Universidad Autónoma de San Luis Potosí
Instituto de Física

Single Molecule FRET Studies of Short and Long RNA Samples

Tesis para obtener el grado de

Doctor en Ciencias
(Física)

Presenta:

Nehemías Leija-Martínez

December 2014

Abstract

In the classic view of the central dogma of biology, ribonucleic acids (RNA) only code for proteins. However, many exceptions to this dogma are now known as a result of genomic studies and spectroscopy and microscopy techniques at the single-molecule level. Now it is well know that some RNA molecules are involved in splicing, signal recognition of proteins and even have catalytic activity *per se*. Such roles have been identified to the shape that the RNA molecule can adopt.

In this thesis we show the design, construction and calibration of an optical setup for measure at the single-molecule level, fluorescence resonance energy transfer (FRET) to study different physical and biological characteristics of short and long RNA molecules. The optical setup was made up of three main parts, a laser light source for excitation of dyes, a commercial epi-fluorescent microscope and an ultra sensitive single photon detectors. The calibration of the optical setup was made by using short DNA-labeled fragments with donor-acceptor dyes attached at opposite ends of the DNA molecules. A calibration curve for FRET was obtained. The calibration curve was used to measure for the first time the end-to-end distances of RNA molecules obtained from different biological sources. Our results prove that the two ends of the RNA molecules studied are in close proximity. These results migth indicate that the close proximity provides a so called "effective circularization" of RNA molecules that should facilitate translation.

We also observed with smFRET that conserved RNA-targets sequences, located in the messenger RNA of the human immunodeficiency virus (HIV-1), adopt different structural changes by punctual mutations and temperature changes. Our results suggest the equilibrium coexistence of different RNA conformations, whose population percentage are easily changed by moderate temperature changes. To study if the observed RNA populations expose its RNA-target sequence, the RNA-Induced Silencing Complex (RISC), consisting in its core of an Argonaute protein (*Thermus thermophilus Argonaute*) and a short single-stranded interference RNA (siRNA) either Pol or Nef, was incubated with the above RNA-targets and FRET measurements were performed. Our results indicate that RNA-punctual mutations may help or affect to the RNA target to escape of cleaving by RISC complex.

To my parents

Acknowledgments

First and foremost, I thanks to my advisers Profs. Jaime Ruiz and Eduardo Gomez, for the opportunity for professional and personal growth as a graduate student. During my time in the Lab both have given me the freedom and the support to explore while providing the subtle guidance.
I am also very grateful with Prof. Sergio Casas for teaching me molecular biology techniques and for giving me the opportunity to work in his laboratory (Laboratorio de Genómica Funcional y Comparativa, IPICyT). I am also grateful with I.A.F. María Isabel Isordia Jasso, laboratory operations manager, in Sergio Casas's laboratory for give me the support and knowledge in the uses of molecular tools and for responding my questions.
I also mention my gratitude to Prof. José Alfredo Mendez for helping me in molecular biology techniques specially for label DNA, RNA and proteins. Moreover for the opportunity he gave me to use his laboratory equipment for carrying out biological tests.
Integral to this environment I mention not only the students in Sergio Casas's laboratory but also in the department of molecular biology at IPICyT for the work environment. They bore the brunt of my initial ignorance in biology. Special thanks to Dr. Mayte Cervantes and Dr. Edith Uresti who provided me with tips in molecular biology.

I would also like to offer my gratitude to all my friends in my working-lab and all my friend which work in different labs in the physics institute at UASLP. I have been fortunate to work with talented and fun people who encouraged and inspired me. Their support, knowledge, and good nature have made this process enjoyable, even in the most stressful times. I would also like to the support in this process and for providing me with many useful comments and advice to Dr. Rubén D. Nava, Dr. Elizabeth Reynaga, Dr. Cesar Maldonado, Dr. Donato Valdez, Dr. Edgar Alvizo and Dr. Roger Vega Acosta.

I cannot imagine how I could have done this without the help of I.E. Emmanuel Vazquez Martínez, laboratory operations manager, which was always attending tech-

nical problems in the lab.

I am thankful to Profs. Vannesa Olivares, José Sampedro, Jorge Arreola and Roberto Sánchez for letting me use test equipment in their laboratories.

Last and certainly not least, I am grateful with Consejo Nacional de Ciencia y Tecnología (CONACyT) for support with a PhD fellowship.

Contents

1

Introduction

A large number of instrumental techniques have been developed to understand the inner workings of the machinery found in living organisms on length scales ranging from the molecular to the macroscopic. For instance, fluorescence based methods such as diffuse optical imaging, magnetic resonance imaging, bio-luminescence imaging, and fluorescence microscopy have allowed to observe biological processes such as metabolic states of the cells at microscopic scale [1]. Microfluidic device technologies have allowed to study the interaction between protein-protein or protein-deoxyribonucleic acids (DNA's) [2]. Other techniques such as light scattering, X-ray, and neutron scattering are also currently employed to study molecular systems [3]. However, the aforementioned generally only provide the average behavior of an ensemble of molecules and thus cannot resolve, for instance, the many different possible conformational changes of the molecules in study [4]. Nevertheless, over the past two decades single-molecule techniques based in fluorescence and force spectroscopy have been developed. These techniques have enabled us to go beyond ensemble average measurements to the ability to observe static properties and dynamics of interconversion between different states of individual bio-macromolecules under biologically relevant conditions [5–8]. These developments in single-molecule detection is one of the most promising applications not only in biological system but also in physical or chemical systems [9]. Among the different single-molecules techniques, Fluorescence Resonance Energy Transfer (FRET) [10, 11], also known as Förster Resonance Energy Transfer in honor to Theodore Förster who in 1948 development the main theory [12], is the most popular and arguably the most general approach for studying single molecules. In biological systems this technique has been directly used to resolve intermediate states, static and dynamic sub-populations in solutions, and multiple process where ribonucleic acids (RNAs), nucleic acids (DNAs) and proteins are involved. For example, by using FRET at single molecule level (smFRET), Margittai and coworkers [13] observed that free syntaxin 1 protein, a small integral membrane

protein that is abundantly expressed in neurons and neuroendocrine cells, switches between an inactive closed and an active open configuration with a relaxation time of 0.8 ms. Their results explain why regulatory proteins such as SNAP 25 are needed to arrest the syntaxin 1 in one conformational state. Yusdi S et al. [14] using smFRET studied the polymerase conformational transitions that precede nucleotide addition by the DNA polymerase I. They clearly distinguish open and closed conformations that predominate in Pol-DNA and Pol-DNA-dNTP complexes, respectively. Moreover, they observed that in the absence of substrate, DNA polymerase I shows high degree of conformational flexibility. Such flexibility was not anticipated on the basis of the available crystallographic structures. Other important and interesting results of protein folding detected by smFRET are described in references [15–20].

Implementation of smFRET to study RNA molecules has revealed a repertoire of dynamic properties of RNA molecules that are difficult to analyze with ensemble experiments, including nonaccumulative folding intermediate states, parallel folding pathways, and equilibrium conformational fluctuations [21–23, 25, 148]. These single-molecule results indicate that RNA molecules fold across a highly rugged energy landscape, along a multitude of folding pathways, and through many intermediate folding states.

The success of FRET relies in that it allows to observe dynamics and molecular interactions at the 1 to 10 nm scale [101], a scale comparable to the size of most bio-molecules. This is achieved by using only two fluorophores, one called donor and another called acceptor. The fluorophores are attached at two different sites of the molecule under study without affecting its biological or physical function. Conformational changes in the host molecule implies fluctuation in the separation distance of the two dyes causing changes in fluorescence intensities of the donor and the acceptor. This variation of fluorescence is used as an indicator of the dynamics of the host molecule. Additionally, because the two fluorophores are on different parts of the molecule, intermolecular motion can be measured in the molecular center of mass.

In this thesis we use FRET to measure the proximity between the ends of some ribonucleic acid molecules obtained from Cowpea Chlorotic Mottle Virus (CCMV), Brome mosaic virus (BMV) as well the fungus *Trichoderma atroviride* (*T. atroviride*). Also, we use FRET to monitoring ensemble and structural changes of HIV-1 based

RNA targets as a function of temperature as well as its interaction with an RNA-Induced Silencing Complex (RISC). In this chapter, we review FRET theory in order to present the fundamental concepts of this technique and we outline the remainder of this dissertation.

1.1. Physical Description of FRET

Fluorescence Resonance Energy Transfer technique is based on the physical processes that happens when an excited fluorophore called donor transfers energy non-radiatively to a fluorophore, called acceptor, via a weak induced electronic dipole-dipole coupling mechanism. This process can be described schematically by the Jablonski diagram [27] as shown in figure (1.1).

