programmed under Labview to achieve an addressable temperature control.

Let me sum up the principle of the PID strategy and its tunability. What a process control algorithm must achieve is the convergence of the system temperature to a predetermined setpoint. One has to turn the Peltier elements into heating if we are below the setpoint, and turn them into cooling if we are above. Thus, the Peltier should deliver a heat power P whose sign is the sign of the difference (actual temperature T - setpoint T_0): the simplest choice is to tune the delivered power

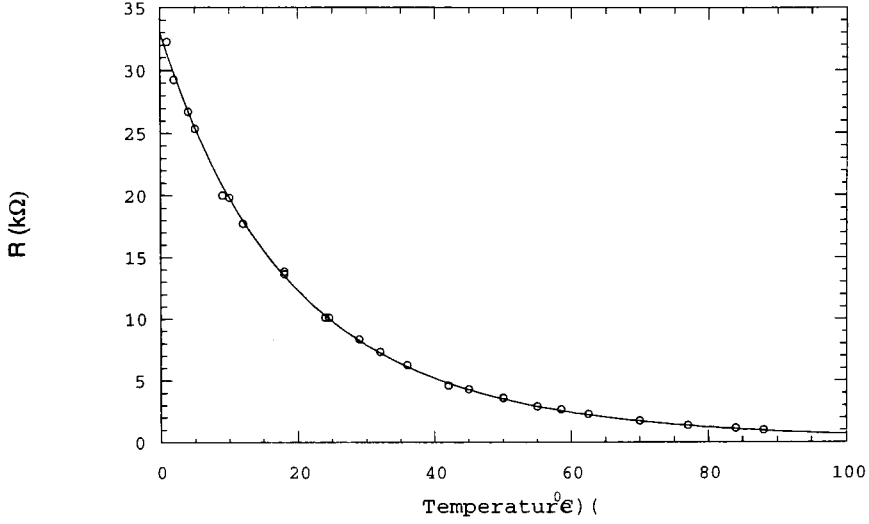


Figure C.1: Temperature calibration of the thermistor. A thermistor is placed inside a NESLAB RTV37 thermal bath. Its resistance is monitored as the temperature is scanned by 5°C-steps every 15min. The continuous line is the fit with Equation C.1.

proportionally to the difference to the setpoint: $P = \alpha(T - T_0)$. The heat equation for the solution is then:

$$\kappa C_p \frac{\partial T}{\partial t} = \alpha(T - T_0) - P_{loss}. \quad (C.2)$$

where P_{loss} is the heat loss in the system (positive if T_0 is above the room temperature, negative in the other case). This equation leads to an exponential relaxation towards a stationary temperature, but this temperature

To correct for the heat loss, one has to introduce an integrative term which will push up the system to converge right at the setpoint:

$$\kappa C_p \left(\frac{\partial T}{\partial t} \right)_t = \alpha(T(t) - T_0) + \beta \int_{t-t_0}^t (T(u) - T_0) du - P_{loss} \quad (C.3)$$

By deriving this equation, one can see how the system converges towards the setpoint:

$$\kappa C_p \frac{\partial^2}{\partial t^2} (T(t) - T_0) = \alpha \frac{\partial}{\partial t} (T(t) - T_0) + \beta (T(t) - T_0).$$

This equation is the classical dampen-oscillator, whose obvious solution is 0 (the system has converged towards the fixed point). One controlling term (a time derivative of the temperature) is sometimes added to damp the fast response of the system.

The difficult part is thus to tune the coefficients α and β to set the system in the critical regime. Figure C.2 shows a mistuned PID controller converging to its setpoint with dampen oscillations. We ultimately implemented the PID control algorithm from Labview (National Instruments), which enabled us to implement a straightforward closed-loop (ultimate gain) tuning procedure.

Closed-loop tuning procedure (*in Libris* [194])

1. Set both the rate and reset times on the PID controller to 0

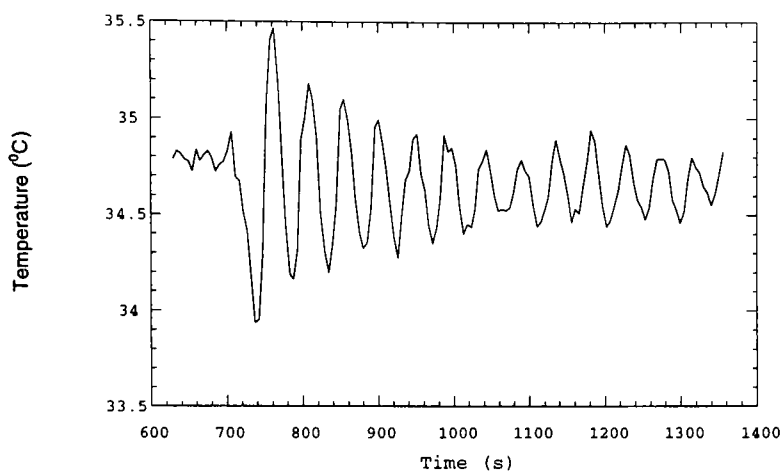


Figure C.2: Temperature recording for a mistuned PID controller. The controller does not converge nicely to the set temperature and produces oscillations slowly dampen towards the set temperature (here 35°C).

2. Reduce carefully the proportional band (PB) in small steps. Disturb the loop after each step by making a small change in the setpoint. The temperature will start oscillating as the PB is reduced. We want to achieve perfectly sustained oscillations of the temperature to be able to record the inner frequency of our system (Figure C.3).
3. When these oscillations are sustained, record the controller PB as in PBU in percent.
4. Record the period of the oscillation as T_u in minutes.
5. Set the proportional band to $1.67P_{Bu}$, the reset time to $0.5 T_u$ min, and the rate to $0.125 T_u$ min.

The result of this manual tuning of the PID controller is a very good stability of the temperature control, as well as fast response time. Note that is difficult to tune perfectly the PID controller for all set temperature (from 5 °C to 98 °C) as the heat loss and the Peltier efficiency are very temperature-dependent. The solution which gave the most consistent temperature ramps consists in tuning the PID controller for a setpoint of 35 °C (Figure C.4). The amplitude of the temperature fluctuations are less than 0.02°C at 35°C, and typically less than 0.1°C at all temperatures. The response time of our temperature controller for a 3ml-solution is of 10°C/min: this allows fast temperature sweeping and good control on the thermal equilibrium of the solution.

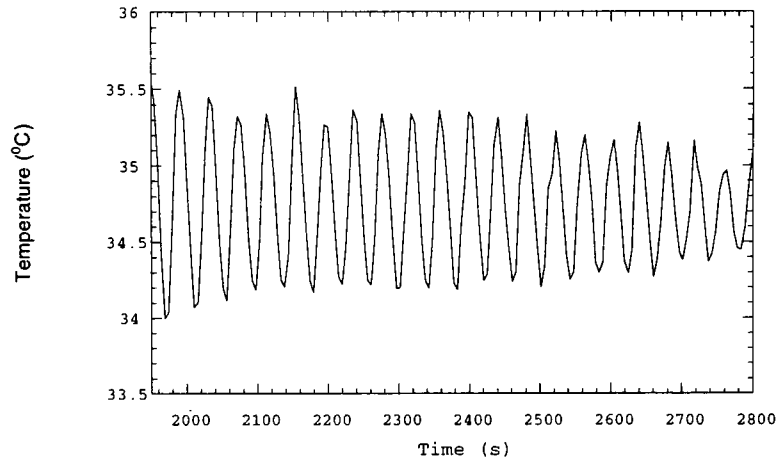


Figure C.3: Temperature recording versus time for a PID controller set to oscillate steadily. We record the period of the oscillation ($T_u=0.66\text{min}$) and the proportional band ($PB_u=1.6$) from this run to finely tune the PID controller.

