SINGLE-MOLECULE STUDIES OF NUCLEIC ACID FOLDING AND NUCLEIC ACID-PROTEIN INTERACTIONS

Daniel Cibrán Pérez González

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2017

Full metadata for this thesis is available in
St Andrews Research Repository
at:
http://research-repository.st-andrews.ac.uk/

Identifiers to use to cite or link to this thesis:

DOI: https://doi.org/10.17630/10023-12039
http://hdl.handle.net/10023/12039

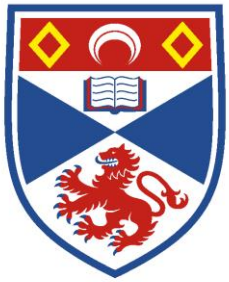
This item is protected by original copyright

This item is licensed under a
Creative Commons License

https://creativecommons.org/licenses/by-nc-nd/4.0

Single-molecule studies of nucleic acid folding and nucleic acid-protein interactions

Daniel Cibrán Pérez González



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of PhD at the University of St Andrews

June 2017

A Anxa, Xosé Manuel e Gloria.

Declaration

---

1. Candidate's declarations:

I, Daniel Cibrán Pérez González, hereby certify that this thesis, which is approximately 37,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September, 2013 and as a candidate for the degree of Doctor of Philosophy in September, 2013; the higher study for which this is a record was carried out in the University of St Andrews between 2013 and 2017.

_____ _____

Date Cibrán Pérez González

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

_____ _____

Date Dr. Carlos Penedo

3. Permission for publication: (to be signed by both candidate and supervisor)

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work

not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

PRINTED COPY

No embargo on printed copy.

ELECTRONIC COPY

Embargo on all or part of print copy for a period of 2 years on the following ground:

- Publication would preclude future publication

ABSTRACT AND TITLE EMBARGOES

- I agree to the title and abstract being published

_____	_____
Date	Cibrán Pérez González
_____	_____
Date	Dr. Carlos Penedo

Abstract

Nucleic acids and proteins, some of the building blocks of life, are not static structures but highly dynamic entities that need to interact with one another to meet cellular demands. The work presented in this thesis focuses on the application of highly sensitive fluorescence methods, both at ensemble and single-molecule level, to determine the dynamics and structure of specific biomolecular interactions with nanometer resolution and in temporal scales from nanoseconds to minutes, which includes most biologically relevant processes. The main aims of my PhD can be classified in three areas: i) exploring new fluorescent sensors with increased specificity for certain nucleic acid structures; ii) understanding how some of these nucleic acids sense the presence of small molecules in the cellular environment and trigger gene regulation by altering their structure; and iii) understanding how certain molecular machines, such as helicase proteins, are able to unwind the DNA double helix by using chemical energy in the form of ATP hydrolysis.

Acknowledgements

First and foremost I would like to acknowledge my supervisor Dr Carlos Penedo Esteiro for his support, advice and patience over these years. Without his work and constant help the completion of this PhD would have never been possible.

I am especially grateful to José Peregrina, Reyes Sanles and Jonathan Martínez, who were always helpful and careful inside and outside the lab to make me feel like at home since the very first day.

My deep appreciation goes to Malcolm White and his group: Clare Rollie, Shirley Graham, Biljiana Petrovic, Christophe Rouillon, Reyes and Jing Zhang for adopting me in the BMS and for their invaluable help and support.

I would also like to greatly appreciate the support, training and advice received from past and present group members of our lab: Alfonso Brenlla, Kaley McCluskey, Michael Morten, José, Steven Quinn, Euan Shaw and Francisco Tenopala.

A very special thank to Daniel Lafontaine and his group for their warm welcome, help and hospitality during my stay in Sherbrooke.

I am grateful for the funding received from the Engineering and Physical Sciences Research Council (EPSRC) and University of St Andrews 600th Anniversary scholarship in concept of tuition fees and stipend. My thanks also go to the School of Physics and the Scottish Universities Physics Alliance (SUPA) for giving me the opportunity to carry out this work.

Finally, I would also like to thank my family, with a special mention to Anxa, Xosé Manuel and Gloria, for their continued support. Without them this thesis would have never been written. *Grazas*.

Table of contents

List of abbreviations **xvii**

1 Introduction **1**

1.1 DNA and RNA as building blocks of life 1

1.2 Principles of fluorescence 6

1.3 Fluorescence-based techniques to investigate nucleic acid structures at ensemble level 10

1.3.1 Fluorescence quenching 11

1.3.2 Anisotropy 12

1.3.3 2-Aminopurine 14

1.3.4 Excited-State Intramolecular Proton Transfer 15

1.4 Biophysical techniques to investigate nucleic acid structure at single-molecule level 17

1.4.1 Fluorescence Resonance Energy Transfer 20

1.5 Current challenges in single-molecule spectroscopy to investigate RNA folding during transcription 25

References 26

2 Materials and methods **33**

2.1 Sample preparation protocols 33

2.1.1 Nucleic acid samples 33

2.1.2 Buffers 34

2.2 Steady-state absorption spectroscopy 34

2.3 Circular dichroism 36

2.4 Steady-state fluorescence spectroscopy 37

2.5 Single-molecule total internal reflection fluorescence spectroscopy 38

2.5.1 Surface-immobilisation methods 42
2.5.1.1 BSA immobilisation 43
2.5.1.2 PEG immobilisation 44
2.5.1.3 DT20 immobilisation 45
2.5.1.4 Vesicle encapsulation 46
2.5.2 Oxygen scavenger systems 48
2.5.2.1 Glucose oxidase and catalase 49
2.5.2.2 PCA/PCD 50
2.5.3 Single-molecule FRET data analysis 51
2.5.3.1 Collection and analysis of raw data 51
2.5.3.2 Manual selection of molecules and representation of population histograms 52
2.5.3.3 Analysis of single-molecule FRET trajectories using hidden Markov modelling 54
References 55

3 Novel fluorescence probes for sensing nucleic acids 57
3.1 Introduction 57
3.1.1 Fluorescence sensing of guanine quadruplexes (G4s): current strategies 62
3.2 Materials and methods 63
3.2.1 Sample preparation 63
3.2.2 Gel electrophoresis 64
3.2.3 Absorption 65
3.2.4 Ensemble fluorescence 65
3.2.4.1 Binding affinity 65
3.2.5 Circular dichroism 66
3.3 Results and discussion 67
3.3.1 Characterisation of G4s by polyacrylamide gel electrophoresis and circular dichroism: monomeric versus multimeric species 67
3.3.2 HPIP: a novel sensor for guanine quadruplexes 69

3.3.2.1 Photophysical behaviour of HPIP in aqueous solution . . 69
3.3.2.2 Steady-state characterisation of fluorescence emission of HPIP in the presence of G4s and single-stranded DNA . . 71
3.3.3 Metal complexes as sensors for guanine quadruplexes 74
3.3.3.1 Characterisation of the $[Ru(bpy)_3]^{2+}$ interaction with G4s using fluorescence methods 74
3.3.3.2 Polyacrylamide gel electrophoresis and circular dichroism of $[Ru(bpy)_3]^{2+}$: G4 complexes 76
3.3.4 Investigation of curcumin as a natural product that binds G4s 77
3.3.4.1 Structure and sensing of G-wires by curcumin 77
3.4 Conclusions 80
References 81

4 Quantitative single-molecule FRET methods to quantify the axial rotation of duplex helical stems in RNA riboswitches 87
4.1 Introduction 87
4.2 Materials and methods 92
4.2.1 Preparation of fluorescently-labelled aptamers 92
4.2.2 Single-molecule FRET imaging 93
4.2.3 Correction of single-molecule FRET data 94
4.2.4 Modelling of the experimental constructs 96
4.2.5 Determination of the theoretical FRET 96
4.3 Results and discussion 97
4.4 Conclusions 108
References 109

5 Single-molecule FRET to investigate the folding of the adenine riboswitch through an intermediate state 113
5.1 Introduction 113
5.2 Materials and methods 117

5.2.1 Preparation of fluorescently-labelled aptamer 117
5.2.2 Single-molecule FRET experiments 118
5.3 Results and discussion 119
5.3.1 Magnesium-induced formation of the loop-loop interaction 119
5.3.2 Sodium-induced formation of the loop-loop interaction 122
5.4 Conclusions 127
References 128

6 Single-molecule FRET studies of nascent RNA within transcriptional complexes 129
6.1 Introduction 129
6.2 Materials and methods 134
6.2.1 Preparation of double-labelled transcription elongation complexes . 134
6.2.2 Single-molecule imaging 136
6.3 Results and discussion 138
6.3.1 Characterisation of surface-immobilised elongation complexes and labelling specificity 138
6.3.2 Observation of ligand binding in nascent transcripts comprising the core of the *tbpA* aptamer 139
6.3.3 Observation of ligand binding in transcriptional pause sites 141
6.3.4 Influence of NTP concentration on the cotranscriptional folding of the *tbpA* riboswitch 146
6.3.5 Regulatory mechanism of the *E.coli tbpA* riboswitch 148
6.4 Conclusions 149
References 150

7 Understanding asymmetric nucleic acid unwinding by helicases 153
7.1 Introduction 153
7.2 Materials and methods 159
7.2.1 Preparation of labelled DNA substrates 159
7.2.2 Expression and purification of TaXPD and PcrA helicases 160

7.2.3 DNA melting assays 160
7.2.4 DNA unwinding kinetics 161
7.2.5 Activity gel assays 162
7.3 Results and discussion 163
7.3.1 Relative thermodynamic stability of DNA substrates calculated by melting analysis 163
7.3.2 Real-time DNA unwinding kinetics 165
7.3.3 DNA unwinding amplitude determined by activity gel assays 169
7.4 Conclusions 171
References 173

8 Conclusions and further directions **177**
References 183

Appendix A Theoretical explanation of fluorescence resonance energy transfer **185**
References 190

Appendix B Publications arising from this work **193**

List of abbreviations

γ	Gamma factor
κ^2	Dipole orientation factor
I_A	Fluorescence intensity of the acceptor
I_D	Fluorescence intensity of the donor
K_d	Dissociation constant
R_0	Förster radius
2AP	2-aminopurine
A	Adenine
APS	Ammonium persulfate
ATP	Adenosine triphosphate
AV	Accessible volume
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
CD	Circular dichroism
CW	Continuous wave
DBCO	Dibenzocyclooctyl group
DDS	Dichlorodimethylsilane
DMPC	1,2-Dimyristoyl-sn-glycero-3-phosphocholine
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
$\mathbf{E_{app}}$	Apparent FRET efficiency

E_{FRET}	Corrected FRET efficiency
EC	Elongation complex
EDTA	Ethylenediaminetetraacetic acid
Egg-PC	L-α-Phosphatidylcholine (Egg, Chicken)
EMCCD	Electron multiplying charge-coupled device
ESIPT	Excited-state intramolecular proton transfer
F	Folded state
F_{Mg}	Magnesium-induced folded state
F_{NS}	Ligand-induced folded state
FCS	Fluorescence correlation spectroscopy
FOG	Femtosecond optical gating
FRET	Fluorescence resonance energy transfer
G	Guanine
G4	Guanine quadruplex
GMP	Guanosine monophosphate
GOC	Glucose oxidase and catalase oxygen scavenger system
GSRPT	Ground-state intramolecular proton transfer
I	Intermediate state
MES	2-(N-morpholino)ethanesulfonic acid
mRNA	Messenger RNA
NA	Numerical aperture
NHS	N-Hydroxysuccinimide
nt	nucleotide
PCA	Protocatechuic acid
PCD	Protocatechuate 3,4-Dioxygenase

PEG	Polyethylene glycol
PMT	Photomultiplier tube
RNA	Ribonucleic acid
RNAP	RNA polymerase
SAH	S-adenosyl-homocysteine
SAM	S-adenosylmethionine
SEM	Standard error of the mean
smFCS	Single-molecule fluorescence correlation spectroscopy
smFRET	Single-molecule fluorescence resonance energy transfer
SNR	Signal-to-noise ratio
ssDNA	Single-stranded DNA
SUV	Small unilamellar vesicles
SVA-PEG	Succinimidyl valerate-poly(ethylene glycol)
T	Thymine
TBE	Tris-Borate-EDTA buffer
TEMED	Tetramethylethylenediamine
TIRF	Total internal reflection fluorescence
TPP	Thiamine pyrophosphate
Tris	Tris(hydroxymethyl)aminomethane
TROLOX	6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid
U	Unfolded state / Uracil

1

Introduction

1.1 DNA and RNA as building blocks of life

The central dogma of molecular biology represents how the information circulates within the cell and describes the function of each of the three main classes of biomolecules: DNA, RNA and proteins. The dogma states that the information is stored as deoxyribonucleic acid (DNA), which is then carried as messenger ribonucleic acid (mRNA) until it is translated into proteins, the elements that carry most of the biological functions:

$$\text{DNA} \xrightarrow{\text{Transcription}} \text{RNA} \xrightarrow{\text{Translation}} \text{Protein}$$

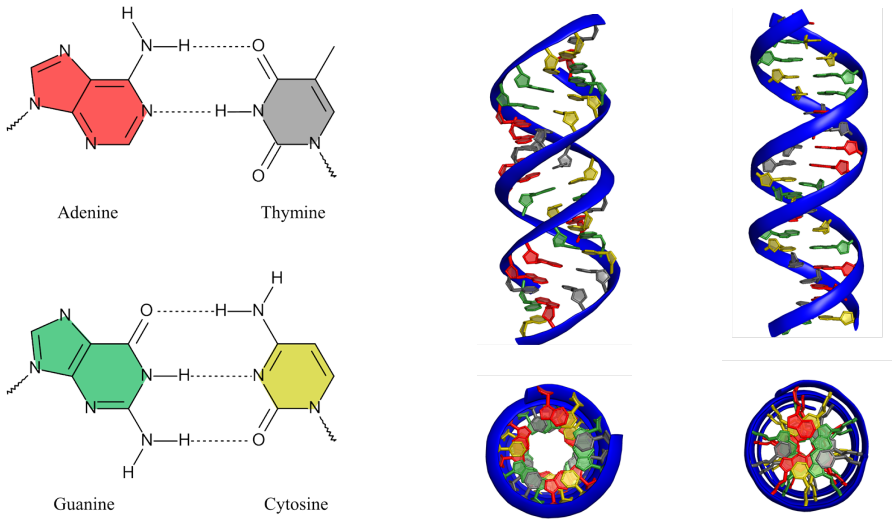
Nucleic acids, both DNA and RNA, are a sequence of repeating units, called nucleotides, that are composed of three elements: i) a phosphate group; ii) a sugar, either deoxyribose or ribose for DNA or RNA, respectively; and iii) a nitrogenous base, which can be adenine (A), cytosine (C), guanine (G), thymine (T) or uracil (U). Adenine and guanine are derivatives of an aromatic structure composed of two fused rings called purine, whilst cytosine, thymine and uracil are derivatives of an aromatic element containing one single ring called pyrimidine. Purine bases can establish hydrogen bonds with their complementary pyrimidine bases, however, the interactions between bases of the same nature is unfavourable. Thus, cytosine binds to guanine by establishing three hydrogen bonds, whereas adenine binds to thymine (DNA) or uracil (RNA) via two hydrogen bonds (Fig. 1.1a).

Introduction

Nucleotide sequences are mainly organised around a single strand of nucleotides (ssDNA or ssRNA) or a double helix formed by two complementary ssDNA or ssRNA strands bound together by base pairing and stacking interactions, which represents the most characteristic structure of nucleic acids. The hydrogen bonds between complementary bases confer stability to the double-stranded DNA, however, the different number of these bonds in the G≡C and A=T/A=U base pairs is reflected as different stabilities in sequences of the same length. Hence, a higher guanine-cytosine content will mean higher stability and thus, higher melting temperature (T_m).

The double helix DNA structure was described for the first time by Watson and Crick in 1953 [Watson and Crick, 1953], who based their finding on unpublished X-ray crystallography data obtained by Rosalind Franklin showing the diffraction pattern of a DNA duplex. This discovery has led to an enormous growth in our understanding of nucleic acids and their biological functions. However, DNA does not only display a single- or double-stranded structure but also a number of different conformations such as i-motif structures, hairpins and guanine quadruplexes. Moreover, other conformations such as Holliday junctions, replication forks and DNA flaps are formed during DNA replication, recombination and repair processes as a result of the unwinding of a DNA duplex in order to read, copy or extend a certain DNA sequence.

The DNA contained in an average human cell would be around 2 meters long if it was extended. This fact highlights the importance of DNA packaging to fit in the 10 µM-diameter size nucleus, although a small portion of DNA can be found outside in the cytoplasm, especially in mitochondria. Nuclear dsDNA in eukaryotes is stored associated to proteins called histones in a highly condensed state. This combination of DNA and histones adopting a coiled form is called chromatin, which also folds into supercoiled structures known as chromosomes. Interestingly, the end of human chromosomes, called telomeres, are guanine-rich single-stranded regions that comprise up to 14 kilobases of $(TTAGGG)_n$ ssDNA repeats, which are thought to adopt G-quadruplex structures to protect telomeres from shortening. The shortening of telomeres produced in each replication cycle in eukaryotes is responsible for cellular aging and for the limited number of cell divisions, thus responsible



(a) **(b)** A-RNA **(c)** B-DNA

Fig. 1.1 Schematics of the A-RNA and B-DNA helices. (a) Chemical structure of the DNA base pairs. Adenine (red)–thymine (gray) with two hydrogen bonds and guanine (green)–cytosine (gold) three hydrogen bonds. (b-c) Side and top views of A-RNA and B-DNA double helix structures.

for cellular senescence. The cellular machinery that compensates telomerase shortening is called telomerase, which performs this maintenance via reverse transcription by using an RNA template to incorporate fragments of single-stranded DNA at the 3' end of telomeres. Interestingly, this mechanism is opposed to the classical view of DNA and RNA as a store and intermediate carrier of information and represents a process of high importance in tumour cells that will be described more in detail in chapter 3. On the other hand, the genetic information of archaea and bacteria is contained in a single circular double-stranded DNA chromosome that folds together with proteins in the cytosol to form the nucleoid. In addition, some genes are often stored in separate circular DNA structures known as plasmids.

In a similar way, the RNA field has experienced a stunning growth in the last decades in which RNA has changed from a mere carrier of information to a biologically active molecule capable of performing catalytic reactions (ribozymes) and participate in gene regulation (microRNA, small interfering RNA, riboswitches, etc.). Most biological processes would take place at an unusually slow rate that would be incompatible with life unless they are performed in the presence of catalysts, normally proteins. In 1989, Thomas R. Cech and

Sidney Altman were awarded with the Nobel Prize in Chemistry for their discovery of non-coding RNAs with catalytic functions, called ribozymes, early in the 1980s [Guerrier-Takada et al., 1983; Kruger et al., 1982; Yurke et al., 2000]. This finding represents the first time that RNA is reported to catalise biochemical processes itself without contribution of any protein or small molecule. For example, ribozymes take part of the ribosomal subunit that links amino acids in the translation of mRNA into proteins.

A new discovery in 2002 reported the first mRNA sequence that is able to sense a small cellular metabolite, coenzyme B12 (adenosylcobalamin), and regulate the expression of an associated gene, in this case a gene encoding a protein involved in the transport of cobalamin compounds [Nahvi et al., 2002]. These mRNAs are called riboswitches and carry out gene regulation functions in bacteria, archaea, plants and fungi without protein mediation. For this reason, riboswitches have emerged as promising genetic regulators that may overcome one of the biggest threats to global health according to the World Health Organization: antibiotic resistance. The metabolite sensing mechanism used to modulate the expression of associated genes makes riboswitches potential antibiotic targets in which ligand analogues would disrupt the normal function of bacteria [Blount and Breaker, 2006; Hermann, 2015]. To date, about 20 classes of riboswitches have been reported and three of them will be studied in this thesis by taking advantage of single-molecule fluorescence-based

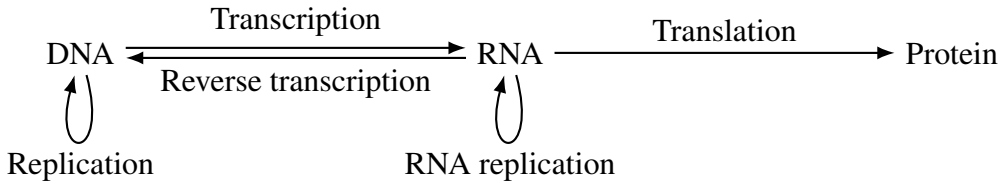
Table 1.1 Parameters of A-RNA and B-DNA double helix structures.

	Helix type	
	A-RNA	B-DNA
Twist per basepair	33.6°	35.9°
Rise per basepair	2.3Å	3.32Å
Helix-packing diameter	25.5°	23.7°
Pitch per turn of helix	24.6Å	33.2Å
Tilt of base normals to helix axis	+19°	−1.2°
Sense of rotation	Right-handed	Right-handed

techniques: S-adenosylmethionine-I (SAM-I) (chapter 4), adenine (chapter 5) and thiamine pyrophosphate (TPP) (chapter 6) riboswitches.

Unlike DNA, RNA mostly exists as a single strand, although some RNA sequences such as ribozymes and riboswitches contain self-complementary sequences and are able to fold into stable three-dimensional structures to carry out specific functions. The complementary sequences are arranged around a double helix A-shape RNA via base pairing. The parameters of A-RNA and B-DNA helices are detailed in Table 1.1 and representations of both shapes are depicted in Figs. 1.1b and 1.1c.

The continuous growth of the DNA and RNA fields in the last decades has radically changed the classic view of nucleic acids and how the genetic information flows within the cell. The wide range of structures and functions of these biomolecules go beyond the concept of simple storage (DNA) and transmission (RNA) of information defined in the classic central dogma of biology, which can now be described as:



The wide range of nucleic acid structures has emphasised the need to improve the existing and develop new methods to sense these nucleic acid architectures. Moreover, the ability of some of these nucleic acids to adopt certain three-dimensional structures such as DNA origami, G-wires and DNA-based molecular machines are finding applications in and outside the biological context (i.e. nanotechnology) [Livshits et al., 2014; Pinheiro et al., 2011; Rothemund, 2006]. In this context, two main limitations of current optical sensing methods need to be overcome for both *in vitro* and *in vivo* applications and also to further progress in the characterisation of technologically relevant nucleic acid motifs: specificity and sensitivity. It is clear that the traditional differentiation between single- and double-stranded nucleic acids is not any longer sufficient to develop fluorescence-sensing methods able to discriminate between nucleic acid conformations. This PhD focuses on exploring new photophysical tools

to detect specific nucleic acids and use these tools at ensemble and single-molecule level to elucidate the structure, folding and function of DNA and RNA structures.

We first investigated three fluorescent probes as potential sensors for different G-quadruplex and G-wire structures. This work is presented in chapter 3, where we evaluate one of these fluorophores as the first ratiometric sensor for the selective and quantitative detection of G-quadruplexes using ensemble techniques. We then introduce single-molecule TIR microscopy and fluorescence resonance energy transfer (FRET) to investigate a series of RNA riboswitches and monitor their folding into specific three-dimensional structures upon addition of their cognate ligand. Chapter 4 describes a new method for the study of a riboswitch sensing S-adenosylmethionine (SAM) in combination with three-dimensional models engineered from the crystal structure, whereas chapter 5 provides insights into the folding pathway of the adenine riboswitch via an obligate intermediate. In chapter 6 we go one step further in the study of RNA riboswitches at the single-molecule level and we present a novel method based on the site-specific labelling of RNAs via click chemistry to investigate, for the first time, the cotranscriptional folding of riboswitches within transcription elongation complexes using smFRET. Lastly, chapter 7 uses ensemble methods to investigate the aspects involved in the asymmetric unwinding of double-stranded DNAs by helicase proteins.

1.2 Principles of fluorescence

Luminiscence refers to the emission of light in all directions by a given compound from an electronic excited state to the ground state, S_0. The mechanism of the absorption and emission of light is illustrated by the Perrin-Jablonski diagram (Fig. 1.2). First, the compound absorbs radiation and reaches its first (S_1) or second (S_2) excited electronic state. In the latter case, the transition from S_2 to S_1, called internal conversion (IC), occur in ~10^{-12} s, much faster than fluorescence emission that occur in the pico- to nanosecond range. In a similar way, the vibrational relaxation from excited vibrational levels to the lowest vibrational level within the same electronic state occur in a timescale of 10^{-13}–10^{-11} s. As a result, the emission of radiation normally takes place from the lowest vibrational level of S_1 [Valeur,

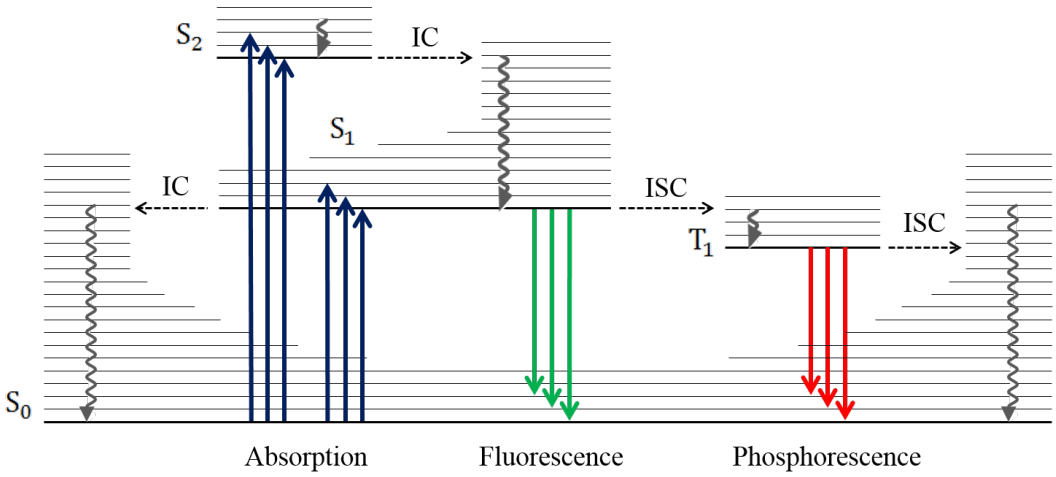


Fig. 1.2 Perrin-Jablonski diagram illustrating the absorption, fluorescence and phosphorescence processes. The absorption of a photon is followed by internal conversion (IC) from S_2 to S_1 and vibrational relaxation to the lowest vibrational level of the first excited singlet state. This is followed by the emission of radiation of fluorescence, by relaxation via non-emissive pathways or by intersystem crossing (ISC) to the first excited triplet state, from which it can relax by emitting phosphorescence or by ISC to the singlet ground state. Once in the ground state, the fluorophore reaches the lowest vibrational level of S_0 non-radiatively.

2001]. However, the excited compound can also relax via non-radiative processes or undergo intersystem crossing (ISC) to a triplet state, T_1. The transition from a triplet state to the singlet ground state is forbidden and, therefore, the radiative relaxation from this step, called phosphorescence, takes place in the 10^{-3} to 10^{0} s timescale.

The chemical compounds that emit fluorescence, named fluorophores, are defined by some of their properties such as maximum absorption and emission wavelengths, Stokes shift, fluorescence lifetime and quantum yield. In order to give a definition of the Stokes shift, it is necessary to define first what is the energy of the fluorescence photon. The energy (expressed in kg m^2/s^2, or J) of the emitted photon, E_{photon}, is the difference between the lowest vibrational level of S_1 and a certain vibrational level of S_0, which is related to its frequency (ν), wavelength (λ) and wavenumber ($\tilde{\nu}$) in SI base units as

$$E_{photon} = E_1 - E_0 = h\nu = \frac{hc}{\lambda} = hc\tilde{\nu} \tag{1.1}$$

where h is Planck's constant (6.626070040 x 10^{-34} kg $m^2 s^{-1}$, or J s) and c is the speed of light in vacuum (299,792,458 m s^{-1}). As represented on the Perrin-Jablonski diagram, fluorescence emission occur at lower energy or higher wavelength than absorption. The gap between the absorption and emission maximum is known as the Stokes shift, which is due to the rapid loss of energy via non-radiative internal conversion or vibrational relaxation in the S_1 and S_0 states. Moreover, fluorophores can undergo some processes such as excited-state proton transfer, interactions with the solvent, complex formation and energy transfer that result in an abnormally large Stokes shift [Lackowicz, 2006].

The fluorescence lifetime, τ, is another of the properties that define a given fluorophore. It represents the average time shift between the absorption and emission of a photon, or how long a molecule lives in the excited state. When a short laser light pulse (i.e. femto- to nanosecond width) excites a certain sample, a number of molecules, A, reach the first electronic excited state and return to the ground state via radiative or non-radiative processes [Valeur, 2001]. The rate of depopulation of the excited state can be expressed as

$$-\frac{d[A]}{dt} = (k_f + k_{nr})[A] \tag{1.2}$$

When assuming $[A]_0$ to be the concentration of molecules in the excited state S_1 at time zero, the integration of the previous expression yields

$$[A] = [A]_0 e^{-(k_f + k_{nr})t} \tag{1.3}$$

The fluorescence lifetime is defined as the time it takes to a population to reach $1/e$ of $[A]_0$. Thus, τ can be inserted in Eq. 1.3 from which we obtain that the magnitude of the fluorescence lifetime is inversely proportional to the radiative and non-radiative deactivation pathways

$$\tau = \frac{1}{k_f + k_{nr}} \tag{1.4}$$

Assuming that the emission signal is proportional to the concentration of molecules, the fluorescence lifetime can be calculated from the fluorescence exponential decay expressed in

Eq. 1.3 or, where more than one lifetime is present, from the sum of decay contributions

$$I = \sum_i a_i e^{-t/\tau_i} \tag{1.5}$$

Lastly, the fluorescence quantum yield is defined as the fraction of molecules that return to the ground state via radiative emission, or the ratio between the number of photons emitted and the number of photons absorbed [Valeur, 2001], given by

$$\phi = \frac{k_f}{k_f + k_{nr}} \tag{1.6}$$

therefore, the value of the fluorescence quantum yield ranges between 0 and 1, where 0 corresponds to a non-emissive compound and 1 to a highly emissive species where each photon absorbed results in the emission of a fluorescence photon. Moreover, it can be deduced from this expression that the quantum yield is proportional to the fluorescence lifetime.

Fluorophores can be classified in two main classes: intrinsic or extrinsic. Intrinsic fluorophores are those that occur naturally, such as the indole group of the amino acid tryptophan, which is extremely sensitive to local changes. For this reason, tryptophan has been widely used as a reporter of protein folding, in which its emission band may be shifted towards longer wavelengths when the protein is unfolded [Lackowicz, 2006; Longworth, 1971]. Other intrinsic sensors comprise the adenine dinucleotide (NADH), flavin mononucleotide (FMN) and flavin dinucleotide (FAD) [Lackowicz, 2006].

On the other hand, extrinsic fluorophores are those synthetic compounds with the desired spectral properties that are attached to the biomolecule of interest. For example, nucleic acids are barely fluorescent and require a labelling with extrinsic probes either via covalent or non-covalent interactions. Two examples of extrinsic fluorophores are the adenine and cytosine analogues 2-aminopurine (2AP) and pyrrolo-dC, which are used to report conformational changes in nucleic acid structures. One of the main advantages of 2AP and pyrrolo-dC is the possibility to insert these nucleotide analogues within the nucleic acid sequence without major perturbation. The properties and applications of 2AP will be discussed in detail in

section 1.3.3. In addition, a list with some of the most common commercial fluorophores used in the study of nucleic acids can be found in section 1.4.1. Other extrinsic fluorophores comprise the dansyl chloride (DNS-Cl) and fluorescein isothiocyanate (FITC), which can react with free amino groups to yield proteins that emit in the visible region of the spectra [Lackowicz, 2006].

The wide range of fluorophores available has contributed to extend the use of fluorescence techniques in the study of biological compounds, either at ensemble or single-molecule level. In this sense, the main characteristics that fluorophores are required to exhibit in order to be used in fluorescence imaging techniques are: i) high molar absorptivity and large quantum yield to obtain the maximum signal possible from a single fluorophore; ii) significant Stokes shift to achieve a significant separation between absorption and emission peaks; and iii) high photostability for real-time observation of fluorophores at the single-molecule level.

1.3 Fluorescence-based techniques to investigate nucleic acid structures at ensemble level

Fluorescence techniques at ensemble level are able to monitor multiple fluorescence properties such as fluorescence intensity, lifetime and anisotropy. Variations in these observables are caused by the interaction of the fluorophore with its environment, including the nucleic acid scaffold and how this reorganises its structure depending on the chemical properties of the solvent (i.e. aqueous solution) and medium (i.e. pH, ionic strength, oxygen concentration). The observation of changes in the fluorescence intensity and anisotropy is often used in binding assays to quantify the binding affinity between a nucleic acid and a ligand or protein. These assays are carried out in conventional fluorimeters that are available in most laboratories. On the other hand, evaluation of fluorescence lifetimes allows us to discriminate between fluorescent probes or nucleic acid conformations according to their different and characteristic lifetime when they are spectrally overlapped. However, techniques such as time-correlated single photon counting (TCSPC) for observation of nanosecond lifetimes

require expensive equipment, including pulsed lasers and fast acquisition detectors, that are normally available in more specialised physics laboratories.

In addition to biophysical studies focused on monitoring the relative variation in fluorescence, either enhancement or quenching, and changes in anisotropy, certain assays include fluorescent sensors that have been tailored to report changes in the environment of the nucleic acid structure at specific positions. This is the case for the 2-aminopurine, an adenine fluorescent analogue, and ESIPT probes, fluorophores that undergo a proton transfer in the excited state.

Other assays explore the interaction between two fluorophores via fluorescence resonance energy transfer (FRET) to extract variations in the distance between them when these are attached at specific domains within the same or different biomolecules. FRET has been used for more than 30 years as a distance reporter to investigate conformational changes in nucleic acids but it was in combination with single-molecule techniques when FRET has shown its huge potential to monitor either conformational changes or interactions between biomolecules. For this reason, the use of FRET will be discussed in detail in section 1.4.

1.3.1 Fluorescence quenching

Fluorescence quenching refers to the decrease in the fluorescence signal of a given compound, which can be classified in a static or dynamic process depending on the deactivation pathway. The dynamic quenching is due to excited-state interactions and refers to the process in which a molecule in close contact with the fluorophore assist the deactivation process via non-radiative pathways, which includes processes such as ESIPT and FRET. In this case, the expression of the fluorescence quantum yield incorporates one additional deactivation term that includes the deactivation rate, k_q, and the concentration of quencher, $[Q]$

$$\phi_Q = \frac{k_f}{k_f + k_{nr} + k_q[Q]} \tag{1.7}$$

The ratio between the quantum yield in the absence (Eq. 1.6) and presence of quencher (Eq. 1.7) yields

$$\frac{I}{I_Q} = \frac{\phi}{\phi_Q} = 1 + K_{SV}[Q] \tag{1.8}$$

which is known as the Stern-Volmer relationship. The Stern-Volmer constant, K_{SV}, is equal to $k_q\tau$, where τ represents the fluorescence lifetime in the absence of quencher.

On the contrary, the static quenching is due to the formation of stable complexes between the fluorophore and another molecule in the ground state. The fluorescence intensity in this case is given by

$$\frac{I}{I_Q} = 1 + K_a[Q] \tag{1.9}$$

where K_a is the association constant of the complex. Since the static quenching affects only to complexes formed in the ground state, the fluorescence lifetime of the compound remains unchanged, which can be used to discriminate between this mechanism and dynamic quenching.

Fluorescence quenching assays have a wide range of applications in biophysics, from binding [van de Weert and Stella, 2011] to kinetic assays. Among these are studies on the exposure of tryptophan residues in proteins [Eftink and Ghiron, 1976] or studies that monitor the aggregation of β-amyloid [Quinn et al., 2014], including our work on the inhibition of these amyloid aggregates by the protein 17β-hydroxysteroid dehydrogenase type 10 (17β-HSD10) that was carried out in collaboration with the Frank Gunn-Moore lab (School of Biology, University of St Andrews) [Aitken et al., 2016].

1.3.2 Anisotropy

Light is an electromagnetic wave characterised by an electric and associated magnetic fields perpendicular to each other and to the direction of propagation. The electric field of light fluctuates with no preferential orientation except for linearly polarised light, in which the electric field is restricted to a certain orientation. The absorption of polarised light depends on the orientation of the transition dipole moment of the fluorophore and the orientation of

the electric field. Thus, the absorption of light is maximum when both dipole and electric field are parallel and zero when they are perpendicular to each other [Valeur, 2001].

Anisotropy assays are performed with excitation of the sample by vertically polarised light and collection of the fluorescence emission through a polariser oriented vertically and perpendicularly to the excitation light. Furthermore, anisotropy can be measured either in steady-state or time-resolved measurements. The main advantage of steady-state over time-resolved anisotropy assays is that these can be carried out in an inexpensive fluorescence spectrophotometer, without the need of using pulsed lasers and a TCSPC setup. The emission anisotropy, r, can be calculated from the expression

$$r = \frac{I_{\parallel} - GI_{\perp}}{I_{\parallel} + 2GI_{\perp}} \tag{1.10}$$

where the G-factor represents a geometry factor that considers the different sensitivity of monochromator and detector to the vertically and horizontally polarised light. The theoretical anisotropy value when the absorption and emission transition dipole moments are parallel, thus the maximum anisotropy value, is 0.4. However, the maximum experimental value ranges from 0.32 to 0.39 due to instrumental artifacts and torsional vibrations of the fluorophore [Valeur, 2001].

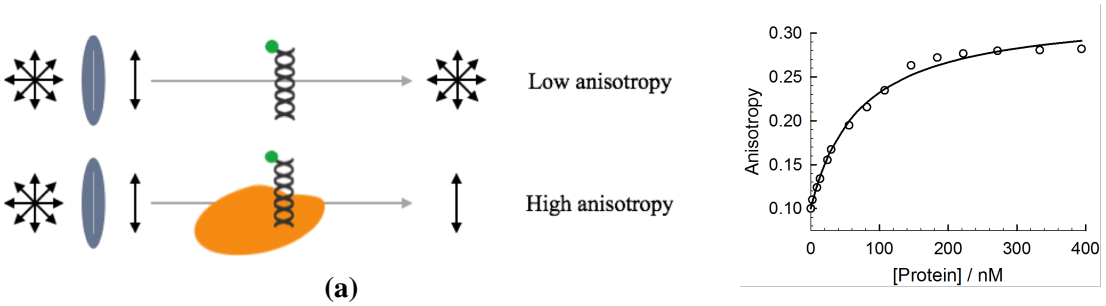


Fig. 1.3 Fluorescence anisotropy-based binding assays. (a) Schematics of an anisotropy-based assay of protein binding to a fluorescently labelled DNA. Small molecules such as DNA rotate fast in solution and efficiently depolarise light, whereas larger compounds such as DNA-protein complexes show a more restricted rotation that is translated in a higher anisotropy value. (b) Example of fluorescence-anisotropy data of DNA-protein binding assays.

Polarisation of light makes anisotropy a powerful tool for the study of binding assays in freely diffusing complexes. A fluorophore or labelled substrate that rotates fast in solution efficiently depolarises light by emitting with a different polarisation than the excitation light, thus displaying a low anisotropy value. When bound to a larger molecule, the restriction in the rotational motion leads to a less efficient depolarisation of light, which results in a high anisotropy (Fig. 1.3a). Where possible, the fluorophore is always placed in the molecule with less molecular weight to achieve the maximum change in the rotational motion of labelled compounds. The emission anisotropy values can be then represented as a function of the concentration of the large molecule to correlate the anisotropy changes and binding, from which obtain the dissociation constant (K_d) of the system (Fig. 1.3b).

1.3.3 2-Aminopurine

The use of the fluorescent adenine analogue 2-aminopurine (2AP), which also forms hydrogen bonds with uracil and thymine (Fig. 1.4a), has been introduced due to the need to develop new fluorescent tools to monitor local changes in nucleic acid structures [Perez-Gonzalez et al., 2016a]. The absorption spectrum of this intrinsically fluorescent adenine analogue presents a band at 305 nm that, despite its low molar extinction coefficient (~6,000 $M^{-1}cm^{-1}$), allows to selectively excite the 2AP without interacting with the RNA or DNA nucleobases. The dependence of its quantum yield, lifetime and anisotropy with the surrounding medium makes this analogue a valuable sensor to study local rearrangements in the structure of DNA or RNA, in which the high quantum yield of 2AP when free in solution (0.68) is dramatically quenched when stacked with other nucleobases [Wilhelmsson, 2010].

The high sensitivity of the 2AP fluorescence to report small changes in the local environment of nucleic acids constitutes a powerful tool in the study of the folding pathway and ligand binding mechanism of RNA riboswitches either as a free ligand analogue in the adenine riboswitch, where the fluorescence emission of 2AP is quenched upon ligand binding (Fig. 1.4b), or by incorporation at specific positions of the aptamer to report base stacking in the SAM-I and adenine riboswitches, among others (Figs. 1.4c) [Gilbert et al., 2006; Heppell et al., 2011; Lemay et al., 2006; Rieder et al., 2007].

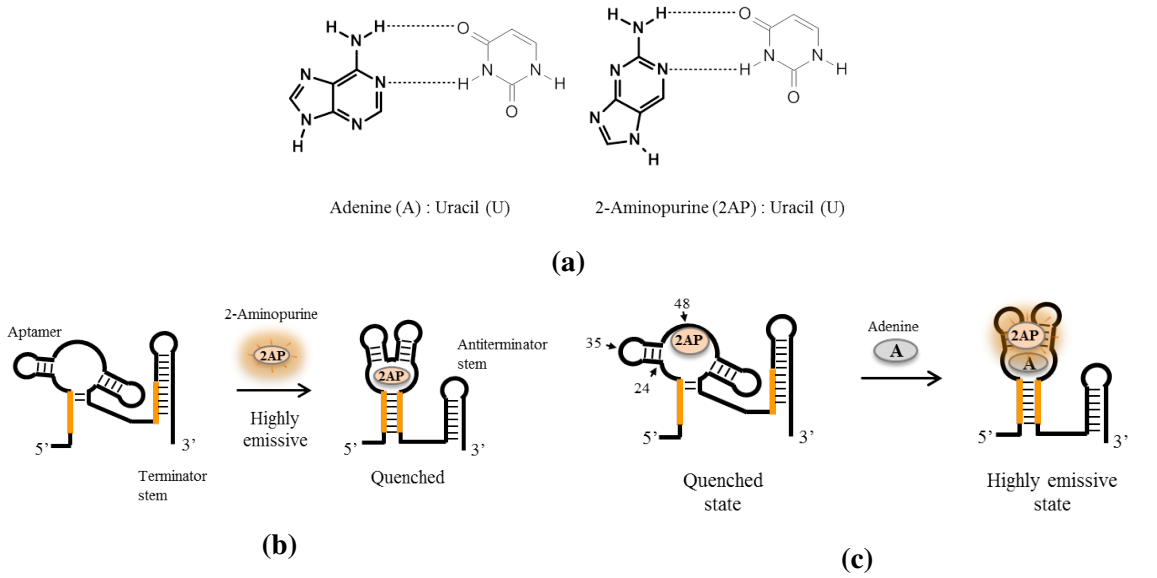


Fig. 1.4 2-aminopurine as a reporter of local structural changes. (a) Schematics of the hydrogen bonding of adenine and its analogue 2-aminopurine bases with uracil. (b) Use of 2AP as analogue of the natural adenine ligand to bind the adenine riboswitch aptamer with similar affinity. (c) Incorporation of 2AP at specific position within the adenine riboswitch to monitor local changes in the aptamer upon ligand binding. The U48AP substitution allows to monitor the increase in the signal of 2AP upon ligand binding and folding of the aptamer into a conformation in which the 2AP is directed towards the solvent [Lemay et al., 2006; Perez-Gonzalez et al., 2016b].

1.3.4 Excited-State Intramolecular Proton Transfer

Absorption of light by a compound carrying a hydroxyl group bound to an aromatic ring motivates a variation in the charge of the -OH, thus altering the acidic properties of the compound. These are called photoacid species and can increase their acidity by 10^{10}-fold in the excited state, as determined from the pK_a values obtained by the Förster cycle [Valeur, 2001]. The presence of a proton acceptor group in close proximity (no further than 2Å) implies the formation of a hydrogen bond between the hydroxyl and the acceptor group, which leads to an ultrafast proton transfer between the two groups in the subpicosecond timescale [Das et al., 1994; Schwartz et al., 1992]. In fact, the short distance between proton donor and acceptor groups is the driven force of this process. Moreover, both hydroxyl and acceptor groups can be part of different molecules (excited-state intermolecular proton transfer) or be part of the same molecule (excited-state intramolecular proton transfer, ESIPT).

ESIPT probes are compounds formed by a heterocyclic ring with a hydrogen bond between a hydroxyl group and a close proton acceptor. In these probes the normal or enol conformer is more stable in the ground state than the tautomer or keto form. However, this situation changes upon excitation of the sample. The absorption of light leads to a rearrangement of the charge around the -OH and the proton acceptor group, altering the acid-base properties of the probe, which results in the phototautomerisation process that leads to the formation of the tautomer (Fig. 1.5). As a result, the emission spectrum shows two different bands corresponding to these two species. The band with higher energy shows a normal Stokes shift and corresponds to the normal or enol species, whilst the band with lower energy and abnormally large Stokes shift is due to the tautomer. Interestingly, the presence of this band in the spectrum makes it ideal to avoid any undesired self-absorption of

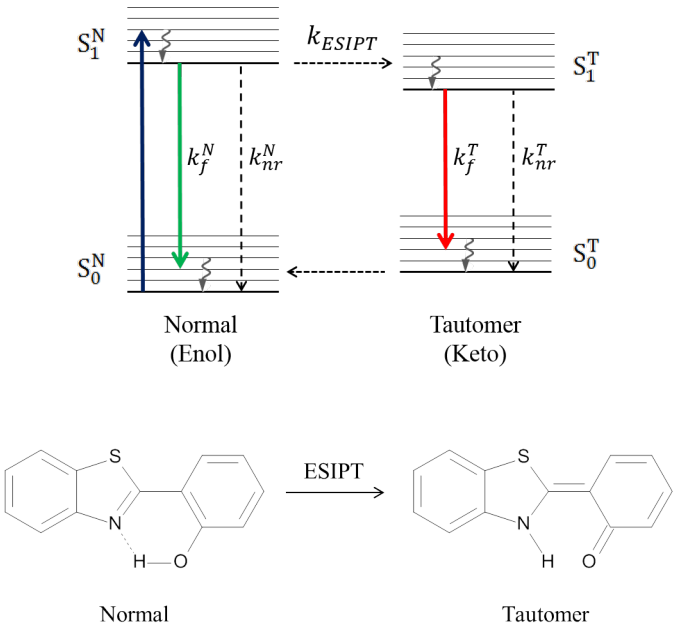


Fig. 1.5 Jablonski diagram illustrating the ESIPT process. The normal form in its first electronic excited state can either emit fluorescence, relax via non-radiative pathways or ungergo ESIPT leading to the formation of the tautomer species in the excited state via proton transfer from the -OH to the imino N in the benzoimidazole ring. Tautomer can then emit radiation with lower energy than the normal form. The tautomer in the ground state undergoes GSRPT and the sensor returns to the normal species in its ground state. A scheme of the ESIPT process is represented at the bottom.

emitted light, or inner filter effect. Once in the ground state, the tautomer undergoes a proton transfer towards the hydroxyl group in a process called ground-state reverse proton transfer (GSRPT) and the fluorophore reaches again the ground state of the normal form.

The behaviour of these probes also depends on the solvent used. For example, protic solvents such as water interact via hydrogen bonding with the hydroxyl group. Hence, these solvents preclude the proton transfer between the proton donor and acceptor groups and have influence in the ratio between the emission of the tautomer and normal species. The applications of this type of sensors are mainly focused on the detection of metal ions such as Cu^{2+} and Zn^{2+} either *in vitro* or *in vivo* (Fig. 1.6) [Ulrich et al., 2008; Yang et al., 2015; Zhao et al., 2012]. However, the presence of two bands makes these probes very attractive for ratiometric studies with other biological applications as has been already reported for β-cyclodextrins [Chipem et al., 2013], β-amyloid [Freire et al., 2015], single-stranded DNA and duplex DNA [Klymchenko et al., 2008].

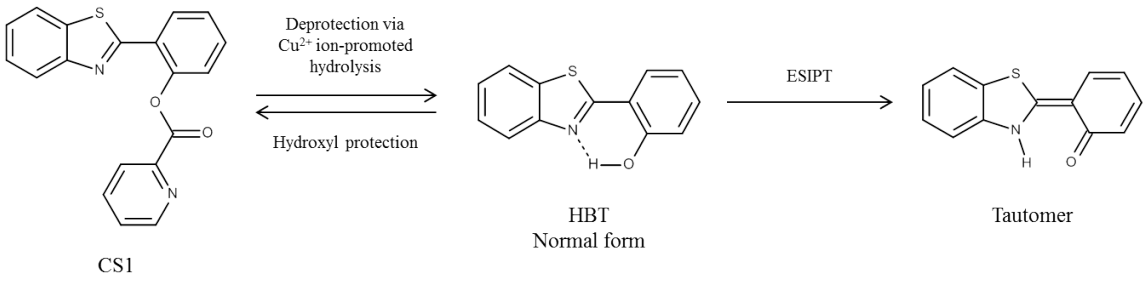
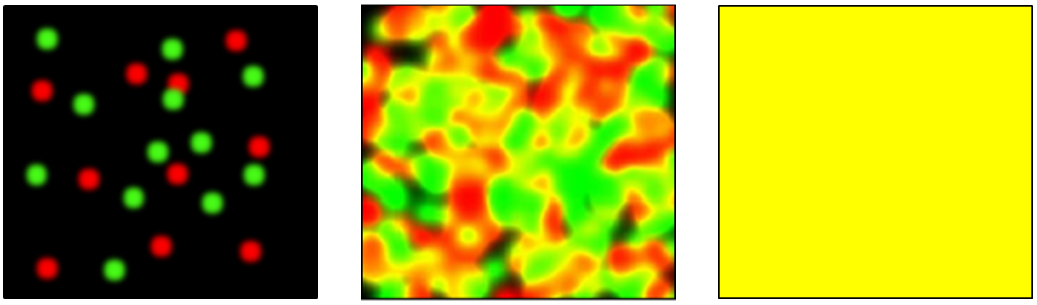


Fig. 1.6 Metal ions detection by ESIPT probes. Schematics of the Cu^{2+} ion-catalysed deprotection of compound CS1 yielding the fluorescent probe 2-(2'-hydroxyphenyl)benzothiazole (HBT), which undergoes ESIPT, as a strategy for the selective detection of Cu^{2+} in living cells by monitoring the increase in the emission of the tautomer species [Yang et al., 2015].

1.4 Biophysical techniques to investigate nucleic acid structure at single-molecule level

The observation of fluorescent processes at ensemble level provides useful information about the system under study, however, recording the signal arising from many molecules



(a) (b) (c)

Fig. 1.7 Schematics showing the static averaging effect in ensemble fluorescence techniques. A compound showing two populations, green and red (a), can be well characterised by observation of discrete molecules that are spatially separated. As the concentration increases, the correct discrimination of the two populations is more limited (b), until they report a unique yellow signal (c) arising from the average of the two populations, which cannot specifically provide information about any of them.

at a time can only report an average signal. This averaging effect is significant in the case of two or more populations coexisting in equilibrium, in which they are indistinguishable (Fig. 1.7). Moreover, not only the information about the different species is lost but also, because molecules are not temporarily synchronised, any information about the dynamics between different nucleic acid structures is hidden in the average (Fig. 1.8). On the contrary, single-molecule techniques provide several possibilities with freely diffusing and surface immobilisation methods to allow real-time observation of the molecular structure and dynamics of discrete molecules.

In the last decades, single-molecule microscopy methods have been developed to overcome the intrinsic limitations of signal averaging. The last technical improvements in fluorescence detectors and the development of new methods to limit the observation volume, and thus improve the signal-to-noise ratio (SNR), have made possible the detection of the low number of photons emitted by a single fluorophorore, providing the ultimate diagnostic and sensitivity limit. The first observation of single fluorescent molecules was carried out by Moerner and Kador for pentacene molecules hosted in a crystal matrix at liquid helium temperatures [Moerner and Kador, 1989]. Shortly after the first observation of isolated molecules at room temperature was reported [Betzig and Chichester, 1993] and the first experiments in aqueous solution were possible only two years later by introducing total-internal

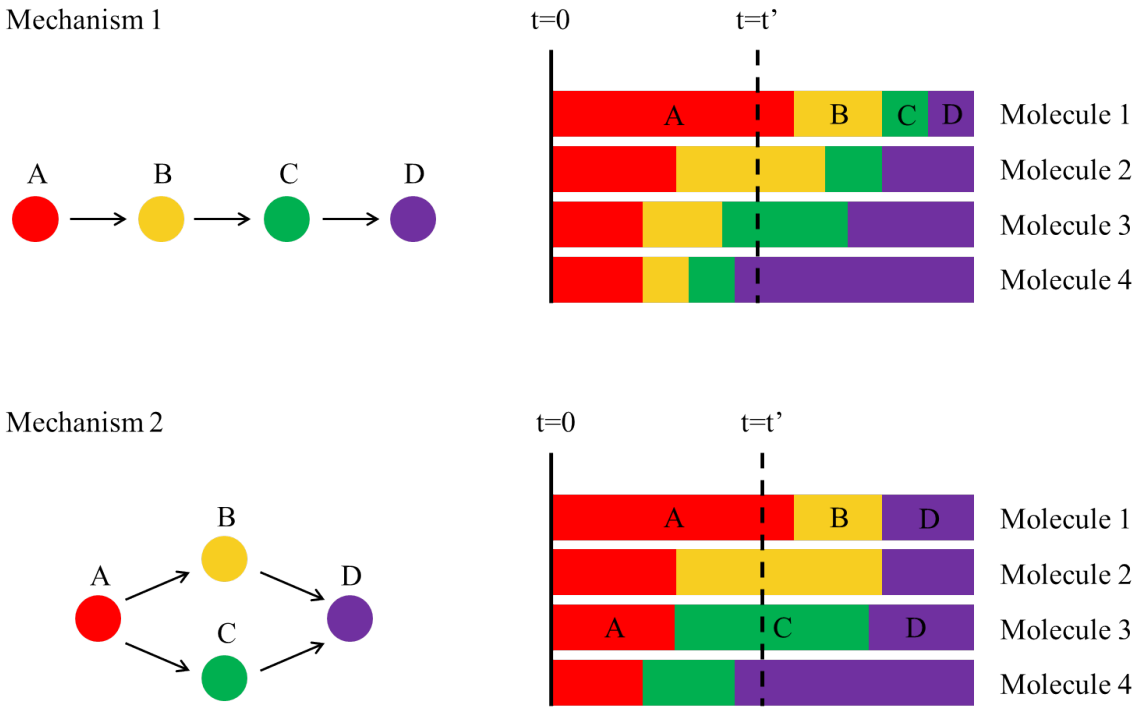


Fig. 1.8 Schematics showing the dynamic averaging effect in ensemble fluorescence techniques. In steady-state techniques, two different mechanisms can display the same signal at two different observation times, t=0 and t=t', which results in the loss of any information about the transitions between the different populations studied. Moreover, the intrinsic signal averaging of ensemble techniques is also unable to provide any detail about the lifetime of each population.

reflection microscopy [Funatsu et al., 1995]. Later, William Moerner's group discovered the possibility to turn on and off the fluorescence emission of a derivative of the green fluorescent protein [Dickson et al., 1997], which was used by Eric Betzig and Harald Hess to develop the photoactivated localisation microscopy (PALM) [Betzig et al., 2006], a wide-field technique based on the activation of a reduced number of photoswitchable fluorescent proteins to allow the observation of spatially separated fluorophores, and thus not restricted by the diffraction limit. The advances in the field of single-molecule microscopy carried out by William Moerner and Eric Betzig were awarded with the Nobel Prize in Chemistry for 2014, shared with Stefan Hell for his work in the development of the stimulated emission depletion (STED) microscopy that allows a selective deactivation of fluorophores to obtain images with super-resolution [Hell and Wichmann, 1994].

Fluorescence-based single-molecule techniques *in vitro* require either single-, in studies such as the observation of diffusion times, certain multi-colour analysis and interactions by particle tethered motion; or multiple-labelled biomolecules, in studies involving structural changes of one biomolecule or changes in the distance between two biomolecules. Because the biological function of many RNA and DNA structures intrinsically involves structural rearrangements of the nucleic acid sequence, single-molecule techniques are commonly combined with FRET in order to monitor conformational changes in nucleic acids and proteins as well as protein-DNA/RNA interactions.

1.4.1 Fluorescence Resonance Energy Transfer

Fluorescence resonance energy transfer (FRET) is a process based on a weak coupling between the oscillating dipole moments of two fluorophores, a donor and an acceptor, that are very close to one another. These energy donor and acceptor fluorophores can be attached to either the same molecule (intramolecular FRET), or different molecules (intermolecular FRET). The non-radiative interaction between them represents an additional deactivation pathway of the donor from the first excited state, along with the emission of fluorescence and other non-radiative pathways (Fig. 1.9a). As a result, FRET is observed as a decrease in the signal of the donor and an enhancement of the acceptor fluorescence. Moreover, FRET quenchers can substitute the acceptor fluorophore to quench the donor. In such case, the close proximity between donor and quencher is only translated in a quenching of the first. The emission spectrum would display a single peak and, therefore, these assays monitor changes in the donor emission intensity as a function of the variation in the inter-dye distance.

The resonant energy transfer, E_{FRET}, depends inversely on the sixth power of the inter-dye distance, r, what makes this technique very sensitive to small distance changes and facilitates its use as a spectroscopic ruler in the 10–75 Å scale [Ha, 2001; Stryer, 1978], within the range for many nucleic acid and protein structures. The derivations needed to obtain the expression that relates the E_{FRET} and r are explained more in detail in Appendix A. This

relationship is given by

$$E_{\mathrm{FRET}} = \frac{1}{1+\left(\frac{r}{R_0}\right)^6} \tag{1.11}$$

where R_0 is the Förster radius, which represents the inter-dye distance at 50% transfer efficiency. The variation of the energy transfer as a function of the donor-acceptor distance is represented in Fig. 1.9b assuming a R_0 of 60 Å for the Cy3-Cy5 pair. This term can be calculated for each pair of dyes from

$$R_0^6 = \frac{9000\,(\ln 10)\,\kappa^2\,Q_D}{128\,\pi^5\,N_A\,n^4}\int_0^\infty F_D(\lambda)\,\varepsilon_A(\lambda)\,\lambda^4\,\mathrm{d}\lambda \tag{1.12}$$

where κ^2 is the dipole orientation factor (normally assumed to be 2/3 for anisotropic orientation of the dyes), Q_D represents the donor quantum yield, N_A is Avogadro's number and n is the refractive index of the medium. The integral represents the overlap between donor emission (F_D) and acceptor absorption spectra (ε_A) in units of $\mathrm{M}^{-1}\mathrm{cm}^{-1}$.