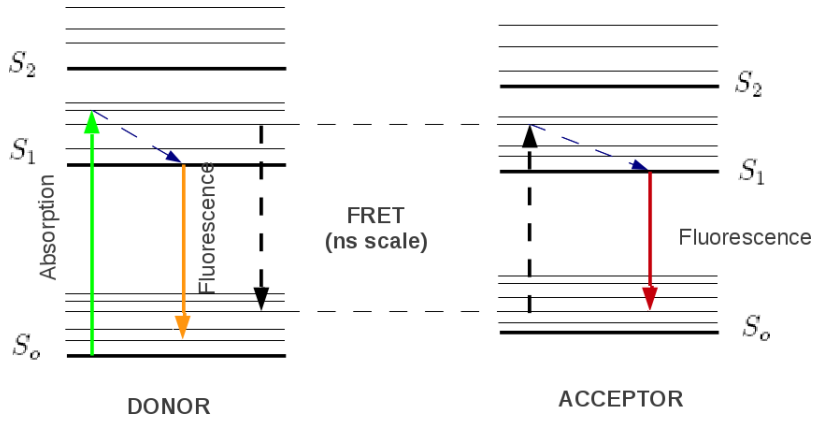

Figure 1.1.: Schematic representation of resonance energy transfer from the donor dye to the acceptor dye through the Jablosnki Diagram.

Jablonski diagram is basically an energy diagram arranged with horizontal lines that represent the limits of electronic energy states. S_o corresponds to the lowest electronic-state while S_1 and S_2 correspond to electronic-excited states. Within each electronic energy state there are multiple vibronic energy states that may be coupled with the electronic state. When a donor dye absorbs a photon from an external source such as a laser, it goes from S_o to either S_1 or S_2 vibronic levels (Fig. 1.1). Thereafter, non-radiative internal conversion brings to the donor to the lowest excited singlet S_1 or S_2 in picosecond scale (dashed blue arrow). Non-radiative relaxation or emission of a photon can take place within nanoseconds and brings back the donor to one of

the vibronic sublevels of the ground state (orange arrow). However the donor can also relax via conversion to a triplet state, or in the presence of a suitable acceptor (Fig. (1.2)), it may transfer excited state energy directly to the acceptor, without emitting a photon, via coupling electronic states (black arrows in Fig. (1.1)). This last scenario is called resonance energy transfer (RET). The result is the excitation of the acceptor to an excited state (S_1 or S_2) for later decaying emitting a photon at a new wavelength. The mechanism of coupling the vibrational levels of the donor and the acceptor can be understood via a virtual photon coupling as it is assumed in the quantum electrodynamic approach for RET processes [31].

1.2. Rate of Energy Transfer in RET

Considering a donor (D) and an acceptor (A) dyes immersed in a homogenous medium (ϵ) with vector of position $\vec{R}_D$ and $\vec{R}_A$ respectively (Fig. 1.2). The distance of separation between D and A is defined as $\vec{R} = \vec{R}_A - \vec{R}_D$.

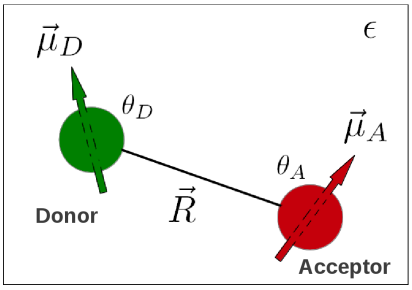

Figure 1.2.: Schematic representation of donor and acceptor dyes immersed in a medium with electric permittivity ϵ and separated a distance $\vec{R}$. The dipole moments of donor and acceptor are denoted as $\vec{\mu}_D$ and $\vec{\mu}_A$ respectively while θ_D and θ_A are the angles between these dipoles and the vector joining the donor and the acceptor.

As initial state, we assume that the donor has absorbed a photon from an external field and it is in the excited state while the acceptor is in the ground state. As final state the donor decays from the excited state emitting a photon or transferring the energy to the acceptor. Quantum electrodynamics (QED) tells us that such excitation transfer is mediated by a virtual photon, a messenger particle that cannot be directly detected and which, by its creation and subsequent annihilation, electro-

magnetically couples the donor decay and acceptor excitation [30]. In this approach the initial $|i\rangle$ and final $|f\rangle$ can be expressed as

$$\begin{aligned} |\, i\rangle &= |0\rangle|\psi_{D^*}, \psi_A\rangle|\phi^n_{D^*}\rangle|\phi^m_A\rangle, \\ |\, f\rangle &= |0\rangle|\psi_D, \psi_{A^*}\rangle|\phi^l_D\rangle|\phi^p_{A^*}\rangle, \end{aligned} \tag{1.1}$$

where $|\phi_k\rangle$ and $|\psi_k\rangle$ $(k = D, A)$ are the electronic and vibrational states [31], and $|0\rangle = |0(\vec{p}, s)\rangle$ is the virtual occupation number state [28] (where $\vec{p}$ denoted mode and s denoted polarization). Then the rate of energy transfer from the donor to the acceptor is obtained through the Fermi's golden rule expression as [28, 29]

$$M_{fi} = \frac{2\pi}{\hbar} \sum_{i,f} \rho(i) |\langle f|T|i\rangle|^2 \delta(E_i - E_f) \tag{1.2}$$

where T is the time-independent portion of the multipolar Hamiltonian operator for the FRET system [30]. Introducing initial and final states (1.1) in Eq. (1.2) it is possible to derive a mathematical expression for the rate of energy transfer [30]

$$W_F = \sum_{n,l,m,p} |T_{DA}(k, \vec{R})|^2 |\langle \phi^{(l)}_D | \phi^{(n)}_{D^*} \rangle|^2 |\langle \phi^{(p)}_{A^*} | \phi^{(m)}_{A^*} \rangle|^2 \delta(\Delta E_D - \Delta E_A) \tag{1.3}$$

where $T_{DA}(k, \vec{R})$ is the coupling tensor

$$T_{DA}(k, \vec{R}) = \frac{k^3 \mu_A \mu_D e^{ikR}}{4\pi\epsilon} \left[\eta_3 \left(\frac{1}{k^3 R^3} - \frac{i}{k^2 R^2} \right) - \eta_1 \frac{1}{kR} \right] \tag{1.4}$$

and η_j is called the orientational factor defined as

$$\eta_j = \hat{n}_D \cdot \hat{n}_A - j(\hat{n}_D \cdot \hat{n}_R)(\hat{n}_D \cdot \hat{n}_R), \;\; j = 1, 3 \tag{1.5}$$

where $\hat{n}_D$, $\hat{n}_A$ and $\hat{n}_R$ being units vectors oriented along the transition dipoles of the donor and acceptor molecules and their separation vector $\vec{R}$ respectively.
From equations (1.3) and (1.4) we can see that the rate of energy transfer from a donor to an acceptor contain three terms; the first one scales with $W_F \sim R^{-6}$, the second scale with $W_F \sim R^{-4}$, and the third with $W_F \sim R^{-2}$. The first term is dominant in the short-range or the near zone [33] region where $kR << 1$, and is

identical to the one deduced by Förster for the physical phenomena of FRET. The second term corresponding to energy transfer in the induction zone and it comes into play at distances beyond the near field, where the $KR << 1$ no longer holds. The third term corresponding to energy transfer in the radiation ($kR >> 1$) zone [33] and take place when a donor emits a photon and the acceptor absorbs the same photon. However the probability for such an event is extremely small [36] because at reasonable fluorophore concentrations the emitted photon is radiated in an arbitrary direction and thus mostly does not encounter an acceptor fluorophore in its path.

1.3. FRET spectral overlap integral

As in the conventional Förster theory, the first term of equation (1.3) can be written in terms of the molar extinction coefficient of the acceptor $\eta(\nu)$ and the normalized emission spectrum of donor $f(\nu)$, by decomposing the δ function over the range of all possible energy values as [32]

$$W_F = \left(\frac{1}{R^6}\right) \frac{9000\ln(10)\phi_D\eta_3^2 c^4}{128\pi^5 n^4 N_A \tau_D} \int_0^\infty f(\nu)\epsilon(\nu)\frac{d\nu}{\nu^4}, \tag{1.6}$$

where n is the index of refraction of the medium, c refers to the speed of light (cm/s), N_A is Avogadro's number, $\tau_D = \tau\Phi_D$ is the total lifetime of the donor excited state in absence of the acceptor, whit Φ_D the quantum yield of donor and τ is the radiative relaxation time of the donor. The parameter η_3^2 is the orientation factor for FRET. It tell us that the FRET interaction between a donor and an acceptor occur if the donor emission dipole moment and the acceptor absorption dipole moment are in favorable orientation (see Fig. 1.2). This orientation factor in the Förster theory is denoted as κ^2, and assumes a numerical value of 2/3 provided that both probes can undergo unrestricted isotropic motion.

The integral in equation (1.6) is know as overlap integral, $J(\nu = \omega/2\pi)$ and express the degree of overlap or resonance between the donor fluorescence spectrum and the acceptor absorption cross section spectrum. It means that the energy lost from excited donor to ground state excite the acceptor group. The figure (1.3) shows an example of the amount of spectral overlap between the dyes Alexa fluor 546 (donor)

and Alexa fluor 647 (acceptor). .