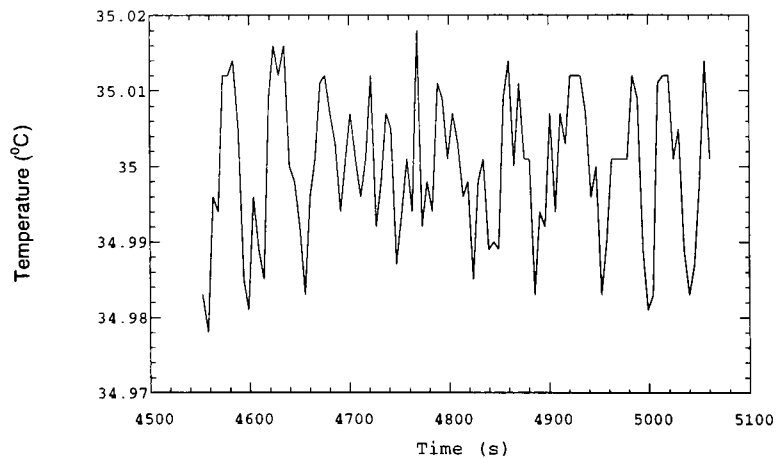


Figure C.4: Temperature recording versus time for a well-tuned PID controller. The amplitude of the temperature fluctuations is $\pm 0.02^\circ\text{C}$.

Appendix D

Anodization of aluminum.

We machined many pieces of the Fluorescence Correlation Spectroscopy setup by ourselves. We used mostly aluminum for robustness and ease of machining. To minimize light reflection from the aluminum surface, we anodized the machined pieces. We present here our small-scale anodization protocol, derived from an industrial protocol [195].

Aluminum oxidizes very quickly in air, and forms a thin protective transparent layer of aluminum oxide that inhibits further oxidation. The goal of the anodization process is to enhance the porosity of the aluminum oxide layer, to make it permeable to a black dye, which is fixed in the layer by further oxidation of the aluminum.

All this protocol must be carried out under a chemical hood, and proper chemistry lab gears should be used.

1. Clean the aluminum pieces with standard detergent solution, and dry.
2. Etch the surface of the piece by dipping in a lye solution (20 pellets of Sodium hydroxide in 500ml distilled for 5min. Then, wash thoroughly under tap water, and bring immediately to the anodization bath.
3. The anodization bath is a 25% sulfuric acid solution, mounted in an ice bucket. Place the piece to be anodized onto teflon holder, and connect it to the positive pin of a power supply with a electrically-shielded aluminum connection (the best is to tap a hole in the piece, and screw the anode pin into it). The cathode of the bath is made out of lead sheet.
4. Apply a current of 1.5Amps *per* square decimeter for 30min (check that the ice bucket remains filled as this anodization reaction overheats easily).
5. Wash the piece under tap water, and immediately dip it in a boiling solution of the black dye (we used a solution of Aldrich nigrosin #19,828-5 at a concentration of 10g/l, which can be used for five to ten anodization). Incubate for 10min while keeping a steady boiling.
6. Wash the piece under tap water and dry in a 60°C oven.

Note that tapped holes and sharp edges tend to anodize too fast: it is better to tap the holes after dyeing is finished.

مہتاب بہ نور دامن شب بشکافت
می نوش دی بہتر ازین نتوان یافت (۴۷)
خوش باش و میندیش کہ مہتاب بسی
اندر سر خاک یک یک خواهد تافت

این بحر وجود آمدہ بیرون ز نہفت
کس نیست کہ این گوہر تحقیق بسفت (۴۸)
ہرکس سخنی از سر سودا گفتہ است
زانروی کہ هست کس نمیداند گفت

می خور کہ بزیر گل بسی خواہی خفت
بی مونس و بی رفیق و بی ہمدم وجفت (۴۹)
زنہار بہ کس مگو تو این راز نہفت
ہر لالہ کہ پژمرد نخواہد بشکفت

XLVII.

*But leave the Wise to wrangle, and with me
The Quarrel of the Universe let be:
And, in some corner of the Hubbub couch'd,
Make Game of that which makes as much of Thee.*

XLVIII.

*For in and out, above, about, below,
'Tis nothing but a Magic Shadow-show,
Play'd in a Box whose Candle is the Sun,
Round which we Phantom Figures come and go.*

XLIX.

*Strange, is it not? that of the myriads who
Before us pass'd the door of Darkness through
Not one returns to tell us of the Road,
Which to discover we must travel too.*

Rubayyat
Omar Khayyam
Translation by Edward FitzGerald

Bibliography

- [1] Tyagi S. and Kramer F.R. Molecular beacons: Probes that fluoresce upon hybridization. *Nature Biotechnology*, 14(3):303–308, 1996.
- [2] Avery O.T. McLeod C.M. and McCarthy M. Transformation of pneumococcal types induced by a deoxyribonucleic acid fraction isolated from pneumococcus type III. *Journal of Experimental Medicine*, 79:137–142, 1943.
- [3] Watson J.D. and Crick F.H.C. A structure for desoxyribose nucleic acids. *Nature*, 171:737, 1953.
- [4] Vologodskii A. *Topology and physics of circular DNA*. CRC Press, 1 edition, 1992.
- [5] Stryer L. *Biochemistry*. Freeman, New York NY, iv edition, 1995.
- [6] Korolev N. Lyubartsev A.P. and Nordenskiöld L. Application of polyelectrolyte theories for analysis of DNA melting in the presence of Na^+ and Mg^{2+} ions. *Biophysical Journal*, 75:3041–3056, 1998.
- [7] Baumann C.G. Smith S.B. Bloomfield V.A. and Bustamante C. Ionic effects on the elasticity of single DNA molecules. *Proceedings of the National Academy of Sciences USA*, 94(12):6185–6190, 1997.
- [8] Hartmann B. and Lavery R. DNA structural forms. *Quaternary Review of Biophysics*, 29(4):309–368, 1996.
- [9] Nucleic acid database (NDB). <http://ndbserver.rutgers.edu>.
- [10] Smith S.B. Cui Y. and Bustamante C. Overstretching B–DNA: the elastic response of individual double-stranded and single-stranded molecules. *Science*, 271:795–798, 1996.
- [11] Bustamante C. Marko J.F. Siggia E.D. and Smith S. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1600, 1994.
- [12] Tinland B. Luen A. Sturm J. and Weill G. Persistence length of single-stranded DNA. *macromolecules*, 30(19):5763–5765, 1997.
- [13] Mills J.B. Vacano E. and Hagerman P.J. Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA). *Journal of Molecular Biology*, 285:245–257, 1999.
- [14] Sinanoğlu O. and Abdunur S. Hydrophobic stacking of bases and the solvent denaturation of DNA. *Photochemistry and Photobiology*, 3:333–337, 1964.
- [15] Friedman R.A. and Honig B. A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophysical Journal*, 71(6):3523–3526, 1995.