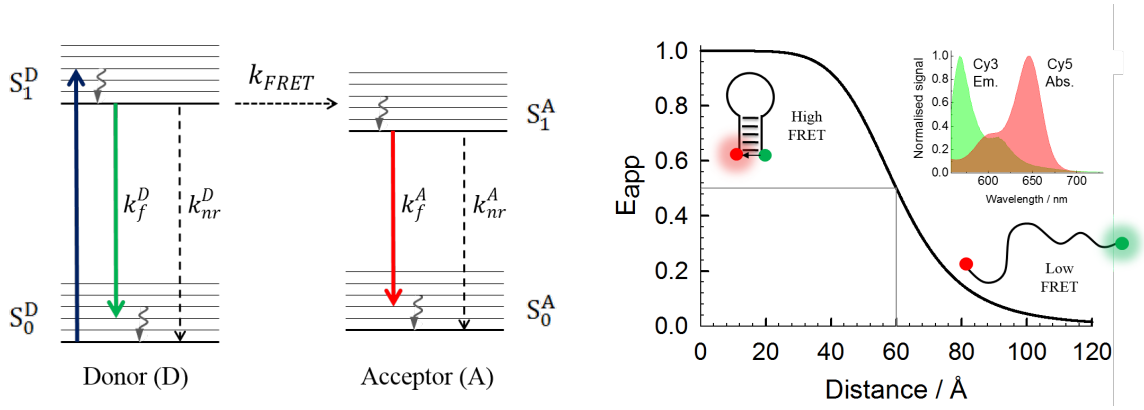


Fig. 1.9 Single-molecule FRET in nucleic acids studies. (a) Jablonski diagram illustrating the FRET process. The donor reaches its first electronic excited state (S_1^D) upon absorption of radiation. The fluorophore can then emit fluorescence (k_f), relax via non-radiative process (k_{nr}) or undergo FRET (k_{FRET}) to an acceptor or quencher in close proximity. The acceptor in an excited state (S_1^A) can finally relax via radiative or non-radiative pathways. (b) The FRET efficiency as a function of the inter-dye distance, r, for the Cy3-Cy5 FRET pair assuming an R_0 of 60 Å. *Right inset:* Spectral overlap between the emission spectrum of Cy3 (green) and the absorption spectrum of Cy5 (red).

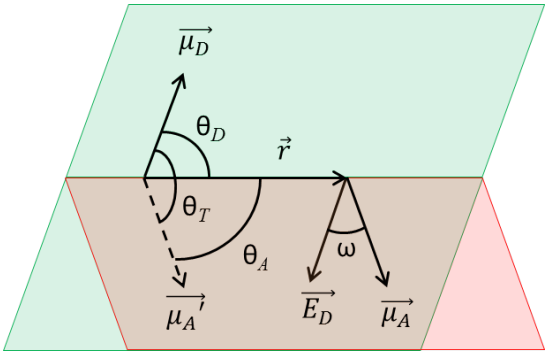
However, the value of the dipole orientation factor, κ^2, may differ from 2/3 and oscillate between 0 and 4 depending on the freedom of rotation and orientation of the donor and acceptor dipole moments [van der Meer et al., 2013]. The orientation factor κ^2 can be defined by

$$\kappa^2 = (\cos\theta_T - 3\cos\theta_D \cos\theta_A)^2 \tag{1.13}$$

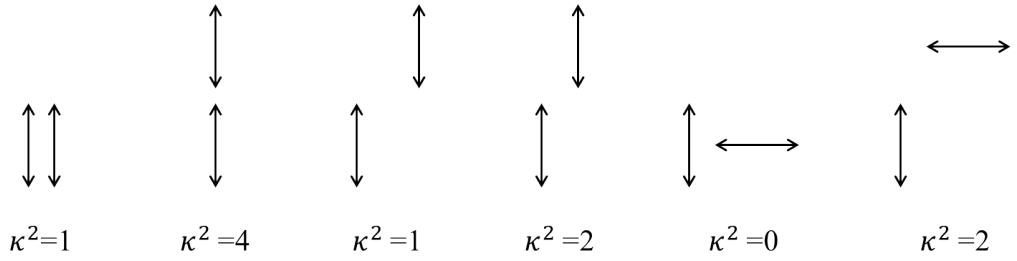
or

$$\kappa^2 = (1+\cos^2\theta_D)\cos^2\omega \tag{1.14}$$

where $\vec{r}$ represents the vector from the center of the donor to the center of the acceptor, θ_T is the angle between donor and acceptor dipole moments, θ_D is the angle between $\vec{r}$ and donor



(a)



(b)

Fig. 1.10 Determination of κ^2. (a) Schematic illustration of the donor ($\vec{\mu_D}$) and acceptor ($\vec{\mu_A}$) dipole moments, distance vector ($\vec{r}$) and summary of angles used to determine the orientation factor between the two fluorophores. The angle between the acceptor dipole moment and the electric field of the donor at the location of the acceptor ($\vec{E_D}$), which is within the $\vec{\mu_D}$-$\vec{r}$ plane (green), is represented by ω. (b) Examples of κ^2 values for different relative orientations of donor and acceptor dipole moments. The donor dipole moment is represented at the bottom or left side of each pair [van der Meer et al., 2013].

dipole, θ_A is the angle between $\vec{r}$ and acceptor dipole and ω represents the angle between the acceptor dipole and the electric dipole field due to the donor at the location of the acceptor, $\vec{E_D}$ (Fig. 1.10a). Thus, the FRET efficiency is maximum ($\kappa^2 = 4$) when both dipole moments are along the same direction and minimum ($\kappa^2 = 0$) when they are perpendicular to each other as shown in Fig. 1.10b.

The commercial Alexa Fluor (Thermo Fisher Scientific, Loughborough, UK), Atto (ATTO-TEC GmbH, Siegen, Germany) and Cyanine (GE Healthcare Life Sciences, Little Chalfont, UK) dyes are among the most commonly used dyes in experimental FRET nowadays with a wide variety of alternatives in the near-UV, visible and near-IR range (Table 1.2). The Förster radii for some of the most used FRET pairs calculated assuming anisotropic orientation of the dyes are shown in Table 1.3. In order to be used in FRET-based techniques, a FRET pair must have, in addition to a high molar absorptivity and high photostability: i) significant spectral overlap between donor emission and acceptor absorption spectra; ii) good separation between both emission spectra (~100 nm); and iii) wide variety of different

Table 1.2 Properties of the most common fluorescent dyes and dark quenchers used in single-molecule FRET.

Dye	λ_{abs}, λ_{em} / nm	ε_{max} / $M^{-1}cm^{-1}$	FRET donors	FRET acceptors
Alexa350	346, 442	35,000		Alexa488
Cy2	489, 506	150,000		Cy3
Alexa488	495, 519	73,000	Alexa350	Alexa555, Alexa647
FAM	496, 516	83,000		Cy3
Cy3	550, 570	150,000	FAM, Cy2	Alexa647, Cy5, Cy5.5
Atto550	554, 576	120,000		Atto647N
Alexa555	555, 565	155,000	Alexa488	Alexa647
Atto647N	644, 669	150,000	Atto550, Cy3	
Cy5	649, 670	250,000	Cy3	Cy7
Alexa647	650, 668	270,000	Alexa488, Alexa555, Cy3	
Cy5.5	675, 694	250,000	Cy3	Cy7
Cy7	747, 776	200,000	Cy5, Cy5.5	
Dabcyl	478, —	32,000	Alexa350, Alexa488, FAM	
QSY9	562, —	85,000	Alexa488, FAM, Cy3, Alexa555	
QSY21	660, —	89,000	Cy5, Alexa647	

Table 1.3 Förster radius (R_0) for the most common FRET pairs used in the study of nucleic acids.

Donor	Acceptor	R_0 / Å	References
Alexa488	Dabcyl	49	[Spence and Johnson, 2010]
Alexa488	Cy5	52	[Woźniak et al., 2008]
Alexa488	Alexa555	70	[Spence and Johnson, 2010]
Alexa488	Alexa647	56	[Spence and Johnson, 2010]
Fluorescein	Cy3	56	[Norman et al., 2000]
Cy3	Cy5	54–61	[Lee et al., 2010b; Murphy et al., 2004]
Cy3	Cy5.5	59	[Hohng et al., 2004]
Cy3	Alexa647	51	[Coban et al., 2006]
Alexa555	Alexa647	51	[Spence and Johnson, 2010]
Alexa555	QSY9	45	[Spence and Johnson, 2010]
Cy5	Cy7	62	[Lee et al., 2010b]

reactive groups to facilitate the site-specific labelling of nucleic acids and proteins. Moreover, to be used in smFRET they are also required to exhibit similar quantum yields in order to obtain clear anti-correlated trajectories without further signal correction [Roy et al., 2008].

The application of FRET to extract distances in biomolecules was pioneered by Clegg and Lilley in the late 1980s-early 90s. Their work on dsDNA and particularly on the structure of 4-way DNA junctions, intermediates in DNA recombination known as Holliday junctions, and its dependency on different salt concentrations has exploited the potential of FRET in the study of nucleic acids at the ensemble level [Clegg et al., 1992, 1993; Murchie et al., 1989]. Since then, FRET has greatly contributed to our understanding of nucleic acids, proteins and their interactions at ensemble level. However, it is at the single-molecule level where we can take advantage of its potential in the observation of conformational changes and dynamics in a timescale ranging from milliseconds to minutes, which includes most biologically relevant processes. The use of FRET in the study of nucleic acids at the single-molecule level was firstly reported by Ha and coworkers [Ha et al., 1996] and has greatly contributed to the understanding of nucleic acids. Interestingly, advances in smFRET and in the study of Holliday junctions have followed parallel paths. The complexity of these 4-way junctions is

ideal to introduce new approaches in the technique, and thus their structure and dynamics have been investigated not only by incorporation of two [McKinney et al., 2003] but also three [Hohng et al., 2004] and up to four different dyes [Lee et al., 2010a], which allows to determine six different inter-dye FRET efficiencies in real time.

1.5 Current challenges in single-molecule spectroscopy to investigate RNA folding during transcription

The mRNA biogenesis includes different cotranscriptional processes such as 5' capping or splicing, in which the RNA folding is directly involved [Perales and Bentley, 2009]. The importance of the cotranscriptional folding is also present in gene regulation carried out by RNA riboswitches, which undergo structural rearrangements upon binding of small cellular metabolites to modulate the genetic expression in certain bacteria, archaea, plants and fungi.

To date, the study of RNA riboswitches using smFRET has been restricted to the observation of synthetic or fully-transcribed sequences and completely folded RNA structures to which donor and acceptor are tethered at specific positions. Chapters 4 and 5 of this thesis are good examples of how smFRET has been used in the study of full-length SAM-I and adenine riboswitches. However, the study of full length sequences is unable to completely understand the ligand binding mechanism and riboswitch folding pathway due to the transient nature of RNA transcripts. It is also important to emphasise that the time window for transcription of RNA riboswitches is only a few seconds since bacterial RNA polymerase incorporates nucleotides to synthesise RNA at a rate of 20–80 nt/s [Pan and Sosnick, 2006], although it would depend on the concentration of nucleotides and experimental conditions, which would allow its observation using single-molecule TIR microscopy.

The inability of RNA polymerase to incorporate labelled nucleotides [Vassylyev et al., 2007] represents the main limitation in the study of cotranscriptional folding using fluorescence-based techniques. A recently reported strategy to visualise RNA synthesis involves incorporation of uridine analogues carrying an alkyne group that reacts with azide-labelled fluorophores via click chemistry to visualise RNA synthesis generated by RNA polymerases I, II and III in

cells *in vivo* [Jao and Salic, 2008]. Although the incorporation of functionalised nucleotides provides a means for the observation of RNA transcription, its use is incompatible with smFRET studies due to the impossibility to selectively label specific positions within the RNA sequence.

We have overcome these limitations in the study of a TPP-binding riboswitch and provide in chapter 7 a new method for the site-specific labelling of nascent mRNA sequences in transcriptionally-paused complexes using a click chemistry approach, which is a crucial step in the study of cotranscriptional RNA folding and dynamics at the single-molecule level using FRET. Moreover, this method provides an alternative for the study of nascent RNAs using different biophysical techniques such as EPR by attaching spin labels at specific positions within the RNA sequence. The promising results presented in this chapter offer a motivation to work in new approaches for the study of nascent RNAs also in combination with force-based techniques, which would provide insights into the cotranscriptional folding of mRNAs with an unprecedented level of detail.

References

Aitken, L., Quinn, S. D., Perez-Gonzalez, C., Samuel, I. D. W., Penedo, J. C., and Gunn-Moore, F. J. (2016). Morphology-specific inhibition of β-amyloid aggregates by 17β-hydroxysteroid dehydrogenase type 10. *ChemBioChem*, 17(11):1029–1037.

Betzig, E. and Chichester, R. J. (1993). Single molecules observed by near-field scanning optical microscopy. *Science*, 262(5138):1422.

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., and Hess, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645.

Blount, K. F. and Breaker, R. R. (2006). Riboswitches as antibacterial drug targets. *Nature biotechnology*, 24(12):1558.

Chipem, F. A. S., Behera, S. K., and Krishnamoorthy, G. (2013). Enhancing excited state intramolecular proton transfer in 2-(2'-hydroxyphenyl) benzimidazole and its nitrogen-substituted analogues by β-cyclodextrin: The effect of nitrogen substitution. *The Journal of Physical Chemistry A*, 117(20):4084–4095.

Clegg, R. M., Murchie, A. I., Zechel, A., and Lilley, D. M. (1993). Observing the helical geometry of double-stranded DNA in solution by fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences*, 90(7):2994–2998.

Clegg, R. M., Murchie, A. I. H., Zechel, A., Carlberg, C., Diekmann, S., and Lilley, D. M. J. (1992). Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction. *Biochemistry*, 31(20):4846–4856.

Coban, O., Lamb, D. C., Zaychikov, E., Heumann, H., and Nienhaus, G. U. (2006). Conformational heterogeneity in RNA polymerase observed by single-pair FRET microscopy. *Biophysical Journal*, 90(12):4605–4617.

Das, K., Sarkar, N., Ghosh, A. K., Majumdar, D., Nath, D. N., and Bhattacharyya, K. (1994). Excited-state intramolecular proton transfer in 2-(2'-hydroxyphenyl) benzimidazole and-benzoxazole: effect of rotamerism and hydrogen bonding. *Journal of Physical Chemistry*, 98:9126–9126.

Dickson, R. M., Cubitt, A. B., Tsien, R. Y., and Moerner, W. E. (1997). On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature*, 388(6640):355–358.

Eftink, M. R. and Ghiron, C. A. (1976). Exposure of tryptophanyl residues in proteins. quantitative determination by fluorescence quenching studies. *Biochemistry*, 15(3):672–680.

Freire, S., Rodríguez-Prieto, F., Ríos Rodríguez, M. C., Penedo, J. C., Al-Soufi, W., and Novo, M. (2015). Towards ratiometric sensing of amyloid fibrils in vitro. *Chemistry - A European Journal*, 21(8):3425–3434.

Funatsu, T., Harada, Y., Tokunaga, M., Saito, K., and Yanagida, T. (1995). Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature*, 374(6522):555–559.

Gilbert, S. D., Stoddard, C. D., Wise, S. J., and Batey, R. T. (2006). Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *Journal of Molecular Biology*, 359(3):754–768.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35(3):849–857.

Ha, T. (2001). Single-molecule fluorescence resonance energy transfer. *Methods*, 25(1):78 – 86.

Ha, T., Enderle, T., Ogletree, D. F., Chemla, D. S., Selvin, P. R., and Weiss, S. (1996). 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*, 93(13):6264–6268.

References

Hell, S. W. and Wichmann, J. (1994). Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters*, 19(11):780–782.

Heppell, B., Blouin, S., Dussault, A.-M., Mulhbacher, J., Ennifar, E., Penedo, J. C., and Lafontaine, D. A. (2011). Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nature Chemical Biology*, 7(6):384–392.

Hermann, T. (2015). Non-coding RNA: antibiotic tricks a switch. *Nature*, 526:650–651.

Hohng, S., Joo, C., and Ha, T. (2004). Single-molecule three-color FRET. *Biophysical Journal*, 87(2):1328 – 1337.

Jao, C. Y. and Salic, A. (2008). Exploring RNA transcription and turnover in vivo by using click chemistry. *Proceedings of the National Academy of Sciences*, 105(41):15779–15784.

Klymchenko, A. S., Shvadchak, V. V., Yushchenko, D. A., Jain, N., and Mély, Y. (2008). Excited-state intramolecular proton transfer distinguishes microenvironments in single-and double-stranded DNA. *The Journal of Physical Chemistry B*, 112(38):12050–12055.

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, 31(1):147–157.

Lackowicz, J. R. (2006). *Principle of Fluorescence Spectroscopy*. New York, NY, third edition edition.

Lee, J., Lee, S., Ragunathan, K., Joo, C., Ha, T., and Hohng, S. (2010a). Single-molecule four-color FRET. *Angewandte Chemie International Edition*, 49(51):9922–9925.

Lee, S., Lee, J., and Hohng, S. (2010b). Single-molecule three-color FRET with both negligible spectral overlap and long observation time. *PLOS ONE*, 5(8):1–9.

Lemay, J.-F., Penedo, J. C., Tremblay, R., Lilley, D. M., and Lafontaine, D. (2006). Folding of the adenine riboswitch. *Chemistry & Biology*, 13(8):857 – 868.

Livshits, G. I., Stern, A., Rotem, D., Borovok, N., Eidelshtein, G., Migliore, A., Penzo, E., Wind, S. J., Di Felice, R., Skourtis, S. S., Cuevas, J. C., Gurevich, L., Kotlyar, A. B., and Porath, D. (2014). Long-range charge transport in single G-quadruplex DNA molecules. *Nature Nanotechnology*, 9(12):1040–1046.

Longworth, J. W. (1971). Luminescence of polypeptides and proteins. In *Excited states of proteins and nucleic acids*, pages 319–484. Springer.

McKinney, S. A., Déclais, A.-C., Lilley, D. M. J., and Ha, T. (2003). Structural dynamics of individual Holliday junctions. *Nature Structural & Molecular Biology*, 10(2):93.

Moerner, W. E. and Kador, L. (1989). Optical detection and spectroscopy of single molecules in a solid. *Physical Review Letters*, 62(21):2535–2538.

Murchie, A. I. H., Clegg, R. M., von Krtzing, E., Duckett, D. R., Diekmann, S., and Lilley, D. M. (1989). Fluorescence energy transfer shows that the four-way DNA junction is a right-handed cross of antiparallel molecules. *Nature*, 341(6244):763–766.

Murphy, M. C., Rasnik, I., Cheng, W., Lohman, T. M., and Ha, T. (2004). Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophysical Journal*, 86(4):2530–2537.

Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L., and Breaker, R. R. (2002). Genetic control by a metabolite binding mRNA. *Chemistry & Biology*, 9(9):1043–1049.

Norman, D. G., Grainger, R. J., Uhrín, D., and Lilley, D. M. J. (2000). Location of cyanine-3 on double-stranded DNA: importance for fluorescence resonance energy transfer studies. *Biochemistry*, 39(21):6317–6324.

Pan, T. and Sosnick, T. (2006). RNA folding during transcription. *Annual Review of Biophysics and Biomolecular Structure*, 35:161–175.

Perales, R. and Bentley, D. (2009). "cotranscriptionality": The transcription elongation complex as a nexus for nuclear transactions. *Molecular Cell*, 36(2):178–191.

Perez-Gonzalez, C., Grondin, J. P., Lafontaine, D. A., and Carlos Penedo, J. (2016a). *Biophysical Approaches to Bacterial Gene Regulation by Riboswitches*, pages 157–191. Springer International Publishing, Cham.

Perez-Gonzalez, C., Lafontaine, D. A., and Penedo, J. C. (2016b). Fluorescence-based strategies to investigate the structure and dynamics of aptamer-ligand complexes. *Frontiers in Chemistry*, 4:33.

Pinheiro, A. V., Han, D., Shih, W. M., and Yan, H. (2011). Challenges and opportunities for structural DNA nanotechnology. *Nat Nano*, 6(12):763–772.

Quinn, S. D., Dalgarno, P. A., Cameron, R. T., Hedley, G. J., Hacker, C., Lucocq, J. M., Baillie, G. S., Samuel, I. D. W., and Penedo, J. C. (2014). Real-time probing of β-amyloid self-assembly and inhibition using fluorescence self-quenching between neighbouring dyes. *Molecular BioSystems*, 10(1):34–44.

Rieder, R., Lang, K., Graber, D., and Micura, R. (2007). Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. *ChemBioChem*, 8(8):896–902.

References

Rothemund, P. W. K. (2006). Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302.

Roy, R., Hohng, S., and Ha, T. (2008). A practical guide to single-molecule FRET. *Nature Methods*, 5(6):507–516.

Schwartz, B. J., Peteanu, L. A., and Harris, C. B. (1992). Direct observation of fast proton transfer: Femtosecond photophysics of 3-hydroxyflavone. *The Journal of Physical Chemistry*, 96:3591–3598.

Spence, M. T. and Johnson, I. D. (2010). *The molecular probes handbook: a guide to fluorescent probes and labeling technologies*. Live Technologies Corporation.

Stryer, L. (1978). Fluorescence energy transfer as a spectroscopic ruler. *Annual review of biochemistry*, 47(1):819–846.

Ulrich, G., Nastasi, F., Retailleau, P., Puntoriero, F., Ziessel, R., and Campagna, S. (2008). Luminescent Excited-State Intramolecular Proton-Transfer (ESIPT) Dyes Based on 4-Alkyne-Functionalized [2,2'-Bipyridine]-3,3'-diol Dyes. *Chemistry - A European Journal*, 14(14):4381–4392.

Valeur, B. (2001). *Molecular Fluorescence*. Wiley-VCH Verlag GmbH.

van de Weert, M. and Stella, L. (2011). Fluorescence quenching and ligand binding: A critical discussion of a popular methodology. *Journal of Molecular Structure*, 998(1-3):144–150.

van der Meer, B. W., van der Meer, D. M., and Vogel, S. S. (2013). Optimizing the orientation factor kappa-squared for more accurate FRET measurements. *FRET–Förster Resonance Energy Transfer: From Theory to Applications. Weinheim, Germany: Wiley-VCH*, pages 63–104.

Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. *Nature*, 448(7150):163–168.

Watson, J. D. and Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171(737-738):3.12.

Wilhelmsson, L. M. (2010). Fluorescent nucleic acid base analogues. *Quarterly Reviews of Biophysics*, 43(2):159–183.

Woźniak, A. K., Schröder, G. F., Grubmüller, H., Seidel, C. A. M., and Oesterhelt, F. (2008). Single-molecule FRET measures bends and kinks in DNA. *Proceedings of the National Academy of Sciences*, 105(47):18337–18342.

Yang, C., Chen, Y., Wu, K., Wei, T., Wang, J., Zhang, S., and Han, Y. (2015). A reactive probe for Cu^{2+} based on the ESIPT mechanism and its application in live-cell imaging. *Analytical Methods*, 7(8):3327–3330.

Yurke, B., Turberfield, A. J., Mills, A. P., Simmel, F. C., and Neumann, J. L. (2000). A DNA-fuelled molecular machine made of DNA. *Nature*, 406(6796):605–608.

Zhao, J., Ji, S., Chen, Y., Guo, H., and Yang, P. (2012). Excited state intramolecular proton transfer (ESIPT): from principal photophysics to the development of new chromophores and applications in fluorescent molecular probes and luminescent materials. *Phys. Chem. Chem. Phys.*, 14(25):8803–8817.

2

Materials and methods

2.1 Sample preparation protocols

2.1.1 Nucleic acid samples

Unlabelled and labelled DNA strands were purchased from IDT (Integrated DNA Technologies, Coralville, USA) and Primetech (Primetech ALC, Minsk, Belarus). Synthetic and transcribed RNA strands were provided either by Daniel Lafontaine's group (Université de Sherbrooke, Canada) or purchased from Dharmacon (GE Dharmacon, Lafayette, USA) with dyes incorporated.

All dsDNA and RNA riboswitches were annealed in Tris base buffer containing 50–100 mM NaCl in water bath at 80 °C for 10 minutes followed by slow cooling overnight to room temperature. DNA samples were further purified by gel electrophoresis on a 8–12% (w/v) polyacrylamide gel in 1 X TBE buffer with a background of 100 mM NaCl. Gels were pre-run in 1X TBE buffer with 10 mM NaCl at 110 V for 45 minutes in cold room (4 °C). Samples containing 3% Ficoll are loaded in the wells and run at 110 V until good separation of the hybridised sample is achieved.

Unlabelled DNAs are detected by UV shadowing using a fluor-coated thin-layer chromatography plate. To extract the sample, bands are cut and covered with Tris-HCl buffer in an eppendorf tube that is left overnight on a rocker in cold room, followed by ethanol precipitation: addition of i) 10% (v/v) sodium acetate (3 M, pH 5.2), which counteracts the

phosphate negative charges of the DNA to make it less water soluble; and ii) 5:2 ratio of cold ethanol ($\geq$ 95%), which has a lower dielectric constant than water and promotes the interaction between the DNA negative charges and the Na^+ cations to precipitate the DNA. This solution is mixed gently, left overnight at -20 °C and then centrifuged at maximum speed for 30 min in cold room to obtain a DNA pellet. Finally, the supernatant is carefully removed and the pellet resuspended in the desired buffer. The final concentration and labelling efficiency is determined by UV-Vis spectroscopy using the molar extinction coefficient of the dyes or, where not possible, the absorption of the dsDNA at ~260 nm.

2.1.2 Buffers

Tris(hydroxymethyl)aminomethane (Tris) buffers were made by addition of 0.12 or 0.30 g Trizma base (MW 121.14) (Sigma Aldrich Co., Gillingham, UK) dissolved in 50 mL of MilliQ water for concentrations of 20 or 50 mM, respectively. The pH was adjusted using a 3-point calibrated pH meter by addition of drops of a HCl solution.

2-(N-morpholino)ethanesulfonic acid (MES) buffers were prepared by addition of 0.20 or 0.49 g MES (MW 195.2) (Sigma Aldrich Co., Gillingham, UK) dissolved in 50 mL of MilliQ water for concentrations of 20 or 50 mM, respectively. The pH was adjusted using a calibrated pH meter by addition of drops of a NaOH solution.

Buffers containing 6% (w/v) glucose were prepared as described above and by addition of 3 g of D-(+)-glucose prior to calibration. These buffers were stored at 4°C.

2.2 Steady-state absorption spectroscopy

Molecules are confined to certain discrete energy levels associated to the orbitals occupied by electrons (electronic excited states) and the nuclear vibrations (vibronic excitated states). Transitions between two energy levels may occur by absorption or emission of radiation with the same energy as the energy difference between the initial and final states. Absoprtion spectroscopy is the technique that measures the amount of unpolarised light that a compound absorbs at an specific wavelength in the ultraviolet (UV) to visible (Vis) region. The principle

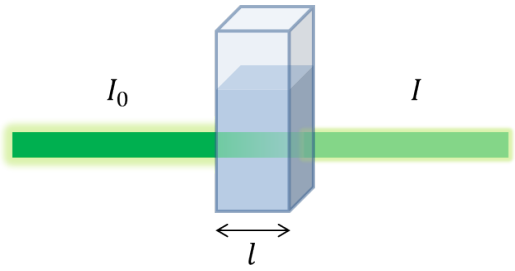


Fig. 2.1 Light absorption process. Schematics of the absorption of monochromatic light by a compound in solution in a cuvette of length l.

behind this technique is the absorption of a photon by a molecule to promote one electron to an unoccupied orbital. The molecule undergoes then a transition from the ground state to an excited electronic state with the same multiplicity. This process occurs in a timescale of ~10^{-15} s [Valeur, 2001].

The amount of light absorbed can be described using two terms, transmittance and absorbance, that are inversely proportional. The transmittance is defined as the ratio between the radiation transmitted through the medium (I) and the incident radiation (I_0) (Fig. 2.1)

$$T = \frac{I}{I_0} \tag{2.1}$$

The absorbance, A, represents the amount of radiation that the sample absorbs and is defined by the negative logarithm of the transmittance

$$A = -\log(T) = \log\frac{I_0}{I} \tag{2.2}$$

The Beer-Lambert law introduces a relationship between absorption and concentration. This law states that the amount of monochromatic radiation absorbed by a compound in solution is directly proportional to the concentration of compound in M, c, and the path length of the light through the medium in cm, l. This is

$$\log\frac{I_0}{I} = \varepsilon l c \tag{2.3}$$

where ε is the molar extinction coefficient of the compound in M^{-1} cm^{-1}, which represents its ability to absorb light at a specific wavelength. The combination of Eqs. 2.2 and 2.3 yields the expression that relates absorbance and concentration

$$A = \varepsilon l c \tag{2.4}$$

showing a linear dependence between the two terms. However, the linear range of the Beer-Lambert equation is limited by the instrumental error and, at high concentrations, due to electrostatic interactions between fluorophores or due to their aggregation.

This technique provides a means to characterise a compound by its distinctive absorption spectrum and to determine its concentration in solution, which can be of great interest to calculate the labelling efficiency of nucleic acids and proteins or to monitor dsDNA unwinding, aggregation and enzymatic processes, among others.

Absorption measurements were performed in a Cary 50 Bio UV-Vis spectrophotometer (Agilent Technologies LDA UK Ltd., Stockport, UK) equipped with a multi cuvette holder using a 1 cm width sub-micro absorption cuvette for volumes from 120 µL (Starna Cells Inc., Atascadero, USA). Small quantities of sample were measured in a NanoDrop 2000 UV-Vis spectrophotometer (Fisher Scientific, Loughborough, UK), which allows to measure 1–2 µL volumes by creating a 1 mm path length. All measurements were performed in a continuous scan mode with prior solvent background subtraction.

2.3 Circular dichroism

Circularly polarised light is composed of two perpendicular electric fields with the same amplitude but with a phase shift of one quarter of a wavelength. The resulting electric vector has a constant amplitude and rotates along the direction of propagation describing a helix, which can be either right- (E_R) or left-handed (E_L) [Woody, 1995]. Circular dichroism (CD) is the technique that measures the absorption of circularly polarised light by an asymmetric compound in the UV-Vis region. This technique is based on the difference in the absorption of left- and right-handed polarised light by a chiral sample, thus called dichroism. The

electric fields of the two beams with opposite polarisation have initially the same amplitudes but behave differently after interacting with a chiral sample because of the different molar extinction coefficients for the two polarised beams, finally adopting an ellipse shape [Woody, 1995]. The result can be expressed as ΔE, the difference between E_L and E_R; in degrees of ellipticity, which are related to ΔE as $\theta = 32.98\Delta E$, where θ is expressed in degrees; and as molar ellipticity, defined as $[\theta]=100\theta/(cl)$, where c is the concentration in M, l is the pathlength in cm and $[\theta]$ is expressed in deg cm^2 dmol^{-1}.

CD has been extensively used to characterise biological samples, such as guanine quadruplexes and proteins, according to the distinctive shape and magnitude of their spectra. It is possible to estimate the parallel, antiparallel or mixed conformation of a guanine quadruplex [Karsisiotis et al., 2011] as well as to assign the secondary structure of a peptide to an alpha-helix, beta-sheet or extended conformation [Greenfield, 2007]. Moreover, circular dichroism can be used to monitor small changes in the structure of a protein or nucleic acid at different buffer, pH, salt, additives or under thermal denaturation.

CD experiments were performed in a Jasco spectropolarimeter J-810 (JASCO UK Ltd., Essex, UK) equipped with a 150 watt Xenon lamp, a Peltier temperature controller and water bath using either circular or rectangular cuvettes. The spectropolarimeter was purged with nitrogen before use to avoid any damage on the mirrors produced by the formation of ozone in the interaction between oxygen and the light emitted by the Xenon lamp. CD spectra are represented as millidegrees of ellipticity as a function of wavelength, averaged for a total of 4 replicates. On the other hand, thermal denaturation assays are collected at a fixed wavlength, then normalised and finally represented as a function of temperature.

2.4 Steady-state fluorescence spectroscopy

Fluorescence spectroscopy is the technique that measures the emission of light from a given compound due to its relaxation from the excited electronic state S_1 to the ground state, S_0. Most fluorimeters can collect both emission and excitation spectra. The emission spectrum is collected for a fixed excitation wavelength, λ_{exc}, and reflects the fluorescence signal

that reaches the detector as a function of the λ_{em}. Although species emit in all directions, the detector is always placed at right angle from the light source in order to minimise the contributions from the scattered and transmitted light. Excitation spectrum collects the signal at a fixed λ_{em} as a function of the excitation wavelength. Since the fluorescence signal is proportional to the absorbed light, the fluorescence excitation spectrum is proportional to the absorption spectrum of the fluorophore, thus allowing us to identify the number of emissive species in solution in the emission spectrum and characterise them according to their unique excitation spectra.

Fluorescence spectroscopy, also referred as bulk or ensemble, has become one of the main techniques used in the study of biological processes for its sensitivity and simplicity, providing an inexpensive alternative that is accessible to many laboratories. Detection of low nanomolar concentrations for most fluorescent dyes permits the observation of high affinity binding events, rotational motions, conformational changes and estimate distances within a biomolecule, requiring only a small amount of sample [Lackowicz, 2006].

Fluorescence ensemble emission and excitation measurements were performed in a Cary Eclipse (Agilent Technologies LDA UK Ltd., Stockport, UK) equipped with a Xenon flash lamp, photomultiplier tube (PMT) detector, multi cuvette holder, internal automated polarisers for measurements of anisotropy and peltier water bath for temperature-controlled experiments between 0 and 100°C. Sample was contained in a ultra-micro fluorescence cuvette (Perkin Elmer Inc., Waltham, USA) with volumes ranging from 130–150 µL.

2.5 Single-molecule total internal reflection fluorescence spectroscopy

The direct observation of single molecules can be affected by a low signal-to-noise ratio (SNR) as a result of the solvent background (i.e. fluorescent impurities, Raman contribution of water) and the limited number of photons emitted by a single dye. Total internal reflection fluorescence (TIRF) spectroscopy has introduced the phenomenon of total internal reflection

to limit the excitation volume to a thin region containing surface-tethered molecules, thus reducing the background produced by excitation of the solvent.

When light travels from a medium with high refractive index (e.g. quartz, $n_1 = 1.459$) to a medium with lower refractive index (e.g. water, $n_2 = 1.333$) there is a certain incident angle, called critical angle (θ_c), above which the light is not refracted but reflected in a phenomenon known as total internal reflection. In refraction, the relation between the refractive indexes of both media and the incident and refracted angles is given by the Snell's law

$$\frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1} \tag{2.5}$$

where θ_c would correspond to the incident angle at which the refracted angle, θ_2, is 90°. Interestingly, incident angles above the critical angle also create an electromagnetic field, called evanescent wave, on the other side of the interface with the same wavelength and whose intensity decays exponentially with the distance, z, according to the expression [Axelrod et al., 1984]

$$I = I_0\, e^{-z/d} \tag{2.6}$$

in which I_0 represents the intensity of the evanescent wave at the interface and d represents the rate of variation of the evanescence signal with the distance, which depends on the parameters shown in

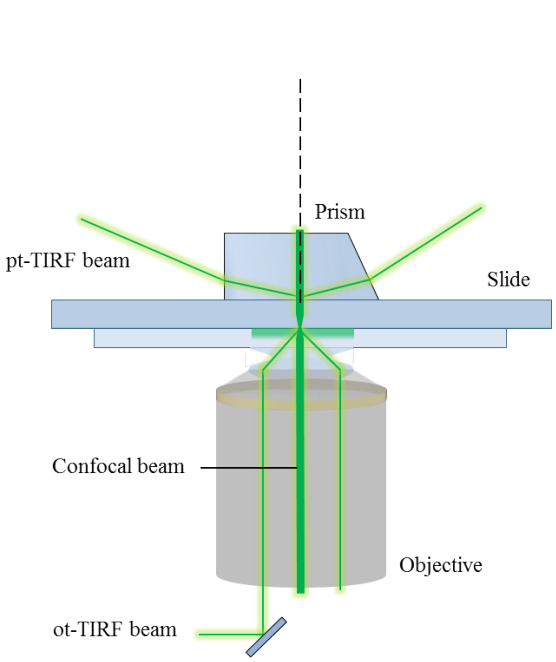
$$d = \frac{\lambda_0}{4\,\pi(n_1^2\,\sin^2\theta_1 - n_2^2)^{1/2}} \tag{2.7}$$

The evanescent wave, with typical penetration depth of about 150–200 nm for a green laser emitting at 532 nm, limits the excitation volume within the sample chamber and allows to illuminate the sample tethered to the surface of a quartz slide with a significant increase of the signal-to-noise ratio.

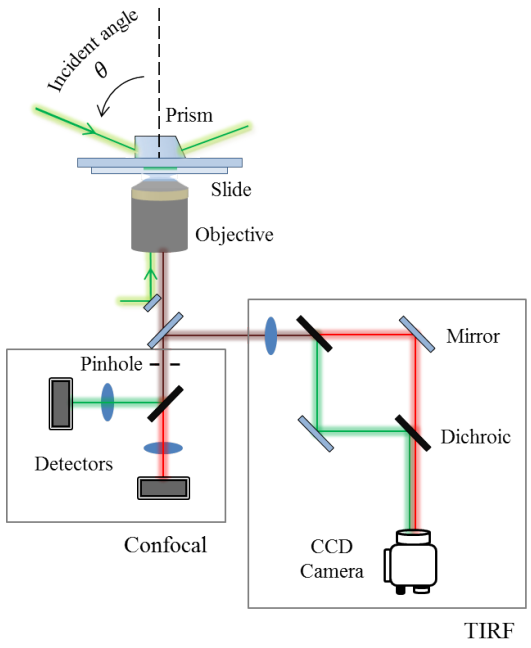
The evanescent wave is the excitation method of this wide-field technique that allows multiple surface-immobilised molecules (150–250) to be excited simultaneously with high temporal (up to ~16 ms in our particular setup when operating with half chip) and spatial resolution. Two main types of microscopes can be differentiated depending on how the light is directed into the slide and the evanescent wave is generated, either through a prism

placed on top of the slide, so-called prism-type (pt-TIR), or through the objective, in an objective-type (ot-TIR) conformation (Fig. 2.2a).

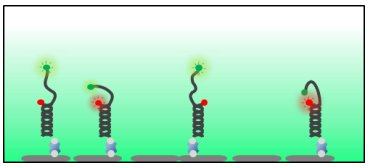
In an objective-type TIR setup, the excitation beam is focused onto the back focal plane of an oil-immersion objective with high numerical aperture (NA $\geq$ 1.4). The sample is easily



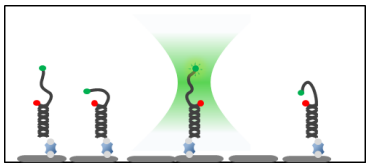
(a) Ot-TIR, pt-TIR and confocal setup



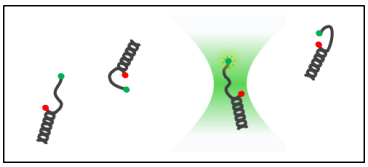
(b) SmFRET excitation and emission paths



(c) TIRF



(d) Confocal



(e) FCS

Fig. 2.2 Schematics of single-molecule confocal and TIRF microscopy. (a) Laser excitation and fluorescence emission pathways for a single-molecule FRET confocal and TIR setup. (b) In pt-TIR the laser beam is coupled into the slide through a Pellin-broca prism placed on top of the slide, whilst ot-TIR and confocal excitation occur through the objective. (c) Surface-immobilised doubly-labelled nucleic acid in pt-TIR. An evanescent field is generated in the sample chamber that illuminates a small volume reducing the excitation of the solvent. The intensity of the evanescent wave decays exponentially with the distance and has a penetration depth of ~150 nm. (d-e) In confocal and FCS the observation of the sample takes place at the focal spot of the laser for either surface-immobilised (d) or freely-diffusing molecules (e).

accessible and can be used to implement other techniques [McCluskey et al., 2014] but, since the excitation and the fluorescence detection are performed through the objective, this method involves a high background signal resulting in a low SNR. In prism-type TIR, the laser beam is coupled into the sample chamber through a Pellin-Broca prism placed on top of the quartz slide over a thin layer of immersion oil. The fluorescence signal is collected through a water-immersion objective with 1.2–1.3 NA. The evanescent wave is easily achievable and, although the signal is lower and the sample is less accessible than for a ot-TIR, prism-type TIR provides a higher SNR and a more reliable real-time injection setup [Lee et al., 2016].

In both ot- and pt-TIR the fluorescence signal is collected into an electron-multiplying charge-coupled device (EMCCD) [McCluskey et al., 2014] cooled to -80 °C for observation of single dyes. Moreover, in smFRET the signals from donor and acceptor are split using dichroic mirrors and directed to two different halves of the EMCCD chip for their simultaneous observation.

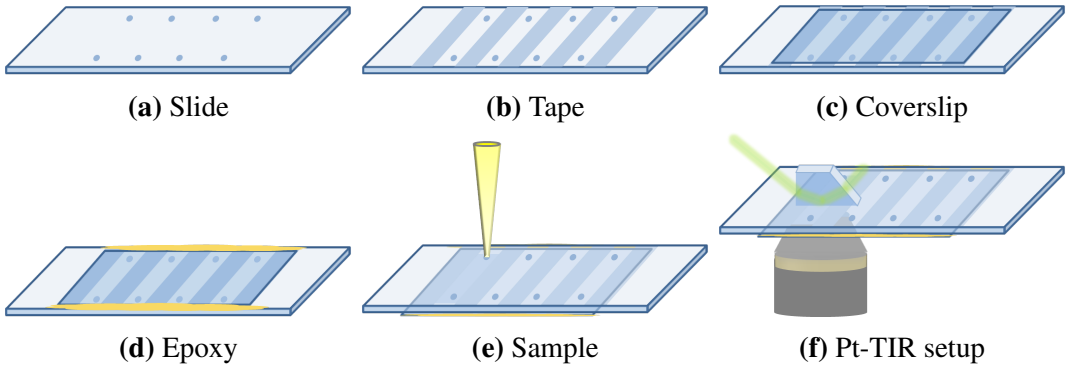


Fig. 2.3 Slide preparation for single-molecule pt-TIR. (a-b) Double-sided tape is attached to a clean pre-drilled slide. (c-d) A glass coverslip is placed on top and the sample chambers are sealed with epoxy. (e-f) Neutravidin and sample can be then flushed through one of the holes of each channel.

Our work has been carried out in a pt-TIRF using quartz slides with 0.75 mm injection holes pre-drilled using diamond drill bits (Fig. 2.3a). These slides were previously cleaned following the cleaning protocol detailed below

1. Microwave slides in water for 1 minute and scrap off coverslips, epoxy and tape. Wipe slides with acetone and methanol.

2. Etching in a Coplin jar with 50:50 H_2O_2:HCl for 2x30 min.

3. Sonicate slides for 15 minutes each time in i) 2% Alconox, ii) acetone, iii) 3 M KOH, iv) methanol and v) 3 M KOH. Rinse and sonicate in ultrapure water for 5 minutes after each of these sonication steps.

4. Slides can be refreshed before use by repeating the last sonication step with 3 M KOH.

Then, the channels were made by attaching double-sided tape across the slide (Fig. 2.3b) and placing a 0.13–0.16 mm or 0.16–0.19 mm thick glass coverslip on top (Fig.2.3c). The edges of the channels were sealed with epoxy resin. When dry, biotin, neutravidin and sample can be flushed into one of the channels.

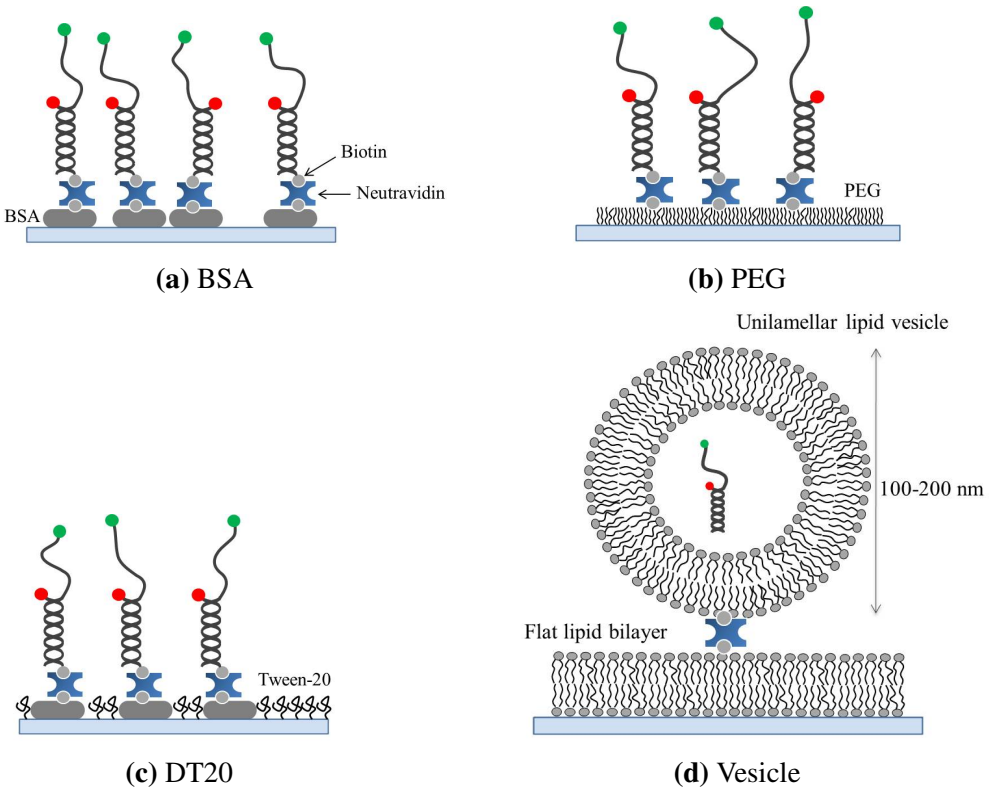
These experiments were performed in a custom-built single-molecule setup including an Olympus IX71 microscope equipped with a 1.2 NA water-immersion objective (UPLAPO60XW, Olympus, Melville, USA). The excitation source is a 50 mW continuous-wave green laser emitting at 532 nm and a 50 mW continuous-wave red laser emitting at 642 nm (Crystalaser, Reno, USA). Prior mapping with carboxylated fluorescent polystyrene beads (Crimson Fluorescent Fluospheres, 0.2 µm) is required for correct correlation of donor and acceptor signals. Fluorescence signals were split in a long-pass dichroic mirror (DCRLP, 645 nm) (Chroma Technology Corp, Bellows Falls, USA) and each is acquired in a 512x512 back-illuminated iXon electron-multiplying CCD camera (Andor, Belfast, UK).

2.5.1 Surface-immobilisation methods

Surface immobilisation of the sample is required for the real-time observation of single molecules in prolonged studies ranging from seconds to minutes, longer than those limited by the diffusion of the sample. The linkage to the surface is performed via biotin-avidin interaction, one of the strongest non-covalent interactions known. In our studies neutravidin is preferred over streptavidin because of its lower non-specific interactions with the surface.

Several immobilisation techniques have been reported to date, providing not only an attachment point to the surface but also a passivation method to avoid non-specific interactions

with the slide and coverslip. We have used four different passivation methods depending on the system used and as a control to evaluate possible artifacts (Fig. 2.4).



(a) BSA **(b)** PEG **(c)** DT20 **(d)** Vesicle

Fig. 2.4 Surface-immobilisation methods for single-molecule fluorescence techniques. (a) Biotin-labelled BSA acts as an anchor for the immobilisation of biotinylated DNA or RNA via biotin-neutravidin interaction. (b) Experiments involving proteins require a passivation of the surface with aminosilane and PEG to prevent non-specific interactions with the surface. (c) Passivation of slide and coverslip with DDS, addition of biotynilated BSA and posterior treatment with Tween 20 provide a good alternative to PEG showing less non-specific binding for studies with proteins. (d) Encapsulation of the sample in SUV provides an immobilisation method free from surface artifacts.

2.5.1.1 BSA immobilisation

This method involves the coating of the surface with biotinylated bovine serum albumin (BSA), which adheres to the surface via non-specific interactions providing the attachment point for biotin-labelled nucleic acids via neutravidin. A 1 mg/mL solution of biotinylated BSA (Sigma Aldrich Co., Gillingham, UK) is incubated for 10 minutes with posterior addition and incubation of 0.2 mg/mL neutravidin (Fisher Scientific, Loughborough, UK).

Diluted sample is injected with concentrations in the 50–250 pM range. BSA surface coating is preferentially used in nucleic acid assays for its simplicity. However, BSA surface coating may present some patches that obligates to adopt a different passivation method for experiments involving proteins to prevent non-specific interactions with the glass coverslip and quartz slide. Two different methods have been reported for their efficient passivation of the surface: PEG and DT20.

2.5.1.2 PEG immobilisation

The most widely used method to prevent the non-specific binding of sticky proteins to the surface is the coating with polyethylene glycol (PEG). Slides and coverslips are first passivated with an amino silane that will react with the succinimidyl valerate (SVA) reactive group of a mix containing 1–2% biotinylated and non-biotinylated PEG (MW 5,000) [Chandradoss et al., 2014; Roy et al., 2008]. PEG slides were prepared as follows

1. Clean slides and coverslips are dehydrated to remove the adsorbed water that precludes the silicate groups to react with the amino silane for passivation of the surface.

   (a) Dry 1 coverslip using nitrogen air gun, burn it very quickly (i.e. milliseconds) through blue flame and cool down using nitrogen.

   (b) Dry 1 slide using nitrogen, burn both sides of the slide in blue flame for 2 mins each side, cool slide down with nitrogen and keep it in a falcon tube to cool down further. A prolonged burning may result in the inactivation of the surface.

2. Slides and coverslips are passivated with amino silane via reaction with the silicate groups of the surface.

   (a) Rinse polypropilene jars with methanol, dispose, refill with methanol and insert slides and coverslips

   (b) Using a glass pipette (Pasteur pipette), add 2.5 mL of acetic acid to each jar containing slides/coverslips and methanol.

(c) Using a glass pipette, add 200 µL of 3-(2-Aminoethyl)3-Aminopropyltrimethoxy silane (Fluorochem Ltd., Hadfield, UK) to each jar.

(d) Put the jars in sonication bath for no longer than 10 minutes.

(e) Empty jars and refill with methanol.

(f) Remove 1 slide/coverslip and rinse both sides with: i) methanol, ii) ultrapure water and iii) methanol. Dry with nitrogen air gun and keep it in a box containing water and close lid.

3. A functionalised SVA-PEG will react with the amino silane to coat the slide and coverslip surface.

(a) Dissolve a mixture containing 1–2 mg biotinylated PEG and 100 mg non-biotinylated SVA-PEG (MW 5,000) (Laysan Bio, Inc., Arab, USA) in 320 µL 100 mM sodium bicarbonate (pH 8.5). This buffer can be prepared by dissolving 84 mg $NaHCO_3$ in 10 mL ultrapure water.

(b) Add the above PEG solution to each slide between the drilled holes and place a coverslip on top.

(c) Leave PEG slides and coverslips in humidity boxes, wrapped in tin foil to avoid light exposure, for 1:30–2:00 h.

(d) Rinse both sides of slides and coverslips with: i) methanol, ii) ultrapure water and iii) methanol. Dry with nitrogen.

4. Mount slide remembering which sides of slide and coverslip was pegylated.

2.5.1.3 DT20 immobilisation

A new passivation protocol based on a coating of the surface with dichlorodimethylsilane (DDS) and subsequent treatment with Tween-20 has been reported by the Ha lab [Hua et al., 2014]. The DDS-Tween-20 (DT20) method provides a shorter and less expensive alternative to PEG and shows less non-specific binding of proteins. Slides and coverslips are firstly coated with DDS, to which biotinylated BSA adheres acting as an anchor for

specific tethering of the nucleic acid via biotin-neutravidin interaction. Lastly, the addition and self-assembly of Tween-20 on the DDS coating provides the passivation required for single-molecule observation of protein-DNA and protein-RNA interactions.

1. Slides and coverslips are coated with DDS.
   (a) Clean slides and coverslips following the cleaning protocol.
   (b) Rinse slides and coverslips with hexane.
   (c) Fill glass jars with hexane and add 50 µL dichlorodimethylsilane (DDS, Sigma, ≥99.5%) using a 1 mL syringe with needle. Place the needle under hexane to avoid air contact since DDS is extremely reactive to moisture.
   (d) Immediately seal glass jars and shake gently at room temperature for 1:30 h. Dispose hexane into a designated hexane-DDS waste.
   (e) Rinse and sonicate the slides and coverslips in hexane for 1 min. Repeat 3 times.
   (f) Dry slides and coverslips are stored under vacuum at -20 °C until the day of use.
2. The treatment with Tween-20 fills the patches between the adsorbed BSA molecules to provide a fully coated surface.
   (a) Incubate the slide with 0.2 mg/mL biotinylated BSA in T50 buffer (20 mM Tris-HCl 50 mM NaCl, pH 8.0) buffer for 10 min.
   (b) Addition of freshly made 0.2% Tween-20 solution in Tris buffer to each channel and incubate for 10 min.
   (c) Addition of 0.2 mg/mL Neutravidin in T50 buffer and incubate for 10 min.
   (d) Addition of 50–250 pM dye-labelled sample containing biotin and incubate for 5 min.

2.5.1.4 Vesicle encapsulation

Vesicle encapsulation in small unilamellar lipid vesicles (SUV) offers a good alternative when immobilisation can alter the structure or function of the system under study. The

encapsulation of the sample is performed following the extrusion method by passing the lipid components and the sample to be trapped inside the vesicle through a polycarbonate membrane filter to obtain homogenous vesicles of the desired diameter (i.e. from 50 to 500 nm). The concentrations of DNA or RNA are in the picomolar range to favour the encapsulation of one molecule per vesicle.

Incorporation of lipids with low transition point, such as DMPC, allow the formation of transient pores at room temperature for experiments that require real-time exchange [Roy et al., 2008]. Both porous and non-porous lipid vesicles carrying the sample and containing $< 0.5\%$ biotin-labelled lipids can be attached via neutravidin to a lipid bilayer or PEG passivated slide to prevent the vesicles to burst in contact with the quartz surface.

1. Allow lipids to reach room temperature. Rinse 3 glass vials with methanol and dry them using nitrogen. Label the vials as rinse (A), sample (B) and bilayer (C). Fill rinse (A) with few hundred µL of chloroform.

2. To avoid contamination of the sample, coat a pipette tip with 250 µL chloroform from rinse (A) and immediately dispose the solvent back to the vial.

3. Non-porous vesicles:

   (a) Transfer 250 µL L-α-Phosphatidylcholine (Egg, Chicken) (Egg-PC) (Avanti Polar Lipids Inc., Alabaster, USA) from stock vial to the second and third vial. Seal lipid stock and store at -20 °C.

   (b) Transfer 166 µL cholesterol solution (20 mg cholesterol in 1666 µL chloroform) to vials B and C. Omit this step if Egg-PC is not being used.

   (c) Add 1.76 µL 1,2-Dipalmitoyl-sn-Glycero-3-Phosphoethanolamine-N-(Cap Biotinyl) lipid (Avanti Polar Lipids Inc., Alabaster, USA) to vials B and C.

4. Porous vesicles:

   (a) Transfer 250 µL 1,2-Dimyristoyl-sn-glycero-3-phosphocholine (DMPC) (Avanti Polar Lipids Inc., Alabaster, USA).

(b) Add 1.76 µL 1,2-Dipalmitoyl-sn-Glycero-3-Phosphoethanolamine-N-(Cap Biotinyl) lipid to vials B and C.

5. Dry the lipid solution by blowing nitrogen over the opening of the vial avoiding splashes. Once a residue is visible at the bottom, place the vial in a dissecator and leave under vacuum for 3–4 hours.

6. Resuspend the lipids in the desired buffer. In order to obtain small unilamellar vesicles (SUV), freeze and thaw 5 to 10 times using liquid nitrogen until the solution turns from cloudy to clear.

7. Load the lipid solution in a glass syringe and pass it through an extruder containing a filter with a pore size corresponding to the desired diameter of the SUV back and forth an odd number of times.

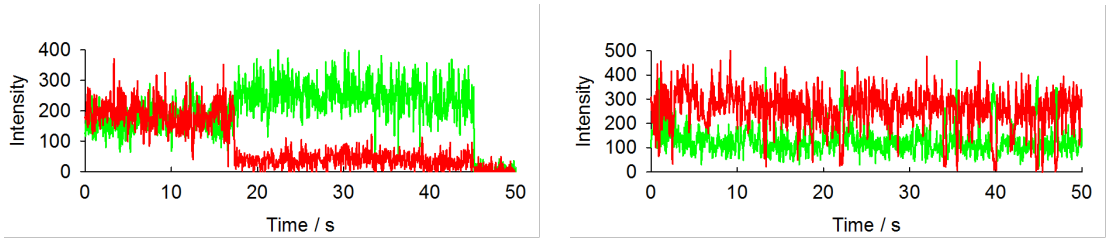
8. Purify the sample using a PD-10 desalting column with Sephadex G-25 resin (GE Healthcare Life Sciences, Little Chalfont, UK). Store sample at 4 °C for no longer than one week.

2.5.2 Oxygen scavenger systems

An excited triplet state has a longer lifetime than excited singlet states and thus, fluorophores that undergo intersystem crossing have a longer time window to react with molecules in the local environment and undergo chemical modifications that may result in a structure that is no longer emissive. The process in which the fluorophore alters permanently its structure and looses its ability to emit fluorescence is called photobleaching (Fig. 2.5a). The number of photons emitted by a fluorophore before photobleaching can range from a few to several millions and is strongly dependent on its structure and environment. For example, molecular oxygen can act as a quencher of the dark triplet state, leading to the formation of the excited singlet oxygen, which is extremely reactive with organic dyes [Renn et al., 2006]. Single-molecule TIR involves a continuous excitation of the dyes and requires the use of oxygen scavenger systems, such as glucose oxidase-catalase and PCA/PCD [Aitken et al.,

2008], to remove the oxygen present in solution and avoid irreversible damage of the dyes. To further improve the stability of the dyes, buffers may be previously degassed. Interestingly, it is possible to identify the presence of either one or multiple dyes by observation of their photobleaching events, which can be of great interest for some studies in single-molecule imaging.

Furthermore, oxygen scavengers are used in an imaging buffer together with triplet-state quenchers such as β-mercaptoethanol and the more efficient 6-hydroxy-2,5,7,8-tetramethyl-chroman-2-carboxylic acid (Trolox), a water-soluble vitamin E analogue. The triplet-state quenchers avoid the dye to undergo intersystem crossing towards the dark triplet state that would result in the observation of photoblinking events. Photoblinking, unlike photobleaching, is a reversible process and can last from milliseconds to several seconds (Fig. 2.5b).



(a) Photobleaching **(b)** Photoblinking

Fig. 2.5 Observation of photoblinking and photobleaching events. The single-molecule FRET traces show the fluorescence signal of donor (green) and acceptor (red) dyes as a function of time. (a) SmFRET trace showing photobleaching of the acceptor (~17 s) and donor (~45 s). (b) SmFRET trace showing several photoblinking events for the acceptor.

To prepare a 50 mM stock of Trolox, 125 mg of powder (Sigma Aldrich Co., Gillingham, UK) is dissolved in 1 mL of methanol, to which 9 mL of ultrapure water is added. The pH is then adjusted to pH 9.5 using a NaOH solution. The imaging buffer contains Trolox in a final concentration of 1 mM.