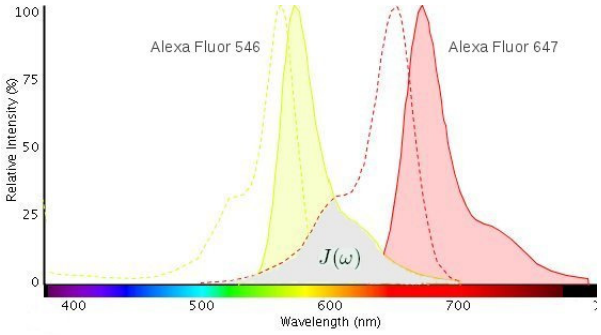

Figure 1.3.: Absorption (continuous line) and emission (dashed lines) spectral properties of alexa fluor 546 and alexa fluor 647. Gray box indicate the overlap integral $J(\nu)$ bewteen these dyes. Imagen obtained from Invitrogen SpectraViewer [34]

1.4. Efficiency of Energy Transfer

Measuring the rate of resonant energy transfer from a donor to an acceptor could be complicated. Instead, most readily accessible measure is obtained by the efficiency of energy transfer E defined as [12]

$$E = \frac{R_o^6}{R_o^6 + R^6} \tag{1.7}$$

where the term R_o has units of length and can be determined directly from the overlap integral, equation (1.6), as

$$R_o^6 \;=\; 8.8 \times 10^{-25} \frac{\kappa^2 \phi_D}{n^4} \int_0^\infty f(\nu)\epsilon(\nu)\frac{d\nu}{\nu^4}. \tag{1.8}$$

R_o is known as Förster's distance and means the distance at which the rate (or probability per unit of time) of transfer efficiency is 50%. Spectroscopy measurements performed by the suppliers of fluorescent dyes provide generally the value of R_o for the pair of dyes chosen. For example, the value of R_o reported by Molecular Probes (Invitrogen) for dye pairs Alexa Fluor 546 and Alexa Fluor 647 has a value of $R_o = 74$Å in TE buffer solution (10 mM Tris-HCl pH=8.0, 1 mM EDTA) and at

room temperature [35].

For FRET experiments the selection of pair of dyes should be taken into account depending on the separations involved in the sample under study. For example the E value at R =5 nm of separation distance between the Tryptophan (donor) and the Dansyl (acceptor) dyes ($R_o = 24$Å) is close to zero. Therefore, for this pair of dyes no information is available at distances large that 5 nm.

1.5. Efficiency of Energy transfer in terms of fluorescent intensities

Due to the relation between fluorescence lifetime and intensity of fluorescence emitted by the donor in absence or in presence of an acceptor, it is possible to express the efficiency of energy transfer from a donor to an acceptor in terms of their fluorescence emission intensities, I_D and I_A respectively, as [36]

$$E = \frac{R_o^6}{R_o^6 + R^6} = \frac{I_A}{I_A + \gamma I_D} \tag{1.9}$$

where the factor $\gamma = \phi_A \eta_A / \phi_D \eta_D$ depends on the collection efficiencies of the fluorescence intensity of the donor (η_D) and the acceptor (η_A) as well as of their quantum yields ϕ_D and ϕ_A respectively.

The FRET efficiency E given in equation (1.9) probably is the most common method for measuring the efficiency of energy transfer from donor to acceptor. However, the FRET efficiency E can also be obtained by measuring the donor fluorescence lifetime as $E = 1 - \tau_{DA}/\tau_D$, where τ_{DA} is the fluorescence lifetime of the donor in presence of an acceptor. FRET efficiency determined from donor lifetimes has the clear advantage that is not subject to background and instrumental corrections of the donor and acceptor intensities [36].

Conclusion

Fluorescence resonance energy transfer (FRET) is a technique that has been performed mostly by biophysicists. It allows to measure distances of a few nanometers by using two fluorescent dyes, one called donor and another called acceptor, making

FRET extremely sensitive to small distances where the intermolecular interaction frequently occurs.

The quantum electrodynamic approach used to describe the physical phenomenon of FRET tell us that the probability of resonance energy transfer from a donor to an acceptor not only depends on distance, but also on the suitability of the dipole pair to resonate *per se*. That is, the selection of suitable FRET pairs has to take into account that the donor possesses a high quantum yield Φ_D, so that sufficient excited state energy is available for transfer, and the acceptor possess a high molar extinction coefficient, so that it can absorb the donated energy with high efficiency. Moreover, the emission spectrum of the donor and the absorption spectrum of the acceptor must overlap. Lower values of this overlap may also lead to FRET interaction but only at distances smaller than $R < 10$ nm as is in the case of Tryptophan and Dansyl dye pairs. Other factor involved in FRET interaction is the orientation dipole moments κ^2 of both dyes, but in case of random dipoles, the κ value is estimated as 2/3. This assumption is true for many but not all situations. In systems where the non-random orientation of dipoles is observed (e.g., biomembranes), the κ value needs correction [37].

2

Optical Setup For Measuring FRET

Different spectroscopy techniques have been implemented for measuring FRET. The most common involve the use of standard fluorescence microscopy techniques such as wide field epi-fluorescence [38], total internal reflection fluorescence (TIRF) [38, 39], two-photon excitation [40], and confocal microscopy [41, 42]. Each of them has specific advantages as well as disadvantages. Nevertheless, a common problem is to maximize the signal from a single donor and a single acceptor while at the same time scrupulously minimize background photons, noise from unwanted sources, and detector dark counts [43]. In this chapter we show the implementation of our optical setup for measuring FRET at the level of single molecule as well as in ensemble average. The FRET setup was made up of three main parts, a laser light source for excitation of fluorophores, a commercial epi-fluorescence microscope for collection of the fluorescence signal through confocal detection and ultra sensitive single photon detectors.

2.1. Illumination system

In order to implement an optical setup for measuring FRET, an illumination light laser centered at 515 nm (Excelsior-515-50 of Spectra Physics) was coupled to an epifluorescent microscopy Nikon Eclipse E800 by using two mirrors, M_1 and M_2, to guide the excitation laser light into the collector lens of the microscope. The wavelength of the laser was chosen to match the properties of the donor dye and reduce direct excitation of the acceptor. In this thesis the fluorescent molecules used as donor were the Alexa Fluor 546 (AF546) and the Cyanine-3 (Cy3).

To increase the incident laser beam diameter into collector lens, such that a tight focus could be achieved at the back focal plane of every objective into the nose-piece of the microscope, a beam expander (10X objective CP-Achromatic 10X/0.25, Zeiss

Optics), mounted on three-dimensional micrometer translation stage (NF15AP25, Thorlabs) to facilitate its alignment with the optical path, was placed between the optical path from M_2 to the collector lens (Fig 2.1).

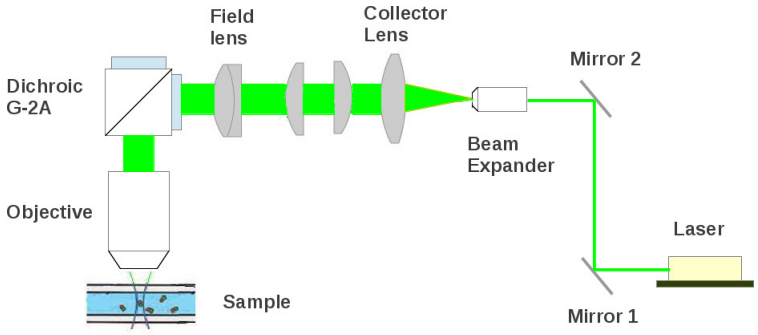

Figure 2.1.: Illumination optical system for FRET experiments. The laser is directed by mirrors M_1 and M_2 to the collector lens of the Nikon Eclipse E800 microscope. The laser is collimated by the illumination optical system of the microscope and focused by the Plan Apo Nikon 100X objective.

To align the laser light with the illumination optical axis of the microscope, the nose-piece of the microscope was rotated to select 10X objective. The mirrors M_1-M_2 were adjusted in absence of the beam expander to align the laser beam parallel to the vertical axis of the microscope. The beam expander was installed and adjusted while ensuring that the beam travels through the center of the optical lenses of the microscope. The laser power was adjusted with different neutral density filters (ND), collimated and directed with the illumination optical system of the microscope towards a filter cube G-2A (Nikon). The filter cube G-2A reflected the laser light into the microscope nose-piece. Mirrors M_1-M_2 and the 10X beam expander were adjusted again until the laser was visible on the ceiling above the 10X objective as a green, speckled circle. After which a mirror was placed upside down on the stage of the microscope to reflect the light back toward the filter cube G-2A. M_1 and M_2 were slightly adjusted again until the incoming beam and the reflected beam overlap completely. To verify that the laser was on the center of field of view, a reticule mounted into the microscope nose-piece was selected to confirm it. Schematic optical diagram of the illumination system is shown in figure (2.1). The beam expander was placed at $L = 45.2$ mm behind of collector lens to achieve conjugated planes with every objective of the microscope.

2.1.1. Spot Size

An oil-immersion objective (Plan Apo 100X, NA=1.4, WD=0.016 mm, Nikon) was used to focus the laser beam onto the sample for both single-molecule and ensemble FRET experiments. The 100X objective has spherical and chromatic corrections so that provides perfect axial and lateral color matching. The high numerical aperture (NA=1.4) of the objective allows to gather light from single molecules.
The laser spot size of the 100X objective was determined by focusing the laser into a test precision target of 0.01 mm scale (Thorlabs) and recording the image with a Cyber-shot DSC-W610 Nikon camera (Fig. 2.2). The spot diameter was measured analyzing the image on free-software ImageJ [44] which provided a spot size of 0.96±0.021 μm. According to diffraction theory for Gaussian diffraction limit [45],

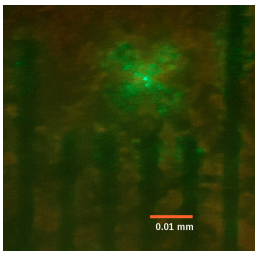

Figure 2.2.: The diameter of the focused spot from 100X objective onto a precision target of 0.01 mm scale.

this diameter should have a value of $d = 1.27\lambda f/D = 0.24\,\mu$m for our 100X objective ($f = 0.16$ mm is the focal length of the objective and, D =5 mm, is the input beam diameter). However, the above calculation assumes a perfectly collimated Gaussian-shaped and in our case possible aberrations in the optical elements of the beam expander may have altered the intensity distribution so that the spot size is slightly larger than the diffraction limit.