- [16] Dewey T.G. and Turner D.H. Laser temperature-jump study of stacking in adenylic acid polymers. *Biochemistry*, 18(26):5757–5762, 1974.
- [17] Holcomb D.N. and Tinoco I.Jr. Conformation of polyriboadenylic acid: pH and temperature dependence. *Biopolymers*, 3:121–133, 1965.
- [18] Arnott S. Chandrasekaran R. and Leslie A.G. Structure of the single-stranded polyribonucleotide polycytidylic acid. *Journal of Molecular Biology*, 106(3):735–748, 1976.
- [19] Cantor C.R. and Shimmel P.R. *Biophysical Chemistry: the Behavior of Biological Macromolecules*, volume III. Freeman, 1980.
- [20] Hall B.D. and Spiegelman S. Sequence complementarity of T2-DNA and T2-specific RNA. *Proceedings of the National Academy of Sciences USA*, 47:137–146, 1961.
- [21] Marmur J. and Doty P. *Journal of Molecular Biology*, 5:109–, 1962.
- [22] Zimm B.H. and Bragg J.K. Theory of the phase transition between helix and random coil in polypeptide chains. *Journal of Chemical Physics*, 31:526, 1959.
- [23] Goldenfeld N. *Lectures on Phase Transitions and the Renormalization Group*, chapter the Transfer Matrix, pages 88–101. Addison-Wesley, New York, 1992.
- [24] Daune M. *Biophysique moléculaire: structures en mouvement*, chapter III, pages 29–37. InterEditions, 1993.
- [25] SantaLucia J.Jr Allawi H.T. and Seneviratne Ananda P. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, 35:3555, 1996.
- [26] SantaLucia J.Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences USA*, 95(4):1460–1465, 1998.
- [27] Bloomfield V.A. Crothers D.M. and Tinoco I.Jr. *Physical Chemistry of Nucleic Acids*. Harper & Row, 1974.
- [28] Isambert H. and Siggia E. Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme. *Proceedings of the National Academy of Sciences USA*, 2000.
- [29] M. Zuker's DNA folding algorithm,. <http://mfold.wustl.edu/folder/DNA/form1.cgi>.
- [30] Kim Y. Geiger J.H. Hahn S. and Sigler P.B. Crystal structure of a yeast TBP/TATA-box complex. *Nature*, 365(6446):512–520, 1993.
- [31] Kim J.L. Nikolov D.B. and Burley S.K. Co-crystal structure of tbp recognizing the minor groove of a TATA element. *Nature*, 365(6446):520–527, 1993.
- [32] Parvin J.D. McCormick R.J. Sharp P.A. and Fisher D.E. Pre-bending of a promoter sequence enhances affinity for the TATA-binding factor. *Nature*, 373(6516):724–727, 1993.
- [33] Leger J.F. Robert J. Bourdieu L. Chatenay D. and Marko J.F. RecA binding to a single double-stranded DNA molecule: a possible role of DNA conformational fluctuations. *Proceedings of the National Academy of Sciences USA*, 95(21):12295–12299, 1998.
- [34] Stasiak A. Di Capua E. and Koller T. Elongation of duplex DNA by RecA protein. *Journal of Molecular Biology*, 151(3):557–564, 1981.

- [35] Cluzel P. Lebrun A. Heller C. Lavery R. Viovy J.L. Chatenay D. and Caron F. DNA: An extensible molecule. *Science*, 271(5250):792–794, 1996.
- [36] Lavery R. and Lebrun A. Modeling DNA stretching for physics and biology. *Genetica*, 106(1-2):75–84, 1999.
- [37] Allemand J.-F. Bensimon D. Lavery R. and Croquette V. Stretched and overwound DNA forms Pauling-like structure with exposed bases. *Proceedings of the National Academy of Sciences, USA*, 95:14152–14157, 1998.
- [38] Strick T.R. Allemand J.-F. Bensimon D. Bensimon A. and Croquette V. The elasticity of a single supercoiled DNA molecule. *Science*, 271:1835–1837, 1997.
- [39] Shivashankar G.V. Feingold M. Krichinsky O. and Libchaber A. RecA polymerization on double-stranded DNA by using single-molecule manipulation: The role of ATP hydrolysis. *Proceedings of the National Academy of Sciences USA*, 96(14):7916–7921, 1999.
- [40] Weiss S. Fluorescence spectroscopy of single biomolecules. *Science*, 283(5408):1676–1683, 1999.
- [41] Turro N.J. *Modern Molecular Photochemistry*. University Science Books, Mill Valley CA, 1 edition, 1991.
- [42] Stryer L. and Haugland R.P. Energy transfer: a spectroscopic ruler. *Proceeding of the National Academy of Sciences, USA*, 58(2):719–726, 1967.
- [43] Morrison L.E. and Stols L.M. Sensitive fluorescence-based thermodynamic and kinetic measurements of DNA hybridization in solution. *Biochemistry*, 32:3095–3104, 1993.
- [44] Mergny J.L. Botorine S.A. Garestier T. Belloc F. Rougée M. Bulychiev N.V. Koshkin A.A. Bourson J. Lebedev A. Valeur B. Thuong N.T. and Hélène C. Fluorescence energy transfer as a probe for nucleic acid structures and sequences. *Nucleic Acids Research*, 22:920–928, 1994.
- [45] Parkhurst K.M. and Parkhurst L.J. Kinetic studies by fluorescence resonance energy transfer employing a double-labeled oligonucleotide: hybridization to the oligonucleotide complement and to single-stranded DNA. *Biochemistry*, 34:285–292, 1995.
- [46] Clegg R.M. Murchie A.I. Zechel A. and Lilley D.M. Observing the helical geometry of double-stranded DNA in solution by fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences USA*, 90(7):2994–8, 1993.
- [47] Bazemore L.R. Takahashi M. and Radding C.M. Kinetic analysis of pairing and strand exchange catalyzed by RecA. detection by fluorescence energy transfer. *Journal of Biological Chemistry*, 272(23):14672–14682, 1995.
- [48] Mao C. Sun W. Shen Z. and Seeman N.C. A nanomechanical device based on the b-z transition of DNA. *Nature*, 397(6715):144–146, 1999.
- [49] Parkhurst K.M. Brenowitz M. and Parkhurst L.J. Simultaneous binding and bending of promoter DNA by TATA binding protein: real time kinetic measurement. *Biochemistry*, 35:7459–7465, 1996.
- [50] Cooper J.P. and Hagerman P.J. Analysis of fluorescence energy transfer in duplex and branched DNA molecules. *Biochemistry*, 29:9261–9268, 1990.