2.5.2.1 Glucose oxidase and catalase

The glucose oxidase and catalase (GOC) system constitutes one of the most used oxygen scavengers in single-molecule techniques [Roy et al., 2008]. A stock of GOC oxygen scavenger can be prepared by addition of 100 µL catalase from bovine liver (Sigma Aldrich

Co., Gillingham, UK) to 10 mg glucose oxidase from *Aspergillus niger* (Sigma Aldrich Co., Gillingham, UK). This solution is then mixed and centrifuged at maximum speed for 1 min. The yellow supernatant constitutes the GOC system.

The GOC oxygen scavenger is used along with Tris-HCl buffer containing 6% (w/v) D-(+)-glucose that will be oxidised into gluconic acid as

$$\text{D-(+)-glucose} + O_2 + H_2O \xrightarrow{\text{glucose oxidase}} \text{gluconic acid} + H_2O_2$$

and the oxygen peroxided produced is decomposed by the catalase

$$H_2O_2 \xrightarrow{\text{catalase}} H_2O + \frac{1}{2}O_2$$

The production of gluconic acid implies a progressive increase in the acidity of the medium that represents the most important disadvantage of the GOC oxygen scavenger. The new enzymatic PCA/PCD system provides an alternative with improved dye stability and negligible pH change under experimental conditions.

2.5.2.2 PCA/PCD

The PCA/PCD oxygen scavenger system was firstly reported by Aitken and coworkers [Aitken et al., 2008]. This system constitutes an enzymatic scavenger in which the substrate 3,4-dihydroxybenzoic acid (PCA) is oxidised by the protocatechuate dioxygenase (PCD) with no pH drop and a five-fold higher decrease in the concentration of O_2 in solution when compared with GOC.

A 50 mM stock of PCA can be prepared by resuspending 77 mg PCA (Fisher Scientific, Loughborough, UK) in 8 mL of ultrapure water and adjust the pH to 9 with NaOH. This solution is topped up with water to 10 mL and 50 µL aliquotes are kept at -20 °C. To make a 25 µM stock of PCD, the content of enzyme from *Pseudomonas sp.* (Sigma Aldrich Co., Gillingham, UK) is resuspended in 450 µL of buffer containing 50% glycerol, 50 mM Tris-HCl (pH 8.0). The enzyme is aliquoted, frozen with liquid nitrogen and stored at -20 °C.

Final concentrations of substrate and enzyme in the imaging buffer are 2.5 mM PCA and 250 nM PCD.

2.5.3 Single-molecule FRET data analysis

2.5.3.1 Collection and analysis of raw data

The mapping and processing of the raw movies are performed using custom-built routines in IDL 6.0 (Exelis, Boulder, USA). The analysis of the data starts with the averaging of the first 10 frames of the movie, which are saved as an image that includes both donor and acceptor channels. The program then discards any spot larger than the diffraction-limited size and any spot with intensities below a given threshold. Then, the software pairs those spots in the donor channel with their counterpart in the acceptor channel (Fig. 2.6) and finally creates a .traces file in which it includes the donor and acceptor intensities for all the molecules detected in that movie.

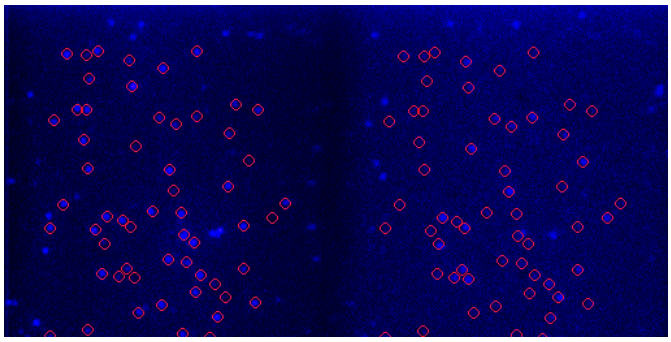


Fig. 2.6 Donor and acceptor pairing by IDL. Example showing how IDL selects those spots that meet the requirements (red circles) and pairs the spots on the donor channel (left) with their correspondent signal on the acceptor channel (right).

However, a calibration of the internal coordinates of the program is required for correct correlation of donor and acceptor signals. This is peformed by using carboxylated fluorescent polystyrene beads (Crimson Fluorescent Fluospheres, 0.2 µm) that are non-specifically bound to the surface of the slide. We manually select three donor-acceptor pairs whose coordinates will be used by the program as a reference for the mapping of the donor and acceptor channels.

Due to optical aberrations, this assignment cannot be achieved by simple translation and requires the use of a polynomial algorithm.

2.5.3.2 Manual selection of molecules and representation of population histograms

The representation of the single-molecule trajectories is done using laboratory-written routines in Matlab (Mathworks, Natick, USA). The TIR2009 tool opens the .traces files saved by IDL in order to represent the donor and acceptor intensities for each molecule as a function of time (Fig. 2.7). Among other plots available from the TIR2009 software are the trajectories

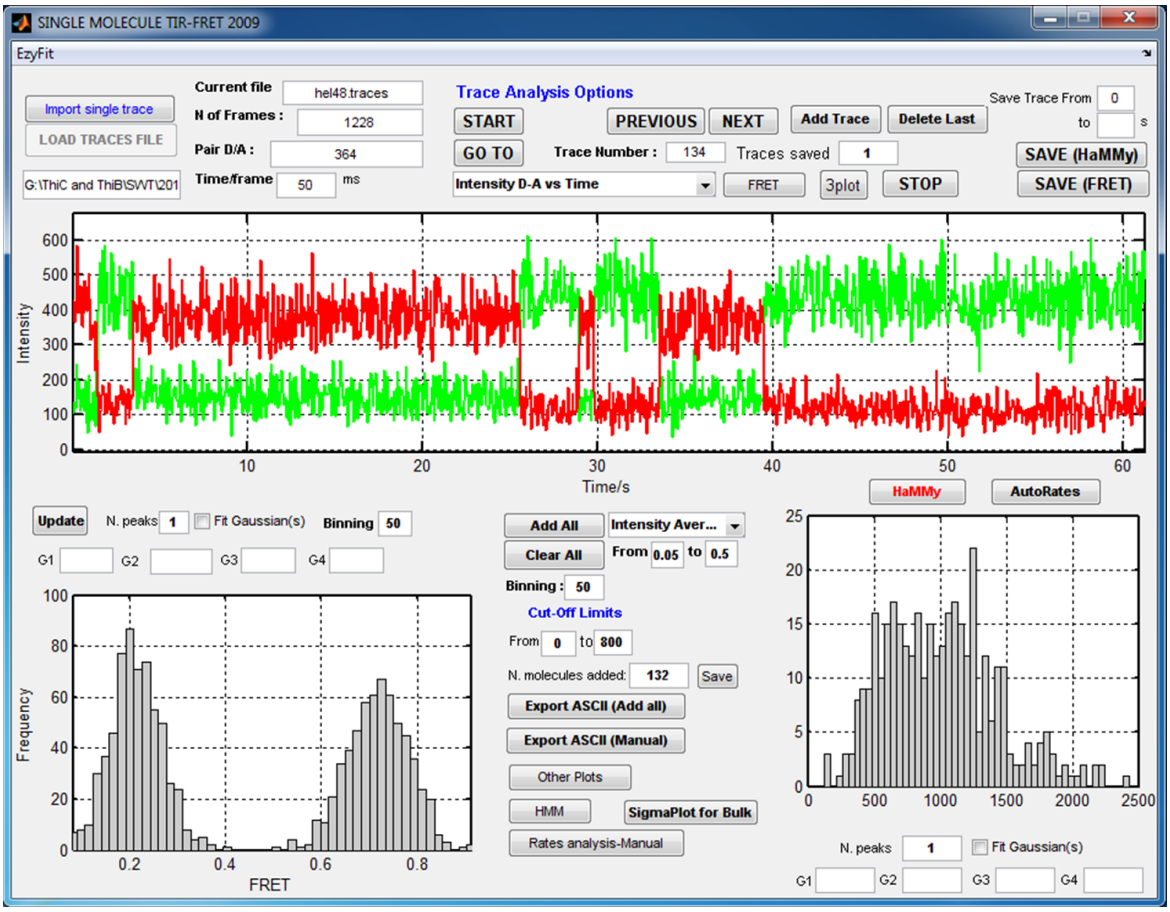


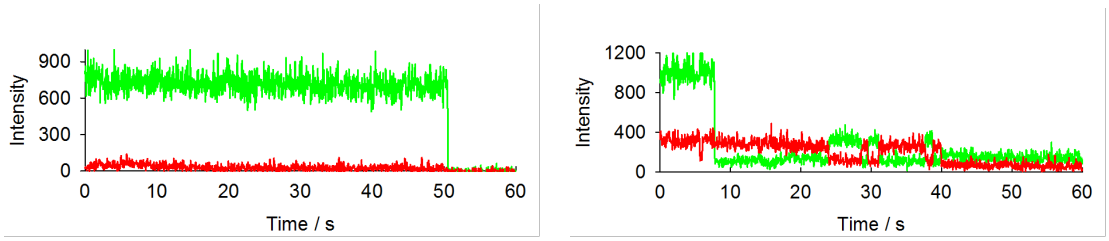
Fig. 2.7 Interface of the TIR2009 visualisation program. Donor (green) and acceptor (red) intensities are plotted on the main panel. The lower-left graph shows the apparent FRET value that results for each individual molecule by considering the entire time range. On the lower-right graph it is represented the intensity average arising from the whole set of molecules of a given movie. However, it is possible to select other plot options such as the E_{app} population histogram built from the full set of molecules.

showing the evolution of the apparent FRET efficiency, E_{app}, which can be calculated from the expression

$$E_{app} = \frac{I_A}{I_A + I_D} \tag{2.8}$$

where I_D and I_A represent the intensities of the donor and acceptor, respectively.

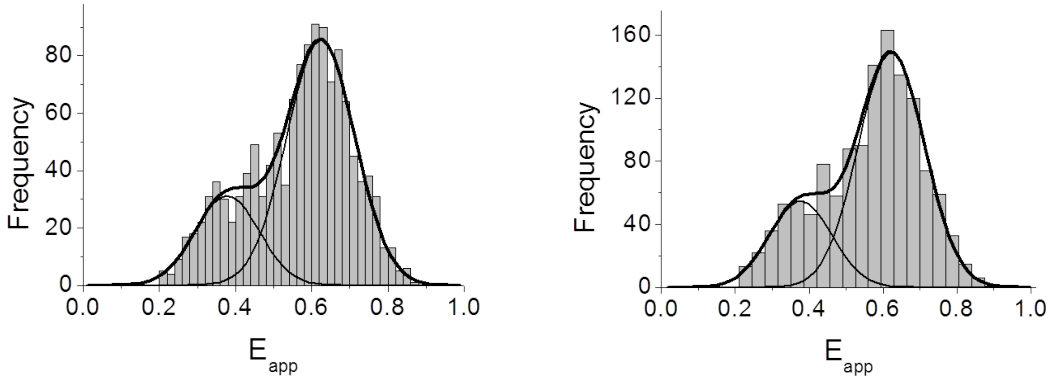
Unfortunately, the labelling efficiency of nucleic acids is never 100% efficient. For example, donor-only labelled molecules would arise as a population with E_{app} ~0.15 that results from the leakage of the donor into the acceptor channel. Then, it is required a manual analysis of the data with the aim to select those traces that clearly correspond to real molecules and discard donor-only labelled molecules, those showing early photobleaching, those without anti-correlated trajectories and signals arising from more than one biomolecule attached to the same neutravidin (Fig. 2.8).



(a) **(b)**

Fig. 2.8 Examples of discarded smFRET traces. (a) Representative trace showing the of a donor-only labelled molecule with photobleaching at ~50 s. (b) Representative smFRET trace arising from a spot containing two donors showing a photobleaching step of one of the two Cy3s at ~8 s.

The results obtained for the apparent FRET efficiencies are then represented as population histograms using Origin 8.0 (OriginLab Corp, Northampton, USA) and fit to a linear combination of Gaussian functions from which the E_{app} value associated to each FRET state is obtained from the center of the Gaussian peak (Fig. 2.9). The number of molecules selected to build one histogram depends on each experiment. We consider that a statistically significant number of molecules is reached when the relative contribution of each population remains constant and is independent of the bin size.



(a) 50 bins **(b)** 28.5 bins

	E_{app1}			E_{app2}		
	Center	Area	Width	Center	Area	Width
50 bins	0.38	25%	0.167	0.62	75%	0.179
28.5 bins	0.38	26%	0.169	0.62	74%	0.179

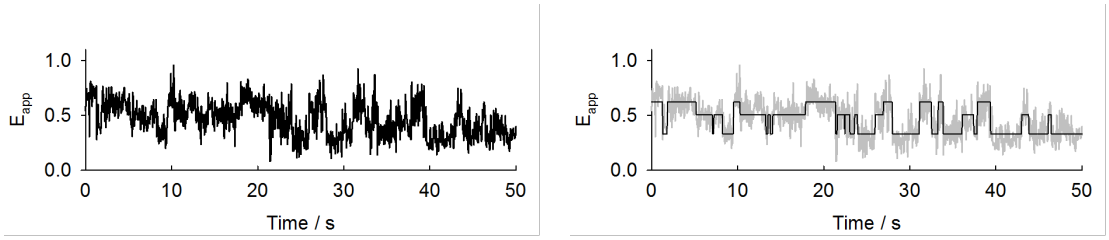
Fig. 2.9 Representative E_{app} population histograms. Population histograms represent the number of molecules observed with a given apparent FRET efficiency. The number of molecules chosen is considered to be statistically relevant when the fit values are not altered by addition of more molecules or by different bin sizes.

2.5.3.3 Analysis of single-molecule FRET trajectories using hidden Markov modelling

In smFRET assays it is important to determine the number of states present in a given system, which sometimes is difficult to achieve by fitting of the population histograms. Moreover, it is also important to define well the limits of each state for the correct determination of the dwell times between transitions. The study of intricate systems can lead to uncertainty and, in such case, a manual analysis may result in a biased interpretation of the data by the researcher.

The use of an automated analysis based on hidden Markov modelling (HaMMy) was introduced to facilitate the correct interpretation of smFRET trajectories. Hidden Markov modelling is a probabilistic method that considers a single-molecule trajectory as a Markov process with an unknown (and thus "hidden") number of states. In a Markov model, a transi-

tion between the current and any other state can occur at each time step with a characteristic probability, which is independent of previous transitions. The HaMMy analysis software was developed and implemented in C++ by McKinney and coworkers at the university of Illinois Urbana-champaign [McKinney et al., 2006]. Since then, the great potential of HaMMy has been widely demonstrated in the study of intricate systems with several states, two states that are too close and in the analysis of large sets of data (Fig. 2.10). Moreover, HaMMy has been also implemented in the analysis of rapid transitions, which can be otherwise overestimated in a manual analysis.



(a) **(b)**

Fig. 2.10 Example of hidden Markov modelling The FRET trajectory of a given molecule in which the transitions between states are not well defined (a) can be analysed using HaMMy to obtain the idealised trajectory (b), which is overlaid (black) on the original (gray).

References

Aitken, C. E., Marshall, R. A., and Puglisi, J. D. (2008). An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical Journal*, 94(5):1826–1835.

Axelrod, D., Burghardt, T. P., and Thompson, N. L. (1984). Total internal reflection fluorescence. *Annual Review of Biophysics and Bioengineering*, 13(1):247–268.

Chandradoss, S. D., Haagsma, A. C., Lee, Y. K., Hwang, J.-H., Nam, J.-M., and Joo, C. (2014). Surface passivation for single-molecule protein studies. *Journal of Visualized Experiments*, (86):e50549.

Greenfield, N. J. (2007). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, 1(6):2876–2890.

References

Hua, B., Han, K. Y., Zhou, R., Kim, H., Shi, X., Abeysirigunawardena, S. C., Jain, A., Singh, D., Aggarwal, V., Woodson, S. A., and Ha, T. (2014). An improved surface passivation method for single-molecule studies. *Nature Methods*, 11(12):1233–1236.

Karsisiotis, A. I., Hessari, N. M., Novellino, E., Spada, G. P., Randazzo, A., and Webba da Silva, M. (2011). Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angewandte Chemie International Edition*, 50(45):10645–10648.

Lackowicz, J. R. (2006). *Principle of Fluorescence Spectroscopy*. New York, NY, third edition edition.

Lee, S., Jang, Y., Lee, S.-J., and Hohng, S. (2016). Single-molecule multicolor FRET assay for studying structural dynamics of biomolecules. In Spies, M. and Chemla, Y. R., editors, *Single-Molecule Enzymology: Fluorescence-Based and High-Throughput Methods*, volume 581 of *Methods in Enzymology*, pages 461 – 486. Academic Press.

McCluskey, K., Shaw, E., Lafontaine, D. A., and Penedo, J. C. (2014). *Single-Molecule Fluorescence of Nucleic Acids*, pages 759–791. Humana Press, Totowa, NJ.

McKinney, S. A., Joo, C., and Ha, T. (2006). Analysis of single-molecule FRET trajectories using hidden markov modeling. *Biophysical Journal*, 91(5):1941 – 1951.

Renn, A., Seelig, J., and Sandoghdar, V. (2006). Oxygen-dependent photochemistry of fluorescent dyes studied at the single molecule level. *Molecular Physics*, 104(3):409–414.

Roy, R., Hohng, S., and Ha, T. (2008). A practical guide to single-molecule FRET. *Nature Methods*, 5(6):507–516.

Valeur, B. (2001). *Molecular Fluorescence*. Wiley-VCH Verlag GmbH.

Woody, R. W. (1995). [4] Circular dichroism. In *Biochemical Spectroscopy*, volume 246 of *Methods in Enzymology*, pages 34 – 71. Academic Press.

3

Novel fluorescence probes for sensing nucleic acids

3.1 Introduction

Guanine-rich DNA and RNA sequences can adopt highly polymorphic four-stranded structures known as guanine quadruplexes (G4s), which play an important role in different biological processes such as replication, transcription, gene regulation and telomere maintenance [Maizels and Gray, 2013; Vummidi et al., 2013]. The primary unit of G4s is the G-quartet, a square planar assembly of four guanines forming Hoogsteen hydrogen bonds (Fig. 3.1). G-quartets form stacks that are mainly stabilised by the presence of monovalent and divalent cations interacting with the negative potential of the central channel due to the guanine O6 atoms, and thus the ionic radius of these cations is an important factor for G4 assembly and stability. Cations such as K^+ and $NH_4{}^+$ are too large to fit in the center of the quartet and adopt a position between two G-quartets, whilst smaller cations such as Li^+ and Na^+ are coordinated in the plane of the G-quartet (Fig. 3.2) [Bhattacharyya et al., 2016].

Pinnavaia and coworkers [Pinnavaia et al., 1975, 1978] have studied the self-assembly of the 5'-guanosine monophosphate (GMP) into G-quartets in the presence of monovalent cations by NMR, reporting the sequence K^+>Na^+,Rb^+≫Li^+,Cs^+, which has been widely accepted over the years, although it may slightly vary depending on the DNA or RNA sequence. Further work on the telomeric sequence of *Saccharomyces* [Venczel and Sen,

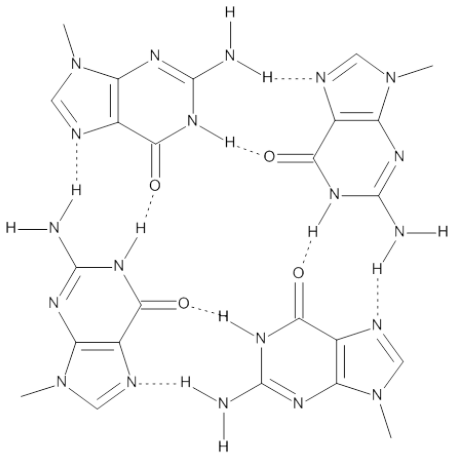


Fig. 3.1 G-quartet structure. Chemical structure of a G-quartet formed via Hoogsteen hydrogen bonding between four guanines.

1993] was not only focused on the influence of monovalent but also divalent cations in the stability of G4s. Gel electrophoresis assays showed an analogue trend in both cases with K^+>Rb^+>Na^+>Li^+,Cs^+ and Sr^{2+}>Ba^{2+}>Ca^{2+}>Mg^{2+}. Despite the large number of metal ions studied, including others such as $NH_4{}^+$ and Tl^+, the most relevant at physiological conditions, and thus the most studied, are Na^+ and K^+.

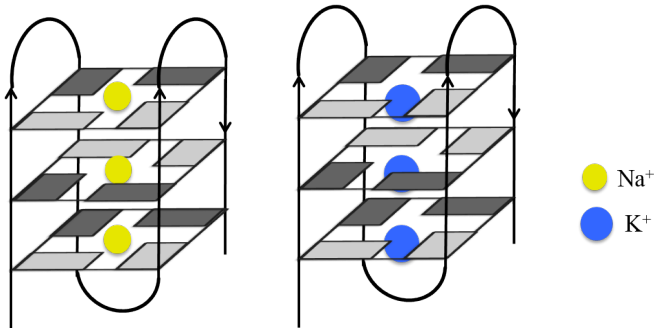
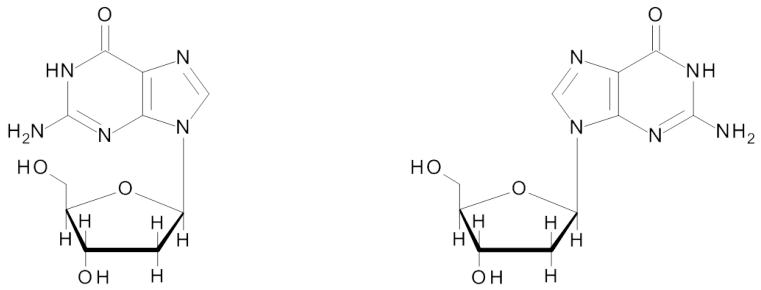


Fig. 3.2 Position of Na^+ and K^+ within G4s. Schematics of the position of Na^+ on the plane of the G-quartet and K^+ between two G-quartets for two anti-parallel G-quadruplexes.

The formation of G-quadruplexes can be predicted by the presence of the motif sequence $G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$, where N represents the loop sequence and can be any base including guanine. The human genome comprises over 350,000 motifs with loops between 1–7 bases and over 700,000 allowing loops up to 12 nucleotides long [Maizels and Gray,

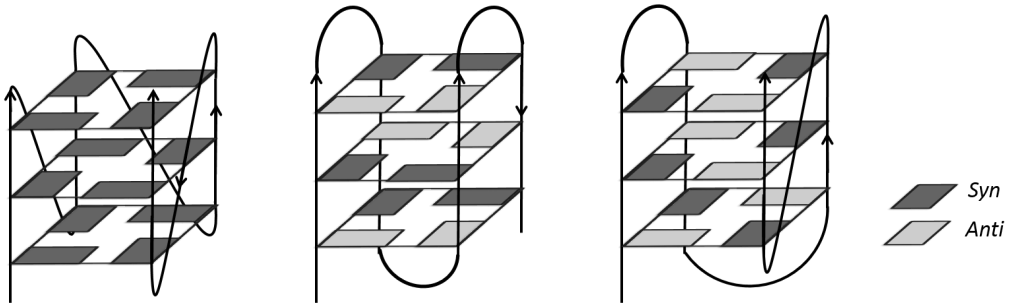
2013]. G4s have been studied for a wide range of different DNA and RNA sequences *in vitro*, specially the formation of G4s in several promoters of oncogenes, which has attracted considerable efforts since the 1990s to characterise and study these structures as potential antitumour targets. The presence of G4s has been reported, among others, in the promoter of the oncogene encoding the B-cell lymphoma 2 (Bcl-2) protein, which functions as an apoptosis inhibitor and is overexpressed in many human cancers [Dai et al., 2006]; and in the *c-kit* [Rankin et al., 2005], *c-myc* [Siddiqui-Jain et al., 2002], and RAS family proto-oncogenes (*HRAS*, *KRAS* and *NRAS*) [Cogoi and Xodo, 2006; Membrino et al., 2010], which are involved in the regulation of cellular growth. Guanine quadruples have also been found in other genes linked to the development of cancer such as the hypoxia-inducible factor 1 (*HIF-1*), which regulates the transcription of over 60 genes involved in cellular homeostasis [De Armond et al., 2005; Semenza, 2003]; in the promoter of the human telomerase reverse transcriptase (*TERT*), which is involved in the regulation of cellular telomerase activity [Palumbo et al., 2009]; and in the promoters of platelet-derived growth factor (*PDGF*) and vascular endothelial growth factor (*VEGF*), which are involved in the formation of new blood vessels [Qin et al., 2007; Sun et al., 2005]. However, the formation of G4s is not restricted to oncogenes and can be observed in other regions within the chromosome such as the non-coding region of the gene *C9orf72*, which is involved in the development of amyotrophic lateral sclerosis and frontotemporal dementia, and the human telomeric sequence, a single-stranded region found at the end of human telomeres that consists of up to 14 kilobases of $(TTAGGG)_n$ repeats [de Lange et al., 1990].



(a) *Syn* **(b)** *Anti*

Fig. 3.3 Schematics of the *syn* and *anti* orientations of the guanosine glycosidic bond.

The wide range of G4 topologies is defined by the relative orientation of the the glycosidic bond of the stacked guanines between the nucleobase and its sugar, whether *syn* or *anti* (Fig. 3.3) [Karsisiotis et al., 2011]. The succession of guanines with the same glycosidic conformation adopt a parallel G4 structure, whereas an opposite orientation of consecutive guanines corresponds to an antiparallel conformation and G4s containing sequences of the same and opposite orientation form a hybrid structure (Fig. 3.4).



(a) Parallel (4) **(b)** Antiparallel (2+2) **(c)** Hybrid (3+1)

Fig. 3.4 Representation of parallel, antiparallel and hybrid G-quadruplexes. (a) The side chains of a parallel G4 structure follow all the same direction. The glycosidic bonds of all guanines are oriented in the same direction, either *syn* or *anti*. (b) Two of the side chains of an anti-parallel follow one direction, whereas the other two chains follow the opposite direction. Guanines display a succession of *syn-anti* orientations (c) Hybrid G4 conformation with a mixed parallel and antiparallel orientation.

Interestingly, the stacking of G-quartets is not restricted to the formation of discrete structures and can also adopt long polymeric chains, called G-wires, resistant to moderate denaturation (Fig. 3.5) [Marsh and Henderson, 1994]. The formation of these higher order structures with parallel orientation is favoured by the presence of Mg^{2+} as observed for the $C_4T_4G_4T_{1-4}G_4$ [Dai et al., 1995], the c-Myc $(AGGGTGGGG)_2$ [Shi et al., 2013] and the human telomeric $(TTAGGG)_4$ [Sharma and Sheardy, 2014] sequences. Moreover the formation of G-wires has also been studied by CD for the human telomeric sequence in the presence of PEG, acting as a molecular crowding agent, and acetonitrile (AcCN), used as a small neutral molecule to cause osmotic stress [Sharma and Sheardy, 2014]. The results show a progressive transition towards a parallel structure due to the elongation of G-wires and suggest that the formation of these macroassemblies at cellular conditions is likely to happen.

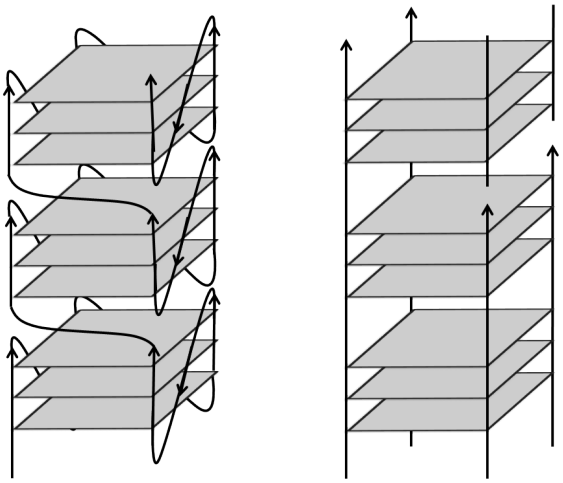


Fig. 3.5 Schematics of different parallel G-wire structures. G-rich sequences can spontaneously assemble into long polymers known as G-wires, superstructures formed by stacking of G-quartets with parallel orientation.

The formation of G-quadruplexes *in vivo* has been controversial for a long time, although clear evidence of their presence has been reported in the last years [Biffi et al., 2013; Duquette et al., 2004; Paeschke et al., 2005; Schaffitzel et al., 2001]. Recent findings have shown the site-specific binding and activity of the telomerase, a telomere specific complex that carries an RNA template to elongate the telomere via reverse transcription, at G4-forming telomeres *in vivo* [Moye et al., 2015]. The telomerase reverse transcriptase (TERT) subunit carries out this function to avoid shortening of the telomere length, which finally leads to cellular death after a number of cellular divisions (senescence). The uncontrolled telomerase activity, which has been found in $\geq$85% of cancers, precludes senescence by continuous maintenance of the telomere length [Moye et al., 2015; Zhou et al., 2011]. For this reason, researchers have focused their efforts to develop G4 binders stabilising these structures with the aim to inhibit the telomerase activity and induce cell death. However, these efforts are not restricted to human telomeres and have been extended to other G4-forming sequences within the genome. The use of G4 as molecular targets to induce apoptosis in tumour cells *in vivo* is exemplified in the work performed by McLuckie and coworkers, in which G4 stabilisers have shown to cause DNA double-strand breaks that lead to cellular death [McLuckie et al., 2013].

The diversity and heterogeneity of G-quadruplexes emphasises the biological importance of these architectures and highlights the need to develop not only G4 selective binders but also fluorescent probes that specifically sense these structures in order to provide valuable information about their folding, conformation and function *in vitro* and *in vivo* [Largy et al., 2013; Maizels, 2006].

3.1.1 Fluorescence sensing of guanine quadruplexes (G4s): current strategies

The wide range of G4 fluorescent probes can be classified in light-up and light-off probes depending on whether they experience an increase or decrease in the fluorescence signal [Largy et al., 2013]. These probes display an affinity for G-quadruplexes with K_d values ranging from 0.2 nM to 100 µM [Vummidi et al., 2013]. A third group would include the permanent probes, which correspond to those designed as covalently-linked fluorescent tags.

Other strategies are based on the double labelling of G4-forming sequences at specific positions to monitor structural changes via FRET at the ensemble or single-molecule level. These assays not only provide insights into the heterogeneity of G4 conformations and their relative populations, as observed for the promoters of proto-oncogenes *c-kit* and *c-myc* [Shirude et al., 2007, 2008] and for the human telomeric sequence [Lee et al., 2005; Ying et al., 2003], but also into their dynamics and stability under different experimental conditions [Aznauryan et al., 2016; Klejevskaja et al., 2016; Noer et al., 2016].

Among all, light-up probes have been the most studied in recent years, where Thioflavin T (ThT) is one of the most versatile G4 sensors. ThT displays an average of ~3-fold discrimination against non-G4 structures [de la Faverie et al., 2014] and experience a 2100-fold increase when bound to the human telomeric sequence [$A(GGGTTA)_3GGG$] in the presence of K^+ with a K_d in the micromolar range [Mohanty et al., 2013]. Moreover, the low fluorescence background provides a means for its use as staining agent for the effective detection of G4s in polyacrylamide gels. The selective binding to G-quadruplexes has also led to the study of a number of iridium, nickel, platinum and ruthenium complexes that

increase their fluorescence when bound to G4 with affinity in the nano- to micromolar range. An example of a selective ruthenium complex would be the $[Ru(bpy)_2(dppz)]^{2+}$, which preferentially binds to G4s over i-motif DNA structures in the presence of Na^+ or K^+ with a K_d of 0.7 µM for the human telomeric sequence [Shi et al., 2010].

However, none of the G4 probes reported can be used in ratiometric sensing of G-quadruplex architectures. This type of analytical method requires the presence of two emission bands which ratio varies as a function of G4 structure or G4 concentration. Ratiometric probes provide a built-in reference in which the two bands represent an absolute observable, free from unwanted photobleaching and background effects. Moreover, the relative variation of two bands can account for effects such as sensor concentration and environmental factors such as polarity, pH and changes in temperature [Freire et al., 2015]. In this chapter we introduce three G4 and G-wire selective probes: i) a ratiometric sensor that undergoes excited-state intramolecular proton transfer (ESIPT), the 2-(2'-hydroxyphenyl)-3H-imidazo[4,5-b]pyridine (HPIP); ii) the water soluble ruthenium complex $[Ru(bpy)_3](PF_6)_2$ as a light-up probe; and iii) a natural product with antioxidant and anti-inflammatory effects, curcumin [Menon and Sudheer, 2007], as a second light-up probe that has been extensively used as amyloid inhibitor and, as a result, as a fluorescent sensor of pre-toxic aggregates.

3.2 Materials and methods

3.2.1 Sample preparation

Oligonucleotides were purchased from IDT. Dry DNA pellets were dissolved in 50 mM Tris buffer pH 7.5 to a concentration of 100 µM and stored at -20 °C.

HPIP was synthesised and purified by the Biophysical Chemistry, Photophysics and Spectroscopy group (Universidade de Santiago de Compostela, Spain) and provided by Flor Rodríguez Prieto. A small amount (spatula tip) of HPIP was dissolved in a few milliliters of 50 mM Tris pH 7.5 and left with shaking overnight. The solution was then filtrated using a

0.22 µm pore size filter membrane. The resulting HPIP solution, in the 5–10 µM range, was stored at room temperature protected from light to avoid degradation.

The $[Ru(bpy)_3](PF_6)_2$ was provided by the group led by Zysman-Colman (School of Chemistry, University of St Andrews). This ruthenium complex was dissolved in 50 mM Tris buffer pH 7.5, filtered using a 0.22 µm pore size filter membrane and stored at room temperature protected from light.

Curcumin was purchased from Sigma (Sigma Aldrich Co., Gillingham, UK) with a content of $\geq 94\%$ curcuminoid and $\geq 80\%$ curcumin. The curcumin solutions were prepared immediately before each experiment to avoid degradation of the sample.

Table 3.1 Sequences of the DNA strands used in this chapter.

Strand	Sequence (from 5' to 3')
ssDNA	ACACTTACACTTACACTTACAC
dsDNA	(a) TTTTTTTTTTATTAAAATTTATAA
	(b) AAAAAAAAAATAATTTTAAATATT
22AG	AGGGTTAGGGTTAGGGTTAGGG
TBA	GGTTGGTGTGGTTGG
Tet7	GGGCAGCGGTGGTGTGGCGGGATCTGGGGTTGTGCGGTGT

3.2.2 Gel electrophoresis

Gel electrophoresis assays were performed in nondenaturing 12% (w/v) acrylamide gels in the absence and presence of 90 mM salt (Na^+, K^+). Samples dissolved in 50 mM Tris 100 mM NaCl or KCl were heat treated at 95 °C in heat block for 1 minute and allowed to cool down on bench. Ficoll was added to a final concentration of 3% before loading on the gels, which were run at 110 V for 5 hours in cold room (4 °C).

Gels were stained by incubation with $[Ru(bpy)_3]^{2+}$ or curcumin solutions for 1 minute before visualisation in a Typhoon FLA 7000 (GE Healthcare Life Sciences, Little Chalfont, UK) gel imaging system equipped with four lasers emitting at 473, 532, 635 and 650 nm.

Gels were finally stained with ThT solution for 1 minute to provide information about the presence of monomeric G-quadruplexes.

3.2.3 Absorption

Absorption spectra were recorded in a Cary 50 Bio spectrophotometer with background subtraction using a 1 cm width sub-micro absorption cuvette.

HPIP concentrations were calculated at the absorption maximum at 332 nm with ε of 22,387 $M^{-1}cm^{-1}$ [Krishnamoorthy and Dogra, 2000]. $[Ru(bpy)_3]^{2+}$ concentrations were calculated at its maximum in the visible region at 450 nm with ε of 14,600 $M^{-1}cm^{-1}$ [Kalyanasundaram, 1982], whilst the concentration of curcumin was determined immediately before each experiment with a ε of 23,800 $M^{-1}cm^{-1}$ at the absorption maximum of 427 nm [Majhi et al., 2010].

3.2.4 Ensemble fluorescence

Ensemble fluorescence measurements were performed in a Cary Eclipse spectrophotometer using ultra-micro fluorescence cuvettes. Excitation of HPIP, $[Ru(bpy)_3](PF_6)_2$ and curcumin was performed at their maxima at 332, 450 and 427 nm, respectively. All experiments were performed at room temperature.

3.2.4.1 Binding affinity

Titration curves were fit to a 1:1 binding model as described by Freire and coworkers [Freire et al., 2015], in which the equilibrium between the free and bound sensor can be represented as

$$HPIP_{free} + G4 \rightleftharpoons HPIP_{bound} \tag{3.1}$$

with a binding constant, K_b, given by

$$K_b = \frac{[HPIP]_{bound}}{[HPIP]_{free}[G4]} \tag{3.2}$$

For experimental conditions with excess of G4, thus [G4]∼[G4]$_0$, and assuming no association or dissociation in the excited state, the equilibrium concentrations for the free and bound dye can be calculated from

$$[HPIP]_{free} = \frac{[HPIP]_0}{1 + K_b\,[G4]_0} \qquad [HPIP]_{bound} = \frac{[HPIP]_0\,K_b\,[G4]_0}{1 + K_b\,[G4]_0} \tag{3.3}$$

Then, the fluorescence intensity at a wavelength λ is equal to the sum of the intensities of the free and bound dyes and can be deduced from the previous equations

$$I(\lambda) = \frac{I_{free}(\lambda) + I_{bound}(\lambda)\,K_b\,[G4]_0}{1 + K_b\,[G4]_0} \tag{3.4}$$

In our experiments we consider the ratio between the fluorescence intensities of the normal and tautomer form as a method to monitor the binding of HPIP to a G4 structure. The variation of this ratio (R) as a function of G4 concentration is given by the equation

$$R = \frac{I(\lambda_1)}{I(\lambda_2)} = \frac{I(N)}{I(T)} = \frac{\dfrac{I_f(\lambda_1)}{I_f(\lambda_2)} + \dfrac{I_b(\lambda_1)}{I_b(\lambda_2)}\dfrac{I_b(\lambda_2)}{I_f(\lambda_2)}\,K_b\,[G4]_0}{1 + \dfrac{I_b(\lambda_2)}{I_f(\lambda_2)}\,K_b\,[G4]_0} = \frac{R_f + R_b\,q_{bf}(\lambda_2)\,K_b\,[G4]_0}{1 + q_{bf}(\lambda_2)\,K_b\,[G4]_0} \tag{3.5}$$

where R_f and R_b are the spectral ratios of the free and bound fluorophore, and $q_{bf}(\lambda_2)$ is the brightness ratio between the bound and free dye at the wavelength λ_2. The binding assays are, however, represented as the inverse of the ratio R for better visualisation of the signal increase when bound to DNA.

3.2.5 Circular dichroism

Circular dichroism measurements were performed in a Jasco J-810 spectropolarimeter equipped with a Peltier temperature controller and water bath using either circular cuvettes or rectangular cuvettes.

Characterisation of 22AG wires and monomeric species was performed for unheated and heated DNA samples. 22AG monomers were obtained after heating of the sample to

95 °C for 1 minute in the presence of 100 mM KCl in heat block and slow cooling on bench. Concentrations of DNA were 4 µM. Experiments are represented as an average of 4 replicates with solvent subtraction and line smooth correction.

Ruthenium complex G-wire denaturation assays were performed for 4 µM DNA and 1.2 µM $[Ru(bpy)_3](PF_6)_2$ in 50 mM Na_2PO_4 buffer at pH 7.5 with a background of 100 mM KCl and 4 µM DNA. Data are represented as the magnitude of the ellipticity at 263 nm as a function of temperature. The temperature slope was set at 1 °C/min.

Curcumin G-wire denaturation assays were performed with 4 µM DNA and excess of curcumin in 50 mM MES at pH 6.5 with a background of 100 mM KCl. Data are represented as the ellipticity at 263 nm as a function of temperature, with a temperature slope of 1 °C/min.

The denaturation curves were normalised and fitted using the Boltzmann sigmoidal equation, from which the melting temperature, T_m, is obtained as the temperature at which the signal reaches an amplitude of 0.5

$$I = I_{min} + \frac{I_{max} - I_{min}}{1 + \exp\left(\frac{T_m - T}{s}\right)} \tag{3.6}$$

where I_{min} and I_{max} represent the minimum and maximum intensity values, respectively, and s represents the slope of the melting curve.

3.3 Results and discussion

3.3.1 Characterisation of G4s by polyacrylamide gel electrophoresis and circular dichroism: monomeric versus multimeric species

The presence of monomeric and multimeric G-quadruplex structures in the human telomeric sequence (22AG) was investigated under different salt concentrations with and without previous heating at 95 °C for 1 min in order to denaturate the DNA. The formation of higher order structures of the 22AG sequence under molecular crowding conditions *in vitro* has shown ambiguous results with studies reporting either its ability [Sharma and Sheardy, 2014]

or inability [Miyoshi et al., 2005, 2012] to adopt G-wire structures. Thus, we have chosen gel electrophoresis and ThT staining to elucidate the formation of long multimeric structures with a background of 90 mM KCl. The unheated sample display a number of higher order structures with different lengths, whereas the heated sample show a single band that may correspond to monomeric G4s. These results suggest that the 22AG sequence is likely to form G-wires in the presence of K^+ and that these can be disassembled at 95 °C, resulting in the formation of monomeric G4 structures.

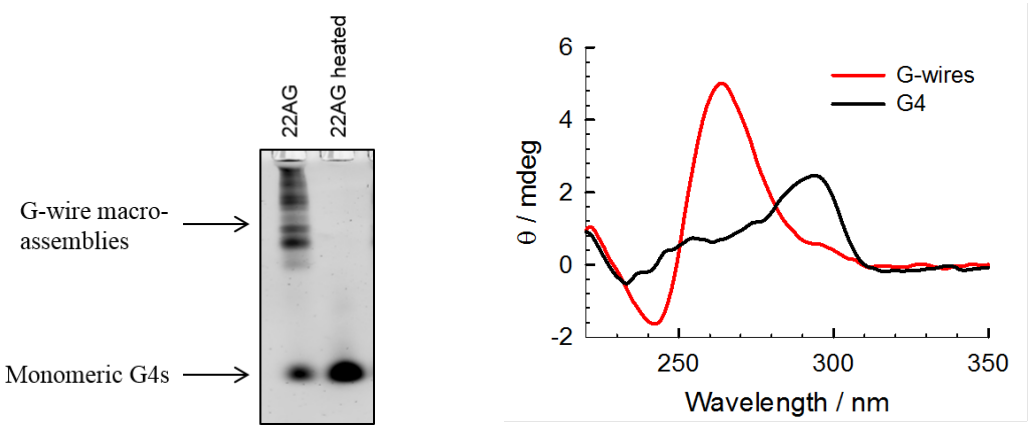


Fig. 3.6 Characterisation of monomeric and multimeric G4 structures. (a) Polyacrylamide gel electrophoresis migration of unheated (22AG) and heated (22AG h) G4 samples. Samples were run in the presence of 90 mM KCl and stained with ThT (b) Circular dichroism assays to study the presence of higher order structures in heated (black) and unheated (red) 22AG samples in the presence of 50 mM KCl.

To further characterise these structures, we have performed CD assays for the study of the unheated and heated 22AG samples at low micromolar concentrations (~4 µM) in 50 mM Tris and a background of 50 mM KCl. The heated structure shows a maximum at 295 nm and a minimum around 230 nm, which may correspond to a mix of anti-parallel and parallel monomeric conformations in a 3:1 or 4:1 ratio as observed in single-molecule FRET assays at 100 mM KCl [Lee et al., 2005; Ying et al., 2003]. On the contrary, the unheated sample displaying higher order conformations on gels shows now a characteristic parallel shape with maximum at 265 and minimum at 240 nm [Karsisiotis et al., 2011].

3.3.2 HPIP: a novel sensor for guanine quadruplexes

3.3.2.1 Photophysical behaviour of HPIP in aqueous solution

The ratiometric sensor HPIP presents two fluorescent species that coexist in equilibrium: the normal conformers N_{inter} and N_{intra}, in which the hydroxyl group is oriented towards the imidazole nitrogen and towards the solvent, respectively (Fig. 3.7). In the excited state, N_{intra} can undergo ESIPT and transfer a proton to a close acceptor group yielding a tautomer species, T, which corresponds to an emission band with an unusually large Stokes shift (>100 nm). It has been observed in the fibrillar aggregation of β-amyloid [Freire et al., 2015] that the restricted rotation of the C–C bond between the hydroxyphenyl and the imidazo[4,5-b]pyridine groups increases the probability of ESIPT. Then, we hypothesise a similar behaviour of this sensor when bound to G4 structures, where the DNA-sensor binding could be followed as an increase of the tautomer band and a decrease of the normal band.

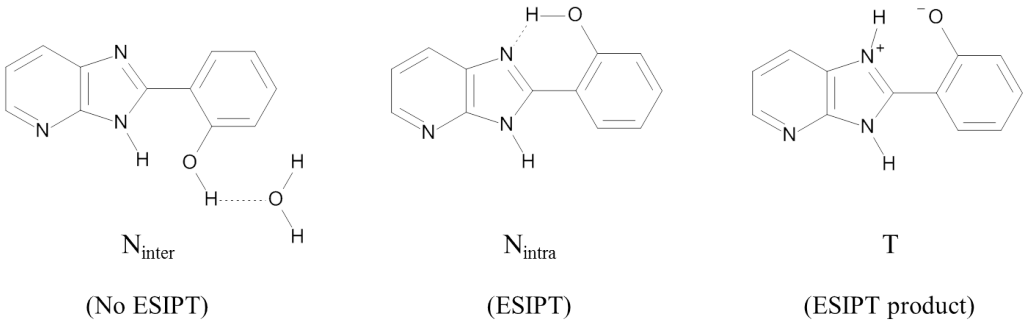
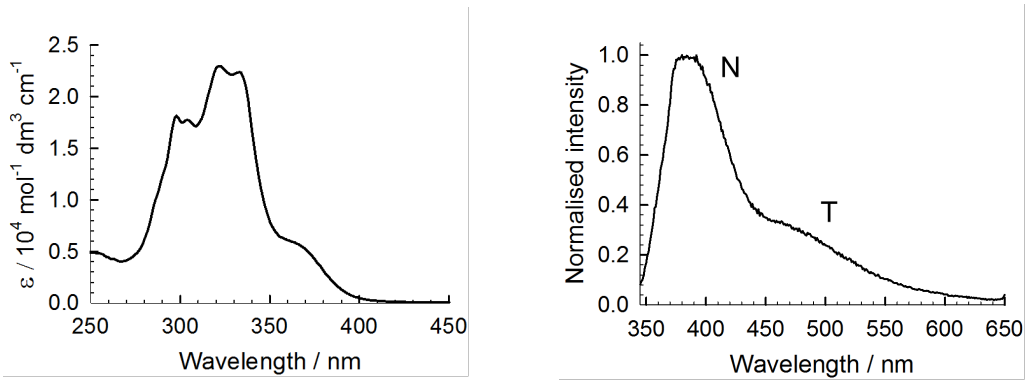


Fig. 3.7 Normal, N_{Inter} and N_{Intra}, and Tautomer, T, species of HPIP.

The absorption spectrum of HPIP shows a characteristic vibronic structure in aqueous solution at pH 7.5 with maximum at 332 nm (Fig. 3.8a) at which the log ε is 4.35 [Krishnamoorthy and Dogra, 2000]. With a pK_a of 8.6, the shoulder centered at ~365 nm corresponds to the contribution of the anion, which is assumed to be non-emissive [Freire et al., 2015].

The emission spectrum at λ_{exc} of 332 nm shows the presence of two bands corresponding to the normal and tautomer species (Fig. 3.8b). The normal N_{inter} is the predominant species in solution and responsible for the emission band at 385 nm. On the contrary, the less



(a) **(b)**

Fig. 3.8 Photophysics of HPIP. (a) Absorption spectrum of HPIP at pH 7.5 showing a shoulder centered at 365 nm corresponding to the anionic species. (b) Emission spectrum for λ_{exc} 332 nm at pH 7.5 showing the presence of the normal (N) and tautomer (T) bands.

favoured N_{intra} conformation undergoes ESIPT yielding the formation of the tautomer, which emits with an unusually large Stokes shift at 485 nm. A detailed model of the behaviour of HPIP in aqueous solution is presented in Fig. 3.9.

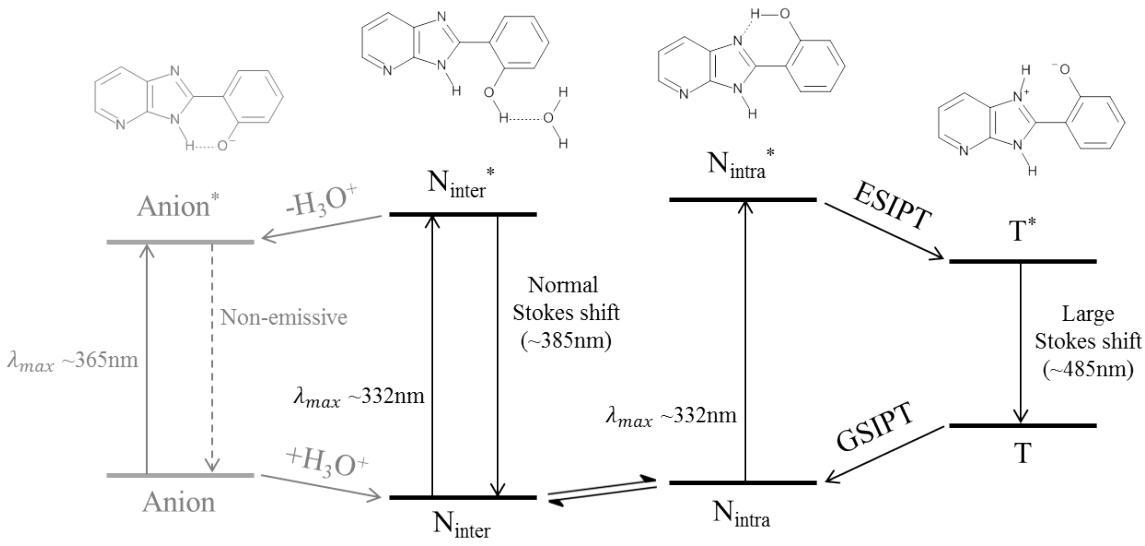
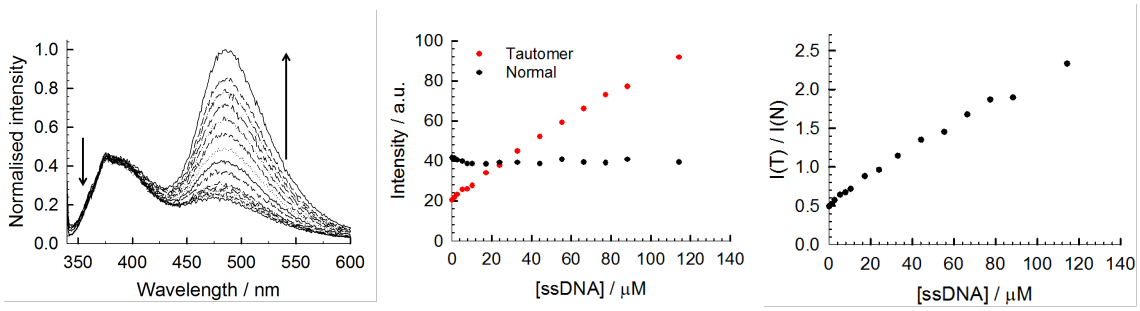


Fig. 3.9 ESIPT model showing the interconversion between normal, tautomer and anionic species in the ground and excited states. The emissive conformers N_{intra}, N_{inter} and T are in equilibrium in the ground state, however, the equilibrium arrows between N_{intra} and T were omitted for clarity due to the equilibrium shift towards the normal species. Moreover, the adoption of the N_{inter} conformation is favoured in aqueous media. The N_{inter}^* conformation is responsible for the emission at 385 nm, whereas the band with an abnormally large Stokes shift at 485 nm is due to T^*.

3.3.2.2 Steady-state characterisation of fluorescence emission of HPIP in the presence of G4s and single-stranded DNA

We have chosen a 22-nt long single-stranded DNA sequence (Table 3.1) that is unlikely to form secondary structures, determined by mfold web server [Zuker, 2003], as a control to investigate the affinity and signal change of HPIP in comparison to that observed in G4-containing samples. HPIP shows a continuous increase in the contribution of the tautomer band with little variation in the intensity of the normal species (Fig. 3.10b). The ratio between the tautomer and normal forms, I(T)/I(N), reaches a value of 2.2 at 100 µM, displaying a ~4.5-fold increase when compared with the unbound sensor (Fig. 3.10c).



(a) **(b)** **(c)**

Fig. 3.10 Characterisation of HPIP binding to ssDNA. (a) Normalised spectra of HPIP binding to ssDNA. (b) Intensities of tautomer and normal species as a function of the concentration of ssDNA. (c) Tautomer/normal ratio as a function of ssDNA describes a straight line within the concentration range of DNA.

The affinity and selectivity of HPIP for G4 monomeric structures was investigated for the human telomeric sequence (22AG), a longer tetracycline binding aptamer (Tet7) and the shorter thrombin binding aptamer (TBA) (Table 3.1) that folds into a chair-like anti-parallel G4 composed of only two G-quartets (Fig. 3.11) [Wang et al., 1993].

The DNA binding can be followed as a decrease in the contribution of the normal species and an increase in the contribution of the tautomer when HPIP is bound to G4 structures, which was then fitted to equation 3.5 as discussed in section 3.2.4.1. Ensemble fluorescence data with a background of 50 mM KCl show a ~18-fold increase of the tautomer/normal band ratio of HPIP for 22AG and Tet7, whilst the ratio when bound to TBA reaches a ~28-fold

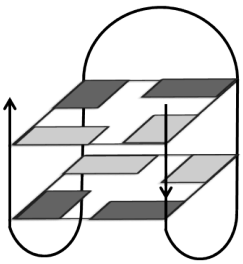
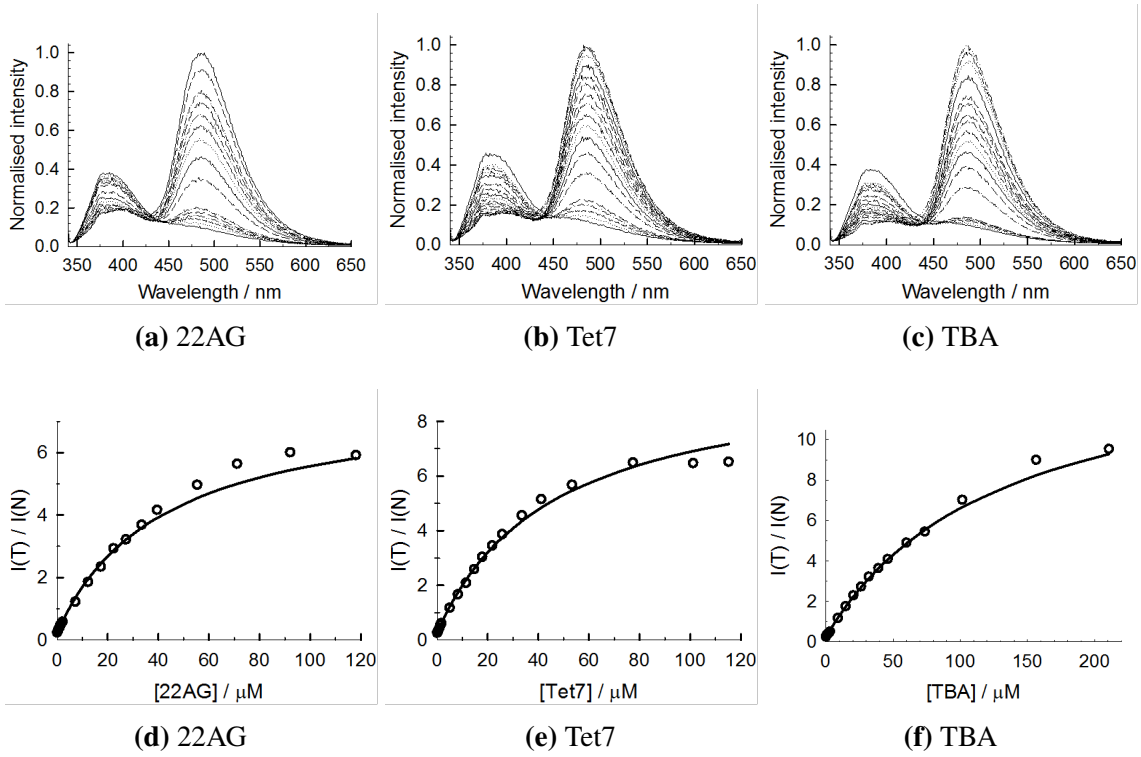


Fig. 3.11 Schematics of a TBA quadruplex. TBA aptamer folds into a chair-like conformation with anti-parallel arrangement of its two G-quartets.

increase (Figs. 3.12d, 3.12e, 3.12f). These results show good discrimination between ssDNA and G4 structures with a I(T)/I(N) ratio that is four to six times greater for G-quadruplexes. The binding affinity is similar in all G4-forming sequences with a K_d of 15.8±0.5 and



(a) 22AG **(b)** Tet7 **(c)** TBA

(d) 22AG **(e)** Tet7 **(f)** TBA

Fig. 3.12 Binding assays of HPIP and the 22AG, Tet7 and TBA G-quadruplexes. (a-c) Normalised spectra representing the variation of the normal and tautomer emissive species as a function of DNA binding. The isoemissive point at 438 nm is present in the three sequences. (d-f) Representation of the tautomer/normal ratio as a function of DNA concentration and fitted curve for a 1:1 binding model reporting the K_d values of 15.8±0.5, 10.1±0.3 and 21.8±1.6 μM for 22AG, Tet7 and TBA, respectively

10.1±0.3 μM for the 22AG and Tet7 G-quadruplexes, respectively, and a K_d of 21.8±1.6 μM for TBA (Table 3.2). Interestingly, the binding affinity of HPIP with these three different G4 topologies is comparable to the K_d reported for ThT (3.5 μM) [Mohanty et al., 2013] or those found for sensors such as thiazole orange (0.3–3 μM) and malachite green (2–35 μM) [Vummidi et al., 2013].

Table 3.2 I(T)/I(N) ratio increase and K_d for each of the strands.

Strand	T/N ratio increase	K_d / μM
ssDNA	4.5-fold*	N.D.
22AG	18-fold	15.8±0.5
TBA	28-fold	21.8±1.6
Tet7	18-fold	10.1±0.3

*Calculated at 100 μM DNA. N.D.: K_d not determined.

Importantly, the G4-forming sequences 22AG, Tet7 and TBA display an isoemissive point at 438 nm (Figs. 3.12a, 3.12b and 3.12c), which indicates that only one new species is formed, the tautomer, and reflects a shift in the equilibrium existing between the two normal forms of HPIP (Fig. 3.13). The lack of isoemissive points in the interaction with the ssDNA strand (Fig. 3.10a) indicates that the equilibrium between the N_{intra} and N_{inter} conformations remains unchanged in the presence of single-stranded DNA, which provides a second reference for the discrimination against these non-G4 forming sequences. The slight shift at high concentrations of DNA (≥100 μM) is due to the intrinsic fluorescence of G4 structures arising at ~410 nm.