2.2. Confocal detection

Fluorescence emitted by the donor and the acceptor dyes were collected through the same 100X objective and transmitted through the filter cube G-2A toward tube lens L_1 (see Fig. (2.3)). The tube lens forms the primary image plane of the 100X objective at a distance of 8 cm measured from the surface platform, situated above

the head of the microscope where is the trinocular tube, to where the image plane is formed. This distance was measured using fluorescent micro-spheres with 0.5 μm of diameter and internally labeled with fluorescent dye TetramethylRhodamine (Bangs Laboratories). Low molar concentration of micro-spheres were introduced into a rectangular capillary tubing (0.05 mm of internal wall and 0.05 mm of external wall, Model 5005, Vitrocom) by capillarity and focused with the 100X objective. The micro-spheres were recorder on the image plane with a CCD camera placed at previous reported distance.

For FRET experiments, a single interference band-pass filter 67-118 BF1 (Technspec 514.5nm OD4, Edmund Optics) was placed before the tube lens to remove stray laser light at 515 nm transmitted from the filter cube G-2A.

To achieve the extremely high signal-to-noise ratio for single-molecule detection, we placed at the image plane position a precision 100 μm pinhole (Thorlabs). The pinhole rejects fluorescence originated from outside the focal region and fluorescence collected from peripheral light. Moreover it decreases sources of background fluorescence from Ramman and Rayleigh scattering [46].

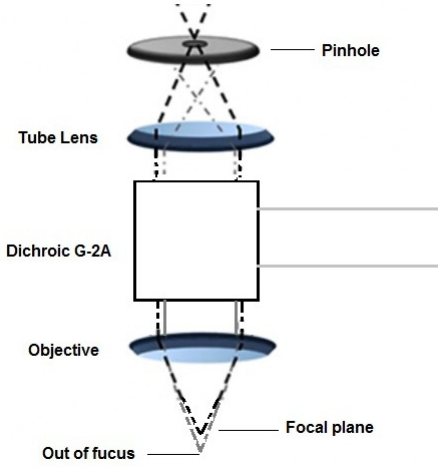

Figure 2.3.: Optical rays from different focal planes are collected by the objective but mainly rays from focal plane (orange rays) pass through the pinhole.

We achieved with the pinhole that a point light source was imaged at the object plane so that the illuminated point and the source were confocal. In other words, only an object in the volume confocal to the source and pinhole will contribute to

the detected light. According to Elisabeth *et al.* [47], the confocal probe volume can be approximated as a cylinder with a radius defined by the diffraction limited waist of a Gaussian beam as

$$V_p \sim (\pi/2)^{3/2} d^2 z_0/4 \tag{2.1}$$

where d is the spot diameter of our 100X objective and $z_0 \sim 1\,\mu$m its depth of focus (data provided by the manufacturer). Then according to equation (2.1) the estimated probe volume of our experimental system has a value of $V_p \sim 2$ fL. Therefore the probability P of detecting a single-molecule, in a concentration between C=90-120 pM, is $P = CV_pN_A$ ∼0.1-0.14 [48] (N_A is Avogadro's number). The low value of single-molecule occupancy probability at above concentrations guarantees that only one molecule is in the volume probed by the laser.

2.2.1. Fluorescence Detection Setup

Fluorescence emitted by the donor and the acceptor that passed the confocal pinhole was collimated by an achromatic doublet lens L_2 (effective focal length $efl_2 = 100$ mm. Edmund Optics) and directed towards a dichroic mirror (BS2) 540DRLP (Omega optical) which separated the donor and acceptor the fluorescence. Since the BS2 is unable to separate completely between the donor and the acceptor light, an optical filter centered at 580 nm with FWHM=30 (580DF30, Omega optical) and other one centered at 670 nm with FWHM=40 (670DF40, Omega optical) were placed in the optical path of the donor and the acceptor signal respectively. Each component (donor and acceptor fluorescence) was focused onto a single photon counting module (SPCM-AQR-14, Perkin-Elmer Optoelectronics), mounted on three-dimensional micrometer translation stages (NF15AP25, Thorlabs), using doubled achromatic lens (L_A efl = 45mm and L_D, efl = 40 mm, Edmund Optics). For each fluorescent-photon burst detected by the single photon counting module in an integration time $t = 1$ ms, stream of bits were sent directly to a SCB-68 card (National Instruments) and stored in a computer through an home-built Labview algorithm. Details of the Labview algorithm can be consulted in the appendix [**?**]. The integration time of 1 ms was selected for our FRET measurements since for biological samples with freely

diffusion motion, the typical residence time of the molecules in the laser volume is around 1 ms [50].

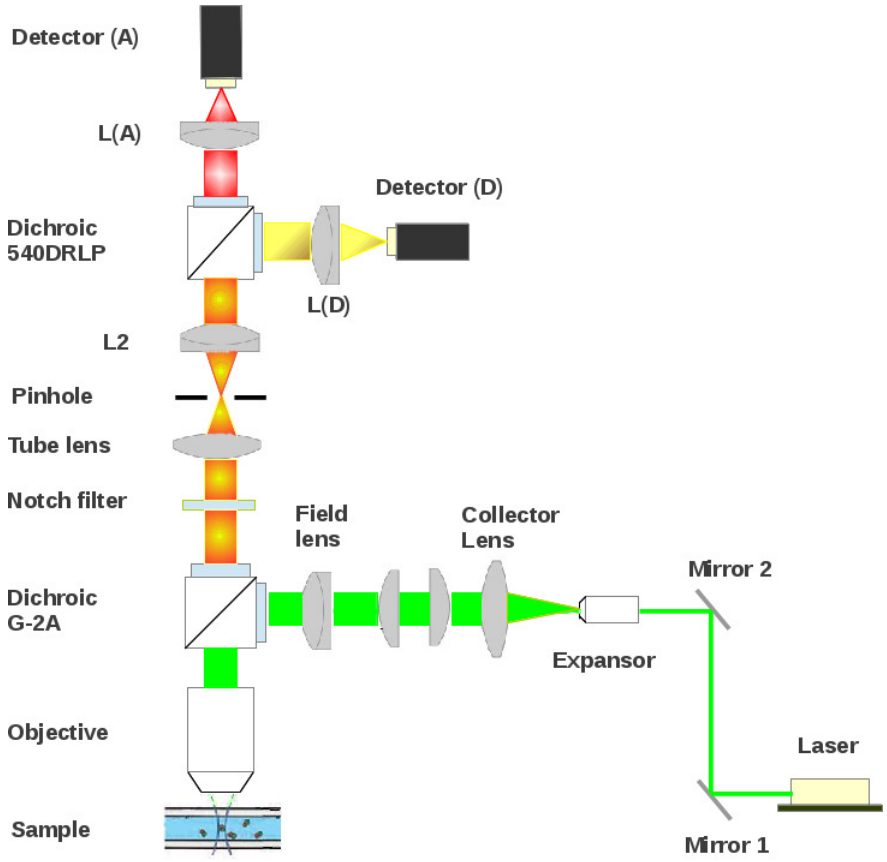

Figure 2.4.: Schematic diagram of the optical setup mounted in the laboratory for single-molecule FRET experiments. The excitation laser light is show in green and the fluorescence emission from the donor and the acceptor is shown in yellow and red respectively. Lenses $L(A)$ and $L(D)$ focus the fluorescent emission of the acceptor and the donor onto separated detectors. The detectors send the counts information to a personal computer through a SCB-68 card from National Instruments.

2.2.2. Detection alignment

The donor channel (D-channel) detector was aligned with the confocal volume by focusing immobilized fluorescent micro-spheres. Fluorescence emitted by micro-spheres was collected by the same objective and D-channel detector was adjusted until maximum intensity was detected in a bin-time of 1 ms. The acceptor channel (A-channel)

detector was aligned by using FRET between donor dye alexa fluor 546 (AF546) and acceptor dye alexa fluor 647 (AF647). Low concentration mixture of above fluorophores was focused at low intensity with the 100X objective. Then, the A-channel detector was adjusted until maximum intensity was detected in a bin-time of 1 ms. Figure (2.5A-B) shows the photon trajectories in the D-channel and A-channel respectively due to its alignment with the confocal volume. Fewer number of counts in the A-channel compared to the D-channel is due to the different samples used for its alignment and not by the optical properties of the system. The total optical design of our setup for measure FRET is schematically described in figure (2.4). The laser and all optical components used in the illumination and the detection systems were mounted over an optical table with rigid spacers coupled to Thorlabs cage system components to ensure that the optical components were free of vibrations.