- [51] Ozaki H. and McLaughlin L.W. The estimation of distances between specific backbone-labeled sites in DNA using fluorescence resonance energy transfer. *Nucleic Acids Research*, 20(19):5205–5214, 1992.
- [52] Deniz A.A. Dahan M. Grunwell J.R. Ha T. Faulhaber A.E. Chemla D.S. Weiss S. and Schultz P.G. Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of Förster distance dependence and subpopulations. *Proceedings of the National Academy of Sciences USA*, 96(7):3670–3675, 1999.
- [53] Seidel C.A.M. Schulz A. and Sauer M.H.M. Nucleobase-specific quenching of fluorescent dyes. 1. Nucleobase one-electron redox potentials and their correlation with static and dynamic quenching efficiencies. *Journal of Physical Chemistry*, 100:5541–5553, 1996.
- [54] Tyagi S. Bratu D.P. and Kramer F.R. Multicolor molecular beacons for allele discrimination. *Nature Biotechnology*, 16(1):49–53, 1998.
- [55] Kielkopf CL. White S. Szewczyk J.W. Turner J.M. Baird E.E. Dervan P.B. and Rees D.C. A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science*, 282(5386):111–115, 1998.
- [56] Mulder C. and Delius H. Specificity of the break produced by restricting endonuclease R₁ in simian virus 40 DNA, as revealed by partial denaturing mapping. *Proceedings of the National Academy of Sciences USA*, 69(11):3215–3219, 1972.
- [57] Sigman D.S. Spassky A. Rimsky S. and Buc H. Conformational analysis of lac promoters using the nuclease activity of 1,10-phenanthroline-copper ion. *Biopolymers*, 24(1):183–197, 1985.
- [58] Inman R.B. A denaturation map of the λ phage DNA molecule determined by electron microscopy. *Journal of Molecular Biology*, 18:464–476, 1966.
- [59] Guéron M. Kochoyan M. and Leroy J.L. A single mode of DNA base-pair opening drives imino proton exchange. *Nature*, 328(6125):89–92, 1987.
- [60] Guéron M. and Leroy J.L. Studies of base pair kinetics by nmr measurement of proton exchange. *Methods in Enzymology*, 261:383–413, 1995.
- [61] Leroy J.L. Charretier E. Kochoyan M. and Guéron M. Evidence from base-pair kinetics for two types of adenine tract structures in solution: their relation to DNA curvature. *Biochemistry*, 27(25):8894–8898, 1988.
- [62] Deng H. Bloomfield V.A. Benevides J.M. and Thomas G.J. Jr. Dependence of the Raman signature of genomic B-DNA on nucleotide base sequence. *Biopolymers*, 50(6):656–666, 1999.
- [63] Movileanu L. Benevides J.M. and Thomas G.J.Jr. Temperature dependence of the Raman spectrum of DNA. Raman signatures of premelting and melting transitions of poly(dA-dT)•poly(dA-dT). *Journal of Raman Spectroscopy*, 30(8):637–649, 1999.
- [64] Ptashne M. and Gann A. Imposing specificity by localization: mechanism and evolvability. *Current Biology*, 8(22):R812–R822, 1998.
- [65] Pribnow D. Nucleotide sequence of an rna polymerase binding site at an early T7 promoter. *Proceedings of the National Academy of Sciences USA*, 72(3):784–788, 1975.
- [66] Siebenlist U. Nucleotide sequence of the three major early promoters of bacteriophage T7. *Nucleic Acids Research*, 6(5):1895–1907, 1979.

- [67] Ponnambalam S. Webster C. Bingham C. and Busby S. Transcription initiation at the Escherichia coli galactose operon promoters in the absence of the normal -35 region sequences. *Journal of Biological Chemistry*, 261(34):16043-16048, 1986.
- [68] Siebenlist U. Gilbert W. Contacts between Escherichia coli RNA polymerase and an early promoter of phage T7. *Proceedings of the National Academy of Sciences USA*, 77(1):122-126, 1980.
- [69] Siebenlist U. Simpson R.B. and Gilbert W. E. coli RNA polymerase interacts homologously with two different promoters. *Cell*, 20(2):269-281, 1980.
- [70] Kierkegaard K. Buc H. Spassky A. and Wang J.C. Mapping of single-stranded regions in duplex dna at the sequence level: single-strand-specific cytosine methylation in rna polymerase-promoter complexes. *Proceedings of the National Academy of Sciences USA*, 80(9):2544-2548, 1983.
- [71] Lefevre J.F. Lane A.N. and Jardetzky O. A temperature dependent transition in the Pribnow box of the trp promoter. *FEBS Letters*, 190(1):37-40, 1985.
- [72] Patel D.J. Kozlowski S.A. Weiss M. and Bhatt R. Conformation and dynamics of the Pribnow box region of the self-complementary d(C-G-A-T-T-A-T-A-A-T-C-G) duplex in solution. *Biochemistry*, 24(4):936-944, 1985.
- [73] Spassky A. and Sigman D.S. Nuclease activity of 1,10-phenanthroline-copper ion. conformational analysis and footprinting of the lac operon. *Biochemistry*, 24(27):8050-8056, 1985.
- [74] Spassky A. Rimsky S. Buc H. and Busby S. Correlation between the conformation of Escherichia coli -10 hexamer sequences and promoter strength: use of orthophenanthroline cuprous complex as a structural index. *EMBO Journal*, 7(6):1871-1879, 1988.
- [75] Magde D. Elson E. and Webb W.W. Thermodynamic fluctuations in a reacting system measurement by fluorescence correlation spectroscopy. *Physical Review Letters*, 29(11):705-708, 1972.
- [76] Ehrenberg M. and Rigler R. *Journal of Chemical Physics*, 4:390-401, 1974.
- [77] Koppel D.E. Statistical accuracy in fluorescence correlation spectroscopy. *Physical Review A*, 10(6):1938-1945, 1974.
- [78] Magde D. Elson E. and Webb W.W. Fluorescence correlation spectroscopy. ii. an experimental realization. *Biopolymers*, 13:29-61, 1974.
- [79] Rigler R. Mets Ü. Widengren J. and Kask P. Fluorescence correlation spectroscopy with high count rate and low background: analysis of translational diffusion. *European Journal of Biophysics*, 33:169-175, 1993.
- [80] Eigen M. and Rigler R. Sorting single molecules - application to diagnostics and evolutionary biotechnology. *Proceedings of the National Academy of Sciences USA*, 91(13):5740-5747, 1994.
- [81] Carl Zeiss Jena GmbH. *Confocor: new insights into molecular interactions*, 1996.
- [82] Einstein A. *Investigations on the theory of the Brownian movement (1905)*. Dover Editions, New York, R.Furth edition, 1956.
- [83] Perrin J. Lottermoser. *The Atoms*. Leipzig & Dresden, 1914.
- [84] Henri V. *Comptes Rendus*, 146:1024, 1908.