Our results highlight the selectivity of HPIP for G4-forming sequences with a K_d in the micromolar range. The high increase in the tautomer/normal band ratio and the presence of an isoemissive point when bound to G4s provide the basis for the development of a new type of fluorescent probes allowing ratiometric sensing of DNA structures. Furthermore, the different tautomer/normal ratio shown for the three structures studied may indicate that HPIP is not only a G4 sensor but can also discern between different G4 topologies, although this hypothesis needs to be confirmed with the observation of a wider set of G4 conformations.

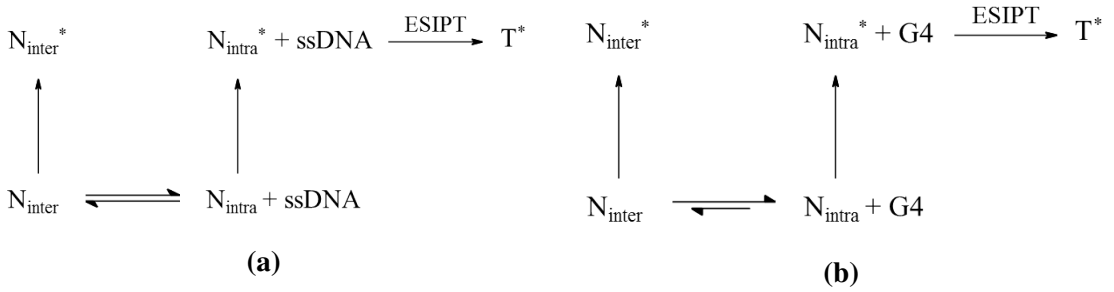


Fig. 3.13 Model proposed for the interaction of HPIP with single-stranded DNA and G4 sequences. (a) In the presence of ssDNA, the equilibrium between the two normal species N_{inter} and N_{intra} remains unchanged, which is translated in a low decrease in the fluorescence of the normal species. (b) In the presence of G4-forming sequences, the equilibrium between N_{inter} and N_{intra} is shifted towards the latter with a clear decrease in the contribution of the normal peak upon G4 binding.

3.3.3 Metal complexes as sensors for guanine quadruplexes

3.3.3.1 Characterisation of the $[Ru(bpy)_3]^{2+}$ interaction with G4s using fluorescence methods

A wide range of different platinum, iridium and ruthenium complexes have been investigated for their potential role stabilising G-quadruplexes. We have studied the interaction between G4s and the tris(bipyridine)ruthenium(II) bis(hexafluorophosphate), $[Ru(bpy)_3](PF_6)_2$, a ruthenium complex arranged around a chiral octahedral geometry that is stable in aqueous solution (Fig. 3.14).

The absorption spectrum of $[Ru(bpy)_3](PF_6)_2$ shows a main high energy band at 285 nm due to the ligand-to-ligand transitions (LLCT) and a second band in the visible region due to a metal-to-ligand charge transfer (MLCT) with maximum at 450 nm and a molar extinction coefficient of 14,600 $M^{-1}cm^{-1}$ (Fig. 3.15a) [Kalyanasundaram, 1982]. The latter band allows us to selectively excite the ruthenium complex far from the absorption region of the DNA nucleobases, which present an unusual high fluorescence in G4 structures [Dao et al., 2011]. The emission spectrum of $[Ru(bpy)_3]^{2+}$ shows a single band with maximum at 616 nm (Fig. 3.15b).

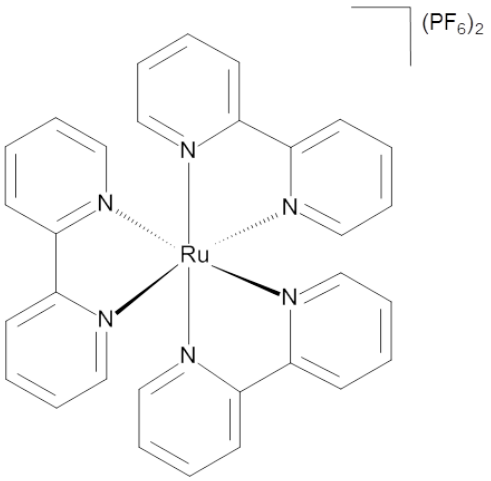


Fig. 3.14 Chemical structure of the $[Ru(bpy)_3](PF_6)_2$ ruthenium complex. Molecular structure of this ruthenium complex comprising three bidentate bipyridine ligands organised around an octahedral geometry.

Despite the low fluorescence enhancement in these type of complexes when bound to G-quadruplexes and their low quantum yield in aqueous media (0.042) [Kalyanasundaram, 1982], we have decided to exploit their use as gel staining agents for the correct discrimination of G4 and G-wire structures.

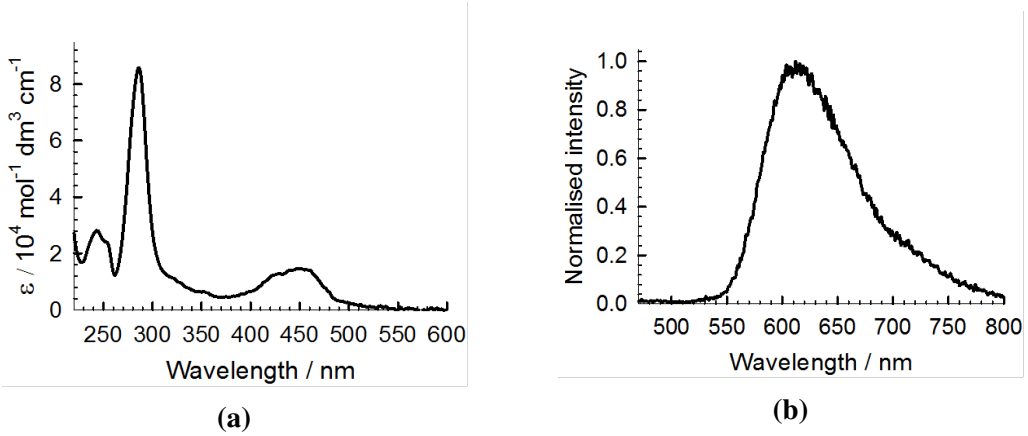


Fig. 3.15 Photophysical properties of $[Ru(bpy)_3](PF_6)_2$. (a) Absorption spectrum in aqueous solution. (b) Emission spectra at λ_{exc} of 450 nm displaying a maximum λ_{em} at 616 nm.

3.3.3.2 Polyacrylamide gel electrophoresis and circular dichroism of $[Ru(bpy)_3]^{2+}$: G4 complexes

We investigated the selectivity of this ruthenium complex for monomeric and multimeric G4 structures formed by the human telomeric sequence using DNA polyacrylamide gel electrophoresis. The results (Fig. 3.16a) show that $[Ru(bpy)_3]^{2+}$ is able to selectively bind G-wires (22AG) but unable to stain the discrete G4 structures of the heated sample (22AG h). The final staining with ThT reveals the presence of the monomeric structures for both unheated and heated samples.

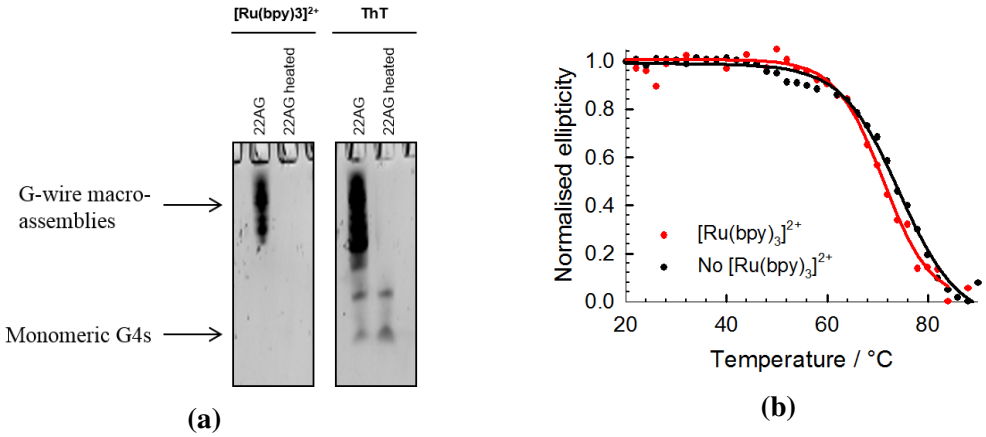


Fig. 3.16 G-wire sensing assays of $[Ru(bpy)_3](PF_6)_2$. (a) Gel staining assays comparing the binding affinity of $Ru(bpy)_3]^{2+}$ and ThT for monomeric and multimeric G4 structures. (b) Melting profile of 22AG G-wires in the absence (black) and presence (red) of $[Ru(bpy)_3](PF_6)_2$.

We then used circular dichroism for the study of the potential role of this complex stabilising G-wires. We followed and normalised the ellipticity value at 263 nm, corresponding to the main positive band that is characteristic of parallel G4s. The magnitude of the ellipticity at this wavelength decreases as the temperature increases due to the denaturation of the G-wires. The data show that the G-wire sample is slightly less resistant to denaturation (~2 °C less) in the presence of $[Ru(bpy)_3]^{2+}$ than the control in the presence of 100 KCl.

This ruthenium complex provides a new tool to discriminate between the G4 macroassemblies and discrete G-quadruplexes in polyacrylamide without altering their structure, and thus $[Ru(bpy)_3]^{2+}$ senses G-wires without affecting their formation. However, the poor

solubility of this compound may restrict its use and this is why we introduced curcumin as a water-soluble natural product that is commercially available in the study of G4 selective sensors.

3.3.4 Investigation of curcumin as a natural product that binds G4s

3.3.4.1 Structure and sensing of G-wires by curcumin

We investigated the use of curcumin as a natural G4 sensor because of its availability and reduced cost. Curcumin coexists as an equilibrium between its keto and enol tautomers (Fig. 3.17). The formation of the latter is favoured at basic pH due to the pK_{a1} of 7.7–8.5 of the enolic proton [Priyadarsini, 2009]. Furthermore, commercial curcumin contains a fraction of two curcumin derivatives: the demethoxy and bisdemethoxycurcumin. Curcumin has been associated with anti-inflamatory activity and, among other processes, it has been shown to reduce β-amyloid levels *in vivo* [Yang et al., 2005], regulate the expression of transcriptional factor HIF-1, inhibit the anti-apoptotic protein Bcl-2 and modulate the activity of several growth factors [Kunnumakkara et al., 2008].

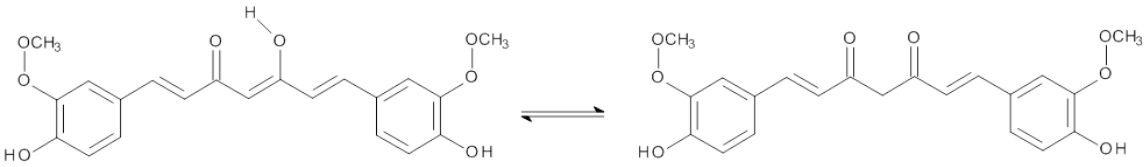


Fig. 3.17 Structures of curcumin enol and keto tautomers that coexist in equilibrium.

Curcumin rapidly degrades at neutral and alkaline pH into vanillin, ferulic acid and other subproducts but its stability substantially increases at pH 6 [Zebib et al., 2010] with the major contribution of the keto form. The photophysical studies of curcumin in slightly acidic conditions show an absorption and emission maxima at 427 and 544 nm (Figs. 3.18a and 3.18b) [Bernabe-Pineda et al., 2004], although these maxima are highly dependent on the solvent and pH [Priyadarsini, 2009].

We investigated the interaction of curcumin with G-wires and dsDNA in 50 mM MES at pH 6.5 with a background of 100 mM KCl using fluorescence kinetic assays. These studies

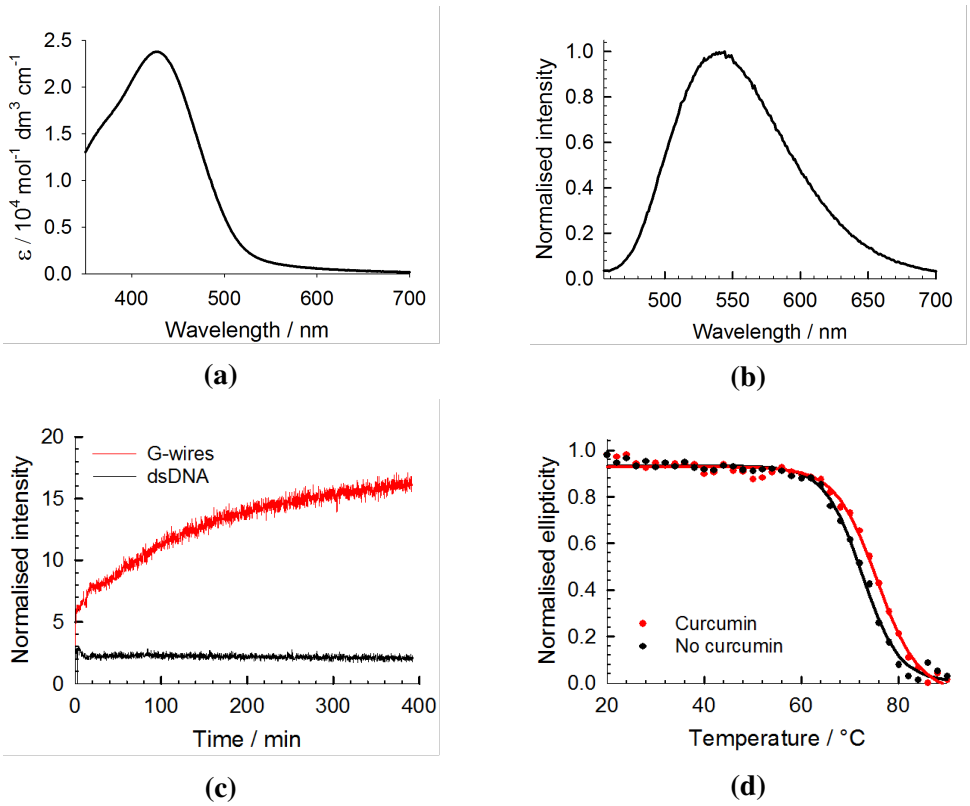


Fig. 3.18 Photophysical properties of curcumin in aqueous solution and its interaction with G-wires. (a) Absorption spectrum with absorption maximum at 427 nm. (b) Emission maximum at and maximum emission at 544 nm. (c) Kinetics of the interaction between 50 nM curcumin with 15 µM 22AG wires and dsDNA in 50 mM MES buffer pH 6.5. (d) CD thermal denaturation assay of 22AG wires in the presence of curcumin.

were performed for the 22AG and Tet7 G-wires as well as for the monomeric 22AG G4s. Interestingly, the addition of 15 µM 22AG wires to 50 nM curcumin leads to a slow increase in a timescale of ~400 minutes, leading to a 16-fold increase in the curcumin fluorescence signal (Fig. 3.18c). On the contrary, curcumin emission displays only a 2-fold increase upon addition of 15 µM 22AG G4s or dsDNA. These results suggest a preferential rearrangement around G-wires that may also stabilise and protect bound curcumin from degradation.

We have then used circular dichroism to investigate the stability of 22AG G-wires in the presence and absence of curcumin with a background of 100 mM KCl. As for the ruthenium complex, we followed the normalised ellipticity value at 263 nm that corresponds to the main positive peak in parallel G4 structures. The data show that the sample with curcumin bound

is more resistant to denaturation with a melting temperature of 75.5 °C, 3.1 °C higher than the 72.4 °C of the control in the absence of sensor (Fig. 3.18d).

After evaluating the stabilisation of G-wires by curcumin we carried out different gel electrophoresis assays to study the selectivity of curcumin for these structures as observed for $[Ru(bpy)_3]^{2+}$. These experiments were performed for unheated and heated samples of both 22AG and Tet7 sequences in the presence and absence of 90 mM salt (Na^+ and K^+). In those experiments with and without Na^+, curcumin can stain one band for 22AG wires and a few bands corresponding to Tet7 wires but none of the monomeric G4s (Fig. 3.19). However, in the presence of K^+, curcumin is able to interact with most G-wire macro-assemblies as well as the monomeric Tet7, whilst the monomeric 22AG remains hidden.

The results obtained by ensemble fluorescence and gel electrophoresis may suggest the potential use of this natural sensor as an alternative or complement to ThT as a staining agent to discriminate between G-wire, G4 and dsDNA structures.

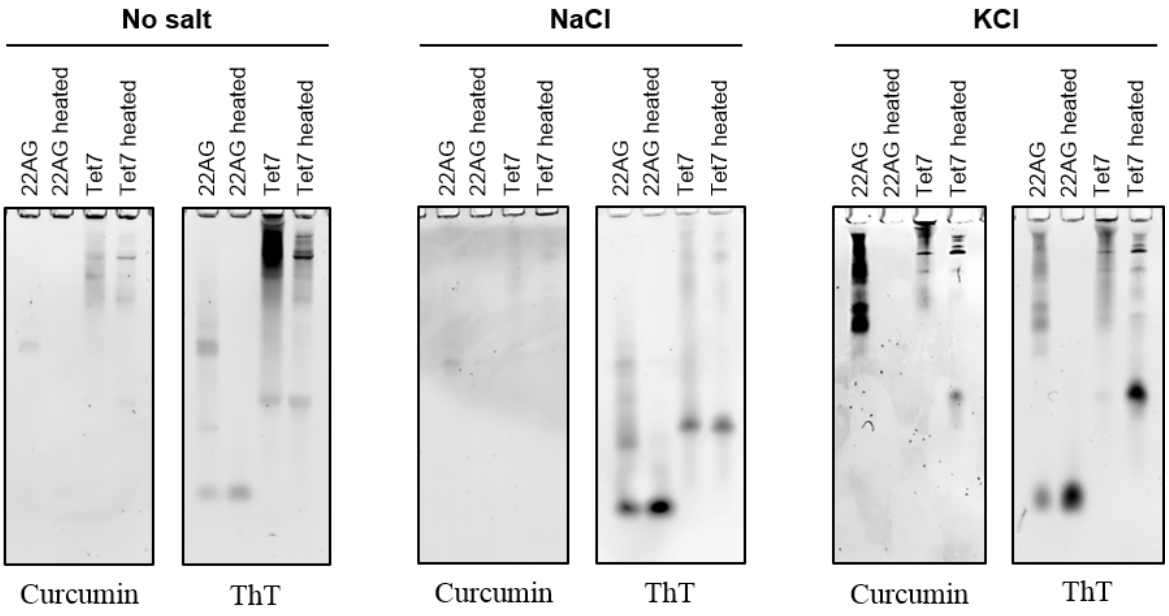


Fig. 3.19 Selectivity of curcumin in the detection of G-wires. Gel electrophoresis assays showing the selective binding of curcumin to multimeric 22AG and Tet7 structures (left panels) and posterior ThT staining (right panels) in the absence and presence of Na^+ and K^+ salts. Curcumin is able to effectively stain 22AG and Tet7 G-wires in all conditions, although it is at 90 mM KCl when curcumin exhibits its high selectivity for higher order structures by staining a wide range of G-wires of different lengths.

3.4 Conclusions

The study of G-quadruplexes as potential therapeutic targets has attracted many efforts in the last decade with special emphasis in the design of novel antitumour treatments. The presence of G-quadruplexes in a wide variety of oncogenes and the $(TTAGGG)_n$ repeats at the end of human telomeres, thus involved in the regulation of the telomerase activity that is responsible for the uncontrolled maintenance of the telomere length, highlights the importance of these structures in tumour cells. Understanding the folding and structure of G4s *in vitro* is crucial for the correct design of selective ligands that are able to stabilise G4s and to provide new tools to, when bound to these structures, regulate or inhibit the activity of the telomerase that is reported in $\geq$85% of cancers. The importance and huge growth of this field results in the need to develop new fluorescent probes that selectively sense G4 structures in order to obtain information regarding their folding, structure and stability under different conditions.

The aim of this chapter is to provide new tools to sense different G4s and G-wires by investigating the signal response of three fluorescent probes when bound to one of these architectures: i) a ratiometric sensor that undergoes excited-state intramolecular proton transfer (ESIPT), the 2-(2'-hydroxyphenyl)-3H-imidazo[4,5-b]pyridine (HPIP), in which the presence of two emission bands with good spectral separation can be used as an absolute observable; ii) the light-up ruthenium coordination compound $[Ru(bpy)_3](PF_6)_2$; and iii) a natural compound with known anti-inflamatory properties, curcumin, which has been investigated because of its availability and reduced cost.

The $[Ru(bpy)_3]^{2+}$ complex and curcumin have shown to be good selective sensors for the detection of G-wires. Although the ruthenium complex shows a small fluorescence increase when bound to G-wire structures, it does show good selectivity for these structures, which can be used for the their detection in gel electrophoresis assays. The good discrimination between G-wires and single- or double-stranded DNA that curcumin provides can be used in either ensemble fluorescence or gel electrophoresis assays. Moreover, curcumin has shown to stabilise G-wire structures against thermal denaturation, which can be of great interest for future work in the development of new sensors with similar binding selectivity and specificity.

On the other hand, the study of HPIP with the 22AG, Tet7 and TBA G-quadruplexes show a good discrimination against ssDNA sequences, thus providing the first ratiometric sensor for the detection of G4s. However, this study find some limitations in the solubility of the sensor, which can form aggregates in solution in the presence of salt, and in the relatively low affinity to G4s, which is in the micromolar range. In order to overcome these limitations we would need to design and synthesise new derivatives with improved solubility and binding affinity, which was beyond the scope of this chapter and would require to be carried out in more specialised organic chemistry laboratories.

References

Aznauryan, M., Søndergaard, S., Noer, S. L., Schiøtt, B., and Birkedal, V. (2016). A direct view of the complex multi-pathway folding of telomeric G-quadruplexes. *Nucleic Acids Research*, 44(22):11024–11032.

Bernabe-Pineda, M., Ramirez-Silva, M. T., Romero-Romo, M., Gonzalez-Vergara, E., and Rojas-Hernandez, A. (2004). Determination of acidity constants of curcumin in aqueous solution and apparent rate constant of its decomposition. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 60(5):1091–1097.

Bhattacharyya, D., Mirihana Arachchilage, G., and Basu, S. (2016). Metal cations in G-quadruplex folding and stability. *Frontiers in Chemistry*, 4:38.

Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chemistry*, 5(3):182–186.

Cogoi, S. and Xodo, L. E. (2006). G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Research*, 34(9):2536–2549.

Dai, J., Dexheimer, T. S., Chen, D., Carver, M., Ambrus, A., Jones, R. A., and Yang, D. (2006). An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human Bcl-2 promoter region in solution. *Journal of the American Chemical Society*, 128(4):1096–1098. PMID: 16433524.

Dai, T.-Y., Marotta, S. P., and Sheardy, R. D. (1995). Self-assembly of DNA oligomers into high molecular weight species. *Biochemistry*, 34(11):3655–3662. PMID: 7893661.

Dao, N. T., Haselsberger, R., Michel-Beyerle, M.-E., and Phan, A. T. (2011). Following G-quadruplex formation by its intrinsic fluorescence. *FEBS Letters*, 585(24):3969–3977.

References

De Armond, R., Wood, S., Sun, D., Hurley, L. H., and Ebbinghaus, S. W. (2005). Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1α promoter. *Biochemistry*, 44(49):16341–16350. PMID: 16331995.

de la Faverie, A., Guédin, A., Bedrat, A., Yatsunyk, L. A., and Mergny, J.-L. (2014). Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Research*.

de Lange, T., Shiue, L., Myers, R. M., Cox, D. R., Naylor, S. L., Killery, A. M., and Varmus, H. E. (1990). Structure and variability of human chromosome ends. *Molecular and Cellular Biology*, 10(2):518–527.

Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., and Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes & development*, 18(13):1618–1629.

Freire, S., Rodríguez-Prieto, F., Ríos Rodríguez, M. C., Penedo, J. C., Al-Soufi, W., and Novo, M. (2015). Towards ratiometric sensing of amyloid fibrils in vitro. *Chemistry - A European Journal*, 21(8):3425–3434.

Kalyanasundaram, K. (1982). Photophysics, photochemistry and solar energy conversion with tris(bipyridyl)ruthenium(ii) and its analogues. *Coordination Chemistry Reviews*, 46:159–244.

Karsisiotis, A. I., Hessari, N. M., Novellino, E., Spada, G. P., Randazzo, A., and da Silva, M. (2011). Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angewandte Chemie*, 123(45):10833–10836.

Klejevskaja, B., Pyne, A. L. B., Reynolds, M., Shivalingam, A., Thorogate, R., Hoogenboom, B. W., Ying, L., and Vilar, R. (2016). Studies of G-quadruplexes formed within self-assembled DNA mini-circles. *Chemical Communications*, 52(84):12454–12457.

Krishnamoorthy, G. and Dogra, S. K. (2000). Excited state intramolecular proton transfer in 2-(2'-hydroxyphenyl)-3H-imidazo[4,5-b]pyridine: effect of solvents. *Journal of Luminescence*, 92(1–2):91–102.

Kunnumakkara, A. B., Anand, P., and Aggarwal, B. B. (2008). Curcumin inhibits proliferation, invasion, angiogenesis and metastasis of different cancers through interaction with multiple cell signaling proteins. *Natural Products Special Issue*, 269(2):199–225.

Largy, E., Granzhan, A., Hamon, F., Verga, D., and Teulade-Fichou, M.-P. (2013). *Visualizing the Quadruplex: From Fluorescent Ligands to Light-Up Probes*, pages 111–177. Springer Berlin Heidelberg, Berlin, Heidelberg.

Lee, J. Y., Okumus, B., Kim, D. S., and Ha, T. (2005). Extreme conformational diversity in human telomeric DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 102(52):18938–18943.

Maizels, N. (2006). Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nature Structure & Molecular Biology*, 13(12):1055–1059.

Maizels, N. and Gray, L. T. (2013). The G4 genome. *PLoS Genet*, 9(4):1003468.

Majhi, A., Rahman, G. M., Panchal, S., and Das, J. (2010). Binding of curcumin and its long chain derivatives to the activator binding domain of novel protein kinase C. *Bioorganic & medicinal chemistry*, 18(4):1591–1598.

Marsh, T. C. and Henderson, E. (1994). G-Wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures. *Biochemistry*, 33(35):10718–10724.

McLuckie, K. I. E., Di Antonio, M., Zecchini, H., Xian, J., Caldas, C., Krippendorff, B.-F., Tannahill, D., Lowe, C., and Balasubramanian, S. (2013). G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. *Journal of the American Chemical Society*, 135(26):9640–9643.

Membrino, A., Paramasivam, M., Cogoi, S., Alzeer, J., Luedtke, N. W., and Xodo, L. E. (2010). Cellular uptake and binding of guanidine-modified phthalocyanines to KRAS/HRAS G-quadruplexes. *Chemical Communications*, 46:625–627.

Menon, V. P. and Sudheer, A. R. (2007). Antioxidant and anti-inflammatory properties of curcumin. In *The molecular targets and therapeutic uses of curcumin in health and disease*, pages 105–125. Springer.

Miyoshi, D., Fujimoto, T., and Sugimoto, N. (2012). Molecular crowding and hydration regulating of G-quadruplex formation. In *Quadruplex Nucleic Acids*, pages 87–110. Springer.

Miyoshi, D., Karimata, H., and Sugimoto, N. (2005). Drastic effect of a single base difference between human and tetrahymena telomere sequences on their structures under molecular crowding conditions. *Angewandte Chemie International Edition*, 44(24):3740–3744.

Mohanty, J., Barooah, N., Dhamodharan, V., Harikrishna, S., Pradeepkumar, P. I., and Bhasikuttan, A. C. (2013). Thioflavin T as an efficient inducer and selective fluorescent sensor for the human telomeric G-quadruplex DNA. *Journal of the American Chemical Society*, 135(1):367–376.

Moye, A. L., Porter, K. C., Cohen, S. B., Phan, T., Zyner, K. G., Sasaki, N., Lovrecz, G. O., Beck, J. L., and Bryan, T. M. (2015). Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nature Communications*, 6:7643.

References

Noer, S. L., Preus, S., Gudnason, D., Aznauryan, M., Mergny, J.-L., and Birkedal, V. (2016). Folding dynamics and conformational heterogeneity of human telomeric G-quadruplex structures in Na^+ solutions by single molecule FRET microscopy. *Nucleic Acids Research*, 44(1):464–471.

Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., and Lipps, H. J. (2005). Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nature Structure & Molecular Biology*, 12(10):847–854.

Palumbo, S. L., Ebbinghaus, S. W., and Hurley, L. H. (2009). Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *Journal of the American Chemical Society*, 131(31):10878–10891. PMID: 19601575.

Pinnavaia, T. J., Marshall, C. L., Mettler, C. M., Fisk, C. L., Miles, H. T., and Becker, E. D. (1978). Alkali metal ion specificity in the solution ordering of a nucleotide, 5'-guanosine monophosphate. *Journal of the American Chemical Society*, 100(11):3625–3627.

Pinnavaia, T. J., Miles, H. T., and Becker, E. D. (1975). Self-assembled 5'-guanosine monophosphate, nuclear magnetic resonance evidence for a regular, ordered structure and slow chemical exchange. *Journal of the American Chemical Society*, 97(24):7198–7200.

Priyadarsini, K. I. (2009). Photophysics, photochemistry and photobiology of curcumin: Studies from organic solutions, bio-mimetics and living cells. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 10(2):81–95.

Qin, Y., Rezler, E. M., Gokhale, V., Sun, D., and Hurley, L. H. (2007). Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Research*, 35(22):7698–7713.

Rankin, S., Reszka, A. P., Huppert, J., Zloh, M., Parkinson, G. N., Todd, A. K., Ladame, S., Balasubramanian, S., and Neidle, S. (2005). Putative DNA quadruplex formation within the human c-kit oncogene. *Journal of the American Chemical Society*, 127(30):10584–10589. PMID: 16045346.

Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J., and Plückthun, A. (2001). In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proceedings of the National Academy of Sciences*, 98(15):8572–8577.

Semenza, G. L. (2003). Targeting HIF-1 for cancer therapy. *Nature Reviews Cancer*, 3(10):721–732.

Sharma, V. R. and Sheardy, R. D. (2014). The human telomere sequence, $(TTAGGG)_4$, in the absence and presence of cosolutes: a spectroscopic investigation. *Molecules*, 19(1):595–608.

Shi, S., Geng, X., Zhao, J., Yao, T., Wang, C., Yang, D., Zheng, L., and Ji, L. (2010). Interaction of [Ru(bpy)2(dppz)]2+ with human telomeric DNA: Preferential binding to G-quadruplexes over i-motif. *Biochimie*, 92(4):370–377.

Shi, Y., Luo, H. Q., and Li, N. B. (2013). A highly sensitive resonance Rayleigh scattering method to discriminate a parallel-stranded G-quadruplex from DNA with other topologies and structures. *Chemical Communications*, 49:6209–6211.

Shirude, P. S., Okumus, B., Ying, L., Ha, T., and Balasubramanian, S. (2007). Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene c-Kit. *J. Am. Chem. Soc.*, 129(24):7484–7485.

Shirude, P. S., Ying, L., and Balasubramanian, S. (2008). Single molecule conformational analysis of the biologically relevant DNA G-quadruplex in the promoter of the proto-oncogene c-MYC. *Chemical Communications*, (17):2007–2009.

Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences*, 99(18):11593–11598.

Sun, D., Guo, K., Rusche, J. J., and Hurley, L. H. (2005). Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Research*, 33(18):6070–6080.

Venczel, E. A. and Sen, D. (1993). Parallel and antiparallel G-DNA structures from a complex telomeric sequence. *Biochemistry*, 32(24):6220–6228.

Vummidi, B. R., Alzeer, J., and Luedtke, N. W. (2013). Fluorescent probes for G-quadruplex structures. *ChemBioChem*, 14(5):540–558.

Wang, K. Y., McCurdy, S., Shea, R. G., Swaminathan, S., and Bolton, P. H. (1993). A DNA aptamer which binds to and inhibits thrombin exhibits a new structural motif for DNA. *Biochemistry*, 32(8):1899–1904.

Yang, F., Lim, G. P., Begum, A. N., Ubeda, O. J., Simmons, M. R., Ambegaokar, S. S., Chen, P. P., Kayed, R., Glabe, C. G., and Frautschy, S. A. (2005). Curcumin inhibits formation of amyloid β oligomers and fibrils, binds plaques, and reduces amyloid in vivo. *Journal of Biological Chemistry*, 280(7):5892–5901.

References

Ying, L., Green, J. J., Li, H., Klenerman, D., and Balasubramanian, S. (2003). Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences*, 100(25):14629–14634.

Zebib, B., Mouloungui, Z., and Noirot, V. (2010). Stabilization of curcumin by complexation with divalent cations in glycerol/water system. *Bioinorganic chemistry and applications*, 2010.

Zhou, W., Brand, N. J., and Ying, L. (2011). G-quadruplexes–novel mediators of gene function. *Journal of Cardiovascular Translational Research*, 4(3):256–270.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13):3406–3415.

4

Quantitative single-molecule FRET methods to quantify the axial rotation of duplex helical stems in RNA riboswitches

4.1 Introduction

Riboswitches are highly structured segments of non-coding mRNA found in bacteria, archaea, plants and fungi. Riboswitches can bind cellular metabolites that induce a change in their conformation to control the genetic expression of associated genes. Currently, about 20 classes of riboswitches have been identified, sensing a diversity of small metabolites such as amino acids, coenzymes, metal ions, nucleic acids and sugars [Perez-Gonzalez et al., 2016]. The structure of a riboswitch is composed of two parts: an aptamer domain and an expression platform. The aptamer domain is a highly conserved sequence that selectively binds the small metabolite, whilst the expression platform is a more variable sequence located downstream and is responsible for modulating the gene expression (Fig. 4.1) [Blouin et al., 2009; Winkler and Breaker, 2005]. Most conformational changes induced in a riboswitch upon ligand binding lead to a repression of the gene expression. When the concentration of ligand is sufficient to bind the aptamer, the riboswitch folds into a conformation that prevents the formation of the transcription anti-terminator stem on the expression platform and promotes the formation of the terminator stem that leads to premature transcription termination and

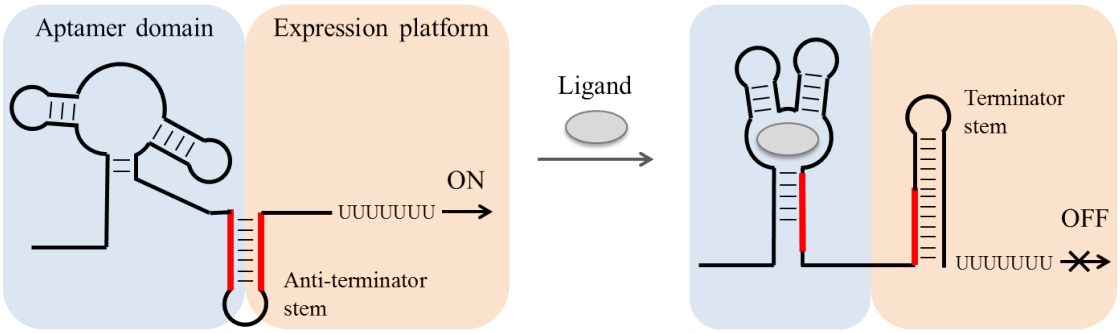


Fig. 4.1 Schematics of the ligand-free and ligand-bound conformations of a negatively-controlled riboswitch. The aptamer domain is responsible for metabolite binding, whilst the expression platform is involved in the control of the genetic expression. In the absence of ligand, sequences from the aptamer and the expression platform form a secondary structure called anti-terminator stem that allows transcription to continue. On the contrary, ligand binding promotes the folding of the aptamer and the formation of a terminator stem that leads to premature transcription termination.

repression of the downstream gene. This fact has increased the interest in the development of metabolite analogues able to bind a specific riboswitch in order to repress the associated gene and disrupt the normal function of bacteria. Thus, riboswitches have emerged as potential targets for a new kind of antibacterial drugs that would help to address the problem of an increasing antibiotic resistance in bacterial pathogens [Blount and Breaker, 2006].

S-adenosylmethionine (SAM) riboswitches are associated to the control of the biosynthesis and transport of different cellular metabolites such as cysteine, methionine, and SAM [Batey, 2011]. To date, six different families of riboswitches binding SAM have been reported. In this chapter I will focus on the SAM-I riboswitch and, more in particular, the *yitJ* aptamer from *Bacillus subtilis*, a negatively controlled riboswitch in which the ligand binding stabilises the aptamer domain and promotes the formation of the terminator stem [Heppell and Lafontaine, 2008; Winkler et al., 2003]. Crystal structure of the *yitJ* aptamer from *Bacillus subtilis* (Fig. 4.2) display a similar arrangement to the previously reported SAM-I aptamer from *Thermoanaerobacter tengcongensis* that revealed a pseudoknot interaction between the P2 stem loop and the J3/4 single-stranded region on a four-way junction in which the four stems are organised as two pairs of stacked helices involving stems P2-P3 and P1-P4 oriented in an angle of ~70° between each other [Montange and Batey, 2006; Lu et al.,

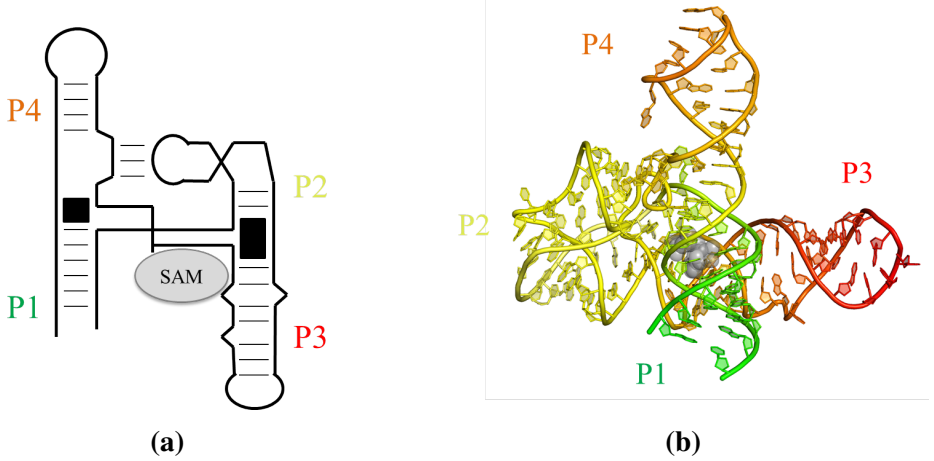


Fig. 4.2 Structure of the SAM-I aptamer. (a) Schematics of the ligand-bound folded conformation, F_{NS}, of the SAM-I riboswitch. (b) Crystal structure of the SAM-I *yitJ* aptamer from *Bacillus subtilis*. Ligand is represented in gray (PDB ID: 4KQY) [Lu et al., 2010].

2010]. The *yitJ* aptamer shows high affinity for SAM with a K_d of ~4 nM and high specificity with a ~100-fold discrimination against its analogue S-adenosyl-homocysteine (SAH) that lacks the methyl group and positive charge at the sulfur atom [Winkler et al., 2003].

Heppell and coworkers [Heppell et al., 2011] have taken advantage of ensemble and single-molecule fluorescence techniques to further investigate the key aspects of the rearrangment of the four stems around the four-way junction in the presence of SAM. Ensemble-FRET experiments with fluorescein and Cy3 attached at the end of two stems were performed to monitor changes in the end-to-end distance in the presence of 25 mM NaCl, 10 mM $MgCl_2$ and 10 mM $MgCl_2$ 5 µM SAM. The construct labelled at stems P1 and P3, named vector 1-3, shows a FRET increase at 10 mM $MgCl_2$ but no change upon addition of 5 µM SAM. Vector 1-4 shows only a FRET change after addition of ligand, whilst vector 3-4 shows a FRET increase at 10 mM $MgCl_2$ and a decrease upon addition of SAM. The rearrangement of the four stems around the four-way junction was also studied by introducing a 2AP at the interface of each pair of stacked stems. The construct where the 2AP was located at the P2-P3 interface shows a decrease in its fluorescence at 10 mM $MgCl_2$ but no change was detected upon addition of 5 µM SAM. On the contrary, the fluorescence signal of the 2AP at the P1-P4 interface remains invariant at 10 mM $MgCl_2$ and decreases at 10 mM $MgCl_2$ 5 µM SAM. The use of these two ensemble fluorescence techniques provided valuable information

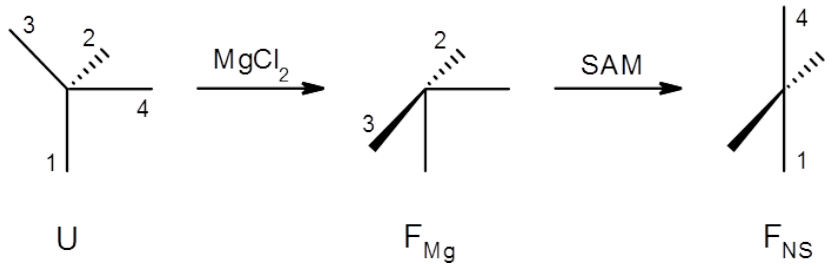
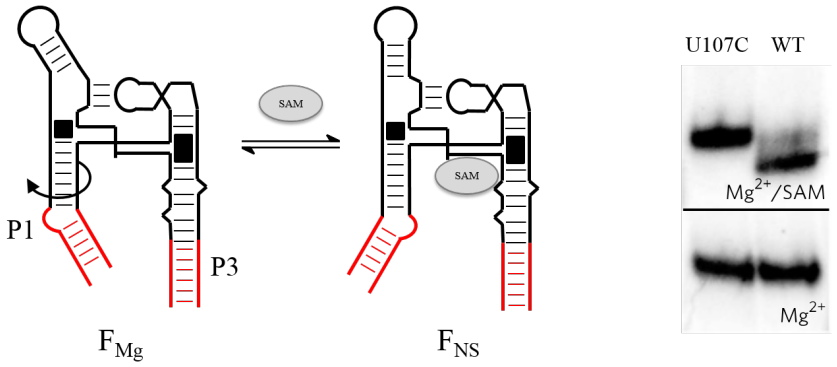


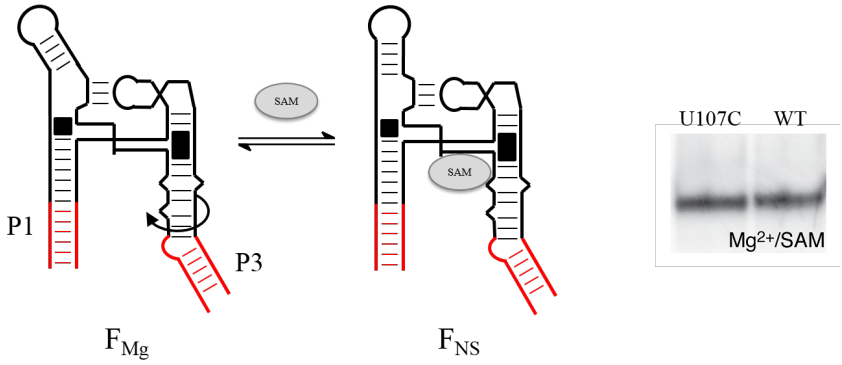
Fig. 4.3 Two-step folding model of the SAM-I riboswitch. In the presence of Mg^{2+}, stems P2 and P3 adopt a stacked position, acompanied by a juxtaposition of P1 and P3 (F_{Mg}). Then, the addition of SAM induces the formation of the P1-P4 stacking in the folded native state (F_{NS}). Adapted from [Heppell et al., 2011].

to monitor the folding of the aptamer in two steps: i) magnesium induces the formation of a pseudoknot interaction between the P2 stem-loop and the P3-P4 junction, the P2-P3 stacking and juxtaposition of P1 and P3 stems; and ii) the addition of SAM induces the formation of the P1-P4 stacking and rearrangement of the J1/2 and J3/4 single-stranded regions (Fig. 4.3) [Eschbach et al., 2012; Heppell et al., 2011]. Furthermore, ligand binding is also associated to the helical rotation of stem P1 along its helical axis as observed by gel electrophoresis (Figs. 4.4a) [Heppell et al., 2011], which is though to play an important role in the control of the genetic expression. On the contrary, stem P3 is stabilised in the presence of Mg^{2+} and no rotation was observed upon addition of SAM (Figs. 4.4b).

Although gel electrophoresis and ensemble-FRET assays provide some evidence about a fixed end-to-end distance [Heppell et al., 2011], we need to further investigate more in detail the relative movement of P1 and P3 stems during ligand binding. In particular, due to the low sensitivity of these assays, we are interested in understanding the relative position of P3 upon ligand binding in relation to P1, which is something that was ambiguous in previous experiments. For this, we have designed experiments phasing stems P1 and P3 by subtraction or addition of base pairs before the labeling site (Fig. 4.5). We specifically positioned the donor at a terminal position of stem P1 and acceptor at an internal position of P3 to allow free rotation, so their orientation does not depend on κ^2, and thus the changes in FRET are only due to changes in the inter-dye distance.



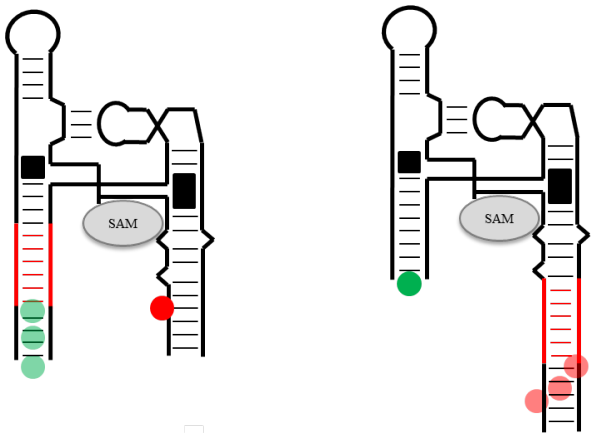
(a) Bulged P1



(b) Bulged P3

Fig. 4.4 Helical rotation of the P1 and P3 stems upon ligand binding. (a) Schematics showing the helical rotation of P1 stem upon ligand binding and gel migration profiles for the wild type (WT) and U107C mutant, which is used as a reporter for the magnesium-induced F_{Mg} state due to its inability to bind the ligand. The bulged group incorporated in P1 shows an increase in the end-to-end distance of the WT aptamer in the folded native state (F_{NS}). On the contrary, no variation is observed in the absence of SAM. (b) Schematics showing the fixed position of P3 upon ligand binding and gel migration profiles for the WT and U107C mutant in the presence of SAM. Adapted from [Heppell et al., 2011].

This project was carried out in collaboration with Daniel Lafontaine's group (Université de Sherbrooke, Canada), who performed the smFRET experiments. Our lab has then analysed the raw data, built the population histograms and designed a three-dimensional model for each construct by elongating P1 and P3 on the crystal structure to obtain the FRET-averaged inter-dye distance. The theoretical distance will be finally compared to the smFRET experimental data to describe the reorientation of P3 in the folding of the folded native state (F_{NS}) conformation.



(a) Phasing of P1 **(b)** Phasing of P3

Fig. 4.5 Induced phasing of stems P1 and P3. (a) Scheme showing the phasing of P1 induced by addition of base pairs (red) before the labelling position. Cy3 is attached at the 3' end of the stem, so it is expected to freely rotate (green circles), whilst Cy5 remains in the same position (red circle). (b) Scheme showing the phasing of P3 by addition of base pairs (red) before the labelling point of the Cy5 that is attached at an internal position. Therefore, the addition of base pairs is translated in a rotation of the position of the dye along the helical axis (red circles).

4.2 Materials and methods

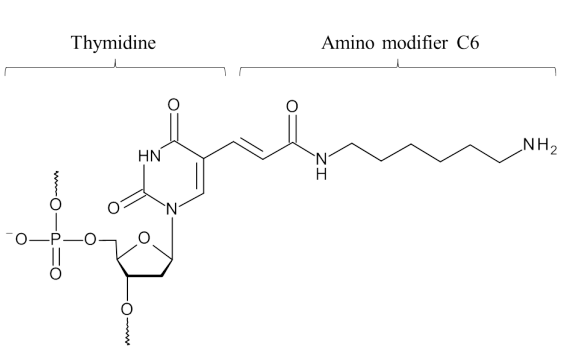
4.2.1 Preparation of fluorescently-labelled aptamers

The *yitJ* wild-type SAM-I aptamer and derivatives were annealed from synthetic and transcribed RNA strands as described in [Boudreault et al., 2015]. The synthetic strands, which carry a primary amino group at an specific position (3' end for P4-P1 and internal for P3), were purchased from IDT. Strands P3-P4 and P1-P3 were transcribed using the T7 RNA

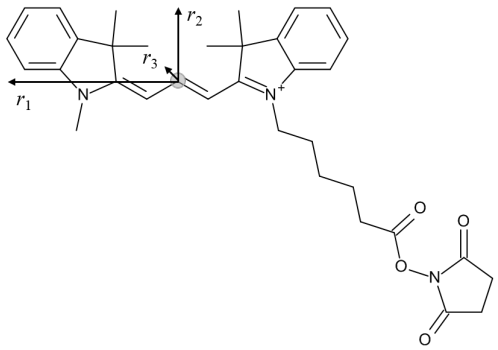
Table 4.1 Synthetic and transcribed strands used in the annealing of the SAM-I riboswitch (WT). Extended sequences are highlighted in bold.

Strand	Sequence (from 5' to 3')
*P*3	Biotin-GCGAGAGAACGC(iAmMC6T)UG
$P3-P4$	GC**GAUUCUGGAC**GGGUGCUAAAUCCAGCAAGCUGCGC
$P4-P1$	GCGCAGCUUGGAAGAUAAGAA**GUACAUCG**CGC(3AmMC6T)
$P1-P3$	GCGUUCUUAUCAAGAGAAGCACAGGGACUGGCCCGACGAAGCUU CAGCAACCUGCCAAGCGUUCUCUCGC

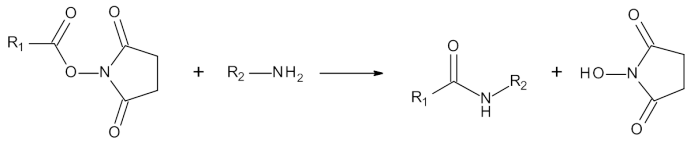
polymerase. The labelling of the synthetic strands was performed as described in [Brenner et al., 2010] using succinimidyl ester (NHS ester) dyes that react with the amino group at the RNA strand for site-specific labelling (Fig. 4.6). The sequences used for the annealing of the aptamers are detailed in Table 4.1. The sequences used for the elongation of stems P1 (up to +8 bp) and P3 (up to +10 bp) are highlighted in bold, with the first bases added starting from the 5' end.



(a) Amino modifier C6 dT



(b) Cy3 NHS ester



(c) Reaction scheme for the NHS ester and primary amino group conjugation

Fig. 4.6 Labelling of synthetic DNA or RNA with a primary amine group and a NHS ester modified dye. (a) Molecular structure of a amino modifier C6 dT at 5', 3' and internal positions. (b) Molecular structure of a Cy3 succinimidyl ester (NHS ester) and three radii parameters used to estimate the dimensions of the dye (r_1, r_2 and r_3). (c) General scheme for the reaction between a NHS ester and a primary amino group used for site-specific labelling or nucleic acids. R1 represents the modified dye. R2 represents the target dT base incorporated at a known position of the DNA or RNA strand.

4.2.2 Single-molecule FRET imaging

Slides were passivated with 1 mg/mL biotinylated BSA and 0.2 mg/mL streptavidin. SAM riboswitch constructs were added in a concentration between 50 and 250 pM. SAM ligand was added to an imaging buffer containing 50mM Tris-HCl (pH 8.1), 50 mM NaCl, 3 mM

$MgCl_2$, 2.5 mM PCA, 250 nM PCD and 1 mM TROLOX. All experiments were performed at room temperature with an integration time of 50 ms.

SmFRET experiments were performed in a home-built prism-type total-internal reflection microscope (Olympus IX71) equipped with a 50mW continuous-wave green laser emitting at 532 nm. Donor and acceptor fluorescence emissions were acquired in a back-illuminated Ixon EMCCD camera. The mapping and process of the raw movies were performed using IDL software and the single-molecule traces obtained were then analysed using laboratory-written routines in Matlab, as described in [McCluskey et al., 2014]. The apparent FRET efficiency was calculated after background correction from the expression $E_{app}=I_A/(I_A+I_D)$, where I_D and I_A represent the fluorescence intensities of donor (Cy3) and acceptor (Cy5), respectively. Single-molecule histograms were obtained by averaging of the first 10 frames of each molecule for the total number of molecules recorded, without discrimination of the the donor-only labelled molecules for calculation of the zero-FRET peak. The smFRET histograms were then fitted to a linear combination of Gaussian distributions to identify the different populations involved in the folding of the aptamer.

4.2.3 Correction of single-molecule FRET data

Intensities of donor and acceptor were corrected as described in [Ouellet et al., 2011] in order to obtain the most accurate distance value from the single-molecule data. This correction takes into account the dye quantum yields, photon detection efficiencies and characteristics of the dichroic mirrors.

The first correction involves the gamma factor (γ), which is a function of the dye quantum yields, different detection efficiencies, pH, temperature and optics [Ouellet et al., 2011]. The gamma factor was calculated from the change in intensity in the donor and acceptor channels after photobleaching of the acceptor

$$\gamma = \frac{\Delta\overline{I_A}}{\Delta\overline{I_D}} \tag{4.1}$$

Raw donor and acceptor intensities were averaged over 50 frames before and after the photobleaching event. The gamma factor for the phasing of P3 data was calculated over 200 molecules within SAM -1 bp to +10 bp, with a mean of 1.493 and a 95% confidence interval of ± 0.082. The value of γ when phasing of P1 was averaged over 100 molecules within SAM -2 bp to +8 bp, yielding a value of 1.805 ± 0.107. The FRET efficiency corrected for the gamma factor is calculated from

$$\mathrm{E_{app}}^* = \frac{I_A}{I_A + \gamma I_D} \tag{4.2}$$

The next step refers to the correction of the zero peak or the leakage of the donor emission into the acceptor channel due to the long-pass dichroic mirror, which results in a value of the FRET efficiency for donor-only molecules different from zero. The value of the zero peak, p, can be calculated and corrected with a Gaussian fit of the single-molecule $\mathrm{E_{app}}^*$ histogram for each construct as shown in Fig. 4.7

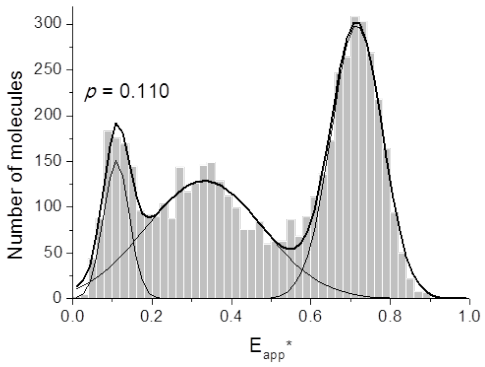


Fig. 4.7 Zero-peak correction. Population histogram showing the $\mathrm{E_{app}}^*$ for the SAM +1 bp construct when phasing P3 and Gaussian fit to obtain the value of the zero peak, p.

Finally, a backreflection ratio r is introduced to take into account the reflection of the acceptor emission into the donor channel. This term was calculated from molecules labelled with Cy5 dye only directly excited with a red laser and corresponds to the difference between the experimental and the theoretical maximum $\mathrm{E_{app}}$ of 1. The data were represented as a single-molecule histogram and fitted to a Gaussian function centered at 0.93647, thus yielding a value of $r = 0.06353$.

The final corrected FRET efficiency is calculated from

$$E_{FRET} = \frac{E_{app}{}^{*} - p}{1 - r - p} \tag{4.3}$$

4.2.4 Modelling of the experimental constructs

The three-dimensional models of all constructs used the crystal structure of the SAM-I *yitJ* aptamer from *Bacillus subtilis* [Lu et al., 2010] as a template to which stems P1, P3 and P4 with the specific sequences were added for a better visual representation of the dimensions of each aptamer. These sequences were modelled using the 'Build Structure' tool of the molecular modelling software UCSF Chimera [Pettersen et al., 2004] following an RNA A-form helix. The two amino modifier C6 dT bases were substituted with two uracil bases to which dye and amino linker were then attached. Furthermore, for a correct orientation of the stems, these sequences incorporate one or two extra base pairs when phasing P1 or P3, respectively, that are overlapped with their counterpart in the crystal structure using Pymol (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC), in which the new stems are aligned to extend the existing in the crystal structure.

4.2.5 Determination of the theoretical FRET

The theoretical inter-dye distance was calculated using the FRET Positioning and Screening (FPS) software developed by Kalinin and coworkers [Kalinin et al., 2012], which simulates the dye accessible volume (AV) by calculating all the possible orientations and assigning them the same probability. The program replicates an ovoid-shaped dye defined by three different radii and a flexible linker defined by its maximum length and width [Sindbert et al., 2011]. The terminal Cy3 and internal Cy5 attached via two amino modifier C6 dTs (Fig. 4.6) are set at the carbon-5 position of the uracil base and described using the parameters provided in [Sindbert et al., 2011] and in the software manual (Table 4.2). The theoretical distance value for the three-dimensional model is then obtained as an integration of all the possible FRET efficiencies with an estimated R_0 of 49.0 Å.

Table 4.2 Parameters of the Cy3 and Cy5 dyes. Linker and dye dimensions for Cy3 and Cy5 attached at carbon-5 of the nucleic acid base with amino modifier C6 as reported in the FPS software manual [Kalinin et al., 2012]. The three radii represent the estimated parameters of the dye when assuming an ovoid shape.

	Linker		Dye		
	Length / Å	Width / Å	Radius 1 / Å	Radius 2 / Å	Radius 3 / Å
Cy3	20	4.5	6.8	3	1.5
Cy5	22	4.5	11	3	1.5

4.3 Results and discussion

We have engineered a total of 22 different constructs (ranging from -2 to +8 bp in the shortening or elongation of P1 and from -1 to +10 bp for P3) that were analysed using single-molecule FRET. The addition of base pairs was designed to induce a phasing of the two stems carrying the donor and acceptor to provide information on the helical periodicity of the rotation of P1 and P3. SmFRET studies were performed at 3 mM $MgCl_2$ in the presence and absence of 1, 10 or 100 µM SAM. At these conditions, the single-molecule histograms representing the apparent FRET efficiency (E_{app}) for the phasing of P1 (Fig. 4.8) and P3 stems (Fig. 4.9) were fitted to a linear combination of three Gaussian populations associated to: i) zero peak (*), due to donor-only labelled molecules; ii) magnesium-induced state (F_{Mg}) corresponding to the formation of the P2-P3 helical stacking and the P1-P3 juxtaposition; and iii) SAM-induced folded native state (F_{NS}) corresponding to the formation of the P1-P4 stacking, rotation of P1 and reorganisation of the single-stranded P1-P2 and P3-P4 junctions, named J1/2 and J3/4. The distance increase between donor and acceptor due to the addition of base pairs to stems P1 or P3 implies a shift of the F_{Mg} and F_{NS} peaks towards lower FRET efficiencies. On the contrary, the position of the zero-peak remains unaffected by the elongation of the stems with an E_{app} of ~0.15.