We estimate the collection efficiency in both detection channels by directing atten-

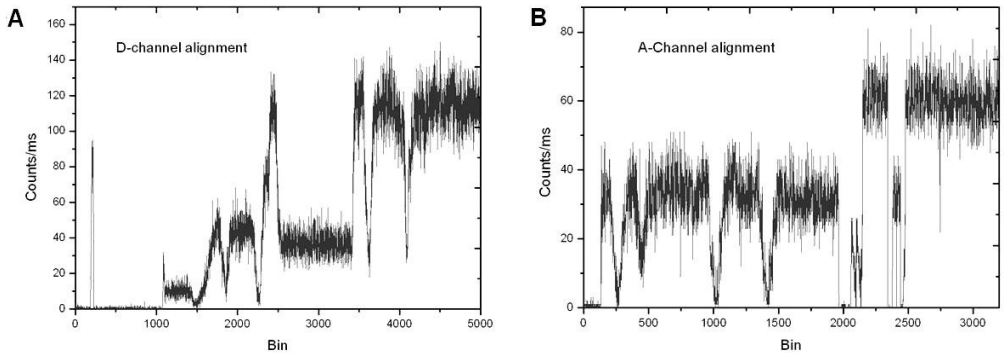

Figure 2.5.: Photon trajectories originates from (A) fluorescent microsphere of 0.5μm of diameter in the D-channel, (B) and mixture of AF546-A647 dyes in the A-channel

uated white light intensity in the condenser lens of the microscope. After, optical filters centered at 580 nm or centered at 670 nm were mounted over the sub-stage of the condenser lens. Light power over the optical filters was measured with the P100 power meter giving a value of $\sim 1\mu$W. A dark disk cardboard with a pinhole of 250 μm on its center, covered the optical filters. The power over the pinhole was estimated to be of the order of 0.12 nW and 0.17 nW when the 580 nm and 679 nm optical filters were mounted onto condenser lens respectively. The 250 μm pinhole was focused with the 100X objective and the collected light by the microscope was

directed toward the detectors. The number of counts from each detectors was converted into power and compared to the collected power. The collection efficiency for D-channel was about $\eta_D \sim 5.3\%$ and for the A-channel $\eta_A \sim 9.8\%$. The difference in collection efficiency could be partially attributed to the photon detection efficiency of the single-photon counting module SPCM-AQR-14 because at 670 nm it has an efficiency of ∼72 % while at 580 nm it has an efficiency of ∼61 %. With the above values, we could estimated the value of the correction factor $\gamma = \eta_A\phi_A/\eta_D\phi_D \approx 0.8$ for our optical setup for the case of dye pairs AF546 and AF647 (Quantum yields for above dyes reported by Invitrogen [51] are ϕ_D=0.79 and ϕ_A=0.33), nevertheless, the accuracy of the measures of $\eta_{A,D}$, variations in the quantum yield of the dyes due to effects of pH solution and temperature variations as well as variations of the quantum yield of the donor in presence of the acceptor, it makes that the correction factor γ estimated using the above method may not be the most appropriated [52,53]. Accurate γ measurements can be obtained by using alternating-laser excitation (ALEX) [52].

2.3. Background Signal

A single detection bin contains counts arising from fluorescence emission (if there is a fluorophore in the confocal volume), from scattering (Rayleigh and Ramman) of the excitation light, and a background signal from the electronic dark counts. To measured these contributions, we built several cleaned sample chambers. Briefly, a sample chamber consisted of a borosilicate glass rectangular tubing (wall diameter 0.05 mm, internal wall 0.05 mm, model 5005, Vitrocom Inc) which at each extreme had attached a sterile tube affixed with epoxy. The rectangular tubing was fixed on a cover glass with epoxy and cleaned with acetone, washed with nanopure water ($18M\Omega cm^{-1}$) and incubated in 1M of Potassium hydroxide (KOH) for 3 min. It was washed again, first with distilled water and after with absolute ethanol. Finally the sample chamber was cleaned again with distilled water. Schematic representation of the sample chamber is shown in figure (2.6).

We measured the background caused by the assembled chamber, dark counts from detectors, immersion oil (refractive index n=1.517, Sigma Aldrich) and different solvents worked in this thesis such as TE buffer (10 mM Tris-HCl pH=8.0, 1 mM EDTA,

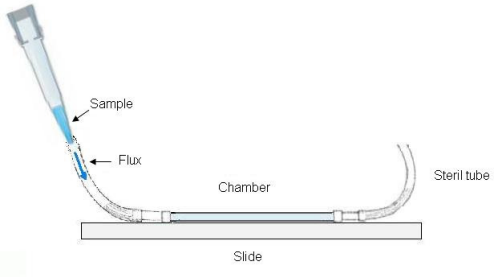

Figure 2.6.: Schematic representation of the sample chamber built for single-molecule FRET experiments

prepared in nanopure water 18 MΩ cm^{-1} and filtered with 2 μm pore membrane Millipore), nanopure water (18 MΩ cm^{-1}, Nanopure System), and PBS 1X buffer (37 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, pH 7.4, prepared in nanopure water and filtered). Solvents were loaded onto a sample chamber and focused at different laser power with the 100X objective a few microns below the glass wall. Figure (2.7) shows the typical photon background counts as a function of time in both channels (Donor, Acceptor) from the sources aforementioned obtained with a laser power of ∼50 μW ($I \sim 6$ KW/cm^2), where acceptor counts are plotted inverted. From this

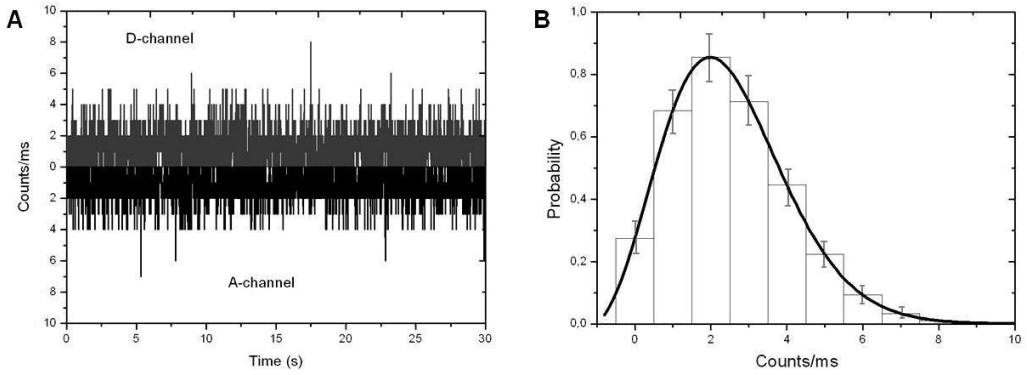

Figure 2.7.: (A) Number of counts detected in both channels in 1 ms time windows from background of TE buffer solution. Acceptor counts is plotted inverted. (B) Background counts from TE buffer solution in D-A-channels follows a Poisson distribution. The background mean in both channels was ∼ 2 counts/ms at 50 μW ($I \sim 6$ KW/cm^2.

data we observed that the background counts occurred randomly at a fixed mean rate. The mean background in both channels was extracted by fitting the background

data with a Poisson distribution. Mean background in D-channel and A-channel was around to 2 counts/ms either in water, TE buffer, or in PBS 1X buffer. For laser powers above of ∼50 μW, the mean background on both detector channels increases linearly. For instance at ∼80 μW the mean background in D- and A-channels increased to 3.1 and 2.7 counts/ms respectively, while at ∼150 μW, it increased to 5.3 and 4.8 counts/ms respectively. For low intensity power (e. g. 20 μW) the mean background in both channels was 1.87 counts/ms, a similar value to the one obtain at 50 μW. For our single-molecule experiments we worked with a laser power of 50 μW.Background counts was similar with and without the Notch filter.

2.4. Crosstalk detection

Crosstalk from the donor AF546 signal to the A-channel was calculated by loading only donor fluorescent dye, at 120 pM in TE buffer solution, into a pre-incubated sample chamber with 0.001% of Tween 20 to prevent surface adhesion of the dye. The laser was focused into sample chamber and photon-burst (intervals during which the count rate increases significantly above the background signal) trajectory were recorded in a 1 ms of integration time. Figure (2.8) shows the time-dependent fluorescent signal observed when a donor transverse the probe volume. Because of the

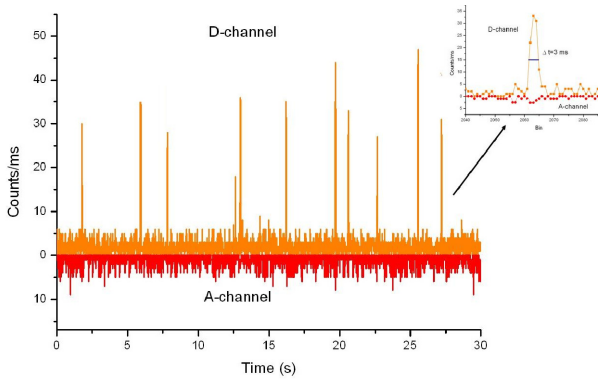

Figure 2.8.: Single-molecule signal from freely diffusing alexa fluor 546. No significant cross-talk is detected at A-detection channel when alexa fluor 546 transverse the probe volume. In the inset is showed a typical burst time due to AF546 molecules transverse the confocal volume

low concentration of the AF546 dye, most of the events are derived from the back-

ground signal. Several things can be observed from the cross-talk photon AF546 trajectory: The first is that the signal from donor clearly is above of the background signal with a mean transit time of 1.3 ms and an average burst sizes of 20 counts/ms which results in an signal-to-noise ratio (SNR) of $(S/N) \sim 14$, sufficient to distinguish single-molecule signals from background. The second, is that the observed jumps in signal intensity obeys a Poisson statistics as previous theoretical works have proved for freely diffusing single-molecule experiments [55, 70]. The third, is that a lower signal contribution (average 0.5 photons/ms) from the donor signal to the A-detection channel is detected. However, at high laser power ≥ 300 μW, ~ 3 photons/ms appears in the acceptor channel. Therefore we decided to work at 50 μW of laser power.
Crosstalk from acceptor signal to D-channel was not measured due to the lack of appropriate excitation light.