- [85] de Groot S.R. and Mazur P. *Non-Equilibrium Thermodynamics*. Dover, Mineola, NY, 1984.
- [86] Berne B.J. and Pecora R. *Dynamic Light Scattering*. Wiley (New York), 1976.
- [87] Qian H. and Elson E.L. Distribution of molecular aggregation by analysis of fluctuation moments. *Proceedings of the National Academy of Sciences USA*, 87(14):5479–5483, 1990.
- [88] Qian H. and Elson E.L. On the analysis of high order moments of fluorescence fluctuations. *Biophysical Journal*, 57(2):375–380, 1990.
- [89] Chen Y. Muller J.D. So P.T. and Gratton E. The photon counting histogram in fluorescence fluctuation spectroscopy. *Biophysical Journal*, 77(1):553–567, 1999.
- [90] Chen Y. Muller J.D. Berland K.M. and Gratton E. Fluorescence fluctuation spectroscopy. *Methods*, 19(2):234–252, 1999.
- [91] Muller J.D. Chen Y. and Gratton E. Resolving heterogeneity on the single molecular level with the photon-counting histogram. *Biophysical Journal*, 78(1):474–486, 2000.
- [92] Kask P. Palo K. Ullmann D. and Gall K. Fluorescence-intensity distribution analysis and its application in biomolecular detection technology. *Proceedings of the National Academy of Sciences USA*, 96(24):13756–13761, 1999.
- [93] Cluzel P. Surette M. and Leibler S. An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. *Science*, 287(5458):1652–1655, 2000.
- [94] Elson E.L. and Magde M. Fluorescence correlation spectroscopy. i. conceptual basis and theory. *Biopolymers*, 13:1–27, 1974.
- [95] Jacobson M.R. and Pederson T. Localization of signal recognition particle RNA in the nucleolus of mammalian cells. *Proceedings of the National Academy of Sciences USA*, 95(14):7981–7986, 1998.
- [96] Kinjo M. and Rigler R. Ultrasensitive hybridization analysis using fluorescence correlation spectroscopy. *Nucleic Acids Research*, 23(10):1795–1799, 1995.
- [97] Politz J.C. Browne E.S. Wolf D.E. and Pederson T. Intranuclear diffusion and hybridization state of oligonucleotides measured by fluorescence correlation spectroscopy in living cells. *Proceedings of the National Academy of Sciences USA*, 95(11):6043–6048, 1998.
- [98] Brock R. and Jovin T.M. Fluorescence correlation microscopy (FCM)-fluorescence correlation spectroscopy (FCS) taken into the cell. *Cellular Molecular Biology*, 44(5):847–856, 1998.
- [99] Brock R. Hink M.A. and Jovin T.M. Fluorescence correlation microscopy of cells in the presence of autofluorescence. *Biophysical Journal*, 75(5):2547–2557, 1998.
- [100] Brock R. Vamosi G. Vereb G. and Jovin T.M. Rapid characterization of green fluorescent protein fusion proteins on the molecular and cellular level by fluorescence correlation microscopy. *Proceedings of the National Academy of Sciences USA*, 96(18):10123–10128, 1999.
- [101] Haupts U. Maiti S. Schwille P. and Webb W.W. Dynamics of fluorescence fluctuations in green fluorescent protein observed by fluorescence correlation spectroscopy. *Proceedings of the National Academy of Sciences USA*, 95(23):13573–13578, 1998.
- [102] Dickson R.M. Cubitt A.B. Tsien R.Y. and Moerner W.E. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature*, 388:355–358, 1997.

- [103] Haugland R.R. Conjugation with amine-reactive probes. *Molecular Probes Technical Note*, MP0143:1–8, 1996.
- [104] Glen Research Corporation. *Catalog of products for DNA research*, 1999. <http://www.glenres.com>.
- [105] Fidanza J.A. and McLaughlin L.W. Introduction of reporter groups at specific sites in DNA containing phosphorothioate diesters. *Journal of the American Chemical Society*, 111:9117–9119, 1989.
- [106] Fidanza J.A. Ozaki H. and McLaughlin L.W. Site-specific labeling of DNA sequences containing phosphorothioate diesters. *Journal of the American Chemical Society*, 114:5509–5517, 1992.
- [107] Haugland R.R. Conjugation with thiol-reactive probes. *Molecular Probes Technical Note*, MP00003:1–4, 1997.
- [108] Elson E.L. and Qian H. Interpretation of fluorescence correlation spectroscopy and photobleaching recovery in terms of molecular interactions. *Methods in Cell Biology*, 30:307–332, 1989.
- [109] Maiti S. Haupts U. and Webb W.W. Fluorescence correlation spectroscopy: Diagnostics for sparse molecules. *Proceedings of the National Academy of Sciences USA*, 94(22):11753–11757, 1997.
- [110] Pawley J.B. *Handbook of biological confocal microscopy*. Plenum Press, New York, 2 edition, 1995.
- [111] Keller H.E. *Objective lenses for confocal microscopy*, chapter 7, pages 111–126. Plenum Press, New York, 2 edition, 1995.
- [112] Cogswell CJ and Larkin KG. *The specimen illumination path and its effect on image quality*, chapter 8, pages 127–138. Plenum Press, New York, 2 edition, 1995.
- [113] Born M. and Wolf E. *Principle of Optics*. Pergamon Press, Oxford, UK, 4 edition, 1970.
- [114] Lipson S.G. Lispon H. and Tannhauser D.S. *Optical Physics*. Cambridge University Press, Cambridge UK, 3 edition, 1995.
- [115] Fundamentals of gaussian optics. Melles Griot Corporation catalog.
- [116] Widengren J. Mets Ü. and Rigler R. Fluorescence correlation spectroscopy of triplet states in solution a theoretical and experimental study. *Journal of Physical Chemistry USA*, 99(36):13368–13379, 1995.
- [117] Lakowicz J.R. *Principle of Fluorescence Spectroscopy*. Kluwer Academics/Plenum Publishers, New York, second edition edition, 1999.
- [118] Testa A.C. Fluorescence quantum yields and standards. *Fluorescence News: Newsletter on Luminescence*, 4(4):1–3, 1969.
- [119] Hell S.W. Lindek S. Cremer C. and Stelzer E.H.K. Measurement of the 4π -confocal point-spread function proves 75nm axial resolution. *Applied Physics Letters*, 64(11):1335–1337, 1994.
- [120] Bonnet G. Tyagi S. Libchaber A. and Kramer FR. Thermodynamic basis of the enhanced specificity of structured DNA probes. *Proceedings of the National Academy of Sciences USA*, 96(11):6171–6176, 1999.