The experimental histograms representing the E_{app} were corrected for the gamma factor, zero-peak and backreflection ratio as described in section 4.2.3 in order to obtain the corrected FRET efficiency (E_{FRET}) (Figs. 4.10 and 4.11). The corrected histograms were then fitted to

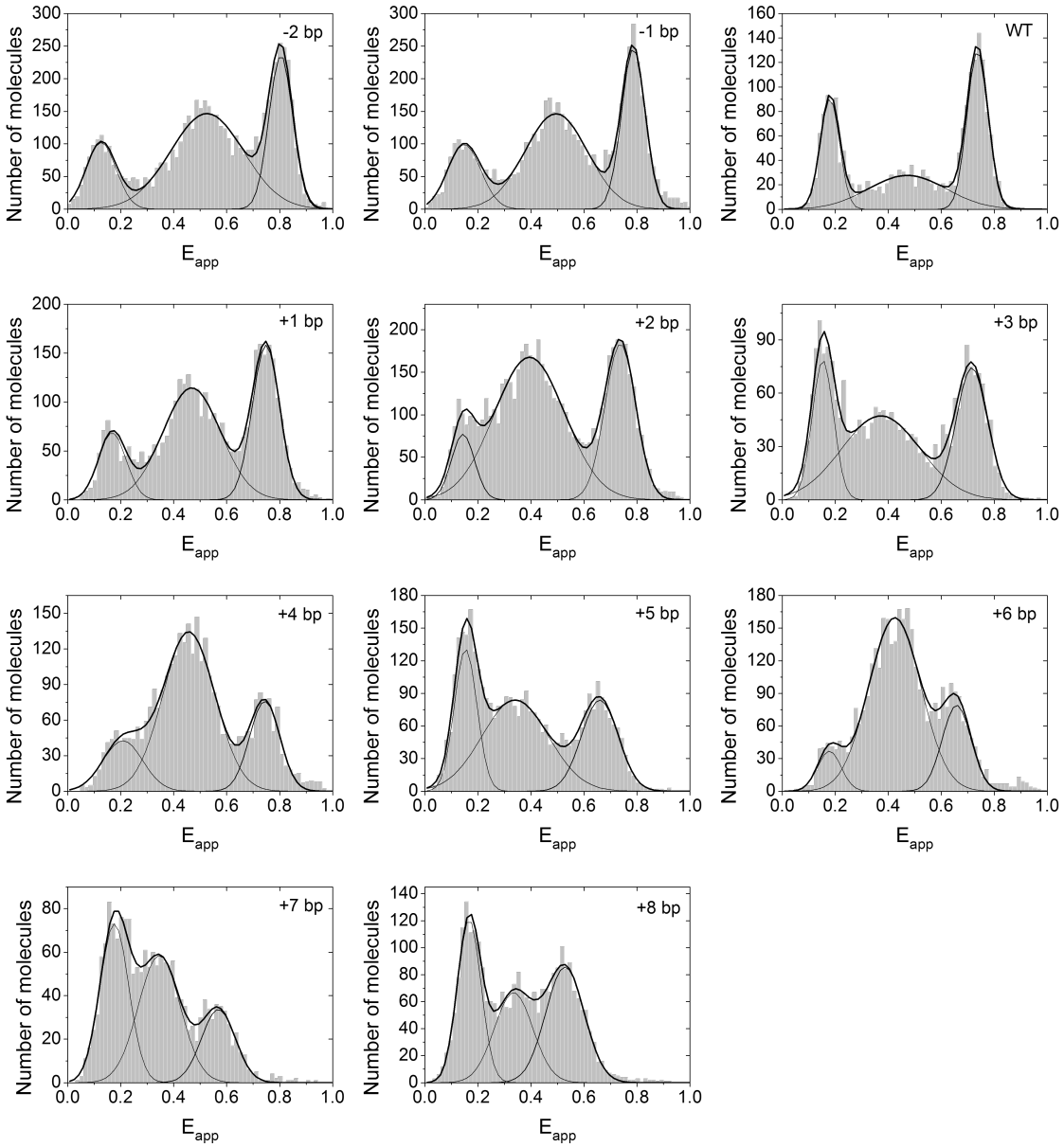


Fig. 4.8 Single-molecule histograms of apparent FRET efficiency (E_{app}) for each of the P1-elongated constructs of the modelled SAM-I aptamer. All experiments were performed at 3 mM $MgCl_2$ and $\geq$ 1 µM SAM. Histograms were fitted to a linear combination of three Gaussian distribution three peaks showing the contribution of the zero-peak (*, E_{app} ~0.15), the ligand-free folded state (F_{Mg}) and ligand-bound state (F_{NS}).

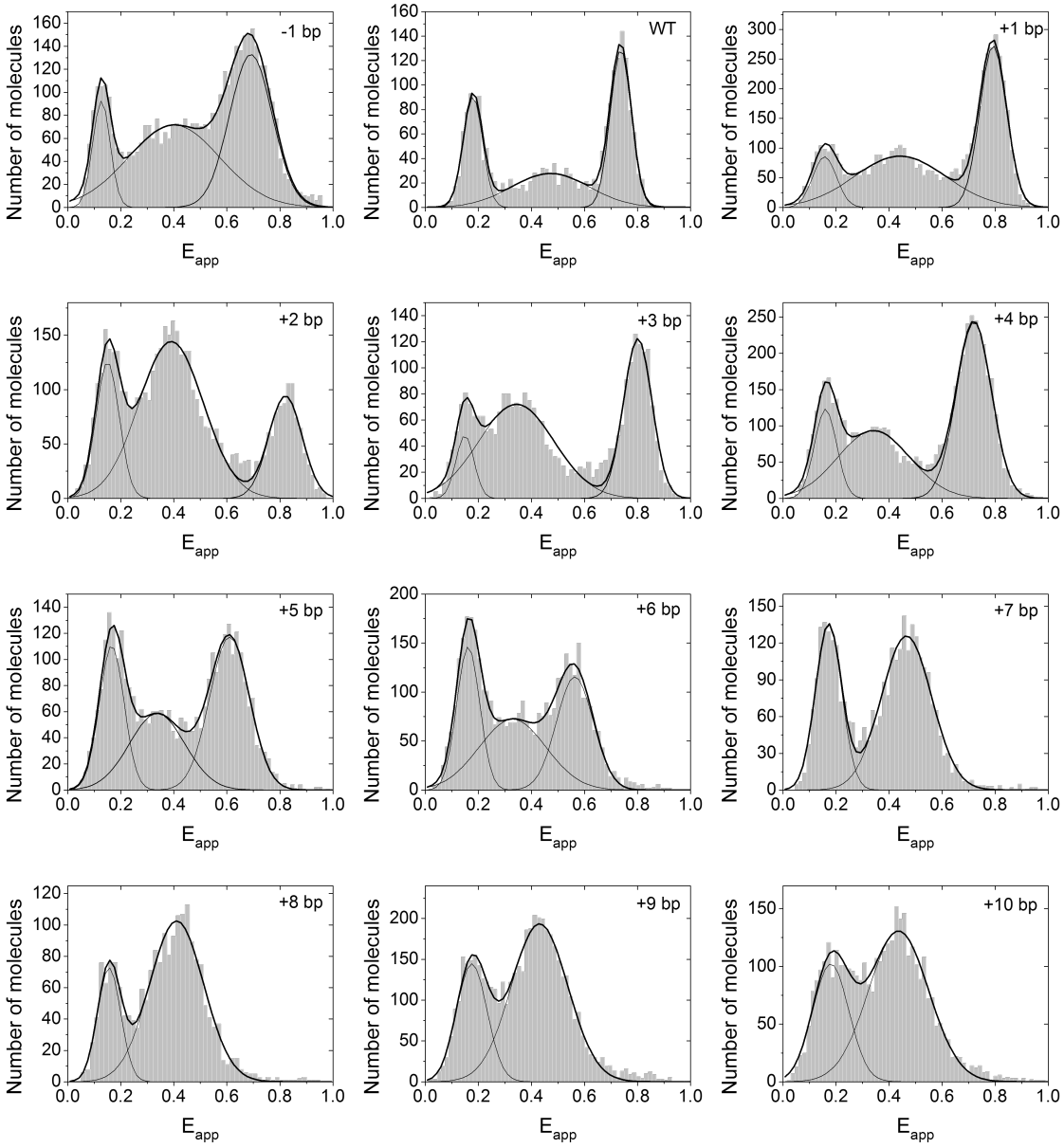


Fig. 4.9 Single-molecule histograms of apparent FRET efficiency (E_{app}) for each of the P3-elongated constructs of the modelled SAM-I aptamer. All experiments were performed at 3 mM $MgCl_2$ and $\geq$ 1 µM SAM. Histograms were fitted to a linear combination of three Gaussian distribution three peaks showing the contributions of the zero-peak (E_{app} ~0.15), the ligand-free folded state (F_{Mg}) and ligand-bound state (F_{NS}).

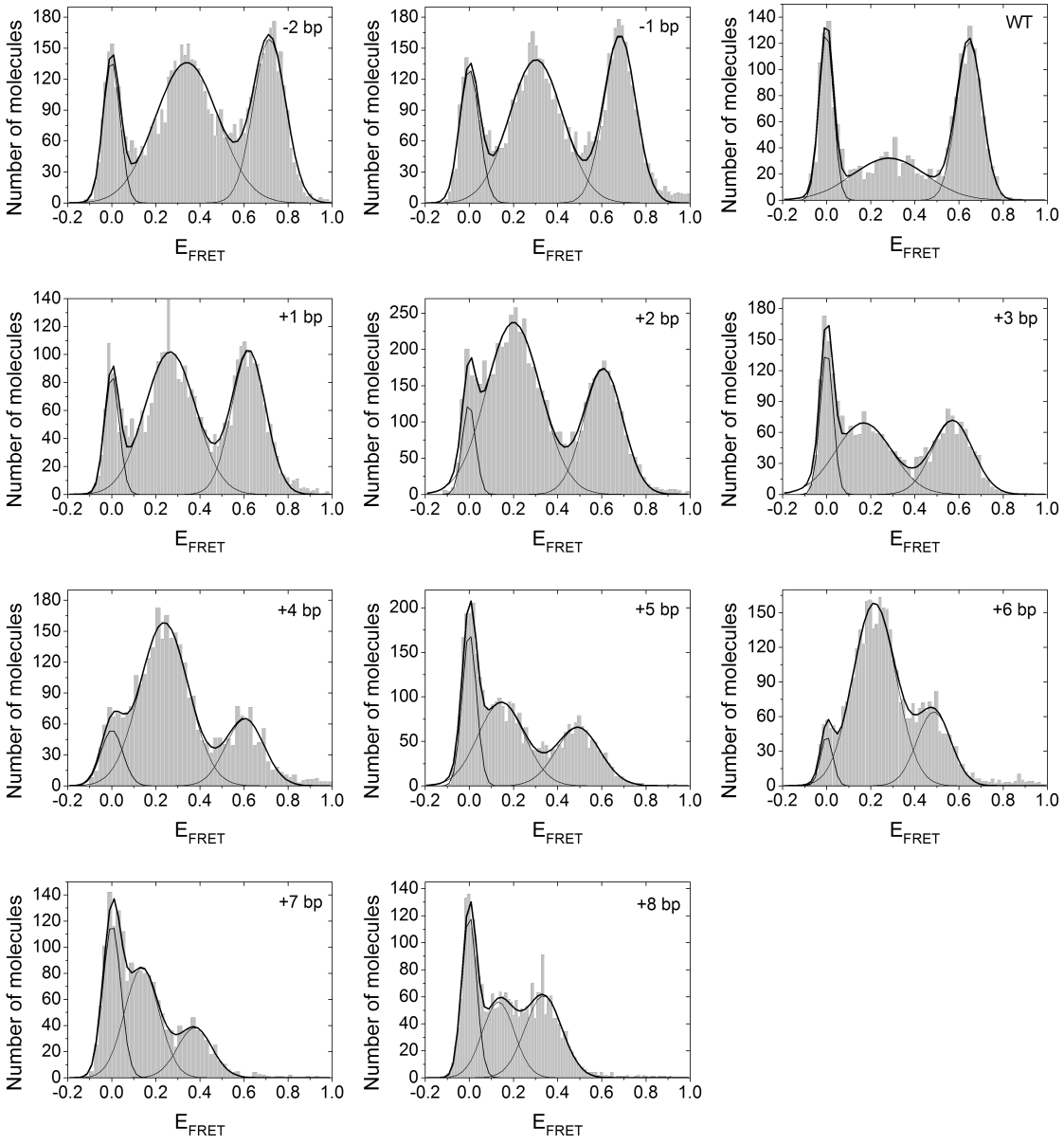


Fig. 4.10 Single-molecule histograms of corrected FRET efficiency (E_{FRET}) for the P1-elongated constructs of the modelled SAM-I aptamer. Experiments were performed in the presence of 3 mM $MgCl_2$ and ≥ 1 µM SAM. Intensities of Cy3 and Cy5 were corrected for the gamma factor (γ), zero-peak (p) and backreflection (r).

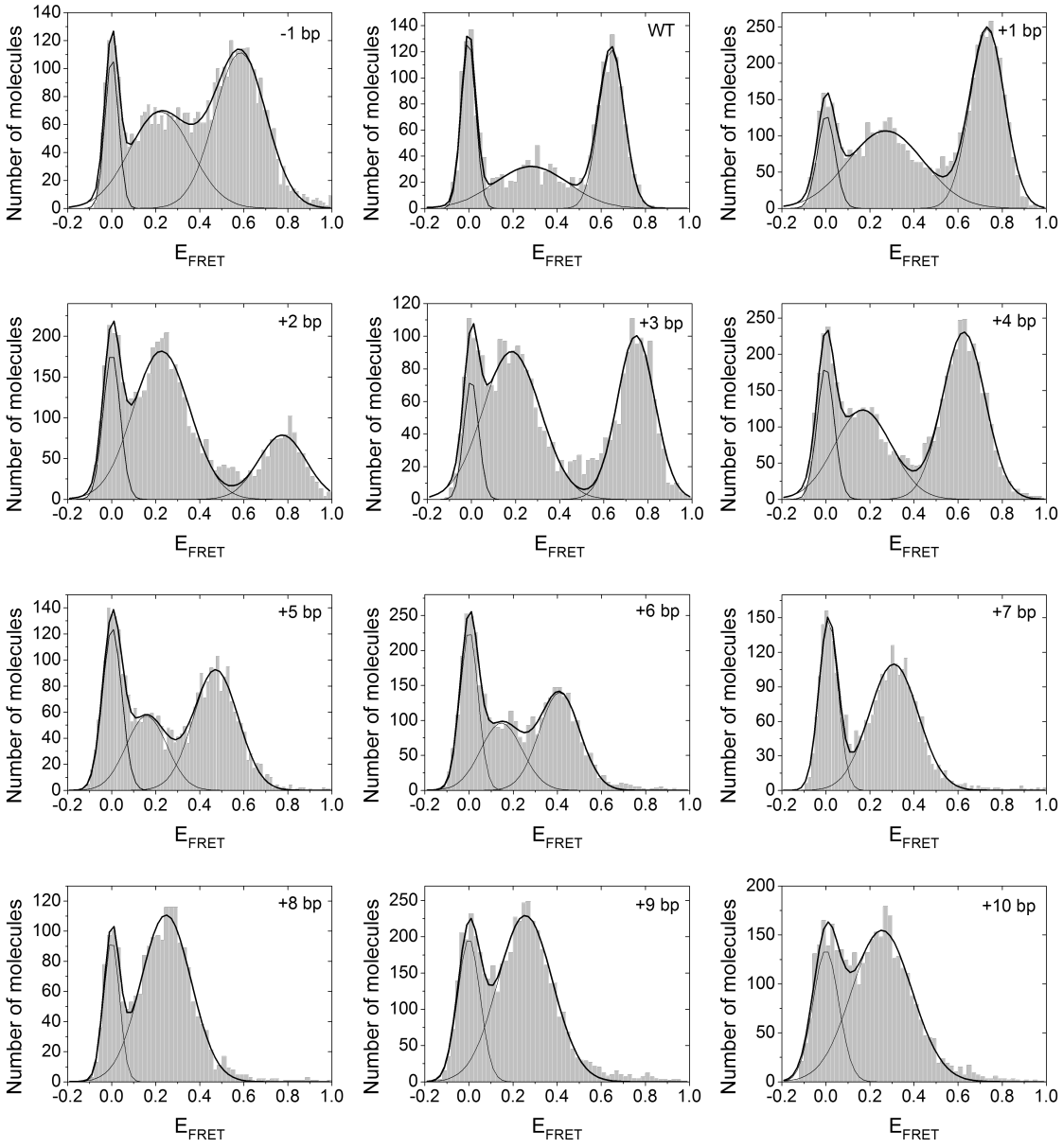


Fig. 4.11 Single-molecule histograms of corrected FRET efficiency (E_{FRET}) for the P3-elongated constructs of the modelled SAM-I aptamer. Experiments were performed in the presence of 3 mM $MgCl_2$ and ≥ 1 µM SAM. Intensities of Cy3 and Cy5 were corrected for the gamma factor (γ), zero-peak (p) and backreflection (r).

Table 4.3 Fitted E_{FRET} values obtained for the phasing of P1 and P3.

Phasing of P1			Phasing of P3		
Construct	F_{Mg}	F_{NS}	Construct	F_{Mg}	F_{NS}
-2 bp	0.340 ± 0.005	0.716 ± 0.003	-1 bp	0.220 ± 0.009	0.583 ± 0.006
-1 bp	0.302 ± 0.004	0.682 ± 0.003	WT	0.283 ± 0.017	0.646 ± 0.002
WT	0.283 ± 0.017	0.646 ± 0.002	+1 bp	0.267 ± 0.010	0.734 ± 0.002
+1 bp	0.265 ± 0.005	0.621 ± 0.004	+2 bp	0.226 ± 0.005	0.775 ± 0.008
+2 bp	0.200 ± 0.004	0.607 ± 0.004	+3 bp	0.180 ± 0.009	0.746 ± 0.004
+3 bp	0.168 ± 0.009	0.571 ± 0.005	+4 bp	0.167 ± 0.009	0.626 ± 0.002
+4 bp	0.237 ± 0.004	0.604 ± 0.007	+5 bp	0.159 ± 0.104	0.473 ± 0.004
+5 bp	0.145 ± 0.007	0.492 ± 0.006	+6 bp	0.147 ± 0.011	0.410 ± 0.008
+6 bp	0.216 ± 0.003	0.486 ± 0.007	+7 bp	0.190 ± 0.004	0.308 ± 0.003
+7 bp	0.156 ± 0.023	0.376 ± 0.012	+8 bp	0.200 ± 0.003	0.247 ± 0.003
+8 bp	0.132 ± 0.008	0.336 ± 0.011	+9 bp	0.192 ± 0.004	0.255 ± 0.003
			+10 bp	0.182 ± 0.006	0.254 ± 0.006

a linear combination of three Gaussian populations to calculate the E_{FRET} values of F_{Mg} and F_{NS} for each construct (Table 4.3).

The single-molecule data obtained in the absence of SAM (data not shown) were used to determine the position of the magnesium-induced state for each construct as this conformation represents the predominant population at 3 mM $MgCl_2$. When comparing the data with and without ligand, a consistent increase of the high-FRET state is seen in the presence of SAM and thus, allowed us to assign the mid-FRET population to the F_{Mg} and the high-FRET to the F_{NS} state for all the aptamers analysed. However, the small contribution of F_{NS} at 3 mM $MgCl_2$ could introduce some error in the determination of the FRET efficiency of both states. Therefore, the FRET efficiencies assigned to F_{Mg} and F_{NS} were calculated, where possible, from the fitting of the single-molecule histograms in the presence of SAM in which the two populations are well defined. The four exemptions are the constructs ranging from +7 to +10 bp for the phasing of P3, in which the two populations are too close to be both identified in the presence of SAM.

The E_{FRET} values of F_{Mg} and F_{NS} for each population were then represented as a function of the base pairs added in the phasing of stems P1 or P3. Both F_{Mg} and F_{NS} display a similar trend when phasing P1 with a continuous decrease in the FRET efficiency as the addition of base pairs increases the distance between donor and acceptor (Fig. 4.12a). This behaviour is analogous to the representation of the FRET efficiency as a function of the inter-dye distance in the case of anisotropic distribution of the dyes, in which $\kappa^2 = 2/3$ and R_0 is constant (Fig. 1.9b). These similarities suggest that the Cy3 terminally attached to P1 can freely rotate, as reported in previous AV simulations [Pan et al., 2014], instead of adopting a stacked position onto the end of the stem [Iqbal et al., 2008; Ouellet et al., 2011], in which a fraction of the Cy3 and Cy5 attached to the phosphate group at the 5' end of DNA and RNA duplexes display a stacked orientation onto the end of the stem as an extra base pair. However, the constructs used in this chapter incorporate an extra base at the 3' end of the P1 to which the dye is attached at the nucleobase, thus supporting the idea that the the donor does not preferentially adopt a stacked conformation.

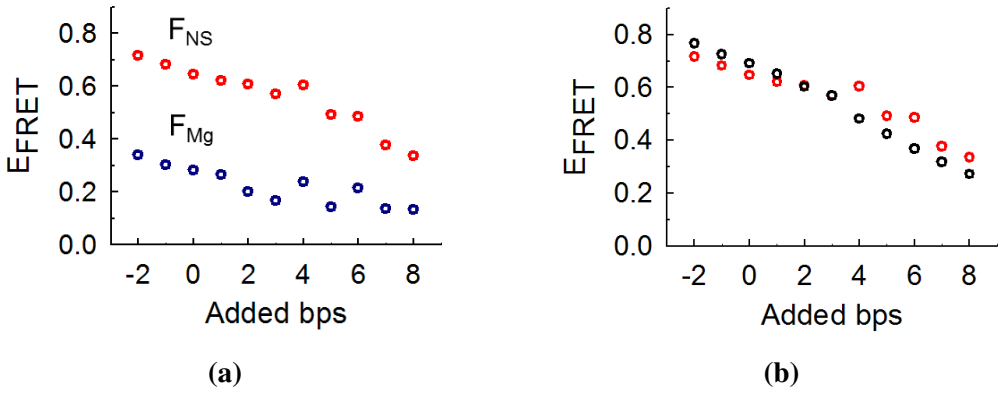


Fig. 4.12 Experimental and theoretical data obtained in P1 elongation constructs. (a) Experimental FRET values of F_{NS} (red) and F_{Mg} (blue) states for each of the constructs obtained by elongation of P1. (b) Experimental (red) and theoretical (black) E_{FRET} values of F_{NS} obtained by elongation of stem P1. The theoretical values were obtained using a R_0 of 49.0 Å.

The phasing of stem P3, in which the dye is attached at an internal position, shows a similar sinusoidal modulation for both ligand-free and ligand-bound conformations (Fig. 4.13a). The two representations exhibit a difference between the maximum (+0 bp for F_{Mg}

and +2 bp for F_{NS}) and minimum E_{FRET} (+6 bp for F_{Mg} and +8 bp for F_{NS}) of 6 base pairs, or ~200° considering an RNA A-form helix (Table 1.1). These results are comparable with the expected shift between maximum and minimum of ~180° corresponding to a model in which the modulation is defined by the helical twist of the attachment nucleobase. Furthermore, the two data sets show a noticeable phase shift between F_{Mg} and F_{NS} of ~2 bp.

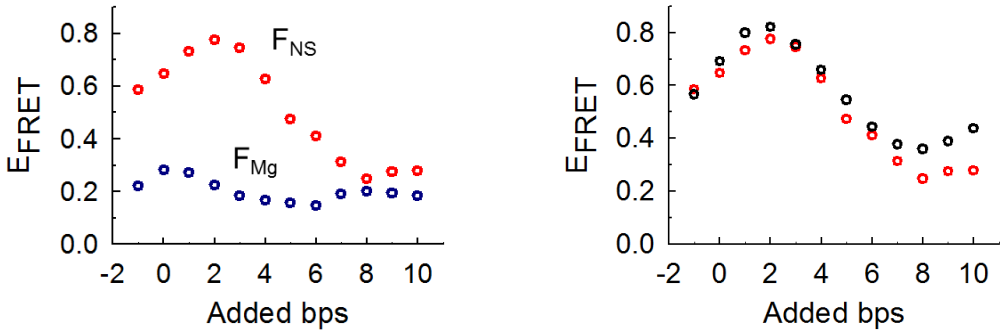


Fig. 4.13 Experimental and theoretical data obtained in P3 elongation constructs.
(a) Experimental FRET of F_{NS} (red) and F_{Mg} (blue) states in the elongation of P3.
(b) Experimental (red) and theoretical (black) E_{FRET} of F_{NS} obtained for $R_0 = 49.0$ Å.

To further characterise the folding mechanism and reorganisation of the SAM-I riboswitch in the presence of ligand, the experimental data were compared with three-dimensional models of the F_{NS} conformations. These models were engineered using the crystal structure of the SAM-I *yitJ* aptamer from *Bacillus subtilis* [Lu et al., 2010] as a template in which stems P1 and P3 are substituted with RNA A-form helices of the same sequences and lengths than those used in smFRET. The dyes were then represented as accessible volume (AV) clouds incorporated at the same specific position as for the experimental constructs.

The FRET-averaged inter-dye distance was calculated from the dye AV clouds for a constant R_0 of 49.0 Å, slightly lower than the Forster radius reported for the Cy3-Cy5 pair ranging from 54 to 61 Å [Lee et al., 2010; Murphy et al., 2004]. This small discrepancy may be due to the labelling positions of donor and acceptor in the SAM-I aptamers. The results of the modelled distances are in good agreement with the experimental data, describing well the continuous decrease for the phasing of P1 (Fig. 4.12b). The phasing of P3 displays an

analogous modulation with maximum at +2 and minimum at +8 bp (Fig. 4.13b), despite the discrepancy between +7 and +10 bp, most probably due to the error introduced by the experimental data. More precisely, due to the impossibility to identify and fit the contribution of F_{Mg} in the E_{FRET} histograms in the presence of SAM leads to the fitting of only one population assigned to the F_{NS} state and, as a result, the FRET values for the ligand-bound conformation may be shifted towards lower FRET efficiencies.

The theoretical FRET results are supported by the representation of the riboswitch models (Fig. 4.14). These show that the freedom of rotation of Cy3 at the terminal position of P1 would imply a minimum change in the FRET efficiency for the phasing of P1. On the contrary, the representation of the Cy5 shows how the dye wraps around the major groove of P3 with a preferential orientation, which changes with the helical twist of the attachment base induced by the addition of base pairs, and explains well the modulation observed in the smFRET data.

Once we have confirmed the validity of the three-dimensional models we now aim to understand, based on experimental and theoretical data, the relative reorientation of stems P1 and P3 upon ligand binding. When phasing P1, the minimal change observed in the FRET efficiencies of F_{Mg} and F_{NS} as we add more base pairs is insufficient to explain the change in FRET associated to the ligand binding, and thus the transition from F_{Mg} to F_{NS}, for a given aptamer (Fig. 4.12a). The experimental phasing of P1 leads to a 53% decrease in the E_{FRET} of F_{NS} after the addition of 10 base pairs in which the Cy3 is shifted down up to 23 Å and rotated up to 336°, but only at that point it reaches a similar FRET efficiency (+8 bp, E_{FRET} ~0.34) than the F_{Mg} for the -2bp aptamer (-2 bp, E_{FRET} ~0.34) (Fig. 4.12a). Furthermore, a change in FRET upon ligand binding associated to a rotation of P1 would imply a switch between the FRET values of F_{Mg} and F_{NS} within the range of constructs that have been studied. This situation, in which the F_{Mg} would be associated to the high-FRET state and F_{NS} to the mid-FRET state, would contradict both experimental and theoretical data.

When phasing P3, the preferential orientation of the Cy5 and the phase shift of 2 bp can indicate a rotation of P3 of ~66° upon ligand binding that would explain the change in FRET between F_{Mg} and F_{NS} (Fig. 4.13a). We have then estimated the change in the inter-dye

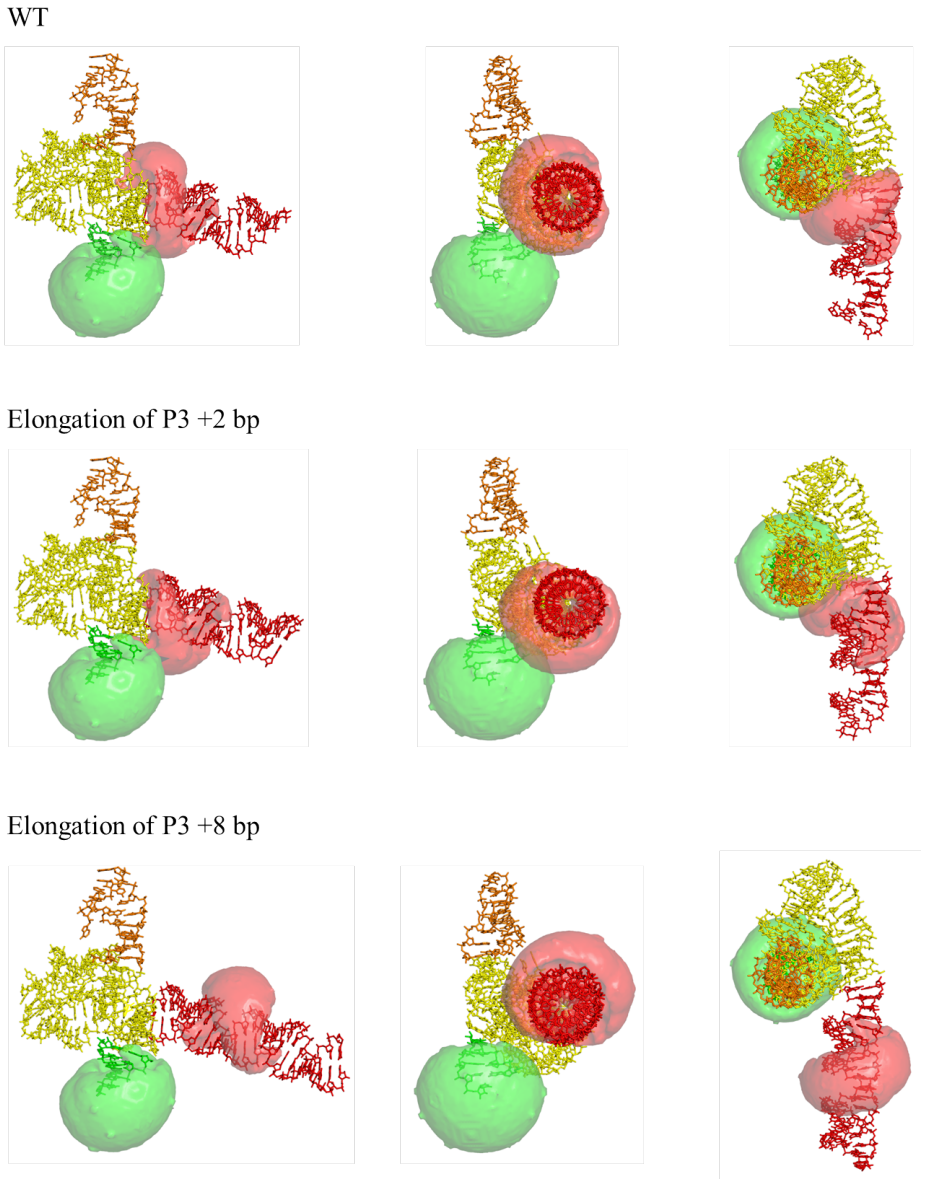


Fig. 4.14 Three-dimensional model of the *yitJ* aptamer from *Bacillus subtilis*. Visual representation of the wild type, +2 bp and +8 bp constructs when phasing P3. The core of the aptamer and stem P2 are represented in yellow, P1 in green, P3 in red and P4 in orange. The AV clouds for both dyes are represented in green (Cy3) and red (Cy5).

distance due to the rotation of P3 along its helical axis but the model is unable to explain this transition, with an estimated distance 2-fold shorter than the experimental distance change. In addition, a rotation of the P3 would mean again a switch between the FRET values of F_{Mg} and F_{NS} that was not observed neither in the experimental nor in the theoretical data.

Thus, our data suggest no rotation of P3 along its helical axis, which is in agreement with the results obtained by gel migration of a P3-bulged aptamer (Fig. 4.4b) [Heppell et al., 2011].

Since the rotation of P1 also fails to explain the change in the FRET efficiency between F_{Mg} and F_{NS}, we hypothesise that P1 moves along the plane perpendicular to the P3 axis to adopt a stacked position with stem P4. The reorientation of P1 and P4 would be translated in a decrease in the distance between Cy3 and Cy5 in the F_{NS} state. Because this rearrangement occur perpendicularly to P3, the end-to-end distance between P1 and P3 stems would not change, as reported by Heppell and coworkers [Heppell et al., 2011]. Furthermore, the phasing of P3 displays a maximum for F_{NS} at +2 bp, whilst the maximum for F_{Mg} appear at the wild type. This means that, due to the preferential orientation of the Cy5, P3 needs to incorporate 2 base pairs (~66°) to adopt a conformation in which the orientation of both dyes is similar (Fig. 4.14). Hence, this 2 bp phase shift observed in the experimental data suggest that the FRET change upon ligand binding in a given construct is due to a rotation of P1 of ~66° along the plane of rotation (Fig. 4.15).

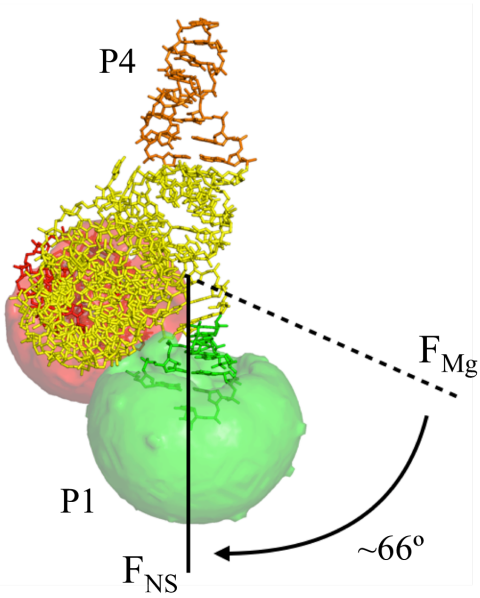


Fig. 4.15 Proposed reorientation of stem P1 upon ligand binding. The position of the stem P1 on F_{Mg} (dashed line) moves ~66° upon ligand binding to the position on the F_{NS} state (solid line) along the plane perpendicular to P3 (sheet plane). The three-dimensional representation corresponds to the +2 bp construct when phasing P3.

4.4 Conclusions

The aim of this work was to provide a closer look into the two-step folding mechanism of the SAM-I riboswitch and, more in particular, the ligand binding and stabilisation of the F_{NS} state. The crystal structure of this conformation displays a four-way junction around which the four stems are organised as two helical stacks. The addition of magnesium induces the formation of the P2-P3 and the presence of SAM induces the posterior formation of the P1-P4 stacking. However, the presence of SAM promotes not only the formation of the P1-P4 stacking, but also a rotation of stem P1 along its helical axis and rearrangement of J1/2 and J3/4 regions in order to stabilise the ligand-bound conformation. Although the rotation of stems P1 has attracted the attention of researchers for playing an important role in the control of the genetic expression, the sense and extent of its rotation remained unknown and the characterisation of the transition from F_{Mg} to F_{NS} was not completely understood.

This chapter has provided insights into the rearrangement of the four-way junction with a reorientation of stem P1 versus P3, introducing for the first time the modelling of three-dimensional aptamers engineered from the crystal structure in combination with smFRET. The experimental results obtained from the elongation and phasing of stems P1 and P3 were analysed and compared with the theoretical FRET efficiencies obtained from the modelled aptamers. The elongation and phasing of the two stems carrying the donor and acceptor was used as an indicator of the helical periodicity for the study of the rearrangement of P1 and P3 upon ligand binding. However, the experiments provided valuable information to suggest an anisotropic orientation of the donor and hypothesise about a potential reorientation of P1 along the plane perpendicular to P3 that prior assays of gel electrophoresis, ensemble-FRET and smFRET were unable to identify. The data presented here demonstrate that smFRET is not only a powerful tool to provide information in real time and discriminate between several populations in an ensemble of molecules, but also has an immense potential in combination with different methods to provide a new approach for the study of RNA riboswitches.

However, further studies need to be done to corroborate the observations presented in this chapter. Such studies would include similar experiments in which the two dyes are

internally and terminally attached to P1 and P3, respectively. These assays would provide more information regarding not only the reorientation of P1 versus P3 due to the formation of the P1-P4 stacking but also regarding the helical rotation of P1 along its helical axis. The main limitation that this method presents is economic, since the need to purchase synthetic strands for 22 different constructs that incorporate functional groups at specific positions for posterior labelling may result in expensive experiments.

References

Batey, R. T. (2011). Recognition of S-adenosylmethionine by riboswitches. *Wiley Interdisciplinary Reviews: RNA*, 2(2):299–311.

Blouin, S., Mulhbacher, J., Penedo, J. C., and Lafontaine, D. A. (2009). Riboswitches: Ancient and promising genetic regulators. *ChemBioChem*, 10(3):400–416.

Blount, K. F. and Breaker, R. R. (2006). Riboswitches as antibacterial drug targets. *Nature Biotechnology*, 24(12):1558–1564.

Boudreault, J., Perez-Gonzalez, D. C., Penedo, J. C., and Lafontaine, D. A. (2015). *Single-Molecule Approaches for the Characterization of Riboswitch Folding Mechanisms*, pages 101–107. Springer New York, New York, NY.

Brenner, M. D., Scanlan, M. S., Nahas, M. K., Ha, T., and Silverman, S. K. (2010). Multivector fluorescence analysis of the xpt guanine riboswitch aptamer domain and the conformational role of guanine. *Biochemistry*, 49(8):1596–1605. PMID: 20108980.

Eschbach, S. H., St-Pierre, P., Penedo, J. C., and Lafontaine, D. A. (2012). Folding of the SAM-I riboswitch. *RNA Biology*, 9(5):535–541.

Heppell, B., Blouin, S., Dussault, A.-M., Mulhbacher, J., Ennifar, E., Penedo, J. C., and Lafontaine, D. A. (2011). Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nature Chemical Biology*, 7(6):384–392.

Heppell, B. and Lafontaine, D. A. (2008). Folding of the SAM aptamer is determined by the formation of a K-turn-dependent pseudoknot. *Biochemistry*, 47(6):1490–1499. PMID: 18205390.

References

Iqbal, A., Arslan, S., Okumus, B., Wilson, T. J., Giraud, G., Norman, D. G., Ha, T., and Lilley, D. M. J. (2008). Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proceedings of the National Academy of Sciences*, 105(32):11176–11181.

Kalinin, S., Peulen, T., Sindbert, S., Rothwell, P. J., Berger, S., Restle, T., Goody, R. S., Gohlke, H., and Seidel, C. A. M. (2012). A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature Methods*, 9(12):1218–1225.

Lee, S., Lee, J., and Hohng, S. (2010). Single-molecule three-color FRET with both negligible spectral overlap and long observation time. *PLOS ONE*, 5(8):1–9.

Lu, C., Ding, F., Chowdhury, A., Pradhan, V., Tomsic, J., Holmes, W. M., Henkin, T. M., and Ke, A. (2010). SAM recognition and conformational switching mechanism in the Bacillus subtilis yitJ S Box/SAM-I riboswitch. *Journal of Molecular Biology*, 404(5):803 – 818.

McCluskey, K., Shaw, E., Lafontaine, D. A., and Penedo, J. C. (2014). *Single-Molecule Fluorescence of Nucleic Acids*, pages 759–791. Humana Press, Totowa, NJ.

Montange, R. K. and Batey, R. T. (2006). Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature*, 441(7097):1172–1175.

Murphy, M. C., Rasnik, I., Cheng, W., Lohman, T. M., and Ha, T. (2004). Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophysical Journal*, 86(4):2530–2537.

Ouellet, J., Schorr, S., Iqbal, A., Wilson, T., and Lilley, D. J. (2011). Orientation of cyanine fluorophores terminally attached to DNA via long, flexible tethers. *Biophysical Journal*, 101(5):1148–1154.

Pan, K., Boulais, E., Yang, L., and Bathe, M. (2014). Structure-based model for light-harvesting properties of nucleic acid nanostructures. *Nucleic Acids Research*, 42(4):2159–2170.

Perez-Gonzalez, C., Grondin, J. P., Lafontaine, D. A., and Carlos Penedo, J. (2016). *Biophysics of Infection*, chapter Biophysical Approaches to Bacterial Gene Regulation by Riboswitches, pages 157–191. Springer International Publishing, Cham.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612.

Sindbert, S., Kalinin, S., Nguyen, H., Kienzler, A., Clima, L., Bannwarth, W., Appel, B., Müller, S., and Seidel, C. A. M. (2011). Accurate distance determination of nucleic acids via Förster resonance energy transfer:

implications of dye linker length and rigidity. *Journal of the American Chemical Society*, 133(8):2463–2480. PMID: 21291253.

Winkler, W. C. and Breaker, R. R. (2005). Regulation of bacterial gene expression by riboswitches. *Annual Review of Microbiology*, 59:487–517.

Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003). An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature Structure & Molecular Biology*, 10(9):701–707.

5

Single-molecule FRET to investigate the folding of the adenine riboswitch through an intermediate state

5.1 Introduction

The purine class of riboswitches represents one of the smallest metabolite-binding domains known to date. They comprise two different types of riboswitches, sensing adenine and guanine, that present a high similarity between them. The aptamer of purine riboswitches is composed of three helices, named P1, P2 and P3, organised around a three-way junction with conserved sequences on the stem junctions and at the P2 and P3 stem loops, named L2 and L3 (Fig. 5.1a) [Batey, 2012; Mandal and Breaker, 2004b].

Crystallographic structures of *xpt-pbuX* guanine riboswitch from *Bacillus subtilis* and adenosine deaminase (*add*) riboswitch from *Vibrio vulnificus* show an aptamer organised around a three-way junction with a parallel arrangement of stems P2 and P3 and a coaxial stacking between P1 and P3 (Fig. 5.1b) [Serganov et al., 2004]. The crystal structures also show a L2-L3 loop-loop interaction through hydrogen bonding between two base pairs at the conserved sequences of each loop. The ligand is located at the core of the three-way junction where it establishes specific hydrogen bonds with four conserved nucleotides, forming a tight pocket where 98% of the ligand is inaccessible to the solvent [Batey et al., 2004].

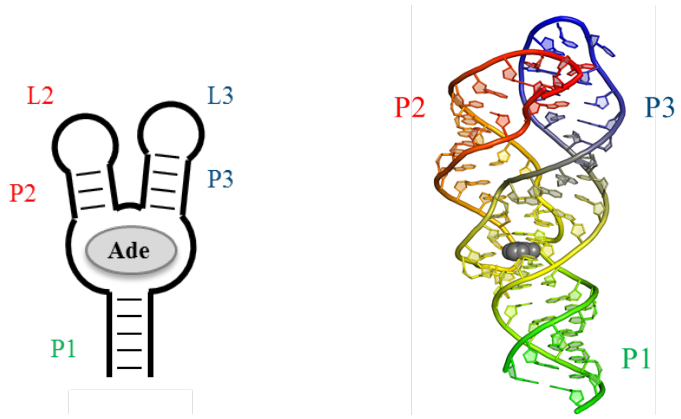


Fig. 5.1 Structure of the adenine aptamer. (a) Schematics of the ligand-bound folded conformation of the adenine aptamer. (b) Crystal structure of the *add* aptamer from *Vibrio Vulnificus*. Ligand is represented in gray (PDB ID: 1Y26) [Serganov et al., 2004].

Interestingly, a single C-to-U mutation at an specific position of the binding site switches the binding preference from guanine to adenine with all the other hydrogen bonds unaltered [Mandal and Breaker, 2004a].

Guanine and adenine riboswitches show high affinity and specificity for their respective ligand. The guanine-binding *xpt* aptamer from *B. subtilis* presents an apparent K_d of ~5 nM with 100,000-fold discrimination against adenine, whereas the adenine-binding *pubE* aptamer from *B. subtilis* presents an apparent K_d of ~300 nM with a ~30-fold discrimination against guanine, although similar affinity for the adenine analogue 2-aminopurine (2AP) as determined by in-line probing [Mandal and Breaker, 2004a].

The *pbuE* adenine riboswitch from *Bacillus subtilis* as well as the *add* from *Vibrio Vulnificus* are a rare case of positively-controlled riboswitch where the absence of ligand promotes the formation of a terminator stem between sequences from P1 and P2 that leads to transcription termination [Serganov et al., 2004]. In the presence of adenine, the aptamer folds into a conformation that favours the formation of the antiterminator stem P1 and allows the transcription to continue (Fig. 5.2).

Wickiser and coworkers have taken advantage of the properties of 2AP as a fluorescent analogue of the adenine ligand to study the binding affinity and folding kinetics of the *pbuE* aptamer [Wickiser et al., 2005]. The fluorescence signal of 2AP free in solution is

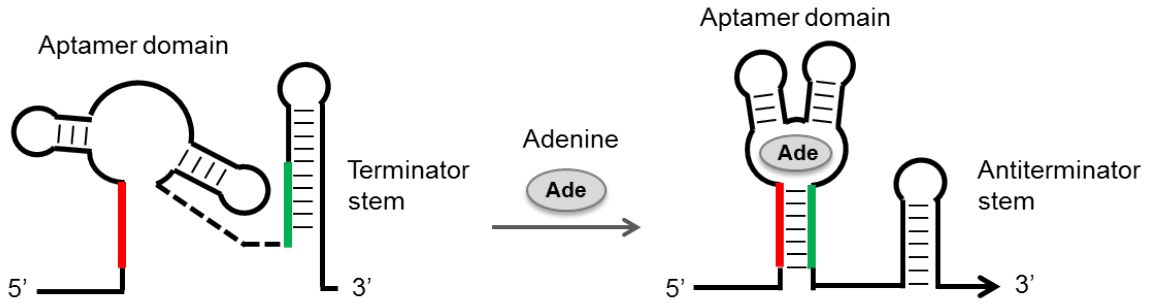


Fig. 5.2 Schematics of the unfolded and folded conformations of the positively-controlled adenine riboswitch. In the absence of ligand the expression platform adopts a repressor terminator stem, called terminator stem, that leads to premature transcription termination. The addition of adenine promotes the formation of a secondary structure, called antiterminator stem, between sequences from the aptamer (switching sequence) and from the expression platform that allows the transcription to continue. Switching sequence forming part of two alternative structures is shown in green. Adapted from [Rieder et al., 2007].

dramatically quenched when bound at the intricate binding pocket of the aptamer, what can be used to monitor the ligand binding to the aptamer. The 2AP affinity was studied using ensemble fluorescence techniques, where the fluorescence signal of 50 nM 2AP was monitored for increasing concentrations of RNA until it is almost totally quenched at 5 µM RNA, reporting an apparent K_d of 764 nM at 25°C for the 2AP-RNA complex formation. On the other hand, the binding kinetics were studied by stopped-flow experiments in excess of 2AP and in excess of RNA. The decay curves of the 2AP signal were fitted to mono-exponential functions indicating that the ligand binding kinetics might occur between two states with no intermediates involved [Wickiser et al., 2005].

Further studies have monitored the folding dynamics and ligand affinity using 2AP, ensemble-FRET and smFRET techniques [Lemay et al., 2006; Rieder et al., 2007]. Ensemble studies by Lemay and coworkers using 2AP are in good agreement with previous in-line probing and ensemble assays, reporting an apparent K_d of 354 nM at 25 °C [Lemay et al., 2006]. The use of 2AP as a binding sensor was extended to study the importance of the loop-loop interaction in the folding mechanism of aptamers with mutations on L2 and L3. The fluorescence signal of 2AP show limited quenching when titrated with these mutants, suggesting that the ligand binding is abolished when the loop-loop interaction is destabilised

and that this interaction not only stabilises the folded conformation but also promotes the formation of the ligand binding site. Further studies by Rieder and coworkers have monitored the folding and ligand binding of the *pbuE* and *add* aptamer by incorporating 2AP substitutions at different positions of the aptamer and the full-length sequences of the riboswitch [Rieder et al., 2007]. The results have shown that the shorter terminator stem of the *add* aptamer allows a certain organisation of the aptamer even when the terminator is formed. However, the aptamer of the *pbuE* remains unfolded when its longer terminator stem, which not only comprises a sequence from P1 but also from P3, is formed.

Ensemble-FRET and smFRET experiments were performed for the *add* adenine riboswitch from *V. Vulnificus* carrying donor and acceptor dyes at the end of the L2 and L3 stem loops to monitor the aptamer folding and formation of the loop-loop interaction at different Mg^{2+} and adenine concentrations [Lemay et al., 2006]. Ensemble-FRET assays were performed at increasing concentrations of Mg^{2+} in the presence and absence of adenine. The data in the absence of adenine show an increase in the FRET efficiency associated with the folding and formation of the loop-loop interaction, whilst the addition of ligand decreases the concentration of Mg^{2+} needed for the folding of the aptamer by 2-fold, suggesting that this is assisted by the ligand. Further smFRET assays have used chemical denaturation of the aptamer to characterise the folding and stabilisation of the ligand-free and ligand-bound folded states, otherwise indistinguishible by FRET [Dalgarno et al., 2013]. They have monitored the undocking rates of both states in the presence of urea reporting a ~50-fold stabilisation of the ligand-bound versus the ligand-free folded state.

SmFRET also displays a transition from a low-FRET state to a high-FRET state corresponding to the transition from the unfolded to the folded state. Interestingly, at low Mg^{2+} concentrations (below 50 µM) a third state emerges with mid-FRET corresponding to an intermediate conformation that was hidden in the 2AP and ensemble-FRET assays [Lemay et al., 2006]. Gilbert and Batey have proposed a structure for the intermediate state corresponding to the formation of the helical stacking between P1 and P3 in the absence of the L2-L3 interaction (Fig. 5.3) [Gilbert and Batey, 2006]. The addition of Mg^{2+} (or Na^{+}) would promote the loop-loop interaction with a rearrangement of P2 parallel to P3.

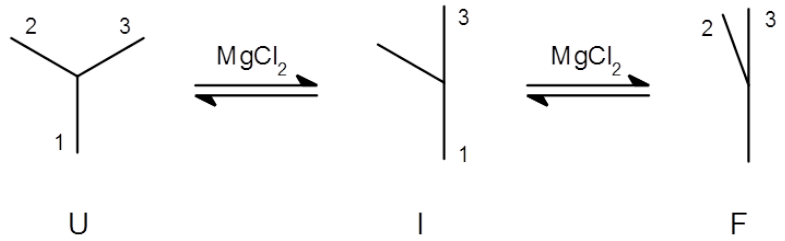


Fig. 5.3 Proposed folding mechanism of the adenine aptamer. Model proposed by Gilbert and Batey [Gilbert and Batey, 2006] based on the study of Lemay and coworkers [Lemay et al., 2006]. The intermediate corresponds to the stacking of P1 and P3, whereas the folded conformation is due to the loop-loop interaction.

The folding mechanism of the adenine aptamer, however, has been proposed using information from a single vector distance. Thus, in order to fully assign the structural changes that lead to the formation of the intermediate, it is necessary to collect information regarding the conformational changes associated to the relative position of P1-P2 and P1-P3. In collaboration with Daniel Lafontaine's group from the Université de Sherbrooke, we have used smFRET to characterise the intermediate state for the vector 1-2 of the aptamer by incorporating the donor and acceptor on stem P1 and stem loop L2, respectively. Importantly, previous smFRET studies carried out by our collaborators have determined a minimum change in the distance between P1 and P3 with a transition from E_{app} ~0.64 to 0.68 with the addition of Mg^{2+} ions and no further change upon addition of ligand. Given that P1-P3 and P2-P3 are not influenced by ligand binding, we restricted our studies to the influence of mono- and divalent metal ions. Their different efficiency in stabilising the P1-P3 stacking will allow us to characterise the folding mechanism of the adenine aptamer.

5.2 Materials and methods

5.2.1 Preparation of fluorescently-labelled aptamer

The adenine-sensing *add* aptamer from *Vibrio Vulnificus* was assembled by ligation of two synthetic RNA strands with T4 RNA ligase and purified as described in [Lemay et al., 2006]. Both strands contain a nucleobase modified with a C6 amino linker at the desired position

for posterior labelling of the RNA strands with NHS ester Cy3 and Cy5 dyes at the P3-P1 and P1-P2 sequences, respectively. This work was carried out by Patrick St-Pierre, from the Lafontaine lab. The sequences used for the ligation of the aptamer are detailed in Table 5.1.

Table 5.1 Synthetic strands use in the ligation of the *add* adenine aptamer from *Vibrio Vulnificus*. Labelling positions are underlined.

Strand	Sequence (from 5' to 3')
$P1-P2$	Biotin-GCGCGAGCGUUGUAUAAUCCUAAUGAUAUGGUUUGGGAGU
$P3-P1$	UUCUACCAAGAGCCUUAAACUCUUGAUUACAAUGCUCGCGC

5.2.2 Single-molecule FRET experiments

Slides were passivated with 1 mg/mL biotinylated BSA and 0.2 mg/mL neutravidin. Fluorescently-labelled aptamer was added in a concentration between 50 and 250 pM. Magnesium and sodium were added to an imaging buffer containing 50 mM Tris-HCl (pH 7.5), 2.5 mM PCA, 250 nM PCD and 1 mM TROLOX.

The smFRET data were collected and analysed as described in section 2.5.3. Traces suspected to present transitions between three states were analysed with automated hidden Markov modelling of the experimental trajectories in order to avoid a biased interpretation [McKinney et al., 2006]. The apparent FRET efficiency was calculated after background correction by $E_{app}=I_A/(I_A+I_D)$, where I_D and I_A represent the fluorescence intensities of donor (Cy3) and acceptor (Cy5), respectively. Single-molecule histograms were obtained by averaging of the first 10 frames of each molecule after manual analysis of the traces. The population histograms were fitted to a linear combination of Gaussian distributions in order to identify the different populations involved in the folding of the aptamer.

The contributions of the folded (F) and unfolded (U) populations present in the magnesium-assisted folding were fitted to a Hill equation

$$F = F_0 + \frac{[Mg^{2+}]^n}{[Mg^{2+}]^n + [K_d]^n} \quad or \quad U = U_0 + \frac{[K_d]^n}{[Mg^{2+}]^n + [K_d]^n} \tag{5.1}$$

where the K_d is the dissociation constant and n corresponds to the Hill coefficient, which describes the cooperativity of the process. If the binding of one ligand molecule enhances the affinity of the others, then $n > 1$. If the affinity of other molecules decreases, then $n < 1$. Moreover, a value of $n = 1$ suggest either a 1:1 stoichiometry or an independent mechanism of association of other molecules.

5.3 Results and discussion

5.3.1 Magnesium-induced formation of the loop-loop interaction

We have used single-molecule FRET to study the folding of the P1-P2 vector in order to fully characterise the folding pathway of the *add* aptamer and, in particular, the nature of the intermediate state (I).

A model of the construct built from the crystal structure representing the folded state of the *add* aptamer displays the Cy3 and Cy5 located at opposite ends of the aptamer (Fig. 5.4), suggesting a transition from higher FRET efficiencies in the unfolded state to lower E_{app} values as the loop-loop interaction stabilises the folded state.

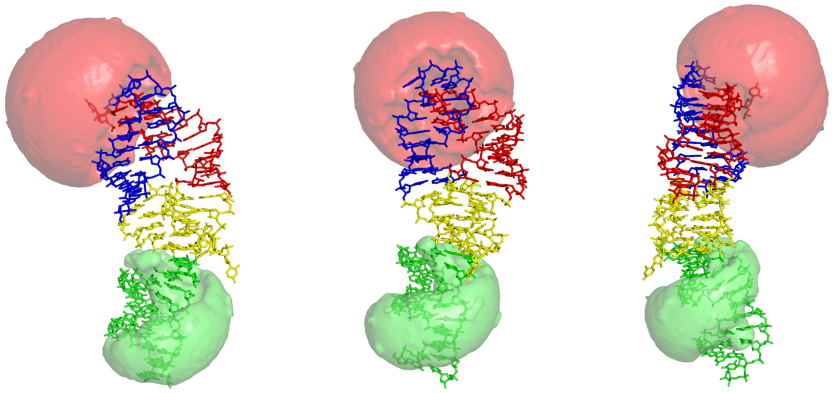


Fig. 5.4 Three-dimensional model of the *add* aptamer from *Vibrio Vulnificus*. Representations of the folded conformation of the labelled *add* aptamer used in the smFRET studies. The core of the aptamer is represented in yellow, stem P1 in green, P2 in red and P3 in blue. Cy3 and Cy5 are represented as green and red AV clouds, respectively (PDB ID: 1Y26) [Serganov et al., 2004].

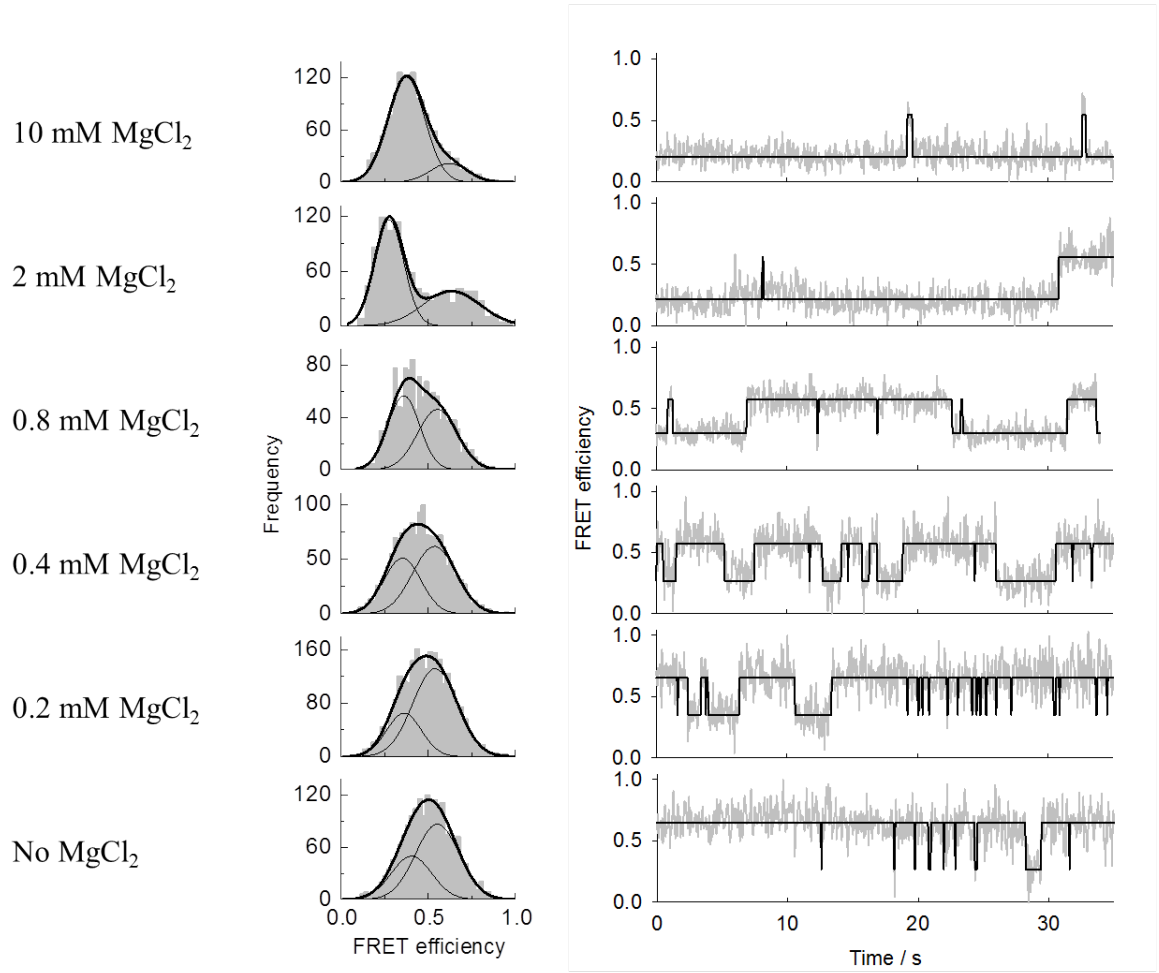


Fig. 5.5 Single-molecule histograms of FRET efficiency and representative trajectories for the *add* aptamer as a function of Mg^{2+} concentration. Experiments were performed with a background of 10 mM NaCl. Histograms were fitted to a linear combination of two Gaussian distributions representing the contributions of the folded (E_{app} ~0.35) and mixture of unfolded and intermediate state (E_{app} ~0.55). Single-molecule traces (gray) were analysed using a hidden Markov modelling (black line). The trajectories show the interconversion between the two populations and progressive stabilisation of the F state.