Conclusion

We present our optical setup used to observe single-molecule signals, specifically from FRET experiments. Our optical setup was built by coupling an illumination and a detection system to a commercial epi-fluorescent microscope. The illumination system contains a laser and a beam expander to create a spot of appropriate diameter at the back focal plane of the 100X objective mounted on the nose piece of the microscope. The 100X objective has a high numerical aperture that allows to maximize the collection efficiency of photons.
The detection system is based in confocal detection in order to limit the observation volume and minimize the background from Rayleigh and Ramman scattering. The detection system also contains single-photon-counting avalanche photo-diodes which shows a high quantum yield as well as a fast temporal response. The detection system is completed by a SSCB-68 card that sends the data collected by the detectors to a personal computer through LabView home-built algorithm. Finally, negligible crosstalk was detected from single-molecule signals from donor into A-channel at 50-80 μW of laser power in the focus.

3

Optical Setup Calibration: FRET Efficiency as a Function of the Inter-dye Distance R_{DA}

Single-molecule FRET has been applied to both immobilized molecules on surface and in free diffusion [59]. With freely diffusing samples the transit time through the excitation region limits the longest time to follow the dynamics of a single molecule. However, it allows the ability to detect a great number of different molecules instead of only one molecule as it is in immobilized FRET experiments [60]. In both types of smFRET experiments, structural and dynamical information is commonly extracted from the FRET efficiency expression

$$E = (1 + (R_{DA}/R_o)^6)^{-1} = I_A/(I_A + \gamma I_D). \tag{3.1}$$

Dynamical effects as well as photo-physical, photochemical, and instrumental factors, are ones of the biggest obstacles for an accurate distance measurement with the FRET technique [61–63]. Therefore accurate measuring requires knowledge of the detection-correction factor γ. Moreover, since fluorescent dyes are typically attached to the bio-molecule via flexible linkers which involve uncertainty in dye position, quenching environment, and even restriction in dye reorientation, it is also necessary to reaffirm the R_o value for the pair of dyes elected [64]. A way to estimated the effects of γ and R_o in FRET measurements, may be obtained by measuring an apparent FRET efficiency with known inter-dye distances R_{DA} as

$$E_{app} = I_A/(I_A + I_D) = 1/(1 + (R_{DA}/R_{eff})^6), \tag{3.2}$$

where the parameter R_{eff} contain the information of R_o and γ as $R_{eff} = R_o\gamma^{1/6}$. These measurements leads to a standard calibration for FRET efficiency as a function of inter-dye distance that can be used as a spectroscopic ruler for single-molecule fluorescence experiments [65].

In this chapter we show the FRET efficiency as a function of distance separation between donor-acceptor dyes measured with our optical setup. We measured FRET efficiency distributions for a donor (AF456) and an acceptor (AF647) dyes attached at opposite ends of freely diffusing DNA molecules of various known lengths. The observed mean FRET efficiencies were plotted as a function of distance for the DNA constructs and fitted by equation (3.2). The adjusted curve give us a value of R_{eff}=8.5±0.9 nm. This calibration curve will serve us as a spectroscopic ruler for distance measurements in nucleic or ribonucleic acids.

3.1. Materials and Methods

We designed short fragments of single-stranded deoxyribonucleic acid (ssDNA) and their complementary strands in order to get doubled-stranded deoxyribonucleic acid (DNA) of lengths 10, 13, 16, 19, 20, 21, 22, 25, 28 and 45 base pairs (bp), where 1 bp=0.34 nm [66]. These lengths are much smaller than the persistence length reported for DNA ($\sim$ 150 bp) [67, 68]. Therefore, our shorts fragments of DNA can be treated as a rigid rods at room temperature and pH between 3-9.

ssDNA primers were designed such that the DNA leaves 3'-recessed ends. Table 1 shows the sequence of each oligonucleotide (for) and its complement (rev). DNA fragments were purchased to Invitrogen Life Technologies

3.2. Hybridization of DNA

To hybridize complementary fragments of ssDNA, each complementary strand was re-suspend in annealing buffer (10 mM Tris-HCl pH 8, 100 mM NaCl and 1 mM EDTA. Filtered with 0.22 μm Millipore membrane GSWP09000) at the same molar concentration. Thereafter, equal number of moles of both complementary oligos were mixed in an eppendorf tube (free of DNases) and heated at 80 ^{o}C for 3 minutes in a thermal bath followed by slow cooling to 4 ^{o}C. To check doubled-stranded DNA

Name	Sequence from 5' to 3' end
For10	agacgtgag
Rev10	gctcacgtc
For13	agacaaggtgag
Rev13	gctcaccttgtc
For16	agacaaattggtgag
Rev16	gctcaccaatttgtc
For19	acacggttcacacgtctg
Rev19	gcagacgtgtgaaccgtg
For20	agacgtgtgacaccgtgag
Rev20	gctcacggtctcacacgtc
For21	agacgtgtgtgaaccgtgag
Rev21	gctcacggttcacacacgtc
For22	agacgtgttgtgaaccgtgag
Rev22	gctcacggttcacaacacgtc
For25	agacgtgcaattgtgaaccgtgag
Rev25	gctcacggttcacaattgcacgtc
For28	agacgtgtgaccgcatttttgagtgag
Rev28	gctcaccaaaaatgcggtcacacgtc
For45	agacgcgcttactagtgcaaattgtgaccgattttgagtgag
Rev45	gctcactcaaaatcggtcacaatttgcactagtaagcgcgtc

Figure 3.1.: Single-stranded DNA comlementary sequences for calibration of FRET equipment

formation gel electrophoresis technique was used. For this purpose, a polyacrylamide gel at 8% of concentration was prepared. Briefly, 3.2 mL Acrylamide:bisacrylamide (29:1, 30% w/v. Invitrogen) were mixed with 2.5 mL H_2O, 2.4 mL 5X TBE buffer (1.1 M Tris; 900 mM Borate; 25 mM EDTA; pH 8.3), 0.2 mL 10% Ammonium persulfate (APS, Invitrogen) and 0.01 mL Tetramethylethylenediamine (TEMED, Invitrogen). The polyacrylamide was poured in an electrophoresis chamber (Bio-Rad) and it was used until solidified. Gel electrophoresis technique allows to separate fragments of DNA or RNA by size or molecular weight by the application of an electrical strength at the top (anodal, negative) end that causes the migration of the DNA or RNA towards the bottom (cathodal, positive) due to negatively-charged of DNA. The rate of migration is proportional to size: smaller fragments move more quickly, and wind up at the bottom of the gel.

A small quantity of each DNA-hybridization reaction was loaded into polyacrylamide gel along with the corresponding single-stranded oligo as control. The DNAs were running around 10 minutes at 4 V/cm. Binding of single-stranded DNAs leads to duplex formation that results in a upper band shift on the gel as shown in figure (3.2) for some DNA aforementioned. The rest of DNA was stored at -20 ^{o}C until use.

Figure 3.2.: Polyacrilamide electrophoresis gel to observed DNA hybridization

3.3. DNA Labeling

The fluorescent dyes chosen for our optical FRET system calibration were the Alexa Fluor 546-14-dUTP (AF546) as donor and the fluorescent dye Alexa Fluor 647-12-OBEA-dCTP (AF647) as acceptor. Both fluorescent molecules were purchased to Invitrogen Life Technologies. The nucleotides deoxyuridine triphosphate (dUTP) and the deoxycytidine triphosphate (dCTP) attached to the dyes are modified at the C-5 position of uridine and cytosine via the unique alkynyl amino flexible linker respectively, which provides a spacer between the nucleotide and the dye to reduce interactions between them. However, the size of the linker causes significant uncertainties in quantitative FRET distance measurements [64]. The size of the above linkers can be determined by the number in the product name, e. g., the "12" in AF647-12-dCTP, indicates the net length of the spacer, in atoms.