- [121] Bonnet G. and Libchaber A. Optimal sensitivity in molecular recognition. *Physica A*, 263(1-4):68-77, 1999.
- [122] Bonnet G. Krichevsky O. and Libchaber A. Kinetics of conformational fluctuations in DNA hairpin-loops. *Proceedings of the National Academy of Sciences USA*, 95(15):8602-8606, 1998.
- [123] Goddard N. Bonnet G. Krichevsky O. and Libchaber A. Sequence-dependent rigidity of single-stranded DNA. *Physical Review Letters*, in press, 2000.
- [124] Abou ela F. Koh D. Tinoco I. Jr and Martin F.H. Base-base mismatches. thermodynamics of double helix formation for dCA₃XA₃G + dCT₃YT₃G (X,Y = A,C,G,T). *Nucleic Acids Research*, 13:4811-4824, 1985.
- [125] Wagner R. Debbie P. and Radman M. Mutation detection using immobilized mismatch binding protein (MutS). *Nucleic Acids Research*, 23:3944-3948, 1995.
- [126] Youil R. Kemper B.W. and Cotton R.G.H. Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. *Proceedings of the National Academy of Sciences USA*, 92:87-91, 1995.
- [127] Nelson N.C. Hammond P.W. Matsuda E. Goud A.A. and Becker M.M. Detection of all single-base mismatches in solution by chemiluminescence. *Nucleic Acids Research*, 24:4998-5003, 1996.
- [128] Marras S.A.E. Kramer F.R. and Tyagi S. Multiplex detection of single-nucleotide variations using molecular beacons. *Genetic Analysis*, 14:151-156, 1999.
- [129] Tibanyenda N. De Bruin S.H. Hassnoot C.A.G. van der Marel G.A. van Boom J.H. and Hilbers C.W. The effect of single base-pair mismatches on the duplex stability of d(T-A-T-T-A-A-T-A-T-C-A-A-G-T-T-G) • d(C-A-A-C-T-T-G-A-T-A-T-T-A-A-T-A). *European Journal of Biochemistry*, 139:19-27, 1984.
- [130] Werntges H. Steger G. Riesner D. and Fritz H.J. Mismatches in DNA double strands: thermodynamic parameters and their correlation to repair efficiencies. *Nucleic Acids Research*, 14:3773-3790, 1986.
- [131] Ikuta S. Takagi K. Wallace R.B. and Itakura K. Dissociation kinetics of 19 base paired oligonucleotide-DNA duplexes containing different single mismatched base pairs. *Nucleic Acid Research*, 15:797-811, 1987.
- [132] Roberval G.P. *Harmonie Universelle*, chapter Traité de Méchanique. Mersenne M, 1636.
- [133] Kostrikis L.G. Tyagi S. Mhlanga M.M. et al. Molecular beacons - spectral genotyping of human alleles. *Science*, 279(5354):1228-1229, 1998.
- [134] Adleman L.M. Molecular computation of solutions to combinatorial problems. *Science*, 266(5187):1021-1024, 1994.
- [135] Magasco M.O. Chemical kinetics is turing universal. *Physical Review Letters*, 78(6):1190-1193, 1997.
- [136] Press W.H. Flannery B.P. Teukolsky S.A. and Vetterling W.T. *Numerical Recipes in C: the Art of Scientific Programming*, chapter 14, pages 517-565. Cambridge University Press, 1988.
- [137] Pörschke D. Thermodynamic and kinetic parameters of an oligonucleotide hairpin helix. *Biophysical Chemistry*, 1:381-386, 1974.

- [138] Muñoz V. Thompson P.A. Hofrichter J. and Eaton W.A. Folding dynamics and mechanism of β -hairpin mechanism. *Nature*, 390:196–199, 1997.
- [139] Wilemski G. and Fixman M. Diffusion-controlled intrachain reaction in polymers. i. theory. *Journal of Chemical Physics*, 60(3):866–877, 1974.
- [140] Wilemski G. and Fixman M. Diffusion-controlled intrachain reaction in polymers. ii. results for a pair of terminal reactive groups. *Journal of Chemical Physics*, 60(3):878–890, 1974.
- [141] Podtelezhnikov A. and Vologodskii A. Simulations of polymer cyclization by brownian dynamics. *Macromolecules*, 30:6668–6673, 1997.
- [142] De Gennes P.G.G. *Scaling Concepts in Polymer Physics*. Cornell University Press, Ithaca, NY, 1979.
- [143] Riley M. and Maling B. Physical and chemical characterization of two- and three-stranded adenine-thymine and adenine-uracil homopolymer complexes. *Journal of Molecular Biology*, 20:359–389, 1966.
- [144] Palmer R.G. Stein D.L. Abrahams E. and Anderson P.W. Models of hierarchically constrained dynamic for glassy relaxation. *Physical Review Letters*, 53(10):958–961, 1984.
- [145] Zazopoulos E. Lalli E. Stocco D.M. and Sassone-Corsi P. DNA binding and transcriptional repression by dax-1 blocks steroidogenesis. *Nature*, 390(6657):311–315, 1997.
- [146] Burd J.F. Wartell R.M. Dodgson J.B. and Wells R.D. Transmission of stability (telestability) in deoxyribonucleic acid. physical and enzymatic studies on the duplex block polymer $d(C_{15}A_{15}) \bullet d(T_{15}G_{15})$. *Journal of Biological Chemistry*, 250(13):109–113, 1975.
- [147] Burd J.F. Larson J.E. and Wells R.D. Further studies on telestability in DNA. the synthesis and characterization of the duplex block polymers $d(C_{20}A_{10}) \bullet d(T_{10}G_{20})$ and $d(C_{20}A_{15}) \bullet d(T_{15}G_{20})$. *Journal of Biological Chemistry*, 250(15):6002–6017, 1975.
- [148] Dauxois T. and Peyrard M. Dynamics and thermodynamics of a nonlinear model for DNA denaturation. *Physical Review E*, 47(1):684–695, 1993.
- [149] Dauxois T. Peyrard M. and Bishop A.R. Thermodynamics of a nonlinear model for DNA denaturation. *Physica D*, 66:35–42, 1993.
- [150] Dauxois T. Peyrard M. and Bishop A.R. Entropy-driven DNA denaturation. *Physical Review E*, 47(1):R44–R47, 1993.
- [151] Cocco S. and Monasson R. Statistical mechanics of torque induced denaturation of DNA. *Physical Review Letters*, 83(24):5178–5181, 1999.
- [152] Peyrard M. Using DNA to probe nonlinear localized excitations? *Europhysics Letters*, 44(3):271–277, 1998.
- [153] Wu M. and Tinoco I.Jr. RNA folding causes secondary structure rearrangement. *Proceedings of the National Academy of Sciences USA*, 95(20):11555–11560, 1998.
- [154] McClure W.R. Mechanism and control of transcription initiation in prokaryotes. *Annual Review of Biochemistry*, 54:171–204, 1985.
- [155] Callaci S. Heyduk E. Heyduk T. Conformational changes of Escherichia coli RNA polymerase sigma 70 factor induced by binding to the core enzyme. *Journal of Biological Chemistry*, 273(49):32995–33001, 1998.