We first analysed the adenine aptamer in the presence of 10 mM NaCl. We observe two Gaussian distributions representing two FRET populations. A major component centered at ~0.55 and a minor population centered at ~0.35 (Fig. 5.5). The addition of Mg^{2+} from 200 µM, which promotes the loop-loop interaction, leads to an increase in the contribution of the low-FRET state, reaching a maximum of a 85% at 10 mM $MgCl_2$. Therefore, this low-FRET population may be associated to the folded state, whereas the mid-FRET state corresponds to the unfolded conformation. The contributions of both populations were fitted to a Hill equation, reporting K_d for Mg^{2+} of 0.86 mM (Fig. 5.6). Moreover, the amount of Mg^{2+} required for saturation is ~2 mM, which is in good agreement with the *in vivo* cellular conditions.

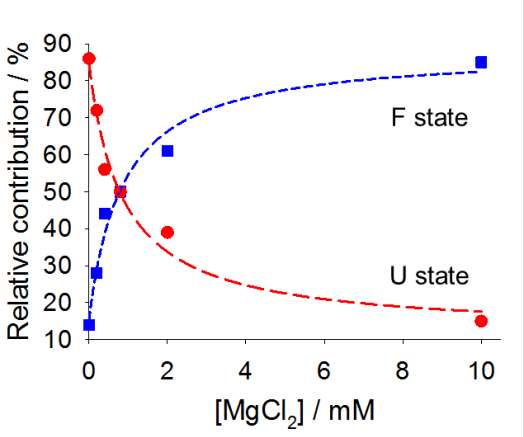


Fig. 5.6 Magnesium-induced folding of the adenine aptamer. Hill fits to the relative contribution of the folded (blue) and unfolded (red) states as a function of magnesium, reporting a K_d =0.86 mM Mg^{2+} for n=1.

Single-molecule trajectories show a stable mid-FRET state with fast transitions from the folded to the unfolded state at 10 mM NaCl and in the presence of 200 µM $MgCl_2$, reflecting the instability of the loop-loop interaction at these conditions (Fig. 5.5). As we increase the concentration of Mg^{2+}, the trajectories display a progressive stabilisation of the low-FRET state with fast transitions from the unfolded to the folded conformation.

It is important to conclude here that neither the population histograms nor the trajectories show three different states. If the unfolded state was a mixture between two states, we should see a progressive shift in the center as the concentration of Mg^{2+} increases. Moreover, the

FRET trajectories fluctuate between two states with identical FRET values across the entire titration range.

5.3.2 Sodium-induced formation of the loop-loop interaction

The folding of the P1-P2 vector in the presence of magnesium shows two clear states where the distance between both stems gets higher as we increase the concentration of divalent cations. This is consistent with a lengthening of the P1-P2 distance from the unfolded state, when the stems display a "Y-shape" to minimise electrostatic repulsions [Dalgarno et al., 2013], to an state similar to that shown in the crystal structure, where the L2-L3 loop-loop interaction is formed and involving a long P1-P2 distance. The crystal structure also shows that stems P1 and P3 are stacked, and therefore it is necessary to resolve whether stacking occur before or after the loop-loop interaction to elucidate the nature of the intermediate state.

Given the information available from P2-P3, where an intermediate was observed, and the minimum change in FRET observed for the P1-P3, we speculate that the observed lack of intermediate in P1-P2 could result from the fact that is a low populated state and Mg^{2+} induces sufficient formation of F via I, even at very low $MgCl_2$ concentrations. For instance, in vector P2-P3 the formation of the I state was only observed at ~50 µM Mg^{2+} [Lemay et al., 2006].

It is known that Na^+ ions are not efficient in stabilising stacking conformations as reported by Hohng *et al.* in their work on 4-way helical junctions [Hohng et al., 2004]. Thus, given that stacking might be involved at some stage during the folding but it is so efficiently stabilised by Mg^{2+} that we are unable to discern this interaction, we decided to carry out a folding titration of vector 1-2 with monovalent Na^+ cations, where the stacking conformation might be not as stable as to promote the quick jump to the loop-loop as observed for Mg^{2+} only.

The single-molecule histograms in the absence of Na^+ display two main peaks centered at ~0.35 and ~0.55, being the latter the main population under these conditions (Fig. 5.7). As the concentration of cations increases, this mid-FRET population decreases. Moreover,

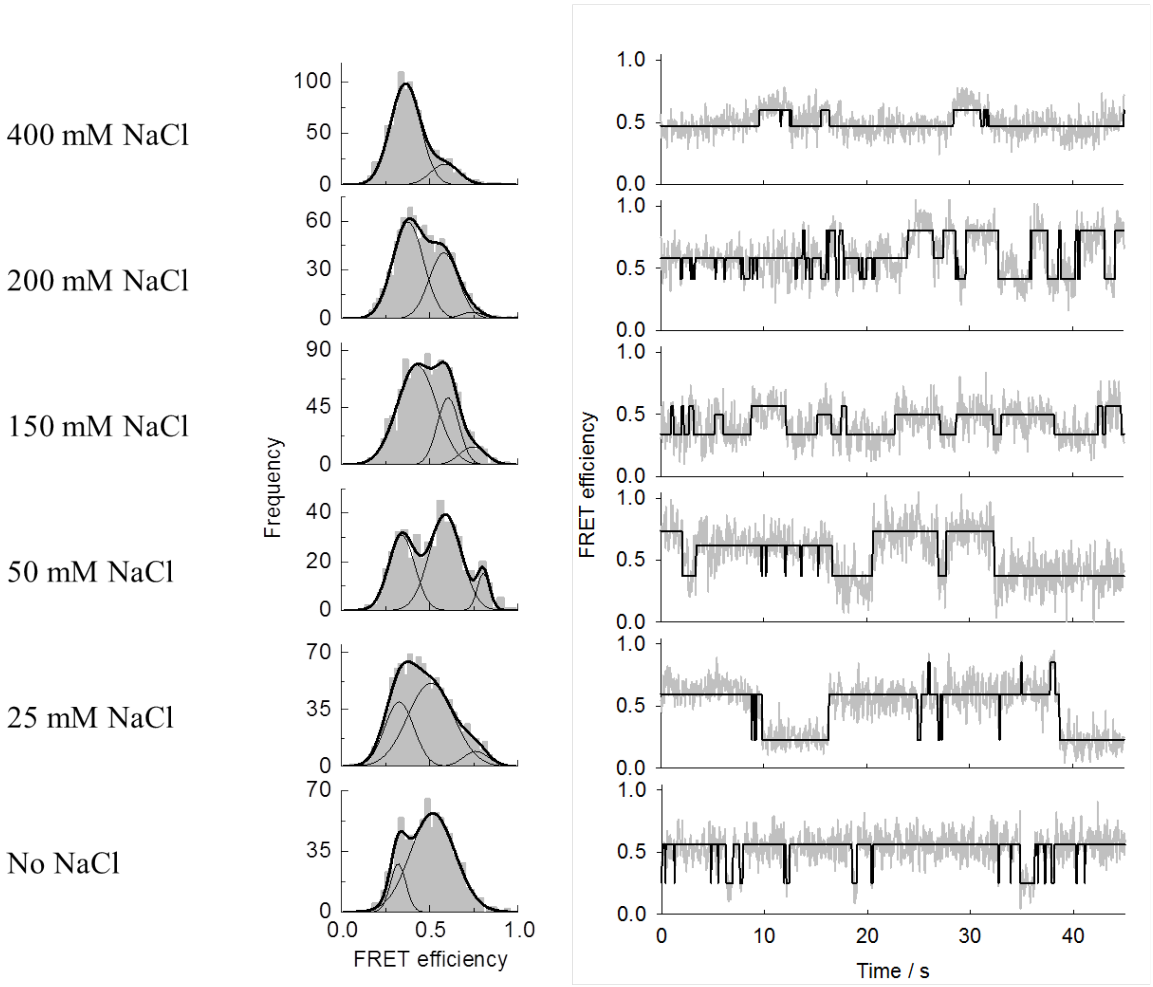
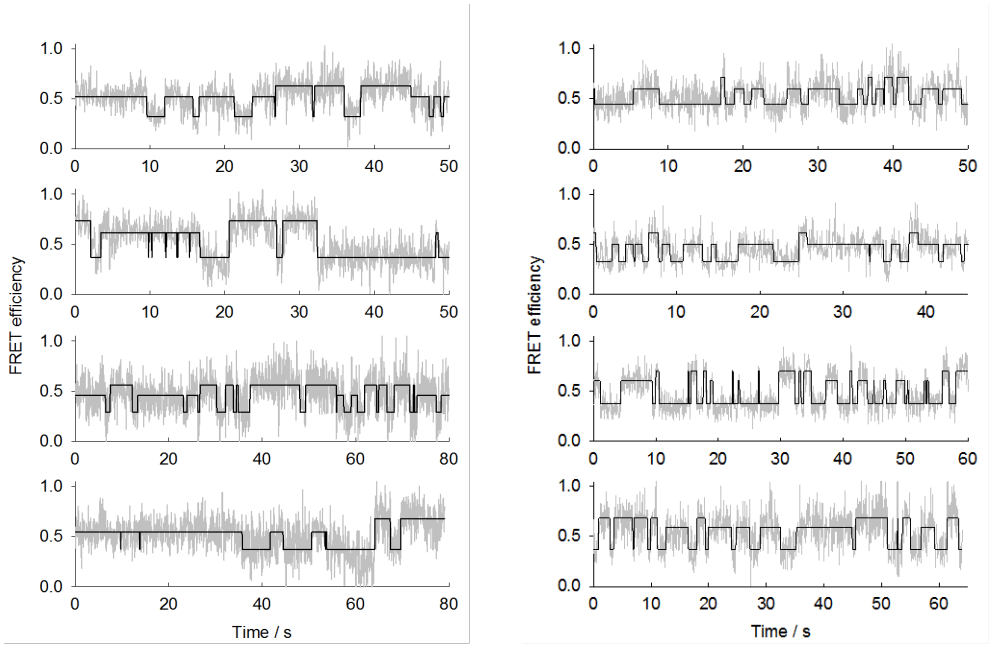


Fig. 5.7 Single-molecule histograms of FRET efficiency and representative trajectories for the *add* aptamer as a function of Na^+ concentration. Histograms were fitted to a linear combination of two or three Gaussian distributions representing the contributions of the folded (F, E_{app} ~0.35), intermediate (I, E_{app} ~0.55 and unfolded state (U, E_{app} ~0.75). Single-molecule traces (gray) were analysed using a hidden Markov modelling (black line). The trajectories show the interconversion between the three populations and progressive folding for increasing concentrations of NaCl.

the histograms in the 25 to 200 mM NaCl range display a third state with E_{app} ~0.75. This high-FRET peak reaches its maximum at 50 mM NaCl, although its contribution is below the 10% in all cases. This state finally disappears at 400 mM NaCl, at which the main population corresponds to the low-FRET state.

The smFRET trajectories display a stable mid-FRET state with fast transitions from mid- to low-FRET in the absence of Na^+ (Fig. 5.7). The addition of NaCl stabilises the P1-P3 stacking and loop-loop interaction, and thus the low-FRET lives for longer periods. The third state showing high-FRET (E_{app} ~0.75) arises at 25 mM NaCl with fast transitions to the mid- and low-FRET states. As the concentration of salt increases, the high-FRET state becomes more stable and the mid-FRET displays a shorter lifetime.

Representative smFRET trajectories collected at 50 mM NaCl, at which the contribution of the high-FRET state reaches its maximum, show the presence of these three states and the interconversion between them (Fig. 5.8a). At 150 mM NaCl the lifetimes of the mid- and



(a) 50 mM NaCl **(b)** 150 mM NaCl

Fig. 5.8 Representative single-molecule traces obtained at (a) 50 mM NaCl and (b) 150 mM NaCl. Solid black line represents the automated hidden Markov modelling of the experimental trace displaying three states.

high-FRET states decrease, whilst the folded conformation increases, what is translated in more dynamics towards the low-FRET state (Fig. 5.8b).

Now it is necessary to assign the high-FRET state to a certain conformation. According to the explanation given at the beginning of this section, the intermediate state may correspond to the stacking of stems P1 and P3, which is very dynamic and unstable in sodium. Our data do not allow us to establish whether this intermediate corresponds exclusively to coaxial alignment or base stacking. Therefore, we have assigned the intermediate as a combination of both (Fig. 5.9).

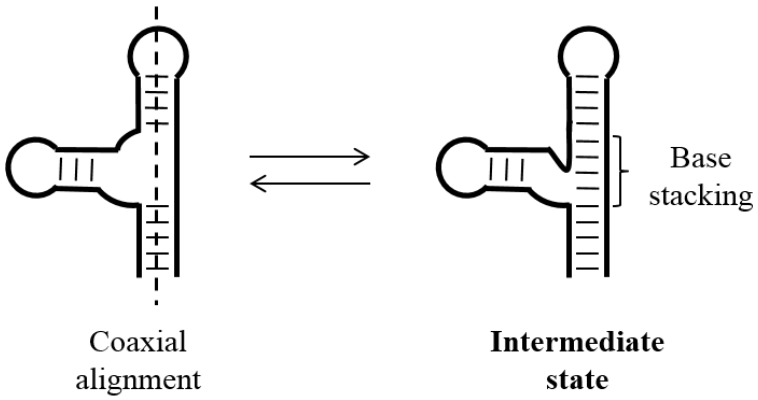
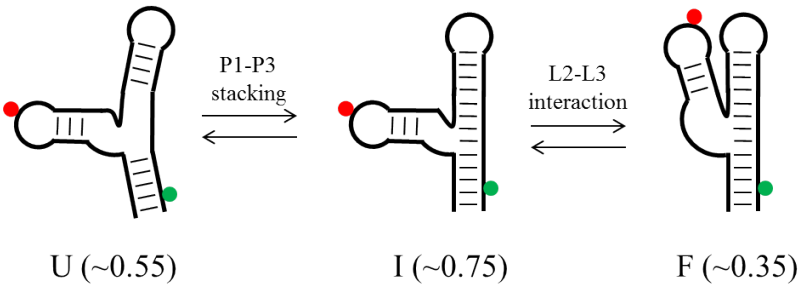


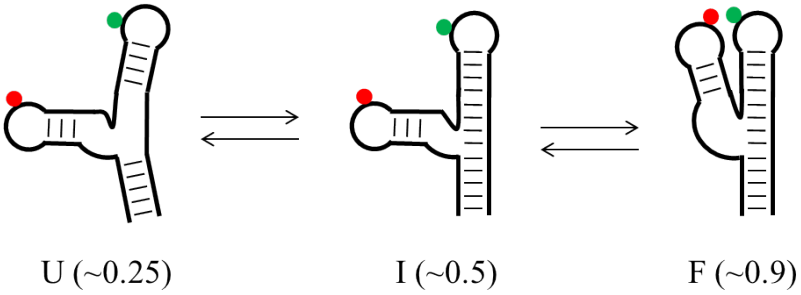
Fig. 5.9 Structure of the intermediate state. Proposed structure assigned to the intermediate state based on single-molecule FRET data in NaCl.

In this scenario, to allow P1-P3 stacking in the absence of the loop-loop interaction, P2 has to transiently increase its dynamics, and thus it gets closer to P1, which results in a high-FRET state. If sufficient concentration of cations, either mono- or divalent, are present at some point during this P1-P3 dynamic stacking/unstacking and movement of P2, the loop-loop interaction is stabilised and results in the formation of the low FRET state. Thus, the results lead us to assign the mid- and high-FRET peaks to the unfolded and intemediate states, respectively (Fig. 5.10).

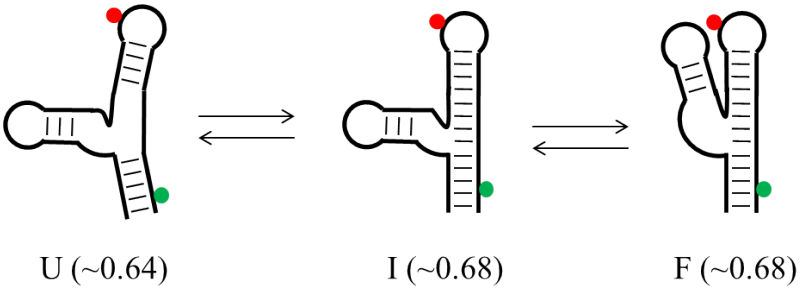
Thus, the initial "Y-shape" state leads to the formation of an intermediate state characterised by a coalignment of P1-P3 and high flexibility of stem P2. Then, increasing concentrations of salt promote the formation of the loop-loop interaction. Importantly, the model we propose based on our results is not only in good agreement with the data from



(a) Vector 1-2



(b) Vector 2-3



(c) Vector 1-3

Fig. 5.10 Folding mechanism of the *add* adenine aptamer. Proposed folding pathway of the adenine aptamer represented for the three vectors studied using smFRET. FRET values associated to each state are shown in parentheses. The three-way junction adopts a "Y-shape" conformation in the unfolded state, which is associated with a mid-FRET state. The addition of salt promotes the stabilisation of the P1-P3 stacking in the intermediate conformation, where P2 presents a high degree of flexibility. Since the P1-P3 stacking shortens the distance between P1 and P2, the intermediate state arises at high-FRET. If the concentration of salt is sufficient to promote the L2-L3 loop-loop interaction, the aptamer folds into a conformation with parallel rearrangement of stems P2 and P3 corresponding to the ligand-free folded state.

Lemay and coworkers [Lemay et al., 2006] but also with the model proposed by Gilbert and Batey [Gilbert and Batey, 2006], in which they hypothesised about an intermediate corresponding to the helical stacking of P1 and P3.

Our data using Na^+ as a folding agent confirmed the presence of an intermediate folding state. Taking into account that this FRET state shows a higher FRET value than the observed for P1-P2 at high Mg^{2+}, we can exclude that this is due to the formation of a state close to that displaying the loop-loop interaction. In this case, we would expect a long P1-P2 distance. Thus, we conclude that the intermediate corresponds to the coaxial alignment and stacking interaction between P1-P3 whether the coaxial alignment is present in the U state already is uncertain at this stage.

5.4 Conclusions

Single-molecule FRET has greatly demonstrated its potential in the observation of discrete molecules in real time. The ability of smFRET to monitor fast events, ranging from milliseconds to minutes, has provided important advantages in the observation of folding dynamics and transitions between states that would remain hidden in traditional ensemble techniques. The aim of the work presented here was to determine the presence of an intermediate state in the folding of the adenine *add* aptamer in order to further characterise the folding pathway of this aptamer as a result of the P1-P3 helical stacking and the L2-L3 loop-loop interaction.

The aptamer was labelled with a donor(Cy3)-acceptor(Cy5) pair at stem P1 and stem loop L2, respectively, to monitor the rearrangement of stem P2 upon addition of metal cations that promote the formation of the P1-P3 stacking. The smFRET data has been analysed as population E_{app} histograms and single-molecule trajectories. Both show a progressive folding of the aptamer in the presence of Mg^{2+}, however, only two states were identified. On the contrary, the presence of Na^+ allowed the discrimination of a third state arising at high-FRET values (~0.75) due to the less efficiency o monovalent cations in stabilising helical stacks.

Thus, the smFRET experiments presented in this work shed light on the nature of the intermediate state, which is assigned to the P1-P3 stacking conformation. These results

have led us to propose a new model describing the folding pathway of the *add* aptamer that experimentally confirms the previous hypothesis [Gilbert and Batey, 2006].

References

Batey, R. T. (2012). Structure and mechanism of purine-binding riboswitches. *Quarterly Reviews of Biophysics*, 45(3):345–381.

Batey, R. T., Gilbert, S. D., and Montange, R. K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, 432(7015):411–415.

Dalgarno, P. A., Bordello, J., Morris, R., St-Pierre, P., Dubé, A., Samuel, I. D. W., Lafontaine, D. A., and Penedo, J. C. (2013). Single-molecule chemical denaturation of riboswitches. *Nucleic Acids Research*, 41(7):4253.

Gilbert, S. D. and Batey, R. T. (2006). Riboswitches: Fold and function. *Chemistry & Biology*, 13(8):805 – 807.

Hohng, S., Wilson, T. J., Tan, E., Clegg, R. M., Lilley, D. M. J., and Ha, T. (2004). Conformational flexibility of four-way junctions in RNA. *Journal of Molecular Biology*, 336(1):69–79.

Lemay, J.-F., Penedo, J. C., Tremblay, R., Lilley, D. M., and Lafontaine, D. (2006). Folding of the adenine riboswitch. *Chemistry & Biology*, 13(8):857 – 868.

Mandal, M. and Breaker, R. R. (2004a). Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol*, 11(1):29–35.

Mandal, M. and Breaker, R. R. (2004b). Gene regulation by riboswitches. *Nat Rev Mol Cell Biol*, 5(6):451–463.

McKinney, S. A., Joo, C., and Ha, T. (2006). Analysis of single-molecule FRET trajectories using hidden markov modeling. *Biophysical Journal*, 91(5):1941 – 1951.

Rieder, R., Lang, K., Graber, D., and Micura, R. (2007). Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. *ChemBioChem*, 8(8):896–902.

Serganov, A., Yuan, Y.-R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., Hobartner, C., Micura, R., Breaker, R. R., and Patel, D. J. (2004). Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chemistry & Biology*, 11(12):1729 – 1741.

Wickiser, J. K., Cheah, M. T., Breaker, R. R., and Crothers, D. M. (2005). The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry*, 44(40):13404–13414. PMID: 16201765.

6

Single-molecule FRET studies of nascent RNA within transcriptional complexes

6.1 Introduction

The process of transcription represents the synthesis of an RNA strand by the RNA polymerase enzyme (RNAP) from a DNA template. In the first step, the polymerase binds a specific region, called promoter, located upstream of the DNA sequence to be transcribed. Before the RNAP escapes the promoter, the RNA transcript can be prematurely released, which produces short (2–8 nucleotides long) RNA products in a process called abortive initiation [Mooney et al., 1998]. Once the polymerase is able to synthesise a transcript of ~10 nucleotides and leave the promoter, the RNAP can proceed to elongate the RNA by adding nucleotides that are complementary to the DNA template at a rate of 20–80 nt/s in bacteria and up to 200 nt/s in bacteriophage [Pan and Sosnick, 2006]. During this stage, the polymerase stops at specific locations associated to pause sites within the elongated RNA in a period from tens of milliseconds to seconds while the RNA waits for an interaction with regulatory molecules [Landick, 1997; Pan and Sosnick, 2006]. Then, the RNAP continues with the elongation step until it reaches the terminator sequence [Pan and Sosnick, 2006]. At this stage, the RNA transcript and RNA polymerase are finally released.

It is during the elongation step when the RNA transcript undergoes conformational changes to adopt specific three-dimensional structures with a wide variety of functions (Fig.

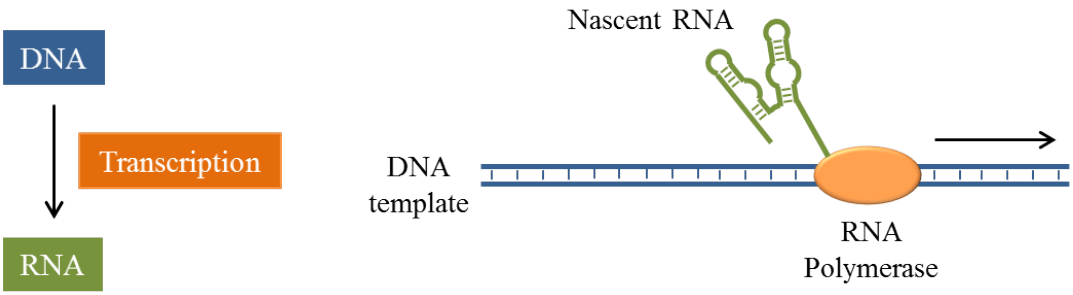


Fig. 6.1 Cotranscriptional folding or RNA. Schematics of the cotranscriptional folding of nascent RNAs as they are being synthesised by the RNA polymerase.

6.1). The folding of mRNA during biogenesis is directly involved in crucial cellular processes such as splicing, translation and bacterial gene regulation [McManus and Graveley, 2011; Pan and Sosnick, 2006]. Riboswitches are examples of how the folding of certain RNA sequences as they are synthesised by the RNAP can be directly involved in the control of genetic expression. These mRNAs bind cellular metabolites during transcription with high sensitivity and specificity to undergo structural changes that directly contribute to regulate the expression of an associated gene. However, the transient nature of these nascent RNA transcripts complicates the cotranscriptional study of the ligand binding mechanism and folding pathway. In fact, our understanding of cotranscriptional events is still in a very early stage.

Most studies on RNA transcription have focused on understanding the mechanism by which the RNAP incorporates nucleotides at the different stages: initiation, elongation and termination. However, there is not much information available regarding the behaviour of nascent RNA (folding, contacts with the RNAP and linking between RNAP pausing and folding). Thus, there is the need to go one step beyond in this field as it has been suggested that RNA folding during transcription may lead to structures that are not present in full-length RNA and it is also known that folding and activity during transcription are biologically relevant for some RNAs that otherwise are not functional. For this reason, we have introduced smFRET to provide a new dimension to these studies by allowing us to monitor polymerase dynamics during transcription.

Although the use of smFRET was effectively introduced to investigate the elongation of mRNA, structural changes of RNAP during transcription and interactions between RNA and RNAP [Andrecka et al., 2008; Chakraborty et al., 2012]; the study of cotranscriptional folding to date has been limited to non-fluorescence techniques. Recently, two cotranscriptional folding studies using selective 2'-hydroxyl acylation analysed by primer-extension sequencing (SHAPE-seq) and force-based techniques have been reported. Whereas the link between the contranscriptional folding and the genetic control of the *crcB* fluoride riboswitch was investigated by SHAPE-seq [Watters et al., 2016], the use of optical trapping allowed to monitor the cotranscriptional folding of the *pbuE* adenine riboswitch in real time [Frieda and Block, 2012; Greenleaf et al., 2008]. More precisely, the use of dual-beam optical tweezers allowed to monitor the aptamer folding and ligand binding that precludes the formation of the terminator stem in the first cotranscriptional study of RNA folding at the single-molecule level. In this case, the RNAP passes through the terminator sequence and stops at the biotin-streptavidin roadblock at the template end. On the contrary, the unbound aptamer is unable to prevent the formation of the terminator hairpin and leads to transcription termination with consequent release of the RNAP [Frieda and Block, 2012].

The inability of bacterial and eukaryotic RNAP to incorporate fluorescent dyes [Vassylyev et al., 2007] has restricted the study of RNA riboswitches using fluorescence-based techniques, either at ensemble or single-molecule level, to pre-folded synthetic or fully-transcribed sequences with the use of the bacteriophage T7 RNAP. To date, smFRET studies of RNA riboswitches use dyes incorporated to the final length RNA at specific positions to measure a particular conformational change during either folding or ligand binding [Boudreault et al., 2015; Perez-Gonzalez et al., 2016a,b; St-Pierre et al., 2014]. However, these studies do not allow to characterise the RNA folding within the transcription complex (Fig. 6.2), which would depend on RNAP pausing [Toulmé et al., 2005; Wickiser et al., 2005; Wong et al., 2007], interactions with the ligand [Blouin et al., 2009], elongation rate [Heilman-Miller and Woodson, 2003; Pan et al., 1999] and interactions between the nascent RNA strand with proteins [Lai et al., 2013; Landick, 1997; Mooney et al., 1998; Pan and Sosnick, 2006].

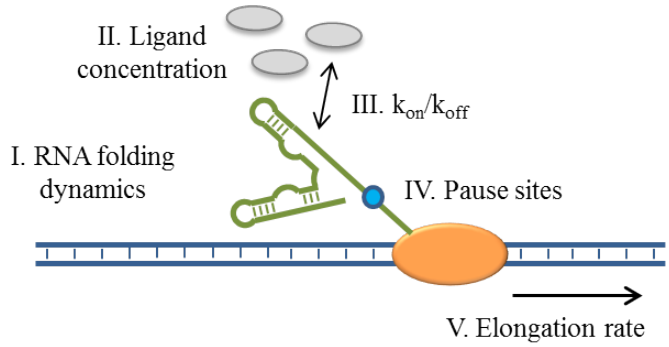


Fig. 6.2 Cotranscriptional folding of RNA riboswitches. Up to five different variables control the gene regulation by riboswitches, from which two of them, RNA pausing sites and transcription rate, can only be studied in cotranscriptional assays.

A recent strategy has partially addressed these limitations to investigate the RNA synthesis by RNA polymerases I, II and III in cells *in vivo* [Jao and Salic, 2008]. Jao and Salic were able to visualise the synthesis of RNA by incorporation of uridine analogues (i.e. 5-ethynyluridine) carrying an alkyne group that can react with azide-labelled fluorophores via click-chemistry. However, the incorporation of 5-ethynyluridine is done at random positions with an average of one analogue every 35 uridine residues, which is incompatible with smFRET studies where the dyes must be attached at known positions within the RNA aptamer.

Furthermore, the nature of the RNA polymerase represents another major limitation in cotranscriptional studies. The importance of using the cognate polymerase relies on their different elongation rates and RNAP pausing, which are highly related to the folding of the RNA. The *E. coli* polymerase provides a pause-dependent and slower transcription rate than the bacteriophage T7 RNAP that has already shown its inability to couple metabolite binding and folding of the AdoCbl riboswitch as opposed to *E. coli* RNAP [Perdrizet et al., 2012]. Thus, the use of the cognate RNAP will provide a more accurate approach to RNA folding due to the differences shown in the recognition of pause sites between polymerases from different species *in vitro* [Artsimovitch et al., 2000].

We developed in collaboration with Daniel Lafontaine's group (Université de Sherbrooke, Canada) a novel experimental approach that allows to monitor, for the first time, the folding and dynamics of nascent RNAs within *Escherichia coli* elongation complexes (ECs). The method we present in this chapter arises as a combination of three main strategies:

click chemistry, stepwise transcription and the use of a biotin-streptavidin roadblock. The site-specific labelling of nascent transcripts is achieved by initial addition of a fluorescent trinucleotide (Cy3-GUU) and the incorporation of an azide-modified UTP at the desired position for posterior attachment of a Cy5 carrying an alkyne group via copper-free click chemistry [Agard et al., 2004; Winz et al., 2012]. We investigated the binding affinity for the coenzyme thiamin pyrophosphate (TPP) of the *tbpA* riboswitch from *E. coli*, which is involved in the control of thiamin biosynthesis and transport of TPP [Rodionov et al.,

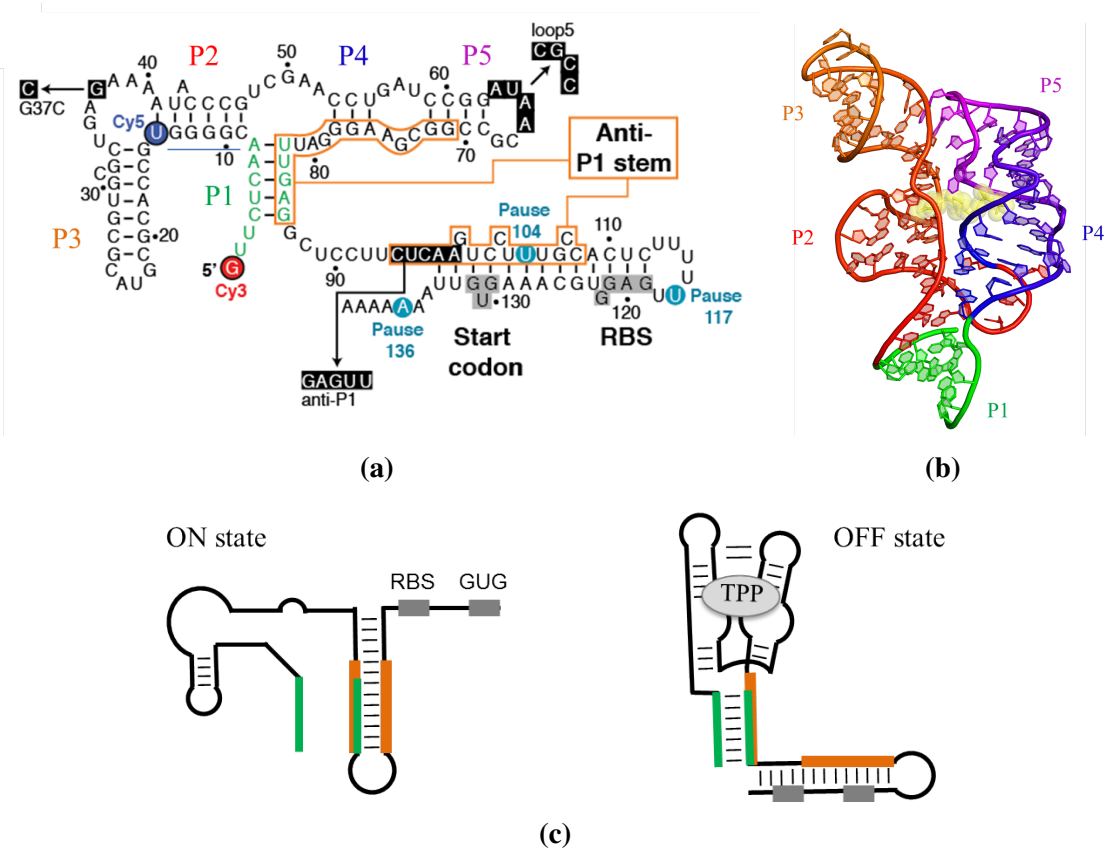


Fig. 6.3 Structure of the TPP riboswitch. (a) Sequence and secondary structure of the TPP-sensing *E. coli tbpA* riboswitch. Transcription pause sites at positions 104, 117 and 136 resported by [Chauvier et al., 2017] are highlighted in blue. The mutations that destabilise ligand binding (G37C), P3-L5 interaction and anti-P1 stem are highlighted in black. (b) Crystal structure of the *E. coli thiM* riboswitch sensing TPP, which is represented in yellow (PDB ID: 2GDI) [Serganov et al., 2006]. (c) Schematics of the ON and OFF states involving the formation of the anti-P1 and P1 stems, respectively. The formation of the P3-L5 interaction that stabilises the ligand-bound conformation is shown in the OFF state. The position of the ribosome-binding site (RBS) and GUG starting sequence is also indicated.

2002]. We have chosen the TPP-sensing riboswitch after a long collaboration with Daniel Lafontaine's group in this riboswitch and because the presence of pausing sites has been confirmed in a previous study (Fig. 6.3a) [Chauvier et al., 2017]. Although no crystal structure of the *tbpA* is available, the *thiM* variant can be used as a model for all TPP-binding riboswitches due to the highly conserved sequence and similar structure. TPP riboswitches are organised around a three-way junction adopting a Y-shape with two parallel domains (P2/J3–2/P3/L3 and J2–4/P4/P5/L5) in which the formation of P3-L5 contacts stabilise the ligand-bound state (Figs. 6.3b and 6.3c) [Serganov et al., 2006; Thore et al., 2006].

This approach provides the first method for the study of cotranscriptional regulation and transcriptional pause sites using smFRET, which will open an entire new field in fluorescence-based single-molecule RNA research: cotranscriptional RNA biology.

6.2 Materials and methods

The development of the protocols for site-specific incorporation of dyes via click-chemistry using stepwise transcription methods was optimised in a collaboration between Daniel Lafontaine's group and our lab. Once this was demonstrated using gel electrophoresis, the single-molecule setup was improved during my stay at Sherbrooke in March-April 2016 which was aimed as a knowledge exchange between both labs to: i) optimise the transfer from the bench to the microscope of the FRET-labelled complexes; ii) perform the first single-molecule experiments in these complexes; iii) analyse the data to determine the FRET change upon ligand binding; and iv) learn how to carry out the labelling steps to continue with the single-molecule assays in St Andrews. Posterior analysis of the smFRET data was carried out by myself.

6.2.1 Preparation of double-labelled transcription elongation complexes

The RNA polymerase (RNAP) and sigma70 factor from *Escherichia coli* were purified as described in [Nudler et al., 2003]. The DNA templates used for transcription were produced by PCR using oligonucleotides containing the *lacUV5* promoter sequence. The elongation

complexes incorporate a biotin at the 5' end of the complementary strand and the roadblock used as stopping site was built by addition of streptavidin in a 5:1 ratio with biotinylated DNA, added 5 minutes before transcription initiation.

Transcription reactions were performed in transcription buffer containing 20 mM Tris-HCl (pH 8.0), 20 mM $MgCl_2$, 14 mM β-mercaptoethanol and 0.1 mM EDTA. The assembly of the complex was performed with 200 nM DNA template, 600 nM sigma70 factor and 300 nM RNAP, which were incubated for 5 min at 37 °C. The transcription was initiated by addition of 25 µM Cy3-GUU trinucleotide and 1.25 µM ATP/CTP/UTP nucleotides, then incubated for 10 min at 37 °C which results in an elongation complex paused at position 9 (EC-9). The free nucleotides were removed from the solution by size exclusion chromatography using G50 columns. The complex was then incubated for 5 min at 37 °C with 2.5 µM CTP/GTP and 2.5 µM azide-modified UTP to allow the formation of EC-17 labelled with Cy5 at position 14, where the only uracil in the 10–17-nt region is found. The elongation complexes were then washed using G50 columns to remove the excess of azide-labelled nucleotides. Finally, these complexes were elongated by addition of 1 mM NTPs for 1 min at 37 °C yielding the paused complexes at the biotin-streptavidin roadblock (Fig. 6.4), which is used for immobilisation of the EC on biotin-functionalised surfaces via biotin-streptavidin interaction.

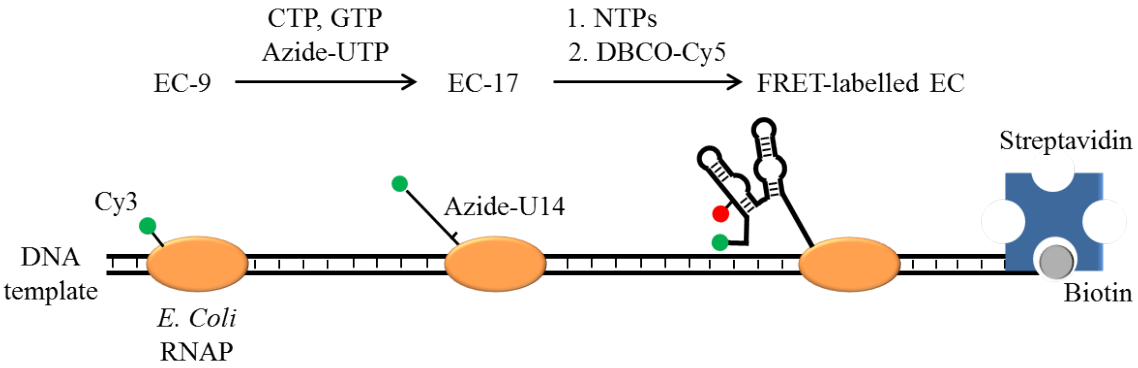


Fig. 6.4 Preparation of double-labelled ECs. Schematics showing the double labelling of elongation complexes for smFRET assays. The RNA synthesis is started using a 5'-GUU trinucleotide with a Cy3 attached, whereas the incorporation of an azide-labelled UTP at U-14 allows the site-specific labelling with Cy5 via copper-free click-chemistry reaction.

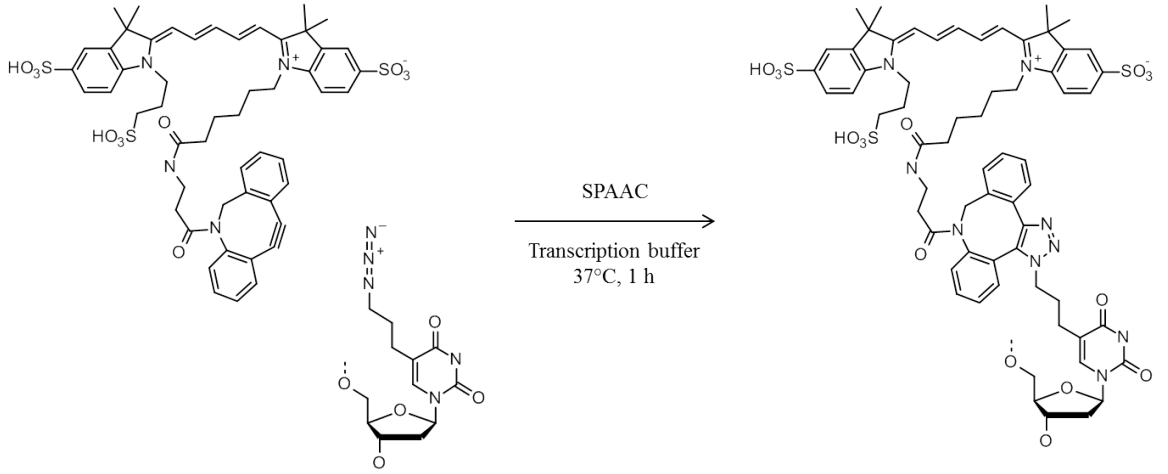


Fig. 6.5 Click-chemistry reaction. General scheme for the copper-free strain-promoted azide-alkyne cycloadition (SPAAC) between the DBCO-functionalised Cy5 dye and azido-uridine base.

The attachment of the Cy5 dye at azide-labelled U14 was performed via copper-free strain-promoted alkyne-azide cycloaddition (Fig. 6.5) [Winz et al., 2012]. The ECs were incubated with 63 µM DBCO-Cy5 (Jena Bioscience, Jena, Germany) in transcription buffer for 1 hour at 37 °C. The transcription complexes were then washed twice in G50 columns to remove unreacted DBCO-Cy5 dye. Double labelled ECs were collected in transcription buffer and incubated for 5 min with TPP or water before injection into the slide.

6.2.2 Single-molecule imaging

In order to avoid non-specific interactions between the protein RNA polymerase and the quartz slide, slides were passivated with a 30:1 mixture of PEG and biotinylated PEG (MW 5,000) and a second passivation step with a short MS4-PEG. FRET-labelled ECs were added to the slide in a concentration of ~250 pM to which they are tethered via biotin-streptavidin interaction (Fig. 6.6). The imaging of paused transcripts was carried out in transcription buffer to which were added 5 mM PCA, 100 nM PCD and 2 mM TROLOX. Experiments were performed at room temperature on a home-built prism-type TIR microscope (Olympus IX71) equipped with a 5 mW CW laser operating at 532 nm. Donor and acceptor fluorescence emissions were separated using longpass dichroic mirrors (DCRLP, Chroma Technology)

and simultaneously imaged onto the left (donor) and right (acceptor) half-chip of a back-illuminated Ixon EMCCD camera. Experimental movies were collected with an integration of 50 ms unless stated otherwise.

Traces were corrected and analysed as described in section 2.5.3. Then, smFRET histograms were fitted to a linear combination of two Gaussian distributions associated to the folded (F) and unfolded (U) conformations using Origin 6.0 and 8.0. The center of each Gaussian was taken as the FRET value of each population. To help visualisation, experimental traces were analysed with automated hidden Markov modelling of the experimental trajectories using HaMMy [McKinney et al., 2006].

Single-molecule kinetic analysis of average rates and dwell times of individual transitions was performed using a custom-built software in MATLAB. Average rates are represented as contour plots, whereas dwell times of individual transitions were plotted as frequency histograms fitted to a mono- or bi-exponential decay function from which obtain the value or values of the transition rates. Each transition is weighted by considering the total number of transitions observed in each molecule in order to avoid a biased interpretation of fast transitions.

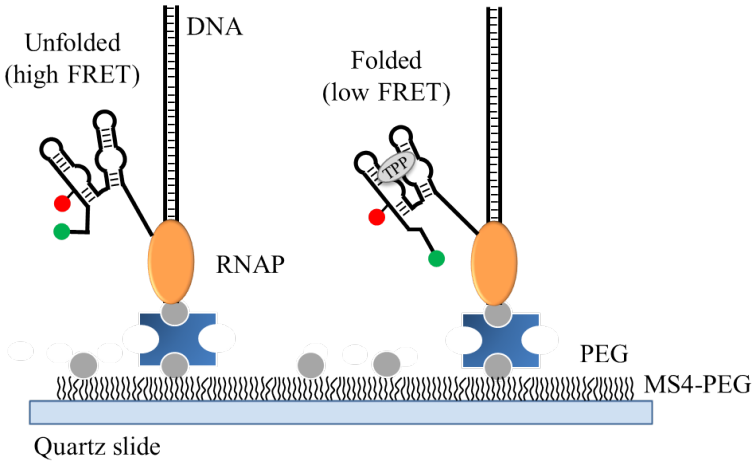
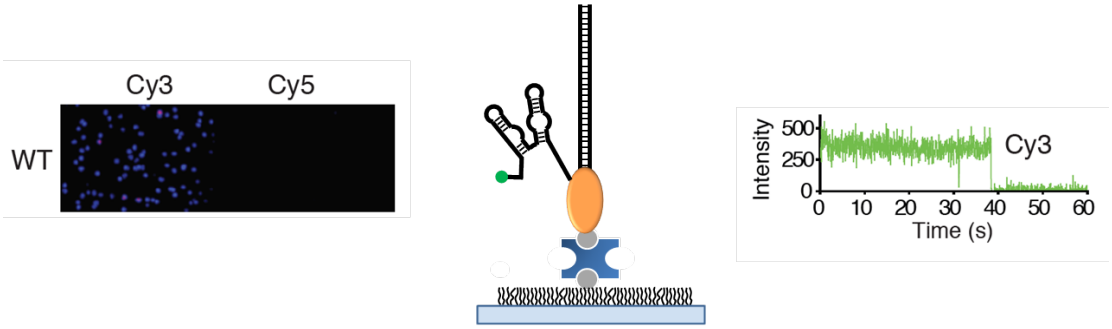


Fig. 6.6 Single-molecule observation of ECs. Schematics of surface-immobilised ECs and representation of the unfolded and folded states of the TPP riboswitch.

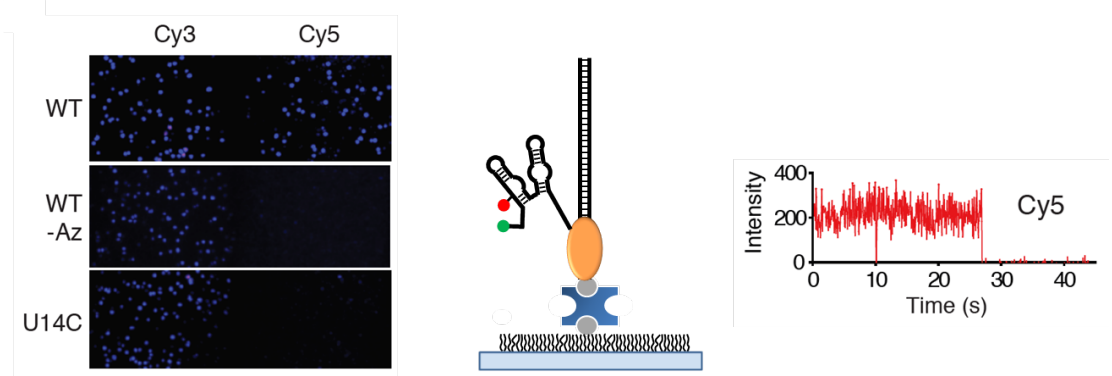
6.3 Results and discussion

6.3.1 Characterisation of surface-immobilised elongation complexes and labelling specificity

The labelling stoichiometry and specific immobilisation on PEG slides was investigated for either single- or double-labelled elongation complexes (Fig. 6.7). We have also tested other passivation methods including DT20, however, the best results were obtained using a two-step pegylation protocol that results from the combination of PEG and biotinylated PEG (MW 5,000) in a 30:1 ratio and posterior addition of the shorter MS4-PEG. In this case, a good number of spots is achieved on the donor channel in the observation of ECs carrying only a Cy3 dye, each showing one single photobleaching step, which corroborates



(a) Single labelled complex



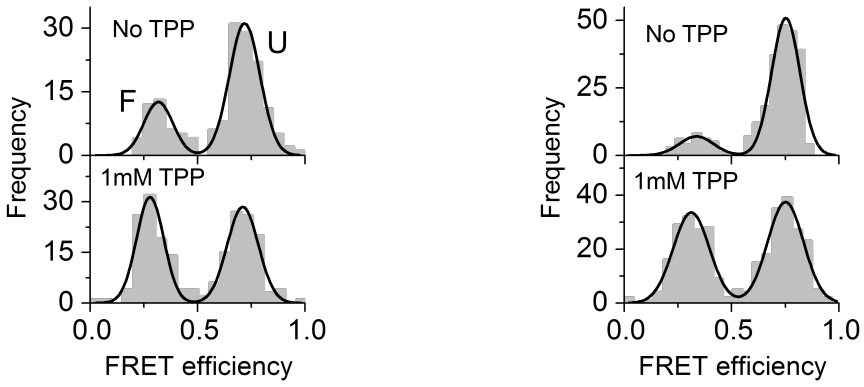
(b) Double labelled complex

Fig. 6.7 Surface immobilisation of EC complexes. Labelling and surface-immobilisation studies showing representative movie frames and Cy3/Cy5 trajectories for single- (a) and double-labelled (b) elongation complexes.

the presence of one donor per complex. Importantly, only significant number of spots was achieved in the presence of all the components of the elongation complex, which includes the biotin, streptavidin, DNA template and RNAP. Hence, the PEG passivation effectively prevents any non-specific interaction with the slide. The observation of double-labelled complexes show significant number of spots on the Cy5 channel that also display a single photobleaching event. On the contrary, the construct lacking the azide-functionalised UTP and the EC-14 complex studied as controls do not display observable spots on the acceptor channel.

6.3.2 Observation of ligand binding in nascent transcripts comprising the core of the *tbpA* aptamer

To prove the concept we studied the elongation complexes stalled at position 88 (EC-88), thus comprising the core of the *tbpA* aptamer, which show in the absence of TPP a main high-FRET population (E_{app} ~0.6) and a smaller contribution at low FRET (E_{app} ~0.3) (Fig. 6.8a). The relative population of both bands notably changes in the presence of 1 mM TPP by shifting the equilibrium towards the low-FRET population, which was assigned to the folded (F) state, whereas the high-FRET population was assigned to the unfolded (U) conformation.

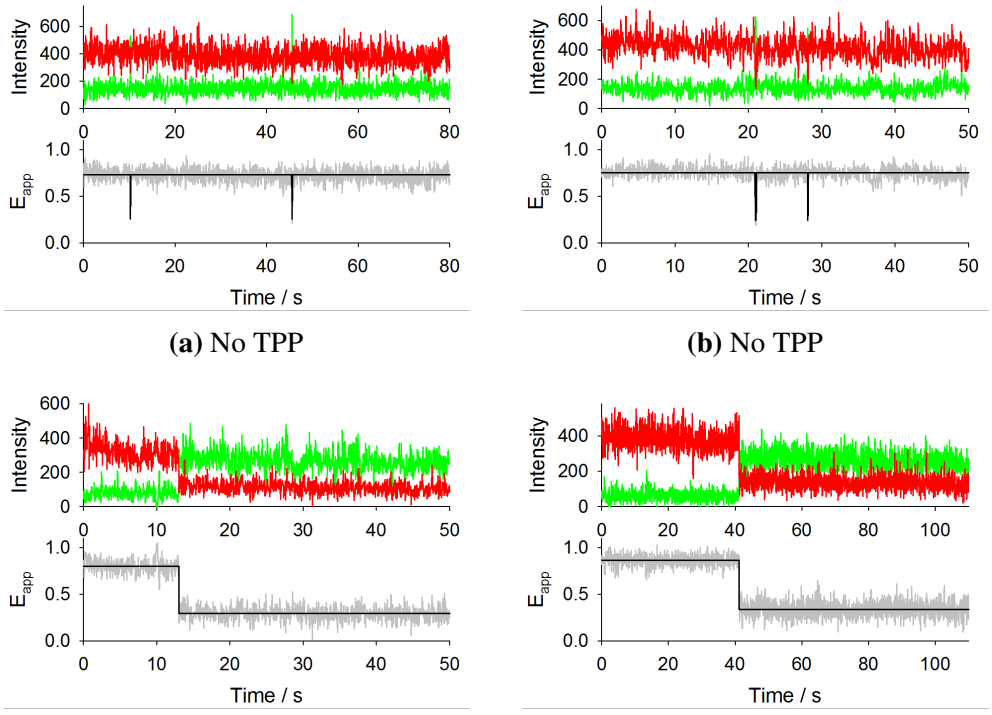


(a) EC-88 WT **(b)** Ligated 88

Fig. 6.8 SmFRET histograms of EC-88 and ligated 88 sequence. Single-molecule FRET histograms in transcription buffer in the absence (top) and presence (bottom) of 1 mM TPP.

This was confirmed by using a semi-synthetic aptamer, which was obtained by ligation of a synthetic strand carrying the Cy3 and Cy5 dyes at the same positions as EC-88 and a T7 RNAP-transcribed sequence. The ligated aptamer also display a main high-FRET peak in the absence of ligand and an increase of low-FRET state at 1 mM TPP (Fig. 6.8b), thus confirming the presence of two populations with apparent FRET efficiencies of ~0.3 and ~0.6. Moreover, these results show that incorporation of Cy3 and Cy5 does not compromise the aptamer functions.

The hidden Markov analysis of the FRET traces show very short transitions from the F to the U states in the absence of TPP, suggesting the ability of the ligand-free aptamer to transiently adopt the folded conformation (Figs. 6.9a and 6.9b). A longer lived F state is observed at 1 mM TPP (Figs. 6.9c and 6.9d).



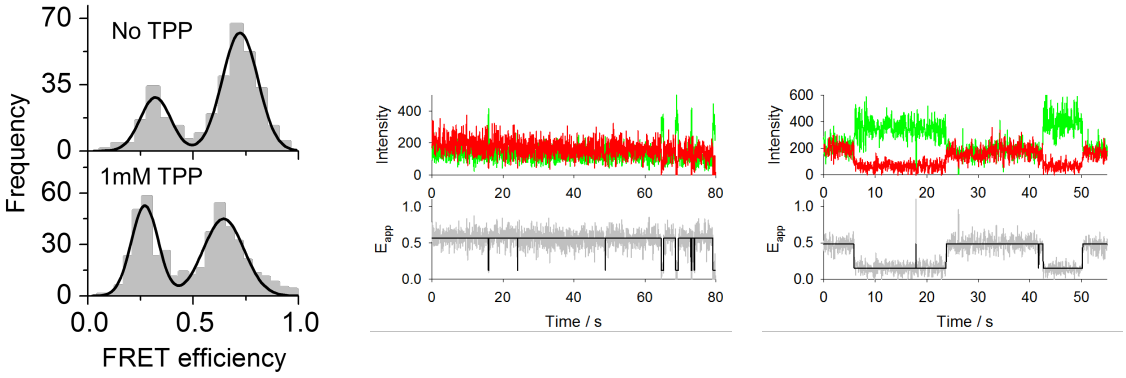
(a) No TPP **(b)** No TPP

(c) 1 mM TPP **(d)** 1 mM TPP

Fig. 6.9 Single-molecule FRET traces of EC-88. SmFRET traces showing the donor (green) and acceptor (red) trajectories as well as hidden Markov analysis of EC-88 in the absence (a-b) and presence (c-d) of 1 mM TPP.

6.3.3 Observation of ligand binding in transcriptional pause sites

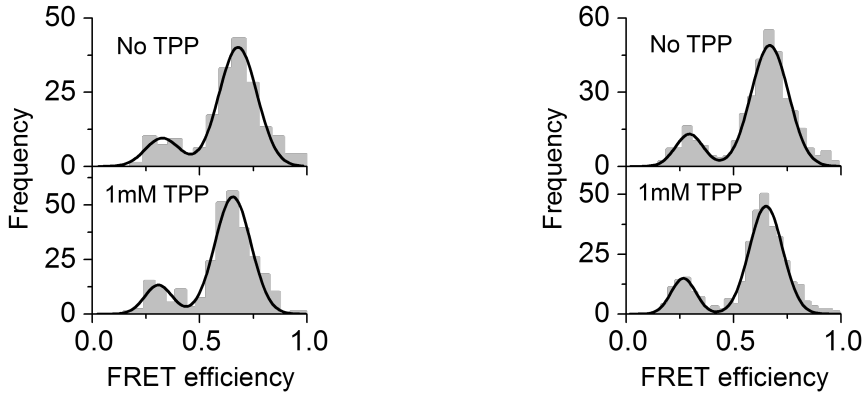
We then investigated the folding of the *tbpA* riboswitch by monitoring the ligand binding in elongation complexes stalled at the transcriptional pause sites 104, 117 and 137 [Chauvier et al., 2017]. EC-104 display similar band distributions than EC-88 in the absence and presence of 1 mM TPP, where the low-FRET state corresponding to the folded conformation significantly increases upon addition of ligand (Fig. 6.10).



(a) EC-104 **(b)** No TPP **(c)** 1 mM TPP

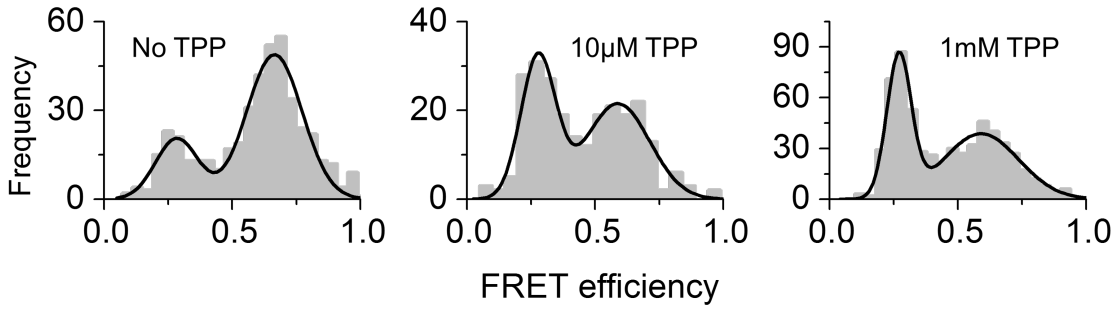
Fig. 6.10 SmFRET data of EC-104. (a) SmFRET population histograms of EC-104 at (top) no TPP and (bottom) 1 mM TPP. (b-c) Representative smFRET traces and hidden Markov analysis of EC-104 in the (b) absence and (c) presence of 1 mM TPP.