Doubled stranded DNA fragments were labeled with above dyes by filling the 5'-overhangs of both complementary strands using the 5'→3' polymerase activity of the Large Klenow Fragment (KLF, Invitrogen). In the presence of nucleotides the KLF can incorporate one nucleotide at a time onto a pre-existing 3'-OH group by using its polymerase activity [69]. The reaction for labeling DNAs contained 0.5 U/μl DNA polI diluted in Klenow buffer (50 mM potassium phosphate pH 7.0, 100 mM KCl, 1 mM Dithiothreitol), 3 μl 10X REact 2 Buffer (500 mM Tris-HCl pH 8.0, 100 mM $MgCl_2$, 500 mM NaCl), 1 μg DNA, 1 μl 0.5 mM Alexa Fluor 546-dUTP, 1.25 μl 0.5 mM Alexa Fluor 647-dCTP. The mixture was carried out to a final volume of 30 μl with nuclease free water (Promega) and incubated at 4 ^{o}C for 20 minutes. Fluorescent dyes not linked at the 3'-recessed ends of each DNA were removed using

gravity columns (Sephadex GE) following manufacturer's protocol. Final product was quantified with Nanodrop 2000c system. The absorption spectrum for some DNA-alexas construction is shown in figure (3.3). The first absorption peak (from

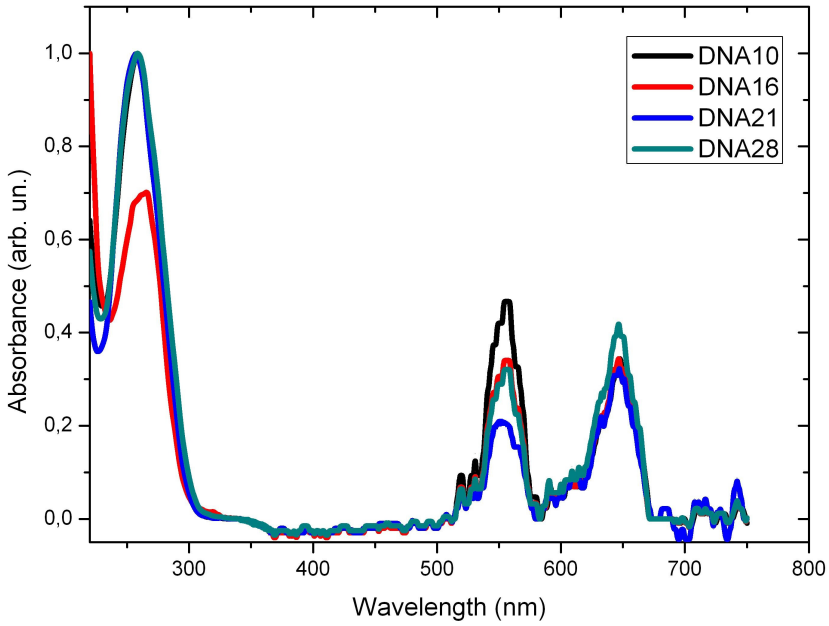

Figure 3.3.: DNA-alexas constructions absorption spectrum. From left to right, DNA, Alexa fluor 546 and Alexa fluor 647 peaks. The number in DNA, e.g. DNA10, indicate the length of DNA in base pairs.

left to right), centered at 260 nm corresponding to DNA while the other two peaks corresponding to the AF546 and the AF647 respectively.

An estimate of the efficiency of labeling for each DNA-alexas construction was extracted from the spectrum. The results give an average of labeling for AF546 of 87±2% while for AF647 of 91 ± 2%. Stock solutions of 10 μM for each DNA-alexas constructs was prepared in TE buffer (Tris-HCl 10 mM pH=8.0, 1 mM EDTA, filter with 2 μm pore membrane Millipore) and stored at -20^{o}C until use. The entire process of DNA-labeling was performed avoiding as much as possible light sources. Figure (3.4) shows a schematic representation of DNA-alexa construction beside the size of the linker of the dye.

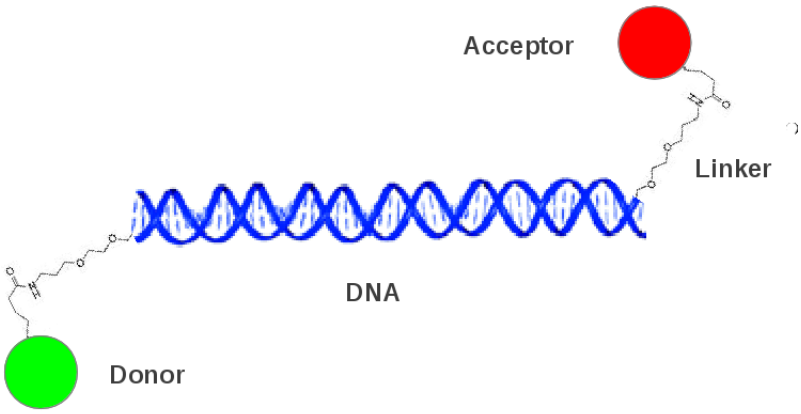

Figure 3.4.: Schematic representation of DNA labeled with fluorescent dyes Alexa Fluor 546 (donor) and Alexa Fluor 647 (acceptor). Dyes were labeled at opposite sites of each DNA.

3.3.1. FRET efficiency as a function of inter-dye distance

Efficiency of energy transfer from the donor AF546 to the acceptor AF647 was measured by using a concentration of labeled DNA molecules between 90 and 150 pM in TE buffer solution. The lowest DNA concentration was obtained by dilution series. A small volume ($\sim$ 80 μl) of sample solution was loaded into the sample chamber (previously described on chapter 2) and focused with the 100X objective. The focal point was placed within the sample and fluorescence burst events on both the donor and the acceptor channels were observed in an integration time of 1 ms. We consider only clear events above the background. Figure (3.5) shows the photon burst trajectories as a function of time as single DNA-alexa construction molecules DNA13, DNA19, DNA28, and DNA45 transverse the confocal detection zone, where the DNA-number indicate the length of the DNA in base pairs. The photons in the D-channel are plotted in green while photons in A-channel are plotted in red and inverted. From these photon trajectories the average amplitude of burst of photons detected in the A-channel appears to increase when acceptor is closer to the donor dye, just as the Förster theory for FRET predicts. On the other hand the reason that the amplitude of photons in each burst changes in amplitude, beside FRET effects, is also influenced by stochastic process like excitation, non-radiative decay,

and even the spatial localization of the donor dye in the probe volume [70].

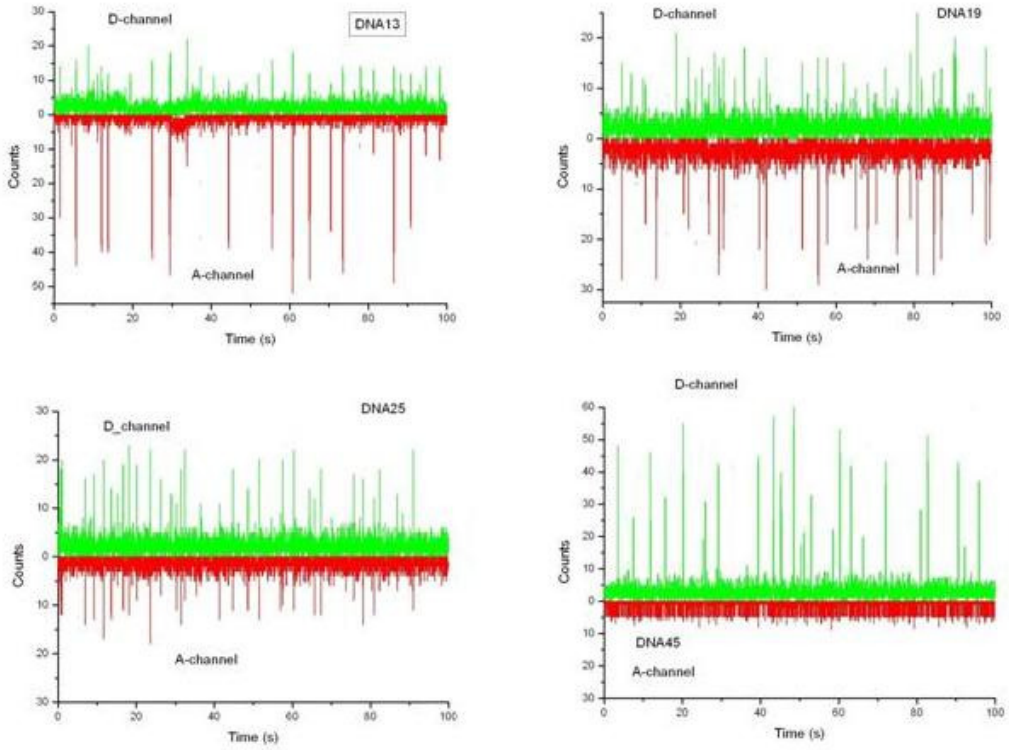

Figure 3.5.: Time fluorescence traces in donor and acceptor detection channels from freely diffusing DNA13, DNA19, DNA28 and DNA45 molecules in solution. Fluorescence bursts above background are clearly visible as molecules traverse the laser beam.

Figure (3.6) shows typical burst duration distribution obtained with our freely diffusing DNA-constructs DNA10, DNA16, DNA28 and DNA45, when transited through the confocal volume. From these data the mean burst duration is on the order of 2.8 ms. Previous experimental studies realized at room temperature with freely diffusing short DNA fragments, in 40 mM Tris-acetate-EDTA, showed that DNA molecules had a diffusion constant of 120×10^{-8} cm^2/s [71]. This value gives a transit time through the excitation volume of about ~ 2.28 ms, similar to the one we observed.