- [156] Chen Y.F. and Helmann J.D. DNA-melting at the *Bacillus subtilis* flagellin promoter nucleates near -10 and expands unidirectionally. *Journal of Molecular Biology*, 267(1):47–59, 1997.
- [157] Buckle M. and Pemberton I.K. The kinetics of sigma subunit directed promoter recognition by *Escherichia coli* RNA polymerase. *Journal of Molecular Biology*, 285(3):955–964, 1999.
- [158] Sullivan J.J. Bjornson K.P. Sowers L.C. and deHaseth P.L. Spectroscopic determination of open complex formation at promoters for *Escherichia coli* RNA polymerase. *Biochemistry*, 36(26):8005–8012, 1997.
- [159] Fedoriw A.M. Liu H. Anderson V.E. and deHaseth P.L. Equilibrium and kinetic parameters of the sequence-specific interaction of *Escherichia coli* RNAPolymerase with nontemplate strand oligodeoxyribonucleotides. *Biochemistry*, 37(34):11971–11979, 1998.
- [160] Helmann J.D. and deHaseth P.L. Protein-nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and dna binding partners. *Biochemistry*, 38(19):5959–5967, 1999.
- [161] Roberts C.W. and Roberts J.W. Base-specific recognition of the nontemplate strand of promoter dna by *E. coli* RNA polymerase. *Cell*, 86(3):495–501, 1996.
- [162] Marr M.T. and Roberts J.W. Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. *Science*, 276(5316):1258–1260, 1997.
- [163] Leff H.S. and Rex A.F. *Maxwell's demon: entropy, information, computing*. Princeton University Press, Princeton, NJ, 1 edition, 1990.
- [164] Szilard L. *On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings.*, volume 1, chapter 3, pages 124–133. Princeton University Press, 1 edition, 1990.
- [165] Qian H. Hopfield J.J. Entropy–enthalpy compensation: Perturbation and relaxation in thermodynamic systems. *Journal of Chemical Physics*, 105(20):9292–9298, 1996.
- [166] Qian H. Entropy–enthalpy compensation: Conformational fluctuation and induced-fit. *Journal of Chemical Physics*, 109(22):10015–10017, 1998.
- [167] Searle M.S. Williams D.H. On the stability of nucleic acid structures in solution: enthalpy–entropy compensations, internal rotations and reversibility. *Nucleic Acids Research*, 21(9):2051–2056, 1993.
- [168] Hopfield J.J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences USA*, 71(10):4135–4139, 1974.
- [169] Ptashne M. and Gann A. Transcriptional activation by recruitment. *Nature*, 386(6625):569–577, 1997.
- [170] Tyagi S. Origins of translation: the hypothesis of permanently attached adaptors. *Origins of Life*, 11(4):343–351, 1981.
- [171] Stuhmeier F. Lilley D.M. and Clegg R.M. Effect of additional unpaired bases on the stability of three-way DNA junctions studied by fluorescence techniques. *Biochemistry*, 36(44):13539–13551, 1997.

- [172] Clegg R.M. Murchie A.I. Zechel A. Carlberg C. Diekmann S. and Lilley D.M. Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction. *Biochemistry*, 31(20):4846–4856, 1992.
- [173] Winfree E. Liu F. Wenzler L.A. and Seeman N.C. Design and self-assembly of two-dimensional DNA crystals. *Nature*, 394(6693):539–544, 1998.
- [174] Mirkin C.A. Letsinger R.L. Mucic R.C. and Storhoff J.J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature*, 382(6592):607–609, 1996.
- [175] Alivisatos A.P. Johnsson K.P. Peng X. Wilson T.E. Loweth C.J. Bruchez M.P. Jr and Schultz P.G. Organization of nanocrystal molecules using DNA. *Nature*, 382(6592):609–611, 1996.
- [176] Braun E. Eichen Y. Sivan U. and Ben-Yoseph G. DNA-templated assembly and electrode attachment of a conducting silver wire. *Nature*, 391(6669):775–778, 1998.
- [177] Bennett C.H. and DiVincenzo D.P. Quantum information and computation. *Nature*, 404:247–255, 2000.
- [178] Adleman L.M. Computing with DNA. *Scientific American*, 279(2):54–61, 1998.
- [179] Miyawaki A. Llopis J. Heim R. McCaffery J.M. Adams J.A. Ikura M. and Tsien R.Y. Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature*, 388(6645):882–887, 1997.
- [180] Zaccolo M. De Giorgi F. Cho C.Y. Feng L. Knapp T. Negulescu P.A. Taylor S.S. Tsien R.Y. and Pozzan T. A genetically encoded, fluorescent indicator for cyclic amp in living cells. *Nature Cell Biology*, 2:25–29, 2000.
- [181] Henderson I.R. Owen P. and Nataro J.P. Molecular switches—the ON and OFF of bacterial phase variation. *Molecular Microbiology*, 33(5):919–932, 1999.
- [182] Collier C.P. Wong E.W. Belohradsky M. Raymo F.M. Stoddart J.F. Kuekes P.J. Williams R.S. and Heath J.R. Electronically configurable molecular-based logic gates. *Science*, 285:391–394, 1999.
- [183] Soukup G.A. and Breaker R.R. Nucleic acid molecular switches. *Trends in Biotechnology*, 17:469–476, 1999.
- [184] Soukup G.A. and Breaker R.R. Engineering precision RNA molecular switches. *Proceedings of the National Academy of Sciences USA*, 96:3584–3589, 1999.
- [185] BarZiv R. and Libchaber A. Condensation of RecA protein on single stranded DNA: onset, finite size effects and disassembly. *To be published*, 2000.
- [186] Bergé P. Pomeau Y. and Vidal C. *L'ordre dans le chaos*. Hermann, Paris, France, 2 edition, 1988.
- [187] Gardner T.S. Cantor C.R. and Collins J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403(6767):339–342, 2000.
- [188] Elowitz MB and Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.
- [189] Tyagi S. Landegren U. Tazi M. Lizardi P.M. and Kramer F.R. Extremely sensitive, background-free gene detection using binary probes and $Q\beta$ replicase. *Proceedings of the National Academy of Sciences USA*, 93(11):5395–5400, 1996.

- [190] Barany F. Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proceedings of the National Academy of Sciences USA*, 88(1):189–193, 1991.
- [191] Mullis K.B. and Faloona F.A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, 155:335–350, 1987.
- [192] Vet J.A.M. Majithia A.R. Marras S.A.E. Tyagi S. Dube S. Poiesz B.J. and Kramer F.R. Multiplex detection of four pathogenic retroviruses using molecular beacons. *Proceedings of the National Academy of Sciences USA*, 96(1):6394–6399, 1999.
- [193] Stolovitzky G. and Cecchi G. Efficiency of DNA replication in the polymerase chain reaction. *Proceeding of the National Academy of Sciences, USA*, 93(23):12947–12952, 1996.
- [194] Ziegler and Nichols. *Labview PID Control Toolkit - Reference Manual*, pages 1–9. National Instruments Inc, 1993.
- [195] Doe J. *Metal Finishing – Guidebook and Directory*. Hackensack NJ, 1991.
- [196] Bertucat G. Lavery R. and Prevost C. A molecular model for recA-promoted strand exchange via parallel triple- stranded helices. *Biophysical Journal*, 77(3):1562–1576, 1999.
- [197] Edman L. Wennmalm S. Tamsen F. and Rigler R. Heterogeneity in single DNA conformational fluctuations. *Chemical Physics Letters*, 292(1-2):15–21, 1998.
- [198] Edman L. Mets Ü. and Rigler R. Conformational transitions monitored for single molecules in solution. *Proceedings of the National Academy of of Sciences USA*, 13(13):6710–6715, 1996.
- [199] Flatters D. and Lavery R. Sequence-dependent dynamics of TATA-box binding sites. *Biophysical Journal*, 75(1):372–381, 1998.
- [200] Gralla J. and Crothers D.M. Free energy of imperfect nucleic acid helices. ii. small hairpin loops. *Journal of Molecular Biology*, 73(4):497–511, 1973.
- [201] Ha T. Ting A.Y. Liang J. Caldwell W.B. Deniz A.A. Chemla D.S. Schultz P.G. and Weiss S. Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proceedings of the National Academy of Sciences USA*, 96(3):893–898, 1999.
- [202] Ha T. Enderle T. Ogletree D.F. Chemla D.S. Selvin P.R. and Weiss S. Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences USA*, 93(13):6264–6268, 1996.
- [203] Haasnoot C.A. Hilbers C.W. van der Marel G.A. van Boom J.H. Singh U.C. Pattabiraman N. and Kollman P.A. On loop folding in nucleic acid hairpin-type structures. *Journal of Biomolecular Structure and Dynamics*, 3(5):843–857, 1986.
- [204] Haasnoot C.A. de Bruin S.H. Berendsen R.G. Janssen H.G. Binnendijk T.J. Hilbers C.W. van der Marel G.A. and van Boom J.H. Structure, kinetics and thermodynamics of DNA hairpin fragments in solution. *Journal of Biomolecular Structure and Dynamics*, 1(1):115–129, 1983.
- [205] Hilbers C.W. Haasnoot C.A. de Bruin S.H. Joordens J.J. van der Marel G.A. and van Boom J.H. Hairpin formation in synthetic oligonucleotides. *Biochimie*, 67(7-8):685–695, 1985.
- [206] Kaplan P.D. Qi O.Y. Thaler D.S. and Lichaber A. Parallel overlap assembly for the construction of computational DNA libraries. *Journal of Theoretical Biology*, 188(3):333–341, 1997.