The study of EC-104 was completed by observation of two mutants that prevent riboswitch folding: G37C and Loop5. TPP ligand intercalates between G37 and A38 and establishes hydrogen bonds with the surrounding bases, which are crucial for ligand recognition [Serganov et al., 2006; Winkler et al., 2002]. A mutation on the guanine residue at position 37 would prevent ligand binding as observed for G37C mutant, in which no change in the distribution of the U and F populations is observed upon addition of 1 mM TPP (Fig. 6.11a). Mutations A63C-U64G-A65C-A66C at stem P5 loop were introduced to prevent the P3-L5 interaction, showing a similar result as for G37C, where the relative population of both low- and high-FRET states remains unchanged upon addition of 1 mM TPP (Fig. 6.11b).

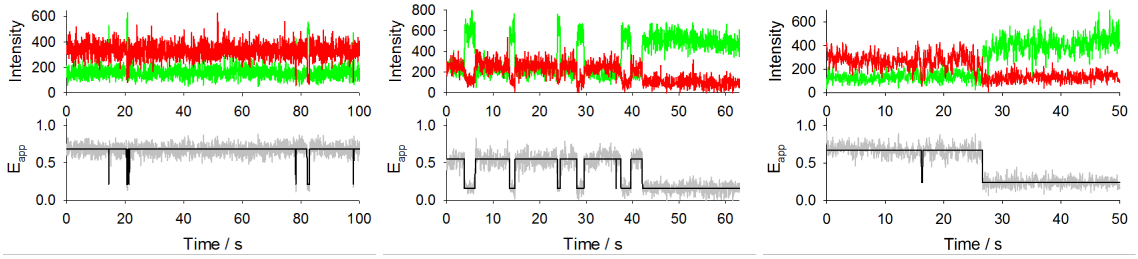


(a) G37C mutant **(b)** Loop5 mutant

Fig. 6.11 SmFRET histograms of the EC-104 G37C and Loop5 mutants. Single-molecule FRET histograms in transcription buffer in the absence (top) and presence (bottom) of 1 mM TPP.



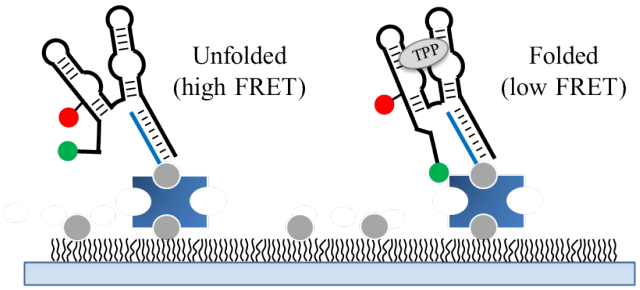
(a) EC-117



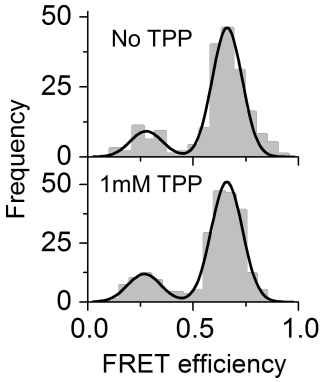
(b) No TPP **(c)** 10 μM TPP **(d)** 1 mM TPP

Fig. 6.12 SmFRET data of EC-117. (a) SmFRET histograms of EC-117 at (left) no TPP, (center) 10 μM TPP and (right) 1 mM TPP. (b-d) Representative smFRET trajectories of EC-117 at the different TPP concentrations.

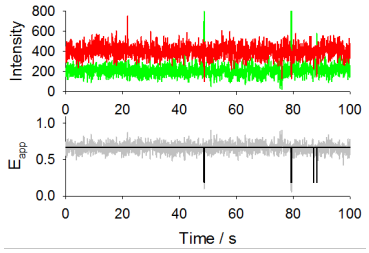
Similarly to EC-104, EC-117 efficiently folds at 1 mM TPP and at even lower concentrations such as 10 µM TPP (Fig. 6.12). This is surprising as the transcript paused at position 117 already includes the anti-P1 sequence and ligand binding should be diminished. The high ligand binding affinity shown in EC-117 may indicate that the position of the RNA polymerase in EC-117 prevents the formation of the anti-P1 stem (Fig. 6.3a). In order to corroborate this possibility we studied a 117-nt RNA transcript outside the transcription complex, Tr-117, which has shown to be unable to bind TPP (Fig. 6.13). This is because the RNA polymerase encloses 8-10 nucleotides, which influences the structure of the riboswitch in cotranscriptional folding events. These results prove the key role played by the RNAP in preventing the formation of the anti-P1 stem and emphasises the need to study the cognate polymerase.



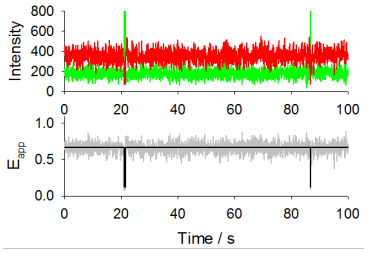
(a) Truncated aptamer



(b) Tr-117



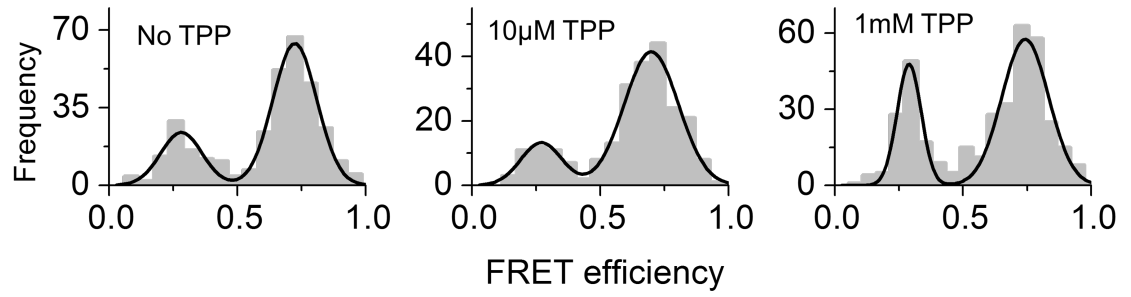
(c) No TPP



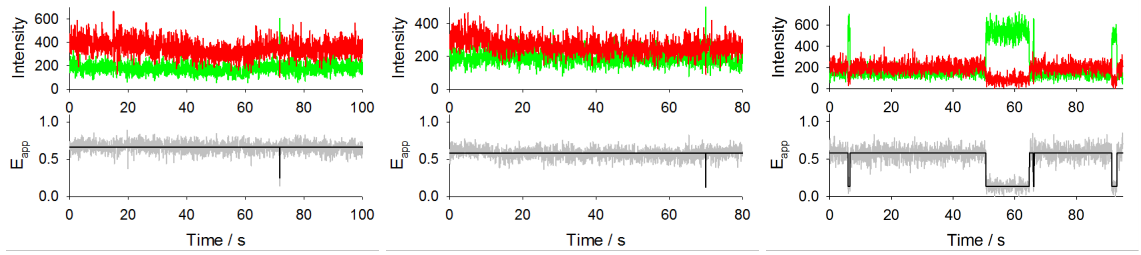
(d) 1 mM TPP

Fig. 6.13 SmFRET data of Tr-117. (a) Schematics of the immobilised run-off transcript Tr-117 attached to the slide via DNA anchor (blue). (b) SmFRET histograms of the 117-nt *tbpA* transcript at (top) no TPP and (bottom) 1 mM TPP. (c-d) Representative smFRET traces of the 117-nt transcript in the absence (c) and presence (d) of 1 mM TPP.

The last complex studied was the EC-136, which remains unchanged in the presence of 10 µM TPP, although this complex shows a small increase of the folded state at 1 mM TPP (Fig. 6.14). Thus, the position of the RNAP at this stage still limits the formation of the anti-P1 stem, although allows a certain degree of ligand binding. Moreover, the ability to transiently adopt the F conformation in the absence of ligand in the EC-88 complex was also observed for EC-104 (Fig. 6.10b), EC-117 (Fig. 6.12b) and EC-136 (Fig. 6.14b), showing short and rare transitions from the high- to the low-FRET state.



(a) EC-136

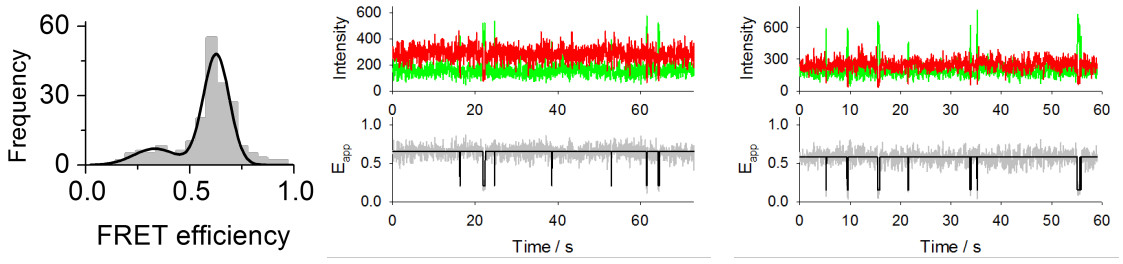


(b) No TPP **(c)** 10 µM TPP **(d)** 1 mM TPP

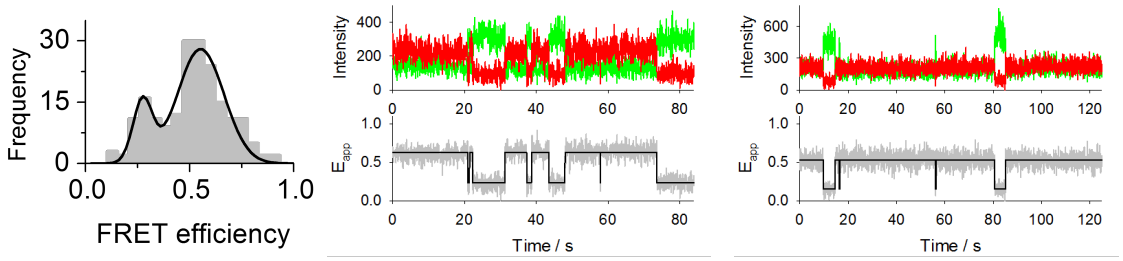
Fig. 6.14 SmFRET data of EC-136. (a) SmFRET histograms of EC-136 at (left) no TPP, (center) 10 µM TPP and (right) 1 mM TPP. (b-d) Representative smFRET traces of EC-136 at the different TPP concentrations.

The elongation complex stalled at position 117 was further investigated at significantly lower Mg^{2+} concentrations (i.e. 2 mM instead of 20 mM $MgCl_2$) to allow more conformational dynamics between the unfolded and folded states within the observation time. In the absence of ligand, EC-117 displays rapid transitions from the U to the F state due to the transient nature of the ligand-free folded conformation (Fig. 6.15a). On the other hand, upon

addition of 10 µM TPP the EC-117 complex shows frequent transitions to a long-lived F state corresponding to the ligand-bound folded conformation (Fig. 6.15b).



(a) 2 mM $MgCl_2$ and no TPP



(b) 2 mM $MgCl_2$ 10 µM TPP

Fig. 6.15 SmFRET data of EC-117 at 2 mM $MgCl_2$. Single-molecule population histograms, representative FRET traces and hidden Markov analysis of EC-117. Experiments were performed at 2 mM $MgCl_2$ in the absence (a) and presence (b) of 10 µM TPP. The addition of ligand increases the lifetime of the low-FRET state corresponding to the F state.

A kinetic study was performed by manual analysis of the those traces displaying transitions between the F and U states for each of the conditions. The results obtained show neither variation in the dwell time of the unfolded state upon addition of TPP, which remains at ~10 s (Figs. 6.16a and 6.16d), nor in the U→F transition rate, which is ~0.08 s^{-1} in the absence and presence of ligand (Figs. 6.16b and 6.16e). On the contrary, the folded state is stabilised at 2 mM $MgCl_2$ 10 µM TPP, showing a ~10-fold increase in the dwell time of the low-FRET state (Figs. 6.16a and 6.16d) as well as ~10-fold decrease in the F→U transition rate (Figs. 6.16c and 6.16f). These results suggest that the mechanism through which the EC-117 recognises TPP is via conformational capture [Haller et al., 2013], where the ligand selectively binds (or "captures") the folded conformation of the aptamer. Thus, the study of the elongation complex stalled at position 117 allowed us not only to identify a late formation

of stem anti-P1 but also to characterise the folding mechanism of the TPP riboswitch at this stage, which otherwise would have never been achievable by observation of full-length sequences.

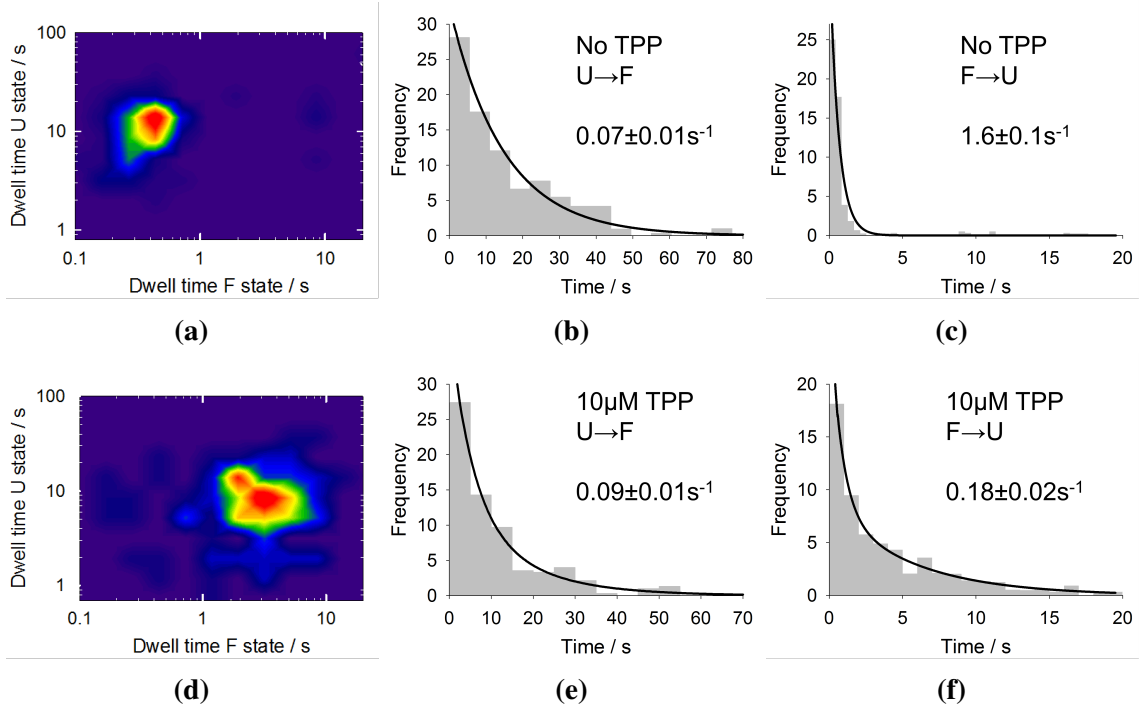
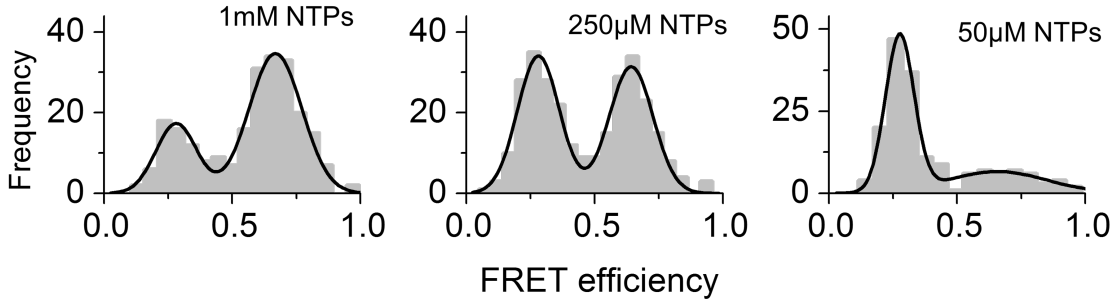


Fig. 6.16 Kinetic analysis of EC-117. (a) Two-dimensional contour plot of the unfolded (U) and folded (F) states at 2 mM $MgCl_2$ in the absence of TPP. (b-c) Dwell-time histograms and exponential decay fit values of the U→F and F→U transitions at 2 mM $MgCl_2$ in the absence of TPP. (d) Two-dimensional contour plot of the U and F states at 2 mM $MgCl_2$ and 10 μM TPP. (e-f) Dwell-time histograms and exponential decay fit values of the U→F and F→U transitions at 2 mM $MgCl_2$ in the presence of 10 μM TPP.

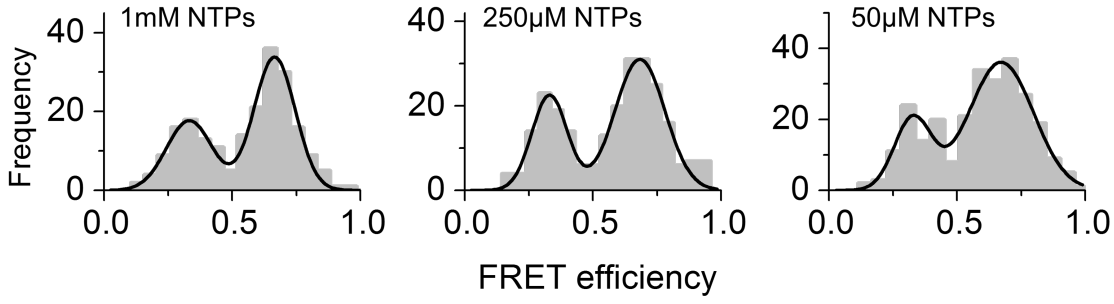
6.3.4 Influence of NTP concentration on the cotranscriptional folding of the *tbpA* riboswitch

The importance of the transcription rate on TPP binding and riboswitch folding was next investigated by examining the folding of a nascent RNA by addition of TPP either co- or post-transcriptionally. We have studied the changes in FRET due to ligand binding by incorporation of the Cy5 to a complex stalled at position EC-28 that was then elongated to position 136 by addition of different concentrations of NTPs. The nascent transcript

shows efficient cotranscriptional folding at low NTP concentrations, which is associated to a longer time window available for efficient TPP binding (Fig. 6.17a). These results were then compared with similar elongation assays performed with addition of TPP after transcription (Fig. 6.17b). Although no change between co- and post-transcriptional studies was observed at 1 mM NTPs due to the higher transcription rate, those at 250 µM and 50 µM do display a higher contribution of the F state when compared with post-transcriptional studies under the same conditions, thus underlining the dependency of riboswitch folding on not only the concentration of ligand but also the transcription rate.



(a) Cotranscriptional folding



(b) Post-transcriptional folding

Fig. 6.17 Co- and post-transcriptional folding of the TPP riboswitch. Single-molecule FRET histograms illustrating riboswitch folding obtained for the EC-28→EC-136 elongation at difference concentrations of NTPs in the presence of 10 µM TPP, which was added either cotranscriptionally (a) or post-transcriptionally (b).

6.3.5 Regulatory mechanism of the *E.coli tbpA* riboswitch

Nascent transcripts sense cellular TPP within a transcriptional window. The RNAP location dictates whether early or late TPP binding occurs, both of which ensuring cotranscriptional riboswitch folding and inhibition of translation initiation.

These results provide insights into the *tbpA* riboswitch cotranscriptional folding pathway and sheds light on its timescale and TPP-binding window. The study of the elongation complexes EC-88, EC-104 and EC-117 comprising the *tbpA* aptamer shows that TPP sensing within this range is possible, which would lead to early transcription termination and, thus, to repression of the downstream gene (Fig. 6.18) [Serganov et al., 2006]. Interestingly, the different response to TPP shown for the 117-nt run-off transcript and EC-117 suggest that the polymerase precludes the formation of the anti-P1 stem and plays a key role in late TPP binding. The formation of the anti-P1 stem does interfere with ligand affinity in EC-136, which showed no response at 10 µM TPP, although a small increase in the F state was observed at 1 mM TPP. In this case, the non-binding of TPP prior to the formation of the anti-P1 stem at EC-136 would lead to translation initiation and, therefore, to gene expression.

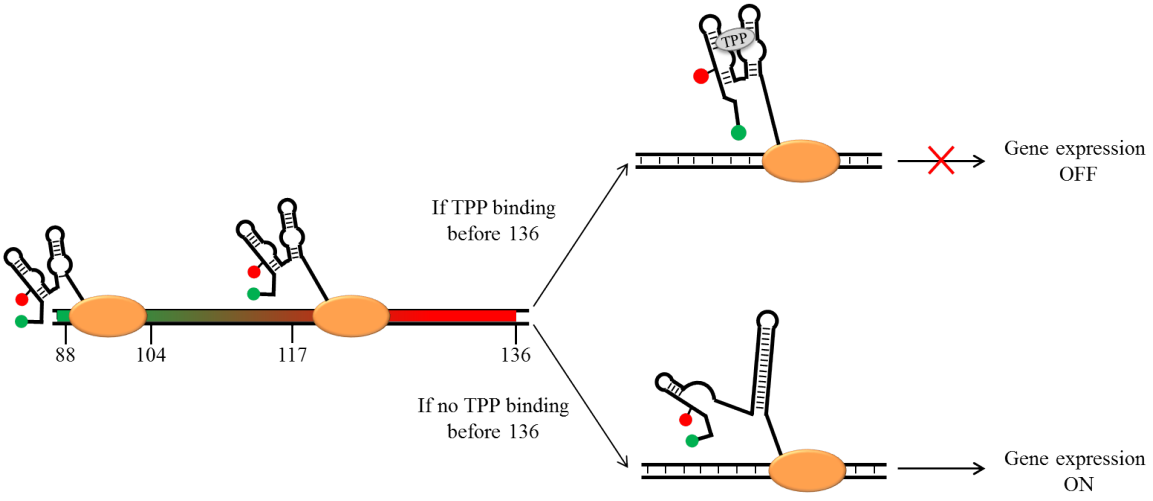


Fig. 6.18 Cotranscriptional folding of the TPP riboswitch. The nascent RNA transcript senses TPP in order to modulate the genetic expression within a limited cotranscriptional window.

6.4 Conclusions

The cotranscriptional folding of RNA is involved in a number of crucial processes such as splicing, translation and gene regulation. Such is the case of RNA riboswitches, regulatory mRNA sequences that undergo conformational changes during transcription upon ligand binding to modulate the expression of an associated gene. The difficulties to monitor the elongation and folding of RNA sequences in real time due to the transient nature of these structures have limited the use of fluorescence techniques, either at ensemble and single-molecule level, to study the ligand binding and folding of full-length transcripts. Thus, these studies do not account for the interactions within the transcription complex and mechanistic processes such as interactions with the RNAP, transcription rate and RNAP pausing.

This chapter presents a novel approach to study the folding and dynamics of nascent RNAs within bacterial elongation complexes using single-molecule FRET. The inability of RNAP to incorporate fluorescent dyes has been resolved by using copper-free click chemistry, which has experienced a stunning growth in the last years for specific labelling of biological samples in aqueous media *in vitro* and *in vivo*. This approach was used to investigate the TPP-sensing *tbpA* riboswitch within *E. coli* elongation complexes, where the use of smFRET allowed to monitor the folding of nascent RNA transcripts in real time. The results show how the nascent transcript undergoes conformational changes upon ligand binding within a reduced transcriptional window, which is delimited by RNAP pause sites, and how this would affect the expression of the downstream gene after finding a point of no return at EC-136, after which the aptamer is unable to bind free TPP due to the formation of the anti-P1 stem.

The method presented in this chapter not only provides the first smFRET approach for the study of cotranscriptional folding of nascent transcripts but also a novel alternative that facilitates the study of nascent transcripts using different biophysical techniques such as electronic paramagnetic resonance (EPR) spectroscopy by allowing to selectively incorporate spin labels within the nascent RNA strand. Furthermore, the use of this fluorescent-based smFRET approach in combination with force-based techniques such as optical trapping would

provide insights into the cotranscriptional folding of riboswitches with an unprecedented level of detail.

References

Agard, N. J., Prescher, J. A., and Bertozzi, C. R. (2004). A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *Journal of the American Chemical Society*, 126(46):15046–15047.

Andrecka, J., Lewis, R., Brückner, F., Lehmann, E., Cramer, P., and Michaelis, J. (2008). Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proceedings of the National Academy of Sciences*, 105(1):135–140.

Artsimovitch, I., Svetlov, V., Anthony, L., Burgess, R. R., and Landick, R. (2000). RNA polymerases from Bacillus subtilisand Escherichia coli differ in recognition of regulatory signals In vitro. *Journal of Bacteriology*, 182(21):6027–6035.

Blouin, S., Mulhbacher, J., Penedo, J. C., and Lafontaine, D. A. (2009). Riboswitches: Ancient and promising genetic regulators. *ChemBioChem*, 10(3):400–416.

Boudreault, J., Perez-Gonzalez, D. C., Penedo, J. C., and Lafontaine, D. A. (2015). Single-molecule approaches for the characterization of riboswitch folding mechanisms. In *DNA-Protein Interactions: Principles and Protocols*, pages 101–107. Springer New York, New York, NY.

Chakraborty, A., Wang, D., Ebright, Y. W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., and Jansen, R. (2012). Opening and closing of the bacterial RNA polymerase clamp. *Science*, 337(6094):591–595.

Chauvier, A., Picard-Jean, F., Berger-Dancause, J.-C., Bastet, L., Naghdi, M. R., Dubé, A., Turcotte, P., Perreault, J., and Lafontaine, D. A. (2017). Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. *Nature Communications*, 8:13892.

Frieda, K. L. and Block, S. M. (2012). Direct observation of cotranscriptional folding in an adenine riboswitch. *Science*, 338(6105):397.

Greenleaf, W. J., Frieda, K. L., Foster, D. A. N., Woodside, M. T., and Block, S. M. (2008). Direct observation of hierarchical folding in single riboswitch aptamers. *Science*, 319(5863):630.

Haller, A., Altman, R. B., Soulière, M. F., Blanchard, S. C., and Micura, R. (2013). Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proceedings of the National Academy of Sciences*, 110(11):4188–4193.

Heilman-Miller, S. L. and Woodson, S. A. (2003). Effect of transcription on folding of the Tetrahymena ribozyme. *RNA*, 9(6):722–733.

Jao, C. Y. and Salic, A. (2008). Exploring RNA transcription and turnover in vivo by using click chemistry. *Proceedings of the National Academy of Sciences*, 105(41):15779–15784.

Lai, D., Proctor, J. R., and Meyer, I. M. (2013). On the importance of cotranscriptional RNA structure formation. *RNA*, 19(11):1461–1473.

Landick, R. (1997). RNA polymerase slides home: Pause and termination site recognition. *Cell*, 88(6):741–744.

McKinney, S. A., Joo, C., and Ha, T. (2006). Analysis of single-molecule FRET trajectories using hidden markov modeling. *Biophysical Journal*, 91(5):1941 – 1951.

McManus, C. J. and Graveley, B. R. (2011). RNA structure and the mechanisms of alternative splicing. *Current opinion in genetics & development*, 21(4):373–379.

Mooney, R. A., Artsimovitch, I., and Landick, R. (1998). Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. *Journal of Bacteriology*, 180(13):3265–3275.

Nudler, E., Gusarov, I., and Bar-Nahum, G. (2003). Methods of walking with the RNA polymerase. In *Methods in Enzymology*, volume Volume 371, pages 160–169. Academic Press.

Pan, T., Fang, X., and Sosnick, T. (1999). Pathway modulation, circular permutation and rapid RNA folding under kinetic control. *Journal of Molecular Biology*, 286(3):721–731.

Pan, T. and Sosnick, T. (2006). RNA folding during transcription. *Annual Review of Biophysics and Biomolecular Structure*, 35:161–175.

Perdrizet, G. A., Artsimovitch, I., Furman, R., Sosnick, T. R., and Pan, T. (2012). Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proceedings of the National Academy of Sciences*, 109(9):3323–3328.

Perez-Gonzalez, C., Grondin, J. P., Lafontaine, D. A., and Carlos Penedo, J. (2016a). Biophysical approaches to bacterial gene regulation by riboswitches. In *Biophysics of Infection*, pages 157–191. Springer International Publishing, Cham.

References

Perez-Gonzalez, C., Lafontaine, D. A., and Penedo, J. C. (2016b). Fluorescence-based strategies to investigate the structure and dynamics of aptamer-ligand complexes. *Frontiers in Chemistry*, 4(33).

Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2002). Comparative genomics of thiamin biosynthesis in procaryotes: New genes and regulatory mechanisms. *Journal of Biological Chemistry*, 277(50):48949–48959.

Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R., and Patel, D. J. (2006). Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature*, 441(7097):1167–1171.

St-Pierre, P., McCluskey, K., Shaw, E., Penedo, J. C., and Lafontaine, D. A. (2014). Fluorescence tools to investigate riboswitch structural dynamics. *Riboswitches*, 1839(10):1005–1019.

Thore, S., Leibundgut, M., and Ban, N. (2006). Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science*, 312(5777):1208.

Toulmé, F., Mosrin-Huaman, C., Artsimovitch, I., and Rahmouni, A. R. (2005). Transcriptional pausing in vivo: a nascent RNA hairpin restricts lateral movements of RNA polymerase in both forward and reverse directions. *Journal of Molecular Biology*, 351(1):39–51.

Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. *Nature*, 448(7150):163–168.

Watters, K. E., Strobel, E. J., Yu, A. M., Lis, J. T., and Lucks, J. B. (2016). Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nature Structure & Molecular Biology*, 23(12):1124–1131.

Wickiser, J. K., Winkler, W. C., Breaker, R. R., and Crothers, D. M. (2005). The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Molecular Cell*, 18(1):49–60.

Winkler, W., Nahvi, A., and Breaker, R. R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, 419(6910):952–956.

Winz, M.-L., Samanta, A., Benzinger, D., and Jäschke, A. (2012). Site-specific terminal and internal labeling of RNA by poly(A) polymerase tailing and copper-catalyzed or copper-free strain-promoted click chemistry. *Nucleic Acids Research*, 40(10):e78–e78.

Wong, T. N., Sosnick, T. R., and Pan, T. (2007). Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proceedings of the National Academy of Sciences*, 104(46):17995–18000.

7

Understanding asymmetric nucleic acid unwinding by helicases

7.1 Introduction

A number of cellular processes involving nucleic acids such as replication, recombination, repair, transcription and translation require the unwinding of the dsDNA or dsRNA sequence in order to read or amend the information encoded. The separation of two annealed strands is done by a group of ATPases, called helicases, that use ATP hydrolysis as the energy source to undergo conformational changes between protein domains that drive the translocation of the protein with a certain polarity, which in duplexes involves the unwinding of the double strand.

Helicases are classified into six superfamilies (SF1–6) according to their conserved sequence motifs. The larger superfamilies SF1 and SF2 are monomeric structures consisting of two RecA-like domains, whose arrangement form the binding site for ATP, whereas SF3–6 form hexameric or double hexameric rings composed of 6 or 12 RecA folds [Singleton et al., 2007]. Helicases are also classified on the basis of their translocation polarity, whether 3'-5' (type A) or 5'-3' (type B), and according to their work on either single- (α) or double-stranded (β) DNA. Interestingly, all SF1 reported belong to type α, whilst SF2 and ring helicases contain examples of both types [Singleton et al., 2007]. Moreover, three families of SF1 and nine families and one group of SF2 have been identified, where a family includes more

than three proteins from a single organism and a group includes two proteins from the same organism [Fairman-Williams et al., 2010].

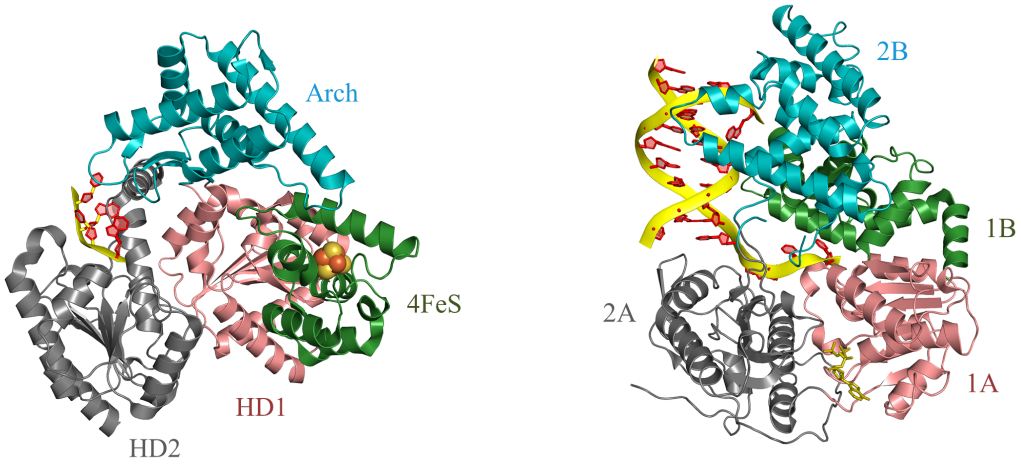
Ensemble and single-molecule fluorescence techniques have long been used for the study of helicase activity [Chisty et al., 2013; Constantinescu-Aruxandei et al., 2016; Dillingham et al., 2000, 2002; Ghoneim and Spies, 2014; Jia et al., 2011; Spies, 2014]. However, these studies were mainly focused on the observation of the enzyme structure and function. The majority of the unwinding models reported to date are focused on macroscopic aspects such as unwinding mechanism, processivity, unwinding rate and substrate specificity; but do not account for microscopic aspects involving the breaking of base pairs and how this is influenced by the nature of the nucleobase (i.e. purine or pyrimidine base). For example, the less stable AT/AU-rich sequences are more readily unwound than the more stable GC-rich segments. Furthermore, the location of purine or pyrimidine bases at the 5' or 3' end of the ssDNA-dsDNA junction may influence the unwinding rate as reported for the NPH-II helicase, in which the activity is modulated by the nucleic acid sequence showing preference for the leading strand carrying purine bases [Taylor et al., 2010].

Recent studies focused their attention in the role of nucleic acid sequences and how their dynamics modulate the binding and unwinding processes [Cheng et al., 2007; Johnson et al., 2007; Jose et al., 2012; Manosas et al., 2010; Phelps et al., 2013; Qi et al., 2013; Taylor et al., 2010]. In 1974, Pörschke reported the zipper model for binding of nucleic acids in which each step implies the addition of one base pair in a synchronised process between both single strands [Pörschke, 1974b,a]. However, some investigations on the structure of nucleic acids report discrepancies with the zipper model and provide evidence of an asymmetric movement of RNA bases that may be of great importance in nucleic acids unwinding by helicases. Mohan and coworkers investigated the folding of RNA duplexes with asymmetric formation of RNA base pairs consisting in the stacking of two bases at the 3' end in a semi-organised single strand and posterior base pairing and stacking of one base at the 5' end [Mohan et al., 2009]. The distinct behaviour of the 3' and 5' ends in RNA unwinding was also observed in molecular dynamics simulations of RNA unwinding [Colizzi and Bussi, 2012] and, more recently, the same authors investigated the asymmetric unwinding by helicases

of two thermodynamically identical duplexes using molecular dynamics, which showed a preferential unwinding of those dsDNAs and dsRNAs carrying purine bases in the leading strand. In order to experimentally support this investigation we carried out work on the asymmetric unwinding of DNA duplexes by two well-characterised model helicases using the xeroderma pigmentosum group D (XPD), a SF2 helicase with 5'-3' polarity that belongs to the Rad3/XPD family, and the plasmid copy number reduction A (PcrA), a SF1 helicase with 3'-5' polarity from the UvrD/Rep family.

XPD helicase is one of the ten subunits of the transcription factor IIH (TFIIH) in eukaryotes, which is required for transcription initiation and nucleotide excision repair (NER), although XPD is only involved in DNA repair [Coin et al., 2007; Lainé et al., 2006; Winkler et al., 2000]. XPD is responsible for DNA unwinding in the repair process of lesions such as photoinduced and oxidative damage. Its crucial role in NER is underlined by the three diseases that result from mutations in the *XPD* gene: xeroderma pigmentosum (XP), which produces several skin abnormalities due to the impossibility to repair DNA damage caused by UV light; and the multisystem disorders trichothiodystrophy (TTD) and XP combined with Cockayne Syndrom (XP/CS) [Lehmann, 2001].

The study of XPD structure and functions has been mostly confined to archaeal XPD because of the high similarities in sequence with its human homologue with which it shares the catalytic core of the enzyme. In particular, the use of X-ray crystallography has provided tremendous insights into the apo and bound crystal structures [Fan et al., 2008; Liu et al., 2008; Wolski et al., 2008]. Archaeal XPD comprises two RecA-like helicase domains (HD1, HD2) and two auxiliary domains (Arch and 4FeS) inserted in the HD1 fold (Fig. 7.1a). The Arch domain is denoted by its characteristic structure and shows strong interactions with the 4FeS domain, which is stabilised by an iron-sulfur 4Fe-4S cluster that plays a key role in DNA recognition and unwinding. The 4Fe-4S cluster is involved in a redox process associated with a charge transport between proteins bound to DNA in which the charge is transferred through the DNA bases and, hence, is influenced by lesions in the DNA that alter the base stacking. The changes in the redox potential associated to lesions in the DNA are used by XPD to locate this regions [Fuss et al., 2015; Mui et al., 2011; Sontz et al., 2012].



(a) XPD **(b)** PcrA

Fig. 7.1 Structures of XPD and PcrA helicases. Crystal structures of (a) XPD SF2Bα (PDB ID:5H8W) [Constantinescu-Aruxandei et al., 2016]; and (b) PcrA SF1Aα helicase substrate complex (PDB ID: 3PJR) [Velankar et al., 1999].

The DNA binding and unwinding mechanism of XPD, however, is not totally understood. The binding mechanism to single-stranded DNA at the repair bubble implies a tight binding to the HD2 [Constantinescu-Aruxandei et al., 2016] followed by coordinated conformational changes between HD2 and HD1 as well as between the Arch, 4FeS and HD1 domains that allows the formation of a pore through which the ssDNA leading strand is directed [Fan et al., 2008; Kuper et al., 2011; Wolski et al., 2008]. The conformational changes leading to DNA loading were studied using smTIR by attaching a Cy3 dye to the Arch domain of *Ferroplasma acidarmanus* XPD (FaXPD) and monitoring of the quenching produced by the 4Fe-4S cluster [Ghoneim and Spies, 2014]. They observed transitions between an open and close state as a result of the spatial separation of the Arch and 4FeS domains in which the dwell time of the closed conformation is 3 times longer than that of the open state. Interestingly, the DNA binding is independent of the structure and thus, independent from the formation of the pore.

The interaction between the Arch and 4FeS domains was investigated using ensemble fluorescence techniques in order to monitor the real time unwinding of a dsDNA labelled with a Cy3 dye and a Dabcyl quencher [Constantinescu-Aruxandei et al., 2016]. These assays were

performed using an intramolecular covalent cross-linking between these two domains to preclude their rearrangement, which resulted in the inability of this construct to translocate DNA through the pore, although DNA binding remained unaffected [Constantinescu-Aruxandei et al., 2016]. The unwinding mechanism of XPD was also investigated by force spectroscopy using optical tweezers with base pair resolution [Qi et al., 2013]. The results obtained by Qi and coworkers show that XPD advances one base pair at a time but also shows one step movements backwards as well as 5-bp backward and forward transitions. Interestingly, the frequency of the backwards steps depends on the availability of ATP and DNA sequence, which underlines the importance of DNA in the unwinding process.

On the other hand, PcrA is an essential protein found in many bacteria with DNA repair and plasmid replication functions [Petit et al., 1998]. Its monomeric structure can readily translocate ssDNA with 3'-5' polarity but its ability to unwind a DNA duplex is limited *in vitro*, although greatly improved in a complex with accessory proteins such as RepD [Chisty et al., 2013; Niedziela-Majka et al., 2007; Slatter et al., 2009; Soultanas et al., 1999]. Furthermore, the dimeric [Yang et al., 2008] and oligomeric [Byrd and Raney, 2005] forms of PcrA also display significant helicase activity.

The crystal structure of PcrA shows two domains, 1A and 2A, each carrying an inserted domain, 1B and 2B, respectively. Velankar and coworkers studied the structure of PcrA complexed with a 10-nt long DNA duplex carrying a $(dT)_7$ single-stranded 3' overhang [Velankar et al., 1999]. The crystal structure of this complex obtained together with the non-hydrolysable ATP analogue ADPNP, which is bound at the binding site located between domains 1A and 2A, mimics the conformation of the "substrate" complex (Fig. 7.1b). The second structure reported contains a sulphate SO_4^{2-} ion at the position that a phosphate resulting from ATP hydrolysis would occupy, thus representing a "product" conformation. Both structures represent two snapshots of different steps along the unwinding pathway. The overlap of the two representations show that domains 1A and 2A adopt an open conformation after ATP hydrolysis. At this point, Velankar and coworkers suggest that domain 2A looses its grip with the single strand whilst 1A remains tightly bound to the ssDNA and pulls it through 2A to undergo translocation. The binding of the ADPNP promotes conformational

changes in which 1A and 2A domains come closer and domains 1B and 2B adopt a structure able to bind to the dsDNA region. In this conformation, the protein destabilises the double helix structure that, combined with the translocation of the single strand, results in the final unwinding of the duplex [Velankar et al., 1999].

A kinetic model for the translocation mechanism of PcrA was proposed based on ensemble fluorescence data in which the release of phosphate groups as products of ATP hydrolysis was monitored in the presence of a fluorescently labelled phosphate binding protein [Dillingham et al., 2000]. After binding to the ssDNA, PcrA advances unidirectionally at single base steps with hydrolysis of 1 ATP molecule per step at a rate of $50\,s^{-1}$. Once PcrA reaches the 5' end, it hydrolyses ATP at a slower rate of $6\,s^{-1}$ without DNA translocation and the complex can dissociate and later reassociate to start again. Further studies investigated the activity of PcrA monomer in the presence of RepD using ensemble and single-molecule TIR fluorescence techniques [Chisty et al., 2013]. Some complexes were able to unwind dsDNA in a similar rate as for the translocation of the ssDNA, thus unwinding in an active way [Manosas et al., 2010], although most complexes show unwinding rates several times slower than translocation of ssDNA, which corresponds to a passive unwinding mechanism. These results suggest that PcrA unwinding, when associated to RepD, is a highly heterogeneous process in which the helicase acts in a partially passive mechanism [Chisty et al., 2013; Manosas et al., 2010].

The experiments presented in this chapter aim to underline the importance of the DNA sequence in helicase activity by monitoring the unwinding of three sets of two thermodynamically identical strands, carrying either purine or pyrimidine bases on the leading strand, by XPD and PcrA. This project was carried out in collaboration with Giovanni Bussi's group (Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy), who performed computational studies on the asymmetric unwinding of DNA by helicases from the superfamilies SF1 and SF2 to investigate the different behaviour of purine and pyrimidine bases, and the Malcolm White lab (School of Biology, University of St Andrews), who performed the expression and purification of the two proteins. We have complemented their work by using ensemble fluorescence kinetic assays to monitor the separation of the dsDNA in real time

by attaching a Cy3 dye and a Dabcyl quencher at the end of the duplex and corroborated by activity gel assays the sequence-specific activity observed for both helicases.

7.2 Materials and methods

7.2.1 Preparation of labelled DNA substrates

Labelled and unlabelled DNA strands were purchased from IDT. Dry DNA pellets were resuspended in 50 mM Tris-HCl (pH 7.5) to a final concentration of 100 µM and stored at -20 °C. The DNA strands used in helicase unwinding assays incorporate a Cy3 dye at 5' and Dabcyl quencher at 3' ends and were designed for quenching of the Cy3 signal in the DNA duplex structure (Table 7.1; Figs. 7.3a and 7.6).

The dsDNA substrates were annealed in Tris-HCl buffer containing a background of 50 mM NaCl. Samples were heated in water bath to 80 °C for 10 minutes followed by slow cooling overnight to room temperature. In order to separate any remaining ssDNA, the sample is purified by gel electrophoresis on a nondenaturing 12% (w/v) polyacrylamide gel with 240 µL ammonium persulfate (APS) and 60 µL tetramethylethylenediamine (TEMED). The double-labelled bands were cut from the gel, chopped in a tube, resuspended in 50 mM Tris-HCl buffer (pH 7.5) and left overnight on a rocker in cold room (4 °C). The dsDNA was then precipitated by addition of 10% (v/v) of 3 M sodium acetate (pH 5.2) and 250% (v/v) of cold absolute ethanol, gently mixed and incubated overnight at -20 °C. This mix was then centrifuged at maximum speed for 30 min in cold room. The supernatant was removed and the DNA pellet was re-suspended in either 20 mM MES pH 6.5, for XPD_S and XPD_L, or 20 mM Tris-HCl pH 7.4 for PcrA DNA substrates.

The final concentration of DNA was determined with the NanoDrop 2000 UV-Vis spectrophotometer using the molar extinction coefficient of the DNA at the absorption maximum due to the spectral overlap of Cy3 and Dabcyl, which precludes the quantification of the concentration by using the molar extinction coefficient of the dye.

Table 7.1 Synthetic strands and labelling positions used in the XPD and PcrA unwinding kinetics and activity gel assays.

	Strand	Sequence (from 5' to 3')
Unwinding kinetics	ssXPD$_S^{Pu}$-Dab	AGCTACCATGCCTGCACGAATGAGGAAGGGAGGAAAGGAAGA-Dab
	ssXPD$_S^{Pu}$-Cy3	Cy3-TCTTCCTTTCCTCCCTTCCTC
	ssXPD$_S^{Py}$-Dab	AGCTACCATGCCTGCACGAATCTCCTTCCCTCCTTTCCTTCT-Dab
	ssXPD$_S^{Py}$-Cy3	Cy3-AGAAGGAAAGGAGGGAAGGAG
	ssXPD$_L^{Pu}$-Dab	CAATCATGCACAATCAGTCACCTCACGTAGAGGAAGGGAGGAAAGGAAGACCAG-Dab
	ssXPD$_L^{Pu}$-Cy3	Cy3-CTGGTCTTCCTTTCCTCCCTTCCTCTACGTGAGG
	ssXPD$_L^{Py}$-Dab	CAATCATGCACAATCAGTCACCTCACGTACTCCTTCCCTCCTTTCCTTCTCCAG-Dab
	ssXPD$_L^{Py}$-Cy3	Cy3-CTGGAGAAGGAAAGGAGGGAAGGAGTACGTGAGG
	ssPcrA^{Pu}-Cy3	Cy3-AGAAGGAAAGGAGGGAAGGAGTAAGCACGTCCGTACCATCGA
	ssPcrA^{Pu}-Dab	CTCCTTCCCTCCTTTCCTTCT-Dab
	ssPcrA^{Py}-Cy3	Cy3-TCTTCCTTTCCTCCCTTCCTCTAAGCACGTCCGTACCATCGA
	ssPcrA^{Py}-Dab	GAGGAAGGGAGGAAAGGAAGA-Dab
Gel assays	ssXPD$_S^{Pu}$	AGCTACCATGCCTGCACGAATGAGGAAGGGAGGAAAGGAAGA
	ssXPD$_S^{Pu}$-Cy3	Cy3-TCTTCCTTTCCTCCCTTCCTC
	ssXPD$_S^{Py}$	AGCTACCATGCCTGCACGAATCTCCTTCCCTCCTTTCCTTCT
	ssXPD$_S^{Py}$-Cy3	Cy3-AGAAGGAAAGGAGGGAAGGAG
	ssPcrA^{Pu}-Cy3	Cy3-AGAAGGAAAGGAGGGAAGGAGTAAGCACGTCCGTACCATCGA
	ssPcrA^{Pu}	CTCCTTCCCTCCTTTCCTTCT
	ssPcrA^{Py}-Cy3	Cy3-TCTTCCTTTCCTCCCTTCCTCTAAGCACGTCCGTACCATCGA
	ssPcrA^{Py}	GAGGAAGGGAGGAAAGGAAGA

7.2.2 Expression and purification of TaXPD and PcrA helicases

The expression and purification of TaXPD and PcrA was performed by Biljana Petrovic Stojanovska, from the Malcolm White lab, as described in [Constantinescu-Aruxandei et al., 2015].

7.2.3 DNA melting assays

DNA melting assays were performed using the built-in thermal application of a Cary Eclipse spectrophotometer with excitation at 547 nm and emission collected at 565 nm. The heating rate was set at 0.2 °C/min with data collection interval of 0.2 °C and an integration time of 1 s for each data point.

Thermal denaturation of 50 nM XPD_S and XPD_L strands was performed in 20 mM MES (pH 6.5) containing at 1 mM $MgCl_2$, whilst PcrA strands were studied at a concentration of 50 nM in 20 mM Tris-HCl buffer (pH 7.4) containing 200 mM NaCl and 1 mM $MgCl_2$. The final volume of all samples was 150 µL. A layer of mineral oil was finally deposited on top to prevent evaporation during the experiment.

The results are represented as the mean of 3 replicates with baseline correction of the dsDNA and ssDNA regions by fitting to a straight line, y_{min} and y_{max}, respectively. The corrected spectrum can be calculated from

$$I = \frac{I^* - y_{min}}{y_{max} - y_{min}} \tag{7.1}$$

with I^* representing the raw intensity.

The final curve is then fitted using the Boltzmann sigmoidal equation from which the melting temperature, T_m, is obtained as the temperature at which the unwinding reaches the 50%

$$I = I_{min} + \frac{I_{max} - I_{min}}{1 + \exp\left(\frac{T_m - T}{s}\right)} \tag{7.2}$$

where I_{min} and I_{max} are the minimum and maximum normalised intensity values, respectively, and s is the slope of the melting curve.

7.2.4 DNA unwinding kinetics

DNA unwinding assays were performed using the kinetics program of a Cary Eclipse spectrophotometer (Agilent Technologies LDA UK Ltd., Stockport, UK) with excitation at 547 nm and emission at 565 nm. All kinetics experiments were carried out at room temperature (20 °C) with continuous excitation and an average time of 0.5 s.

Unwinding kinetic assays of XPD_S and XPD_L (Table 7.1) were performed for DNA concentrations of 50 nM and 500 nM XPD protein in XPD helicase buffer (20 mM MES pH 6.5, 0.1 mg/mL bovine serum albumin (BSA) and 1 mM $MgCl_2$). This mix was incubated for

10 minutes and the reaction was started by addition of either 100 µM or 1 mM ATP with a final volume of 150 µL.

Unwinding kinetic assays of PcrA dsDNA strands (Table 7.1) were performed for DNA concentrations of 50 nM and 500 nM PcrA protein in PcrA helicase buffer (20 mM Tris-HCl pH 7.4, 0.1 mg/mL BSA, 1 mM $MgCl_2$ and 200 mM NaCl). This mixture was incubated for 10 min. before addition of either 10 or 100 µM ATP, with a final volume of 150 µL.

Helicase buffer contains $MgCl_2$, which coordination to ATP plays an important role in ATP binding and hydrolysis, and BSA, which prevents non-specific hydrophobic interactions. DNA strands were incubated together with 10 X excess of protein to allow efficient helicase activity. Addition of NaCl to PcrA unwinding assays was necessary to avoid precipitation of the protein, despite its effect in preventing DNA binding.

The unwinding curves for XPD were then fitted to the expression

$$y = A_1(1 - e^{(-k_1 t)}) + A_2(1 - e^{(-k_2 t)}) \tag{7.3}$$

where A represents the amplitude and k is the rate of unwinding. However, the unwinding rates of PcrA were obtained by fitting to a single exponential rise.

7.2.5 Activity gel assays

Helicase activity determined by activity gel assays was performed with DNA substrates carrying the Cy3 dye but lacking the Dabcyl quencher (Table 7.1).

The reaction mix containing DNA and either XPD or PcrA helicase was incubated in an eppendorf for 10 minutes as described in section 7.2.4. This solution was then split in three different tubes. One tube was separated as a control and ATP was added to the other two samples with final concentrations of 100 µM and 1 mM ATP for XPD, or 10 and 100 µM ATP for PcrA. Moreover, as a control for the migration of ssDNA strands labelled with Cy3, ssDNAs were diluted to a final concentration of 50 nM in either XPD or PcrA helicase buffer.

The XPD unwinding assay was stopped at plateau conditions (25 min for 1 mM ATP and 50 min for 100 µM ATP) by addition of 10 µL of the reaction mix to 10 µL of stop solution

(10 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.5% SDS, 1 mg/mL Proteinase K). The samples were then incubated for 15 min at room temperature. The PcrA unwinding assay was stopped after 1:30 hours for both 10 and 100 µM ATP conditions and stopped by addition of 10 µL of the reaction mix to 10 µL of stop solution. The mix was then incubated for 15 minutes at room temperature. Furthermore, 10 µL of control mixtures were also added to 10 µL of stop mix with further incubation at RT for 15 minutes. Finally, 5 µL Ficoll 15% were added to 10 µL of the previous solutions for a final DNA concentration of 0.25 pM.

The unwinding rate was investigated by gel electrophoresis in a nondenaturing 12% (w/v) polyacrylamide gel with 240 µL APS and 60 µL TEMED. Gels were pre-run for 1 h in Tris/Borate/EDTA buffer (TBE) at 110 V before DNA was loaded. Gels were visualised in a Typhoon FLA 7000 (GE Healthcare Life Sciences, Little Chalfont, UK) gel imaging system with exctitation at 532 nm. The ssDNA and dsDNA bands were associated by comparison with the migration profile of the controls. The intensity of each band was then extracted using the 1D tool software provided with manual background subtraction.

7.3 Results and discussion

7.3.1 Relative thermodynamic stability of DNA substrates calculated by melting analysis

The stability under experimental conditions of the three pairs of strands was investigated by DNA melting assays. The theoretical melting temperature, or temperature at which the unwinding reaches 50%, is given by the length and content of AT and GC base pairs. GC-rich sequences display higher stability due to the three hydrogen bonds between both bases, as opposed to the two hydrogen bonds between AT bases. However, the strands used in our unwinding assays carry the same number of AT and GC base pairs because they are equal in terms of sequence, although on opposite strands and, therefore, with different polarity. We have monitored the increase in the Cy3 signal of the Cy3-Dab labelled strands as a function of temperature, a method that does not require the labelling with fluorescent DNA intercalators

that would interact with the nucleobases, and thus stabilise or destabilise the duplex. To improve the accuracy of the data, these assays were performed with a background of 1 mM $MgCl_2$ as well as 200 mM NaCl for PcrA strands to match the experimental conditions of ATPase unwinding assays.

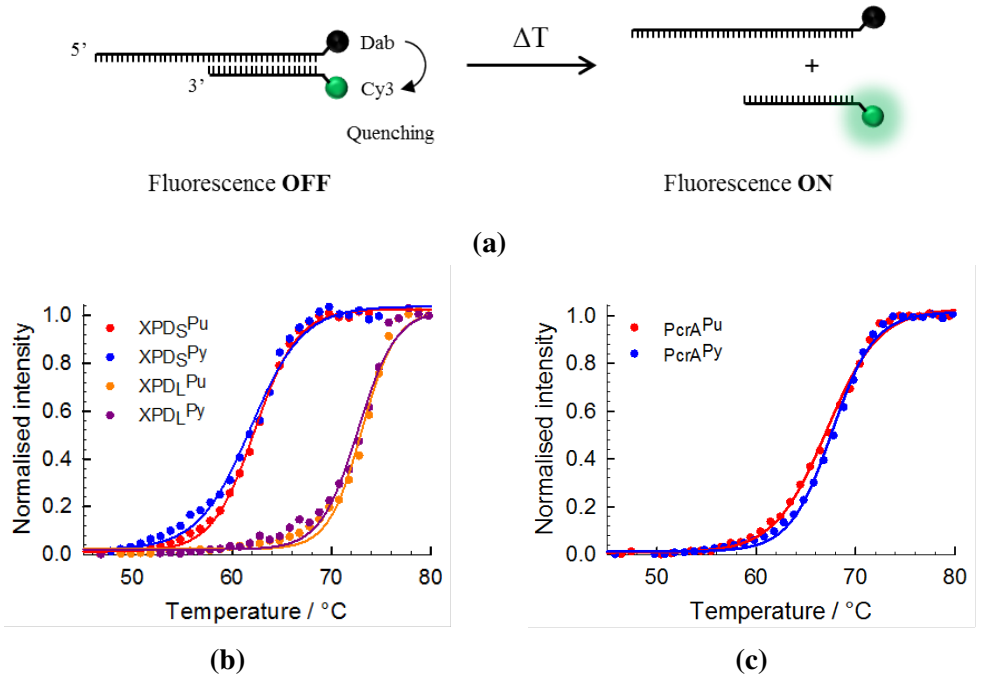


Fig. 7.2 Melting curve analysis of DNA strands. (a) Schematic representation of the melting assay for a Cy3-Dab labelled substrate. (b) Thermal denaturation assays of XPD_S and XPD_L dsDNA strands. Melting profiles were fitted to Eq. 7.2 (solid lines). (c) Thermal denaturation profiles of $PcrA^{Pu}$ and $PcrA^{Py}$. Melting curves were fitted to Eq. 7.2 (solid lines).

The melting curves obtained were then corrected and fitted as described in section 7.2.3. The melting profiles of the short $XPD_S{}^{Pu}$ and $XPD_S{}^{Py}$ strands show the same melting temperature in both cases. Similar results were obtained for the long XPD_L strands, which confirms the expected equal thermal stability of both pairs of sequences (Fig. 7.2b). However, the dsDNAs with 3' overhang used for PcrA unwinding do show little variation in which the strand with pyrimidine bases on the leading strand $PcrA^{Py}$, or strand that carries and is translocated by the helicase, is slightly more stable (T_m=67.6 °C) than the the $PcrA^{Pu}$ strand (T_m=67.0 °C) (Fig. 7.2c). Although the shift is ~0.6 °C, it suggest a slight difference in the

unwinding mechanism of purine- and pyrimidine-rich sequences under these conditions, in which the leading strand containing purine bases is more readily unwound.