- [207] Kern D. Volkman B.F. Luginbuhl P. Nohaile M.J. Kustu S. and Wemmer D.E. Structure of a transiently phosphorylated switch in bacterial signal transduction. *Nature*, 402:894–898, 1999.
- [208] Meseth U. Wohland T. and Rigler R. Resolution of fluorescence correlation measurements. *Biophysical Journal*, 76(3):1619–1631, 1999.
- [209] Molecular Probes, editor. *Handbook of fluorescent probes and research chemicals*, chapter 1, pages 7–40. Haugland RP, Eugene, OR, 6 edition, 1996.
- [210] Monia B.P. Johnston J.F. Ecker D.J. Zounes M.A. Lima W.F. and Freier S.M. Implication of RNA structure on antisense oligonucleotide hybridization kinetics. *Journal of Biological Chemistry*, 267:19954–19962, 1992.
- [211] Ninio J. Kinetic amplification of enzyme discrimination. *Biochimie.*, 57(5):587–595, 1975.
- [212] Ouyang Q. Kaplan P.D. Liu S.M. and Libchaber A. DNA solution of the maximal clique problem. *Science*, 278(5337):446–449, 1997.
- [213] Parkhurst K.M. and Parkhurst L.J. Donor-acceptor distance distributions in double-labeled fluorescent oligonucleotide both as a single strand and in duplexes. *Biochemistry*, 34:293–300, 1995.
- [214] Piatek A.S. Tyagi S. Pol A.C. Telenti A. Miller L.P. Kramer F.R. and Alland D. Molecular beacon sequence analysis for detecting drug resistance in mycobacterium tuberculosis. *Nature Biotechnology*, 16(4):359–363, 1998.
- [215] Roberts R.W. and Crothers D.M. Specificity and stringency in DNA triplex formation. *Proceedings of the National Academy of Sciences, USA*, 88:9397–9401, 1991.
- [216] Shivashankar G.V. Stolovitzky G. and Libchaber A. Backscattering from a tethered bead as a probe of DNA flexibility. *Applied Physics Letter*, 73(3):291–293, 1998.
- [217] Strick T.R. Croquette V. and Bensimon D. Homologous pairing in stretched supercoiled DNA. *Proceeding of the National Academy of Sciences, USA*, 95:10579–10583, 1998.
- [218] Thomas G.J.Jr. Raman spectroscopy of protein and nucleic acids assembly. *Annual Review of Biophysics and Biomolecular Structures*, 28:1–27, 1999.
- [219] Topal M.D. and Fresco J.R. Base pairing and fidelity in codon-anticodon interaction. *Nature (London)*, 263:289–293, 1976.
- [220] Viovy J.L. Heller C. Caron F. Cluzel P. and Chatenay D. Ultrafast sequencing of DNA by mechanical opening of the double helix a theoretical investigation. *Comptes Rendus de l'Academie des Sciences France*, 317(9):795–800, 1994.
- [221] Wang S. Friedman A.E. and Kool E.T. Origins of high sequence selectivity: a stopped-flow kinetics study of DNA/RNA hybridization by duplex- and triplex-forming oligonucleotides. *Biochemistry*, 34:9784–9794, 1995.
- [222] Watson J.D. and Crick F.H.C. The structure of DNA. *Cold Spring Harbor Symposium on Quantitative Biology*, 18:123, 1953.
- [223] Wennmalm S. Edman L. and Rigler R. Non-ergodic behaviour in conformational transitions of single DNA molecules. *Chemical Physics*, 247(1):61–67, 1999.
- [224] Wennmalm S. Edman L. and Rigler R. Conformational fluctuations in single DNA molecules. *Proceedings of the National Academy of Sciences USA*, 94(20):10641–10646, 1997.

- [225] Wiedmann M. Wilson W.J. Czajka J. Luo J. Barany F. and Batt C.A. Ligase chain reaction (LCR)—overview and applications. *Genome Research*, 3(4):S51–64, 1994.
- [226] Yoon K. Turner D.H. and Tinoco I.Jr. The kinetics of codon-anticodon interaction in yeast phenylalanine transfer rna. *Journal of Molecular Biology*, 99:507–518, 1975.
- [227] Stefano J.E. Ackerson and J.W. Gralla J.D. Alterations in two conserved regions of promoter sequence lead to altered rates of polymerase binding and levels of gene expression. *Nucleic Acids Research*, 8(12):2709–2723, 1980.
- [228] Nussinov R. Promoter helical structure variation at the Escherichia coli polymerase interaction sites. *Journal of Biological Chemistry*, 259(11):6798–3805, 1984.
- [229] Russell D.R. Miller P.D. and Bennett G.N. In vitro characterization of hybrid promoters and altered tryptophan operon promoters. *Biochemistry*, 24(6):1410–1417, 1985.
- [230] Horwitz M.S. and Loeb L.A. Promoters selected from random dna sequences. *Proceedings of the National Academy of Sciences USA*, 83(19):7405–7409, 1986.
- [231] Bubkle M. and Buc H. Fine mapping of DNA single-stranded regions using base-specific chemical probes: study of an open complex formed between RNA polymerase and the *lac* UV₅ promoter. *Biochemistry*, 28(10):4388–4396, 1989.
- [232] Malhotra A. Severinova E. and Darst S.A. Crystal structure of a sigma 70 subunit fragment from E. coli RNA polymerase. *Cell*, 87(1):127–136, 1996.
- [233] Zhang G. Campbell E.A. Minakhin L. Richter C. Severinov K. and Darst S.A. Crystal structure of thermus aquaticus core RNA polymerase at 3.3 Å resolution. *Cell*, 98(6):687–690, 1999.



THE LIBRARY



19010000304089