Table 7.2 Melting temperature of the strands used in the XPD and PcrA unwinding assays. Experiments were carried out under ATPase experimental conditions.

	dsDNA length	[NaCl]	T_m
XPD_S^{Pu}	21 nt	-	62.3 ± 0.3 °C
XPD_S^{Py}	21 nt	-	61.9 ± 0.2 °C
XPD_L^{Pu}	34 nt	-	73.0 ± 0.3 °C
XPD_L^{Py}	34 nt	-	72.7 ± 0.1 °C
$PcrA^{Pu}$	21 nt	200 mM	67.0 ± 0.1 °C
$PcrA^{Py}$	21 nt	200 mM	67.6 ± 0.0 °C

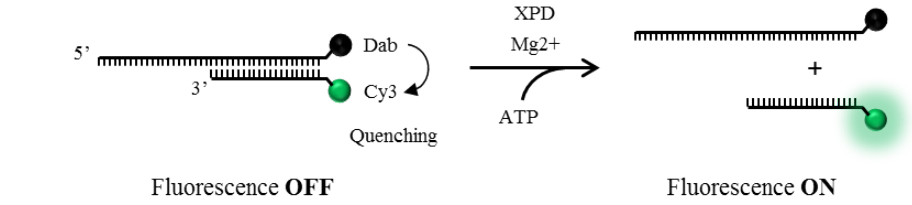
Data are mean±SEM (n=3).

7.3.2 Real-time DNA unwinding kinetics

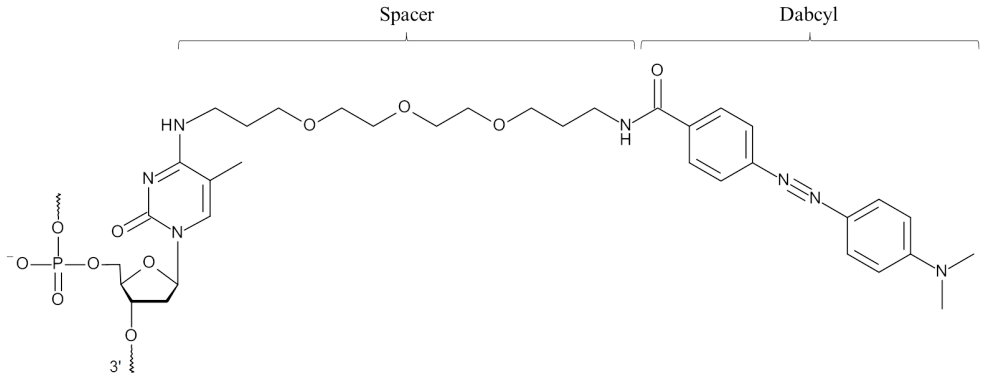
The ability of TaXPD to unwind a set of dsDNAs was firstly investigated for the two short Cy3-Dab labelled strands carrying either purine, XPD_S^{Pu}, or pyrimidine, XPD_S^{Py}, bases in the leading strand (Fig. 7.3). Similar fluorescence-based assays were used to investigate the mechanism of XPD by monitoring the unwinding amplitude of the wild-type, an inactive mutant and an Arch-4FeS cross-linked complex [Constantinescu-Aruxandei et al., 2016].

We monitored the activity TaXPD by addition of 100 µM ATP in a slow process that shows an analogue unwinding curve for XPD_S^{Pu} and XPD_S^{Py} with an amplitude of ~7 and ~7.6 after 50 minutes, respectively (Fig. 7.4a).

The unwinding rate of TaXPD was then investigated by addition of 1 mM ATP, saturating ATP conditions for XPD [Constantinescu-Aruxandei et al., 2016], in a fast process that may depend on the unwinding mechanism of XPD and nucleic acid dynamics. The unwinding amplitude for XPD_S^{Pu} increased ~2.2-fold in relation to the curve obtained at 100 µM ATP, reaching a plateau at 15.5 in 25 minutes (Fig. 7.4b). On the other hand, the amplitude for XPD_S^{Py} only increased 1.4-fold, reaching a plateau at 10.5. The similar amplitude obtained for 100 µM ATP differs from the 48% difference shown at 1 mM ATP that underlines the



(a)



(b)

Fig. 7.3 Fluorescence-based unwinding assays by XPD. (a) Schematics representation of the 5'→3' unwinding by XPD upon addition of ATP. The DNA substrates are composed of a either a 21-nt dsDNA with 21-nt 5' overhang (XPD_S) or 34-nt dsDNA with 20-nt 5' overhang (XPD_L) at which ends are attached a Cy3 dye and Dabcyl quencher. (b) Molecular structure of a Dabcyl quencher incorporated via spacer at the 3' end of a nucleic acid strand.

dependency of TaXPD activity on the DNA sequence with preferential unwinding of the dsDNA carrying purine bases in the leading strand as reported by Taylor and coworkers [Taylor et al., 2010]. Interestingly, the fastest rate of unwinding for $XPD_S{}^{Pu}$ experiences a 64% increase at 1 mM when compared with the data at 100 µM (from, whilst the rate for $XPD_S{}^{Py}$ barely increases. As regards to the slowest rate, the two strands show similar values at both ATP concentrations (Table 7.3).

In order to further investigate the sequence-dependent unwinding of TaXPD we performed ATPase assays for two 34-nt dsDNA strands with 21-nt 5' overhangs. The duplex contains the same 21-nt sequences investigated in XPD_S dsDNAs, which are inserted within a segment containing both purine and pyrimidine bases. Addition of 1 mM ATP leads to 11-fold increase in the intensity of $XPD_L{}^{Pu}$, whilst the strand containing a sequence of pyrimidine bases

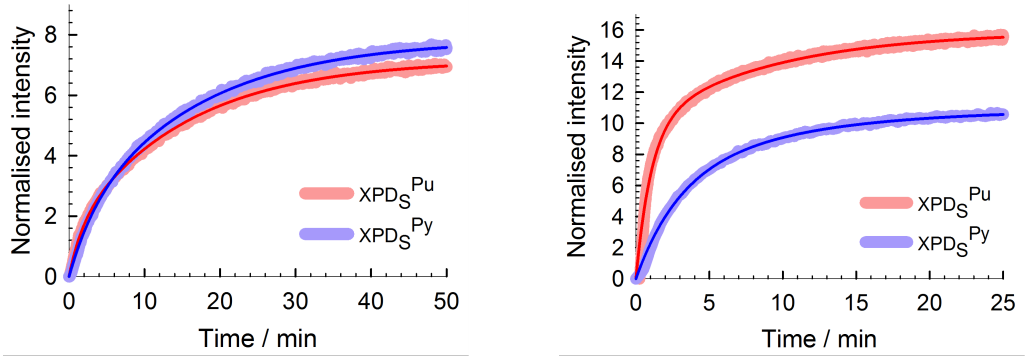


Fig. 7.4 Asymmetric unwinding kinetics of XPD_S strands by XPD. Time course of $XPD_S{}^{Pu}$ (red) and $XPD_S{}^{Py}$ (blue) unwinding by XPD in the presence of (a) 100 µM and (b) 1 mM ATP. The plots were obtained by averaging of 3 replicates. The data were then fitted using Equation 7.3 in order to obtain the unwinding rates (red and blue lines).

in the leading strand, $XPD_L{}^{Py}$, reaches a maximum unwinding amplitude of 7.5 after 25 minutes (Fig. 7.5). Although the unwinding rates show some differences, the amplitudes of both strands display an analogue behaviour as for the short dsDNA strands that showed remarkable evidence for sequence-dependent unwinding at 1 mM ATP. The decrease in the unwinding amplitude for XPD_S and XPD_L within the same timescale may be associated to the longer sequence of the latter strands, which is also responsible for the initial lag phase

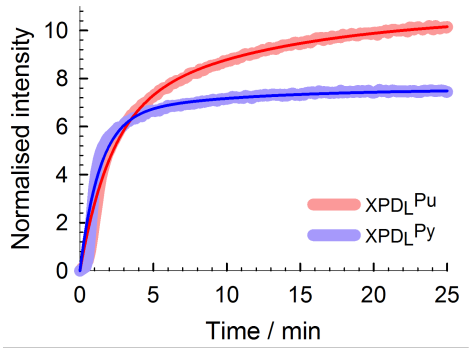


Fig. 7.5 Asymmetric unwinding kinetics of XPD_L strands by XPD. Time course of $XPD_L{}^{Pu}$ (red) and $XPD_L{}^{Py}$ (blue) unwinding by XPD at 1 mM ATP. The plots were obtained by averaging of 3 replicates. The data were then fitted to Eq. 7.3 in order to obtain the unwinding rates (red and blue lines).

Table 7.3 Unwinding rates and amplitudes obtained for XPD and PcrA helicases.

	[ATP]	A_1	k_1 / s^{-1}	A_2	k_2 / s^{-1}
$XPD_S{}^{Pu}$	100 µM	1.462 ± 0.021	0.584 ± 0.015	5.715 ± 0.018	0.066 ± 0.001
$XPD_S{}^{Py}$	100 µM	1.613 ± 0.045	0.311 ± 0.009	6.246 ± 0.040	0.062 ± 0.001
$XPD_S{}^{Pu}$	1 mM	9.826 ± 0.049	0.957 ± 0.008	6.081 ± 0.038	0.111 ± 0.001
$XPD_S{}^{Py}$	1 mM	6.245 ± 0.542	0.345 ± 0.018	4.651 ± 0.495	0.106 ± 0.010
$XPD_L{}^{Pu}$	1 mM	6.412 ± 0.332	0.483 ± 0.026	4.265 ± 0.246	0.084 ± 0.001
$XPD_L{}^{Py}$	1 mM	6.372 ± 0.252	0.755 ± 0.041	1.179 ± 0.213	0.114 ± 0.032
$PcrA^{Pu}$	100 µM	5.606 ± 0.003	0.00568 ± 0.00001	-	-
$PcrA^{Py}$	100 µM	N.D.	N.D.	-	-

Data are mean±SEM.
N.D.: not determined.

observed. Since the unwinding of the first base pairs does not affect the fluorescence signal, longer sequences show a longer initial phase.

To corroborate the generality of the asymmetric unwinding on other helicases we extended these ATPase assays to the study of PcrA, a helicase protein with 3'-5' polarity that belongs to the SF1 family (Fig. 7.6). The two 21-nt dsDNA strands with 21-nt 3' overhangs used in these studies, $PcrA^{Pu}$ and $PcrA^{Py}$, are analogues of $XPD_S{}^{Pu}$ and $XPD_S{}^{Py}$, respectively. Furthermore, we expected that the reduced unwinding activity of PcrA reported using a 18-nt dsDNA [Niedziela-Majka et al., 2007] would be translated in a slower processivity of that strand in which the opening of the replication fork is more restricted due to interactions with the DNA bases.

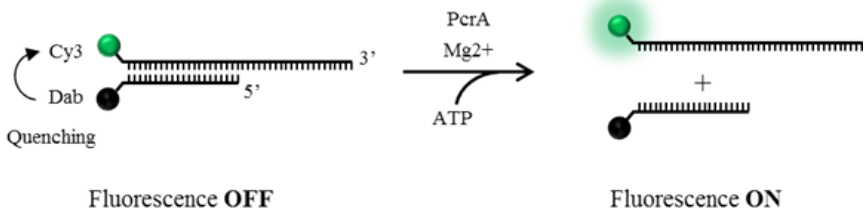


Fig. 7.6 Fluorescence-based unwinding assays by PcrA. Schematics of 3'→5' unwinding by PcrA helicase upon addition of ATP. DNA substrates are composed of a 21-nt dsDNA with a 21-nt 3' overhang.

The fluorescence increase of PcrA^{Pu} at 100 µM reaches an amplitude of ~5.2 after 450 minutes, whilst PcrA^{Py} show a significantly lower ~1.4 increase in which the reaction has not yet reached a plateau (Fig. 7.7). The timescale may reflect the low processivity of PcrA on dsDNA substrates as well as a weaker protein-DNA binding in the presence of 200 mM NaCl. The data for PcrA show a clear distinct rate and amplitude for both strands in which, as observed for TaXPD and for the thermal denaturation of the DNA substrates, the duplexes carrying purine bases in the leading strand are heavily preferred.

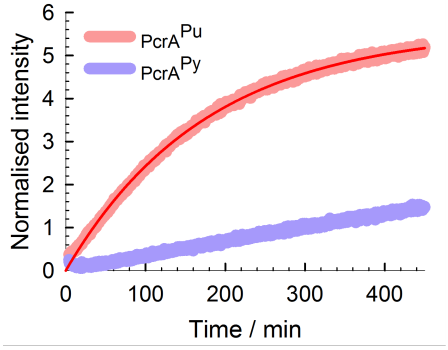


Fig. 7.7 Asymmetric unwinding kinetics by PcrA. Time course of PcrA^{Pu} (red) and PcrA^{Py} (blue) unwinding by PcrA at 100 µM ATP. The plots were obtained by averaging of 3 replicates. The data for PcrA^{Pu} were then fitted to a single exponential equation in order to obtain the unwinding rates (red line).

7.3.3 DNA unwinding amplitude determined by activity gel assays

Helicase activity was further investigated by activity gel assays to support the different unwinding amplitudes of XPD and PcrA helicases displayed for purine- and pyrimidine-rich leading strand sequences in unwinding kinetic assays. Since the results obtained for XPD$_S$ and XPD$_L$ were similar, we limited these assays to the XPD$_S$ and PcrA strands lacking the Dabcyl quencher for the correct quantification of the ssDNA and dsDNA band intensities arising from the Cy3 fluorescence signal, which is correlated to the amount of DNA present.

XPD gel assays were performed under the same conditions as the kinetic experiments and the reactions were stopped at the plateau (25 min for 1 mM ATP and 50 min for 100 µM ATP). Under these conditions, the data show a comparable unwinding amplitude at 100 µM

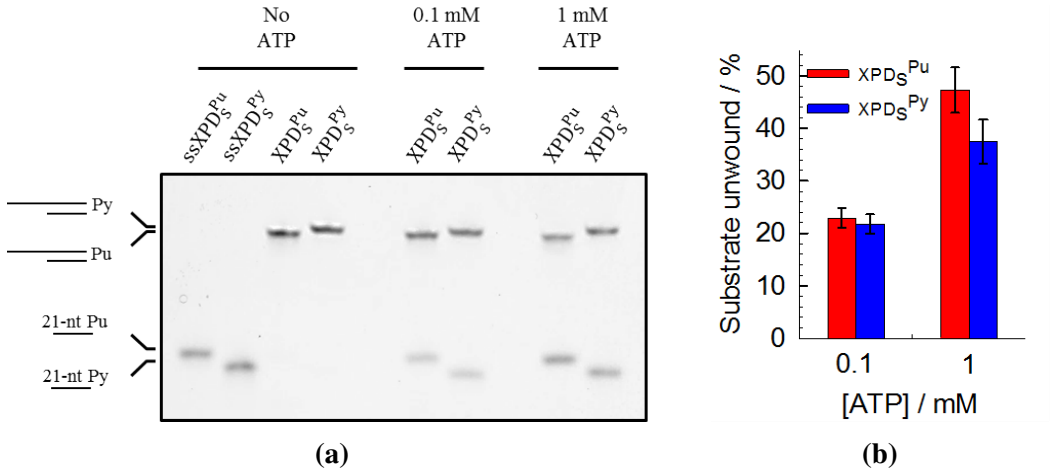


Fig. 7.8 Asymmetric unwinding gel assays by XPD. (a) Representative activity gel assay for determination of the unwinding amplitudes of XPD_S^{Pu} and XPD_S^{Py} strands by XPD at 100 µM and 1 mM ATP. Reactions were stopped at plateau conditions (25 min for 1 mM ATP and 50 min for 100 µM ATP) by addition of stop mix. (b) Unbound substrate obtained at each of the conditions. Data represent the mean for three replicates, whilst error bars show the SEM.

for XPD_S^{Pu} and XPD_S^{Py} with ~23% and ~22% of substrate unwound, respectively (Fig. 7.8). At 1 mM ATP, the unwinding amplitude reaches a ~47% for XPD_S^{Pu}, whilst XPD_S^{Py} shows a ~37% of unwound substrate (Fig. 7.8). In this case, TaXPD continues to show a preferential unwinding of those purine-rich leading strands with a ~26% higher amplitude at 1 mM ATP, as opposed to the ~48% obtained by kinetic assays.

PcrA gel activity experiments were carried out by addition of 10 µM and 100 µM ATP and stopped after 90 minutes. To avoid rewinding of the DNA strands in the presence of Na^+ and SDS, these assays were performed in the absence of NaCl, which increases the affinity for the DNA substrate and, presumably, would also increase the unwinding amplitude. At 10 µM ATP, $PcrA^{Pu}$ shows a ~16% unbound substrate, whereas $PcrA^{Py}$ shows a ~11% of unbound DNA (Fig. 7.9). The difference of ~47% under these conditions is increased up to ~100% at 100 µM ATP, when the unbound substrate reaches a total of ~44% and ~22% for $PcrA^{Pu}$ and $PcrA^{Py}$, respectively (Fig. 7.9). This difference is less than the ~270% expected according to kinetic ATPase assays at 200 mM NaCl, which suggest that the presence of Na^+

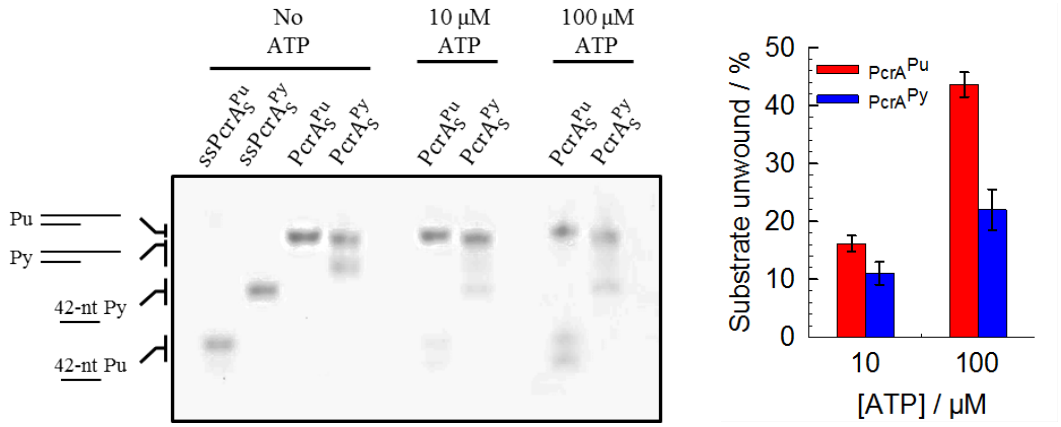


Fig. 7.9 Asymmetric unwinding gel assays by PcrA. (a) Representative activity gel assay for the determination of the unwinding amplitudes of PcrA^{Pu} and PcrA^{Py} strands by PcrA at 10 and 100 µM ATP. Reactions were stopped after 1:30 h by addition of stop mix. (b) Unbound substrate obtained at each of the conditions. Data represent the mean for three replicates, whilst error bars show the SEM.

salt may accentuate the differences between the processivity of both strands as already seen in the thermal denaturation profiles (Fig. 7.2c).

Gel activity data are in good agreement with those obtained in ATPase kinetics assays. The separation of a double-stranded DNA by either SF1 or SF2 helicases show a strong dependence on the nucleic acid sequence, despite the different unwinding mechanism of each superfamily. The rates and amplitudes observed in both cases by either kinetics or gel activity assays suggest that the helicase unwinding mechanism go beyond the view of DNA as a mere substrate to become an active compound that plays a crucial role in the process. The importance of the thermodynamic stability of the substrates may be analysed together with the stability and conformation of the intermediates generated at the replication fork, which would elucidate the possible role of nucleic acid sequences in gene regulation.

7.4 Conclusions

The cellular machineries involved in replication, recombination, repair, transcription and translation of nucleic acids require the opening of the double-stranded DNA or RNA structure

in order to read the information stored or repair damaged bases. The separation of the two complementary strands is performed by helicases, a group of ATPase proteins that use the energy released by ATP hydrolysis to translocate single-stranded or unwind double-stranded structures. The structure of helicase proteins has long been studied using X-ray crystallography, whilst the use of fluorescence ensemble and single-molecule techniques, among others, has greatly contributed to our understanding of helicase activity and their role in these essential cellular processes.

The vast majority of the studies are focused on the enzymatic process and do not account for the structure and stability of specific nucleic acid sequences. However, the increasing evidence on the importance of nucleic acid sequence and dynamics in helicase activity and how they modulate the enzymatic functions has attracted the attention of researchers in the last decade with the aim to provide insights into the unwinding mechanism.

The work presented in this chapter takes advantage of gel electrophoresis assays and ensemble fluorescence techniques to investigate the different unwinding rates and amplitudes of thermodynamically identical strands by XPD, a helicase from superfamily SF2 with 5'-3' polarity, and PcrA, a SF1 helicase that translocates DNA in the 3'-5' direction. The differences shown in both cases may be associated to the presence of either purine or pyrimidine bases in the leading strand. The results obtained for XPD at 100 μM ATP were unable to discriminate between the purine- or pyrimidine-rich leading strand, showing a similar unwinding amplitude in both duplexes. On the contrary, at high ATP concentrations the experiments show a significant preferential unwinding for DNA substrates carrying purine bases on the leading strand over those carrying pyrimidine bases (i.e. up to ~42% greater in XPD at 1 mM ATP and ~260% in PcrA at 100 μM ATP), in agreement with previous studies in the field.

The data emerged from this project represents a huge step forward in the characterisation of the biased helicase unwinding of purine- and pyrimidine-rich leading strand sequences. We demonstrate that the zipper model, which only depends on the thermodynamic stability of the nucleic acid sequences, should be amended to consider the differences that arise as a result of an asymmetric unwinding of double-stranded sequences and to represent the

different dynamics shown in both types of nucleobases. This work exhibits the influence of nucleic acid sequence in the unwinding process, which may suggest biological implications of the asymmetric behaviour of the 3' and 5' ends as proposed by Colizzi and Bussi with data based on molecular dynamics simulations of RNA [Colizzi and Bussi, 2012] and DNA unwinding [Colizzi and Bussi, pers. comm]. Thus, our results motivate the need to further investigate the mechanism of duplex unwinding and how this could be used to control the genetic expression by geometry-dependent unwinding of an specific sequence.

References

Byrd, A. K. and Raney, K. D. (2005). Increasing the length of the single-stranded overhang enhances unwinding of duplex DNA by bacteriophage T4 Dda helicase. *Biochemistry*, 44(39):12990–12997.

Cheng, W., Dumont, S., Tinoco, I., and Bustamante, C. (2007). NS3 helicase actively separates RNA strands and senses sequence barriers ahead of the opening fork. *Proceedings of the National Academy of Sciences*, 104(35):13954–13959.

Chisty, L. T., Toseland, C. P., Fili, N., Mashanov, G. I., Dillingham, M. S., Molloy, J. E., and Webb, M. R. (2013). Monomeric PcrA helicase processively unwinds plasmid lengths of DNA in the presence of the initiator protein RepD. *Nucleic acids research*, 41(9):5010–5023.

Coin, F., Oksenych, V., and Egly, J.-M. (2007). Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Molecular cell*, 26(2):245–256.

Colizzi, F. and Bussi, G. (2012). RNA unwinding from reweighted pulling simulations. *Journal of the American Chemical Society*, 134(11):5173–5179.

Constantinescu-Aruxandei, D., Petrovic-Stojanovska, B., Penedo, J. C., White, M. F., and Naismith, J. H. (2016). Mechanism of DNA loading by the DNA repair helicase XPD. *Nucleic acids research*, page gkw102.

Constantinescu-Aruxandei, D., Petrovic-Stojanovska, B., Schiemann, O., Naismith, J. H., and White, M. F. (2015). Taking a molecular motor for a spin: helicase mechanism studied by spin labeling and PELDOR. *Nucleic acids research*, page gkv1373.

Dillingham, M. S., Wigley, D. B., and Webb, M. R. (2000). Demonstration of unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed. *Biochemistry*, 39(1):205–212. PMID: 10625495.

References

Dillingham, M. S., Wigley, D. B., and Webb, M. R. (2002). Direct measurement of single-stranded DNA translocation by PcrA helicase using the fluorescent base analogue 2-aminopurine. *Biochemistry*, 41(2):643–651.

Fairman-Williams, M. E., Guenther, U.-P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Nucleic acids / Sequences and topology*, 20(3):313–324.

Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. S., Hammel, M., Roberts, V. A., Cooper, P. K., and Tainer, J. A. (2008). XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell*, 133(5):789–800.

Fuss, J. O., Tsai, C.-L., Ishida, J. P., and Tainer, J. A. (2015). Emerging critical roles of Fe–S clusters in DNA replication and repair. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1853(6):1253–1271.

Ghoneim, M. and Spies, M. (2014). Direct correlation of DNA binding and single protein domain motion via dual illumination fluorescence microscopy. *Nano Letters*, 14(10):5920–5931.

Jia, H., Korolev, S., Niedziela-Majka, A., Maluf, N. K., Gauss, G. H., Myong, S., Ha, T., Waksman, G., and Lohman, T. M. (2011). Rotations of the 2B sub-domain of E. coli UvrD helicase/translocase coupled to nucleotide and DNA binding. *Journal of Molecular Biology*, 411(3):633–648.

Johnson, D. S., Bai, L., Smith, B. Y., Patel, S. S., and Wang, M. D. (2007). Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. *Cell*, 129(7):1299–1309.

Jose, D., Weitzel, S. E., and von Hippel, P. H. (2012). Breathing fluctuations in position-specific DNA base pairs are involved in regulating helicase movement into the replication fork. *Proceedings of the National Academy of Sciences*, 109(36):14428–14433.

Kuper, J., Wolski, S. C., Michels, G., and Kisker, C. (2011). Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation. *EMBO Journal*, 31(2):494.

Lainé, J., Mocquet, V., and Egly, J. (2006). TFIIH enzymatic activities in transcription and nucleotide excision repair. *Methods in enzymology*, 408:246–263.

Lehmann, A. R. (2001). The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes & Development*, 15(1):15–23.

Liu, H., Rudolf, J., Johnson, K. A., McMahon, S. A., Oke, M., Carter, L., McRobbie, A.-M., Brown, S. E., Naismith, J. H., and White, M. F. (2008). Structure of the DNA repair helicase XPD. *Cell*, 133(5):801–812.

Manosas, M., Xi, X. G., Bensimon, D., and Croquette, V. (2010). Active and passive mechanisms of helicases. *Nucleic Acids Research*, 38(16):5518–5526.

Mohan, S., Hsiao, C., VanDeusen, H., Gallagher, R., Krohn, E., Kalahar, B., Wartell, R. M., and Williams, L. D. (2009). Mechanism of RNA double helix-propagation at atomic resolution. *The Journal of Physical Chemistry B*, 113(9):2614–2623.

Mui, T. P., Fuss, J. O., Ishida, J. P., Tainer, J. A., and Barton, J. K. (2011). ATP–stimulated DNA–mediated redox signaling by XPD, a DNA repair and transcription helicase. *Journal of the American Chemical Society*, 133(41):16378.

Niedziela-Majka, A., Chesnik, M. A., Tomko, E. J., and Lohman, T. M. (2007). Bacillus stearothermophilus PcrA monomer is a single-stranded DNA translocase but not a processive helicase in vitro. *Journal of Biological Chemistry*, 282(37):27076–27085.

Petit, M.-A., Dervyn, E., Rose, M., Entian, K.-D., McGovern, S., Ehrlich, S. D., and Bruand, C. (1998). PcrA is an essential DNA helicase of Bacillus subtilis fulfilling functions both in repair and rolling-circle replication. *Molecular Microbiology*, 29(1):261–273.

Phelps, C., Lee, W., Jose, D., von Hippel, P. H., and Marcus, A. H. (2013). Single-molecule FRET and linear dichroism studies of DNA breathing and helicase binding at replication fork junctions. *Proceedings of the National Academy of Sciences*, 110(43):17320–17325.

Pörschke, D. (1974a). A direct measurement of the unzippering rate of a nucleic acid double helix. *Biophysical chemistry*, 2(2):97–101.

Pörschke, D. (1974b). Model calculations on the kinetics of oligonucleotide double helix coil transitions. evidence for a fast chain sliding reaction. *Biophysical chemistry*, 2(2):83–96.

Qi, Z., Pugh, R. A., Spies, M., and Chemla, Y. R. (2013). Sequence-dependent base pair stepping dynamics in XPD helicase unwinding. *eLife*, 2:e00334.

Singleton, M. R., Dillingham, M. S., and Wigley, D. B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annual Review of Biochemistry*, 76:23–50.

Slatter, A. F., Thomas, C. D., and Webb, M. R. (2009). PcrA helicase tightly couples ATP hydrolysis to unwinding double-stranded DNA, modulated by the initiator protein for plasmid replication, RepD. *Biochemistry*, 48(27):6326–6334.

References

Sontz, P. A., Mui, T. P., Fuss, J. O., Tainer, J. A., and Barton, J. K. (2012). DNA charge transport as a first step in coordinating the detection of lesions by repair proteins. *Proceedings of the National Academy of Sciences*, 109(6):1856–1861.

Soultanas, P., Dillingham, M. S., Wigley, D. B., Papadopoulos, F., Phillips, S. E. V., and Thomas, C. D. (1999). Plasmid replication initiator protein RepD increases the processivity of PcrA DNA helicase. *Nucleic Acids Research*, 27(6):1421–1428.

Spies, M. (2014). Two steps forward, one step back: determining XPD helicase mechanism by single-molecule fluorescence and high-resolution optical tweezers. *Single Molecule Approaches: Watching DNARepair One Molecule at a Time*, 20:58–70.

Taylor, S. D., Solem, A., Kawaoka, J., and Pyle, A. M. (2010). The NPH-II helicase displays efficient DNA·RNA helicase activity and a pronounced purine sequence bias. *Journal of Biological Chemistry*, 285(15):11692–11703.

Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S., and Wigley, D. B. (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell*, 97(1):75–84.

Winkler, G. S., Araújo, S. J., Fiedler, U., Vermeulen, W., Coin, F., Egly, J.-M., Hoeijmakers, J. H. J., Wood, R. D., Timmers, H. T. M., and Weeda, G. (2000). TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair. *Journal of Biological Chemistry*, 275(6):4258–4266.

Wolski, S. C., Kuper, J., Hänzelmann, P., Truglio, J. J., Croteau, D. L., Houten, B. V., and Kisker, C. (2008). Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. *PLOS Biology*, 6(6):e149.

Yang, Y., Dou, S.-X., Ren, H., Wang, P.-Y., Zhang, X.-D., Qian, M., Pan, B.-Y., and Xi, X. G. (2008). Evidence for a functional dimeric form of the PcrA helicase in DNA unwinding. *Nucleic acids research*, 36(6):1976–1989.

8

Conclusions and further directions

The development of more sensitive optical methods and specific fluorescence-based probes to investigate the structure-relationship in biologically relevant nucleic acid structures constitutes an active area of research in both academic and industrial environments. The discovery of a diverse range of nucleic acid structures (both RNA and DNA) that perform complex functions beyond the mere storage (DNA) and transport (RNA) of the genetic code has emphasised the need to improve the existing methods to sense nucleic acid structures [Perez-Gonzalez and Penedo, 2015; Perez-Gonzalez et al., 2016]. Moreover, some of these nucleic acid motifs have the ability to form three-dimensional structure in solution that are not present in the cell and are finding applications outside the biological context (i.e. nanotechnology).

In this context, two main limitations of current optical sensing methods need to be overcome for both *in vitro* and *in vivo* applications and also to further progress in the characterisation of technologically relevant nucleic acids motifs: specificity and sensitivity. Nowadays, it is clear that the traditional differentiation between single-stranded and double-stranded nucleic acids is no longer sufficient to develop fluorescence-sensing methods capable of discriminating between nucleic acid conformations. The aim of this PhD is dual: i) explore new photophysical tools to detect specific nucleic acids, and II) use these tools to elucidate the structure and function of DNA and RNA structures.

The main advantages of ensemble fluorescence techniques rely on their high sensitivity and wide range of applications allowing the observation of small amounts of sample (i.e. in the micro- to nanomolar range) either *in vitro* or *in vivo*. Moreover, these techniques allow

the simultaneous observation of different observables associated to each fluorophore such as emission intensity, fluorescence lifetime and polarisation. Variations in these observables are caused by the interaction of the fluorophore with its surrounding environment, including the nucleic acid scaffold. Although these techniques provide valuable information, they are limited by the intrinsic averaging of the signal reported, which arises from an ensemble of molecules. Hence, the information about the relative population and dynamics of different nucleic acid structures remains hidden. Moreover, not only the distribution of the different species is lost but also, because molecules are not temporally synchronised, any dynamics due to the switching between different conformations is hidden. On the contrary, single-molecule fluorescence spectroscopy avoids the averaging of the signal by allowing the observation of discrete molecules in real time, which represents the ultimate sensitivity limit in the field of fluorescence. In addition, the lower concentration of sample required ranges in the nano- to picomolar scale.

The work presented in this thesis has taken advantage of the potential of both ensemble and single-molecule fluorescence techniques to investigate the folding, dynamics and conformations of a variety of DNA and RNA structures. Chapter 3 provides a description and characterisation of three fluorescent sensors for the selective detection of G-quadruplex structures. These can be found in a wide range of oncogenes as well as at the end of human telomeres, where the formation of these architectures is involved in the regulation of the telomerase-mediated elongation of telomeres in tumour cells [Maizels and Gray, 2013]. The biological importance of G4s and G-wires has attracted the attention of researchers and their study has experienced a stunning growth in the last decade [Vummidi et al., 2013]. The aim of this chapter is to provide new probes and methods to selectively sense specific G4 structures. For this reason we introduced a ratiometric sensor that undergoes excited-state intramolecular proton transfer (ESIPT), the 2-(2'-hydroxyphenyl)-3H-imidazo[4,5-b]pyridine (HPIP). ESIPT probes are extremely sensitive to the local environment but their use as fluorescence sensors has been mostly limited to the field of metal sensing and its application in a biological context is under-explored. The main advantages of ESIPT probes over conventional fluorescent probes is the presence of two emission bands with good spectral separation, which

provides an absolute observable free from unwanted background effects and independent from the concentration of sensor. Then, we investigated the selective sensing of G-wire structures by using the light-up ruthenium complex $[Ru(bpy)_3](PF_6)_2$ and the natural probe curcumin, which has been studied because of its high availability and reduced cost. Curcumin and $[Ru(bpy)_3]^{2+}$ were able to effectively discriminate between the monomeric G4s and the multimeric G-wires. On the contrary, HPIP provided good discrimination against ssDNA sequences and has shown promising results as the first ratiometric alternative not only in the quantitative detection of G4s but also in the sensing of different G4 conformations. The results obtained underline the capabilities of ratiometric sensing, where HPIP represents the first step in the design of novel probes specifically tailored for the detection of different G4 topologies. Thus, ESIPT-based ratiometric probes provide a new and powerful tool in the recognition and differentiation of nucleic acid structures *in vitro*, although the sensor properties of the HPIP used in this chapter need to be improved. The main aspect to be optimised is the solubility of the compound in water, which is currently too low. Also, a theoretical and experimental analysis of the docking modes of the each species present in the ground state to both ssDNA and G4 quadruplexes will be useful.

Then, we introduced single-molecule TIR spectroscopy to investigate the FRET-labelled SAM-I and adenine riboswitches, which were discussed in detail in chapters 4 and 5, respectively. Riboswitches are mRNA sequences found in the 5' untranslated region of relevant genes found in bacteria, archaea, plants and fungi. The biological importance of these segments and their potential role as antibiotic targets rely on their ability to adopt characteristic tertiary structures upon binding of small cellular metabolites in order to modulate the genetic expression of the associated gene, which is involved in the biosynthesis or transport of metabolite-related compounds [Blouin et al., 2009]. The two-step folding mechanism of the SAM-I aptamer, a four way junction arranged around the P2-P3 and P1-P4 helical stacks [Heppell et al., 2011], has attracted our attention due to the need to further characterise the rearrangement of stems P1 and P3 upon ligand binding. For this purpose, in collaboration with the Daniel Lafontaine lab (Université de Sherbrooke, Canada) we designed a number of SAM constructs in which the addition or subtraction of base pairs in stems P1 and P3 provides

valuable information about the helical periodicity of both stems in order to be associated with their helical rotation and rearrangement upon ligand binding. These constructs were studied using smFRET that, for the first time, was combined with the use of three-dimensional models of the SAM-I aptamer engineered from its crystal structure. The results obtained led us to hypothesise about a reorientation of stem P1 along the plane perpendicular to P3, which has proven the huge potential of this method in the study of RNA aptamers to investigate folding steps that otherwise would remain hidden. However, further work needs to be done by elongating an internally-labelled P1 and a stem P3 carrying a dye attached at its end. These assays should provide more information about the conformational changes observed and about the helical rotation of P1 along its helical axis reported by Heppell and coworkers [Heppell et al., 2011].

The study of RNA aptamers was continued with the observation of the folding and dynamics of the adenine riboswitch, which is, along with the guanine riboswitch, one of the smallest metabolite-binding RNA segments reported to date. The adenine aptamer is organised around a three-way junction with parallel arrangement of stems P2 and P3 and a coaxial stacking of P1 and P3 in the folded state. Interestingly, previous work exploited the potential of smFRET to report a stable intermediate state in the folding mechanism of the aptamer carrying the donor and acceptor dyes on loop stems L3 and L2, respectively [Lemay et al., 2006]. Moreover, this state was characterised as an obligate intermediate in the folding pathway of the adenine aptamer. In this chapter we investigated the presence of the three discrete states involved in the folding pathway of the adenine aptamer by incorporating the donor and acceptor dyes on stem P1 and loop L2 with the aim to provide insights into the P1-P3 stacking and the L2-L3 interaction. This was done in the presence of either monovalent (Na^+) or divalent (Mg^{2+}) cations due to their different capabilities in stabilising helical stacks. The smFRET data suggest an initial "Y-shape" structure in the unfolded state that, upon addition of salt, leads to the formation of an intermediate state that corresponds to a P1-P3 stacked conformation. At this stage, stem P2 displays high flexibility and, at higher salt concentrations, is able to adopt a conformation parallel to P3 stabilised by the loop-loop interaction. Interestingly, the model we proposed in this chapter corroborates not

also previous smFRET studies carried out by Lemay and coworkers [Lemay et al., 2006] but also the folding mechanism proposed by Gilbert and Batey [Gilbert and Batey, 2006].

In chapter 6 we aimed to go one step further in the study of RNA riboswitches and provide, for the first time, insights into the cotranscriptional folding mechanism within elongation complexes using smFRET. To date, the study of riboswitches using fluorescence-based techniques has been limited to the use of double-labelled full-length transcripts due to the inability of bacterial and eukaryotic RNA polymerases to incorporate fluroescent dyes. However, ligand binding and mRNA folding are processes that occur as the RNA transcript is being manufactured and is at this stage when these regulatory segments can modulate the genetic expression of the downstream gene [Pan and Sosnick, 2006]. The transient structure of the mRNA transcripts and their distinct behaviour depending on the nature of the RNAP present some challenging aspects in the observation of the elongation and dynamics of nascent RNAs that limited the use of fluorescence-based techniques in the field. Here, in collaboration with Daniel Lafontaine's group (Université de Sherbrooke, Canada), we addressed these limitations and introduced a novel approach based on smFRET that uses copper-free click chemistry for site-specific incorporation of fluorescent dyes. This new method was used to investigate the TPP-sensing *tbpA* riboswitch within *E. coli* elongation complexes. We first investigated the riboswitch folding in the absence and presence of TPP for elongation complexes stalled at reported transcriptional pause sites [Chauvier et al., 2017]. The importance of monitoring riboswitch conformational changes within transcriptional complexes is underlined by the study of the elongation complex EC-117 and the corresponding run-off transcript Tr-117. Whereas the steric hindrance provided by the RNAP at position EC-117 allows the aptamer to fold in the presence of TPP, the Tr-117 transcript is unable to bind TPP due to the formation of the anti-P1 stem in the absence of RNAP. Moreover, a kinetic analysis of EC-117 suggest that TPP folds via conformational capture mechanism, in which the ligand selectively binds to the folded conformation of the aptamer. The data show that binding occur in a narrow transcriptional window delimited by RNAP pause sites, where EC-136 is the limit point after which the formation of anti-P1 precludes the aptamer to bind free TPP and marks the limit before which the riboswitch

is able to regulate the expression of the downstream gene. The approach described in this chapter provides a novel alternative for the study of RNA folding during biogenesis that will allow the use of other biophysical techniques such as electronic paramagnetic resonance (EPR) and, in collaboration with force-based techniques such as optical trapping, will provide insights into cotranscriptional RNA folding as never seen before.

We finally aimed to extend the use of fluorescence techniques to the study of protein-nucleic acid interactions and monitor the activity of helicase proteins, which was investigated in chapter 7. Helicases are enzymes that are responsible for the unwinding of double-stranded DNA sequences in the cell, thus playing a crucial role in cellular processes such as replication, recombination, repair, transcription and translation that require the reading of the information stored in the DNA bases. The structure and unwinding mechanism of helicases has been widely studied over the years, however, the vast majority of these studies do not account for the key role of specific nucleic acid sequences in their unwinding rate and amplitude. In contrast to the zipper model, a few studies report an asymmetric unwinding of duplexes where leading strands carrying purine bases are more readily unwound [Colizzi and Bussi, 2012; Taylor et al., 2010]. In order to further investigate this possibility we performed unwinding ATP-ase assays to compare the asymmetric behaviour of thermodynamically identical strands using ensemble fluorescence techniques. This work was carried out in collaboration with Giovani Bussi's group (Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy), who studied the asymmetric unwinding of DNA duplexes by SF1 and SF2 helicases using molecular modelling methods, and Malcolm White's group (School of Biology, University of St Andrews), who performed the expression and purification of the proteins used. More specifically, these studies were performed for two model helicases: i) XPD, a SF2 helicase with 5'-3' polarity that is responsible for nucleotide excision repair (NER) in the transcription factor IIH (TFIIH) in eukaryotes; and ii) PcrA, a SF1 helicase with 3'-5' polarity involved in DNA repair and plasmid replication functions. The results at high concentrations of ATP show significant preferential unwinding for DNA substrates carrying a purine-rich leading strand (i.e. up to ~48% greater in XPD and ~270% in PcrA), which was confirmed by gel electrophoresis assays. The reported biased unwinding mechanism

represents an step forward in the characterisation of the asymmetric behaviour of purine- and pyrimidine-rich sequences, which may provide a mechanism for the control of the genetic expression. Further investigation of the preferential unwinding will involve characterising the structure of partially unwound duplexes and intermediate states along the process. This could be achieved by introducing fluorescent nucleotide analogues such as 2AP and pyrrolo-dC, which are extremely sensitive to base pairing and stacking. Also, time-resolved studies at the single-molecule level would provide important insights into the specific role of these sequences as expression regulators.

In summary, this thesis provides insights into the huge potential of fluorecence-based techniques, either at ensemble or single-molecule level, in the study of biological samples ranging from discrete architectures, such as G-quadruplexes and RNA riboswitches, to more intricate structures, such as elongation and protein-DNA complexes. However, these techniques have only shown their full potential by introducing FRET as a spectroscopic ruler to report structural changes in the nanometer-size scale. Ensemble- and single-molecule FRET have demonstrated to provide key information that would be unavailable from other techniques and, as reported in chapter 6, will greatly benefit from the development of new methods to investigate more challenging aspects of crucial biological processes.

References

Blouin, S., Mulhbacher, J., Penedo, J. C., and Lafontaine, D. A. (2009). Riboswitches: Ancient and promising genetic regulators. *ChemBioChem*, 10(3):400–416.

Chauvier, A., Picard-Jean, F., Berger-Dancause, J.-C., Bastet, L., Naghdi, M. R., Dubé, A., Turcotte, P., Perreault, J., and Lafontaine, D. A. (2017). Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. *Nature Communications*, 8:13892.

Colizzi, F. and Bussi, G. (2012). RNA Unwinding from Reweighted Pulling Simulations. *Journal of the American Chemical Society*, 134(11):5173–5179.

Gilbert, S. D. and Batey, R. T. (2006). Riboswitches: Fold and function. *Chemistry & Biology*, 13(8):805 – 807.

References

Heppell, B., Blouin, S., Dussault, A.-M., Mulhbacher, J., Ennifar, E., Penedo, J. C., and Lafontaine, D. A. (2011). Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nature Chemical Biology*, 7(6):384–392.

Lemay, J.-F., Penedo, J. C., Tremblay, R., Lilley, D. M., and Lafontaine, D. (2006). Folding of the adenine riboswitch. *Chemistry & Biology*, 13(8):857 – 868.

Maizels, N. and Gray, L. T. (2013). The G4 Genome. *PLoS Genet*, 9(4):1003468.

Pan, T. and Sosnick, T. (2006). RNA folding during transcription. *Annual Review of Biophysics and Biomolecular Structure*, 35:161–175.

Perez-Gonzalez, C., Lafontaine, D. A., and Penedo, J. C. (2016). Fluorescence-based strategies to investigate the structure and dynamics of aptamer-ligand complexes. *Frontiers in Chemistry*, 4:33.

Perez-Gonzalez, D. C. and Penedo, J. C. (2015). *RNA and DNA Diagnostics*, chapter Single-Molecule Strategies for DNA and RNA Diagnostics, pages 297–332. Springer International Publishing, Cham.

Taylor, S. D., Solem, A., Kawaoka, J., and Pyle, A. M. (2010). The NPH-II helicase displays efficient DNA·RNA helicase activity and a pronounced purine sequence bias. *Journal of Biological Chemistry*, 285(15):11692–11703.

Vummidi, B. R., Alzeer, J., and Luedtke, N. W. (2013). Fluorescent probes for G-quadruplex structures. *ChemBioChem*, 14(5):540–558.

A

Theoretical explanation of fluorescence resonance energy transfer

The first classical attempt to describe the energy transfer in solution was developed by J. Perrin [Perrin, 1924, 1927]. This work proposes that the energy transfer is associated to the interaction between two oscillating dipoles in close proximity. This explanation was extended by J. Perrin's son, F. Perrin, who developed a quantum mechanical theory only a few years later [Perrin, 1932, 1933]. However, both contributions report that the energy transfer depends on the third power of the distance between the two dipoles, which results in longer theoretical distances than those obtained experimentally [Clegg, 2006].

In 1941, Oppenheimer reported in a short abstract at the American Physical Society meeting the first theory of FRET that predicts the dependence of the energy transfer on the sixth power of the distance between the two dipoles [Oppenheimer, 1941], but it was not until 1950 when the full description of this work was published [Arnold and Oppenheimer, 1950]. Before that, Förster reported a classical [Förster, 1946] and quantum mechanical [Förster, 1948] theories of FRET, which improved the explanation supported by J. and F. Perrins on the energy transfer in solution. At this point, Förster also states the dependence of the energy transfer on the sixth power of the distance between both dipoles and introduces, for the first time, the term of the overlap integral.

Despite Oppenheimer's theory was published years before, the popularity of Förster's publication is the reason why it is commonly considered as the most relevant contribution

that relates the theoretical explanation of FRET and the experimental data. An in-depth view of all the contributions that lead to the current concept of FRET are well covered by Robert Clegg in [Clegg, 2006].

In this appendix, we will give an explanation of the FRET efficiency by reviewing previous work from Clegg [Clegg, 1992] and Kuhn [Kuhn, 1970].

As described in section 1.2, a fluorophore in its electronic excited state can relax via radiative or non-radiative processes, among which are included the loss of energy due to collisions, the quenching by another molecule and the intersystem crossing to the triplet state T_1. Thus, the rate of decay of the excited state of the donor can be expressed as a function of the rate constants of these processes as

$$\tau_D^{-1} = k_c + k_{isc} + k_q + k_f \tag{A.1}$$

where k_c represents the relaxation due to collisions, k_{isc} is the intersystem crossing, k_q is the quenching and k_f is the fluorescence rate constant. However, the rate of decay of the excited state of the donor in the presence of an acceptor, termed τ_{DA}, must include the term k_{FRET} to describe the additional deactivation pathway via non-radiative energy transfer

$$\tau_{DA}^{-1} = k_c + k_{isc} + k_q + k_f + k_{FRET} = \tau_D^{-1} + k_{FRET} \tag{A.2}$$

Hence, the quantum yield of the non-radiative energy transfer from the donor to the acceptor, also termed E, can be expressed as

$$\phi_{FRET}^D = E = \frac{k_{FRET}}{k_c + k_{isc} + k_q + k_f + k_{FRET}} = k_{FRET}\,\tau_{DA} \tag{A.3}$$

By rearranging equation A.2, the rate constant of the energy transfer can be expressed as

$$k_{FRET} = \tau_{DA}^{-1} - \tau_D^{-1} = \frac{1}{\tau_D}\left(\frac{\tau_D}{\tau_{DA}} - 1\right) \tag{A.4}$$

The combination of equations A.3 and A.4 results in the expression of the FRET efficiency as a function of τ_D and τ_{DA}

$$E = 1 - \frac{\tau_{DA}}{\tau_D} \tag{A.5}$$

In addition, the ratio between τ_D and τ_{DA} can be expressed by merging equations A.1 and A.2 as

$$\frac{\tau_D}{\tau_{DA}} = 1 + \frac{k_{FRET}}{k_D} \tag{A.6}$$

Since the acceptor only interacts with the component of the electric field that oscillates in the same direction as its transition dipole moment, which is represented by $E_{\|}$, the rate of energy transfer is

$$k_{FRET} = a\,|E_{\|}|^2 \tag{A.7}$$

where a is the proportionality constant associated to the extinction coefficient of the acceptor at the frequency of the donor oscillating dipole. The value of the constant a can be calculated for a thin layer of acceptor molecules of thickness z and concentration C. The intensity of light, I, that passes through this layer is given by the Beer-Lambert law

$$I = I_0\, 10^{-\varepsilon_A C z} \tag{A.8}$$

where ε_A is the molar extinction coefficient of the acceptor. The derivation of equation A.8 yields

$$-\mathrm{d}I = \mathrm{d}I_0\, \varepsilon_A\, C\, 10^{-\varepsilon_A C z} (\ln 10)\, \mathrm{d}z = -I\, \varepsilon_A\, C\, (\ln 10)\, \mathrm{d}z \tag{A.9}$$

However, the energy transfer defined by k_{FRET} only involves the absorption of radiation by one acceptor molecule. Therefore, the energy absorbed by a single molecule may be expressed as

$$k_{FRET} = \frac{-\dfrac{\mathrm{d}I}{\mathrm{d}z}}{C N_A} = \frac{3\, I\, \varepsilon_A\, (\ln 10)}{N_A} \tag{A.10}$$

where N_A is Avogadro's number (6.0221409 x 10^{23}) and the factor of 3 is due to the alignment between the electric field of the donor and the transition dipole moment of the acceptor.

The expression of the transmitted intensity is also given by

$$I = \frac{c\,n}{8\,\pi}|E_{\parallel}|^2 \tag{A.11}$$

where c is the speed of light in vacuum (299,792,458 m/s). By inserting this expression in equation A.10, the expression of k_{FRET} results in

$$k_{FRET} = \frac{3\,c\,n\,\varepsilon\,(\ln 10)}{8\,\pi N_A}|E_{\parallel}|^2 \tag{A.12}$$

After that, we may define the expression of the orientation factor κ as

$$\kappa = \vec{\mu_D}\,\vec{\mu_A} - 3\,(\vec{\mu_D}\,\vec{r})(\vec{\mu_A}\,\vec{r}) \tag{A.13}$$

where $\vec{r}$ is the inter-dye distance vector, whereas $\vec{\mu_D}$ and $\vec{\mu_A}$ are the donor and acceptor dipole moments, respectively (Fig. 1.10a). Then, the component of the electric field in the direction of the acceptor dipole moment is given by

$$E_{\parallel} = \vec{\mu_A}\,\vec{E_D} = \frac{\mu_D}{n^2\,r^3}[\vec{\mu_D}\,\vec{\mu_A} - 3\,(\vec{\mu_D}\,\vec{r})(\vec{\mu_A}\,\vec{r})] = \frac{\mu_D\,\kappa}{n^2\,r^3} \tag{A.14}$$

where n is the refractive index of the medium.

The combination of equations A.14 and A.12 relates the rate of energy transfer and the distance between the two dipoles, r, as

$$k_{FRET} = \frac{3\,c\,n\,\varepsilon\,\mu_D^2\,\kappa^2\,(\ln 10)}{8\,\pi n^4\,N_A\,r^6} \tag{A.15}$$

which defines the dependence of the rate of energy transfer on the sixth power of the distance between two fluorophores.

The rate of absorption by the acceptor represents the amount of donor luminiscence that is effectively quenched. This can be expressed as the power loss of the donor due to its absorption by the acceptor, which is given by

$$P = \frac{\mu_D^2 \omega^4 n}{3 c^3} \tag{A.16}$$

where ω is the frequency of the donor oscillating dipole and μ_D is the amplitude of the donor dipole moment. This equation can be also expressed as the product of the rate of energy loss, k_D, and fluorescence quantum yield of the donor, Q_D,

$$P = Q_D k_D \tag{A.17}$$

The combination of equations A.16 and A.17 yields

$$k_D = \frac{\mu_D^2 \omega^4 n}{3 c^3 Q_D} \tag{A.18}$$

Then, by inserting equations A.15 and A.18 into equation A.6 and by substituting $2\pi c/\omega$ with λ we obtain

$$\frac{\tau_D}{\tau_{DA}} = 1 + \frac{k_{FRET}}{k_D} = 1 + \frac{9000 (\ln 10) \varepsilon \lambda^4 \kappa^2 Q_D}{128 \pi^5 n^4 N_A r^6} \tag{A.19}$$

The spectral overlap integral between the normalised fluorescence emission spectrum of the donor, $F_D(\lambda)$, and the normalised absorption spectrum of the acceptor, $\varepsilon_A(\lambda)$, is defined by

$$J(\lambda) = \int_0^\infty F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 \, d\lambda \tag{A.20}$$

which can be inserted in equation A.19 to yield

$$\frac{\tau_D}{\tau_{DA}} = 1 + \frac{k_{FRET}}{k_D} = \frac{9000\,(\ln 10)\,\kappa^2\,Q_D}{128\,\pi^5\,n^4\,N_A\,r^6} \int_0^\infty F_D(\lambda)\,\varepsilon_A(\lambda)\,\lambda^4\,\mathrm{d}\lambda \tag{A.21}$$

where the Förster radius, R_0, can be defined as

$$R_0^6 = \frac{9000\,(\ln 10)\,\kappa^2\,Q_D}{128\,\pi^5\,N_A\,n^4} \int_0^\infty F_D(\lambda)\,\varepsilon_A(\lambda)\,\lambda^4\,\mathrm{d}\lambda \tag{A.22}$$

Hence, equation A.21 can be finally expressed as

$$\frac{\tau_D}{\tau_{DA}} = 1 + \left(\frac{R_0}{r}\right)^6 \tag{A.23}$$

which, in combination with equation A.5, yields the most commonly known expression that defines the efficiency of the energy transfer and its dependence on the sixth power of the distance between donor and acceptor dipoles

$$E = 1 - \frac{\tau_{DA}}{\tau_D} = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \tag{A.24}$$

References

Arnold, W. and Oppenheimer, J. R. (1950). Internal conversion in the photosynthetic mechanism of blue-green algae. *The Journal of general physiology*, 33(4):423.

Clegg, R. M. (1992). [18] fluorescence resonance energy transfer and nucleic acids. *DNA Structures Part A: Synthesis and Physical Analysis of DNA*, 211:353–388.

Clegg, R. M. (2006). The history of FRET. In *Reviews in Fluorescence 2006*, pages 1–45. Springer.

Förster, T. (1946). Energiewanderung und fluoreszenz. *Naturwissenschaften*, 33(6):166–175.

Förster, T. (1948). Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der physik*, 437:55–75.

Kuhn, H. (1970). Classical aspects of energy transfer in molecular systems. *The Journal of Chemical Physics*, 53(1):101–108.

Oppenheimer, J. R. (1941). Internal conversion in photosynthesis. *Phys. Rev*, 60(2):158.

Perrin, F. (1932). Théorie quantique des transferts d'activation entre molécules de même espèce. cas des solutions fluorescentes. In *Annales de Physique*, volume 10, pages 283–314. EDP Sciences.

Perrin, F. (1933). Interaction entre atomes normal et activé. transferts d'activation. formation d'une molécule activée. *Ann. Institut Poincaré*, 3:279–318.

Perrin, J. (1924). Fluorescence et radiochimie Conseil de Chemie. *Solvay, 2iem, Paris*, pages 322–398.

Perrin, J. E. A. N. (1927). Fluorescence et induction moleculaire par resonance. *CR Hebd. Seances Acad. Sci*, 184:1097–1100.

B

Publications arising from this work

Articles, chapters and reviews

Aitken, L., Quinn, S. D., Perez-Gonzalez, C., Samuel, I. D. W., Penedo, J. C., and Gunn-Moore, F. J. (2016). Morphology-specific inhibition of β-amyloid aggregates by 17β-hydroxysteroid dehydrogenase type 10. *ChemBioChem*, pages 1029–1037.

Boudreault, J., Perez-Gonzalez, D. C., Penedo, J. C., and Lafontaine, D. A. (2015). *DNA-Protein Interactions: Principles and Protocols*, chapter Single-Molecule Approaches for the Characterization of Riboswitch Folding Mechanisms, pages 101–107. Springer New York, New York, NY.

Perez-Gonzalez, C., Grondin, J. P., Lafontaine, D. A., and Carlos Penedo, J. (2016a). *Biophysics of Infection*, chapter Biophysical Approaches to Bacterial Gene Regulation by Riboswitches, pages 157–191. Springer International Publishing, Cham.

Perez-Gonzalez, C., Lafontaine, D. A., and Penedo, J. C. (2016b). Fluorescence-based strategies to investigate the structure and dynamics of aptamer-ligand complexes. *Frontiers in Chemistry*, 4:33.

Perez-Gonzalez, D. C. and Penedo, J. C. (2015). *RNA and DNA Diagnostics*, chapter Single-Molecule Strategies for DNA and RNA Diagnostics, pages 297–332. Springer International Publishing, Cham.

Pérez-González, D. C., Rodríguez-Prieto, F., and Penedo, J. C. (2015). High-affinity fluorescence sensing of G-quadruplexes. *Biophysical Journal*, 108(2):393a–.

Manuscripts in preparation

Chauvier, A., St-Pierre, P., Nadon, J. F., Grondin, J. P., Perez-Gonzalez, C., Eschbach, S. H., Lamontagne, A. M., Penedo*, J. C., , and Lafontaine*, D. A. (2017a). Single-molecule FRET imaging of nascent RNA folding within bacterial transcription complexes.

Chauvier, A., St-Pierre, P., Nadon, J. F., Grondin, J. P., Perez-Gonzalez, C., Eschbach, S. H., Lamontagne, A. M., Penedo*, J. C., and Lafontaine*, D. A. (2017b). Single-molecule FRET mapping of riboswich structure at co-transcriptional pausing sites.

Colizzi*, F., Perez-Gonzalez*, C., Levy, Y., White, M. F., Carlos Penedo, J., and Bussi, G. (2017). Asymmetric double helix opening drives helicases unwinding dynamics.

Perez-Gonzalez, C., Rodriguez Prieto, M. F., and Penedo*, J. C. (2017). Excited-state proton transfer probes as ratiometric sensors of DNA structures.

St-Pierre, P., Jacques, S., Perez-Gonzalez, C., Picard-Jean, F., Penedo, J. C., and Lafontaine, D. A. (2017). The folding pathway of the adenine riboswitch involves a structural intermediate pre-organizing the binding domain.