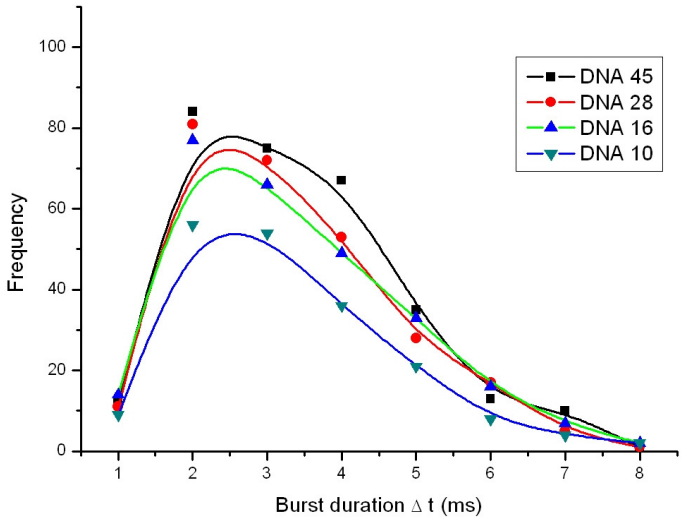

Figure 3.6.: Burst duration distribution from DNA-alexas 45, 28, 16 and 13 bp. The solid curve is only a guide of the data.

Photon traces from FRET trajectories (Fig. (3.5)) contain photon counts from fluorescent molecules (N_A and N_D), background counts (B_A and B_D), and cross-talk between the two detection channels (α_D and α_A. Not detected). Therefore, the total fluorescence intensity from each ith burst is

$$I_A^i = N_A^i + < B_A > + < \alpha_A I_D >, \quad I_D^i = N_D^i + < B_D > + < \alpha_D I_A > \tag{3.3}$$

where I_j^i, $(j = A, D)$ are the fluorescence intensity detected into donor and acceptor channel respectively, $< B_j >= 2$ photons/ms and $< \alpha_j I_j >= 0$ are the mean background and cross-talk signal in the above detectors. For FRET analysis, after single-molecule burst were identified, a threshold in each detector channel was selected to reject signals from no FRET events. The threshold, T, in each detector channel was selected by a Bayes sense theory for signal analysis [72] as

$$T = \frac{\ln\left(\frac{1-P}{P}\right) + \langle N \rangle}{\ln\left(\frac{\langle B \rangle + \langle N \rangle}{\langle B \rangle}\right)}, \tag{3.4}$$

where P is the probability of having a DNA molecule in the probe volume at time t, $(P = N_A C V_p \sim 0.1$. See chapter 2), $\langle B \rangle = 2$ counts/ms is the mean of background

signal (see Chapter 2) and $\langle N \rangle$ is the mean of counts from molecular signal in the $j-$detection channel. For our FRET experiments the total threshold was derived as the sum of individual detection channel-threshold given by equation (3.4). The total threshold was between 10-12 counts for our DNA-alexas constructs. Therefore only the ith bins above the threshold were accepted for FRET analysis. For the accepted i burst events a proximity ratio or apparent FRET efficiency E_i [73]

$$E_i = \frac{I_A^i}{I_A^i + I_D^i}, \tag{3.5}$$

was computed, where the correct donor and acceptor counts, I_j^i, $(j = A, D)$, are given in equation (3.3). The proximity ratio assume a value of $\gamma = 1$ in FRET efficiency expression, as such, any fluctuation in quantum yield will be reflected in the proximity ratio distribution [74].
Each FRET efficiency E_i computed from photon trajectory was stored on a personal computer and a statistical analysis was performed with a home-written Octave algorithm. Details of Octave algorithm can be consulted in the Appendix B. From this analysis, FRET efficiencies for each DNA-alexas constructs were plotted in an histogram. Figure (3.7) shows the results for the set of the DNA-alexas constructions worked in this thesis. The FRET histograms were binned with 0.035, and fitted by a Gaussian. The additional peak at a transfer efficiency close to zero, in DNA-alexas FRET histograms, is thought to be caused by DNA molecules lacking an active acceptor dye and was not included in the calculation of mean transfer efficiencies.

The corresponding width ΔE of FRET histograms is a consequence of the photon statistics and is governed by the shot noise in the system. It is a function of the mean energy transfer E, the detection factor γ, and the detected photon number $N = N_A + N_D$ [73, 75]. An estimate of the effect for $\gamma = 1$ is approximate by the sum of variances in the donor and acceptor counts leading to [70]

$$\sigma_E = \sqrt{\frac{E(1-E)}{N}}. \tag{3.6}$$

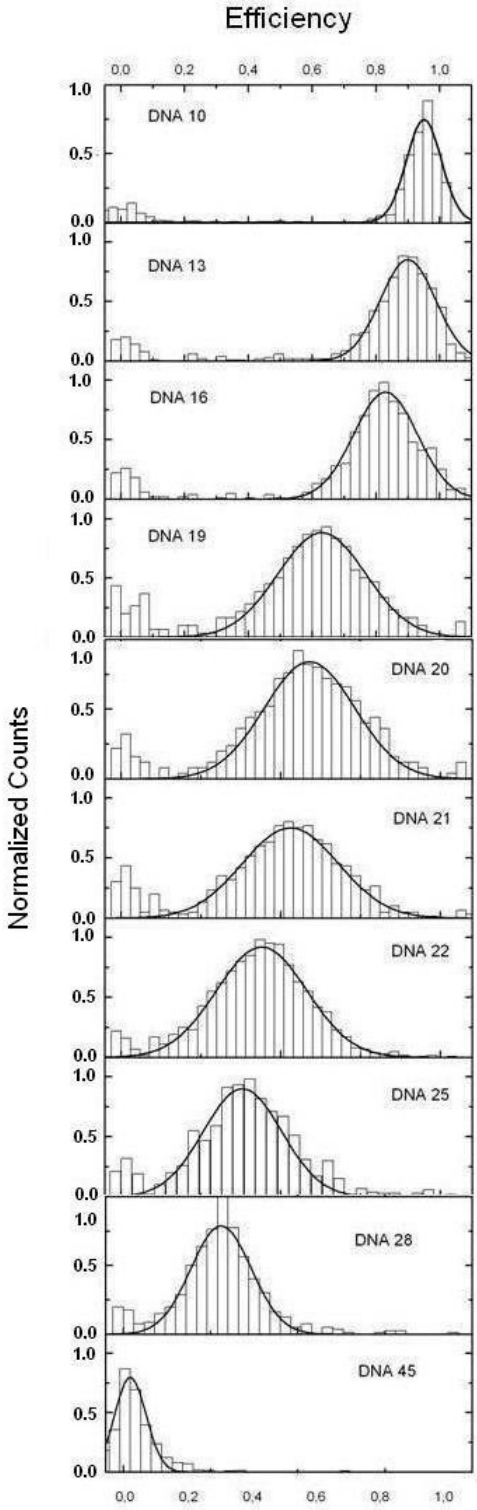

Figure 3.7.: Single-molecule FRET efficiency histograms measurements from different constructions of DNA-alexas. The acceptor and donor dyes are linked to the opposite ends of the DNA chain.

This result indicated that at $E = 0.5$ ($\sigma_E = 1/2\sqrt{2N}$) the shot-noise fluctuations in the non-correlated variables I_A and I_B have the largest effect. In contrast, the width diminishes for $E = 0$ and $E = 1$ as we can see in our FRET DNA-alexas histograms. But so far, ΔE is referred to as the "width" of the FRET distribution but not related to a conventional statistical measure, because complete description of the width of the FRET distribution also requires the knowledge of fluctuations in FRET energy transfer due to the motions of the dye-linkers [65].

3.4. FRET Efficiency: Curve of Calibration

The mean FRET energy transfer obtained from each DNA-alexa construction can not be associated directly with the distance of separation between the donor and the acceptor, because as we mentioned above, the dyes are modified with a flexible linker of length l that introduces an additional contribution to the separation. Therefore for quantitative FRET measurements it should be taken into account [11,73]. Different methods such as accessible volume (AV) [76] and Langevin molecular dynamics simulations [65] have been used to solve it. In the AV method it is assumed that the dye, which is connected with a flexible link l_{eff} can move in positions, that do not cause steric clashes with DNA. A geometric algorithm is used to find all allowed positions. S. Sindbert and coworkers [64] using the AV method and molecular dynamic simulation conclude that for dyes with longer linkers, as it is in the case of our dyes, the effective length was ~ 0.75 nm with a deviation of 0.22 nm. Adding these lengths to the length of each DNA, the mean FRET efficiencies extracted from FRET DNA-alexas histograms were plotted as a function of inter-dye distance as is shown in figure (3.8), where the error bars represent one standard deviation ($\pm\sigma$) from multiple measurements. The corresponding solid line is a fit using the equation (3.2), giving $R_{eff} = R_o\gamma^{1/6} = 8.5 \pm 0.9$nm. A change in of 0.5 nm in the mean position of the dye-linkers, cause a similar change of 0.5 nm in R_{eff}. Therefore, the uncertainty in the R_{eff} will be the same of the linker. A conservative value for the uncertainty is then at the 10% level.

The discrepancy between the value R_{eff} and R_o, estimated from spectroscopic data for AF546 and AF647, may be partly caused by the correction factor γ. The FRET

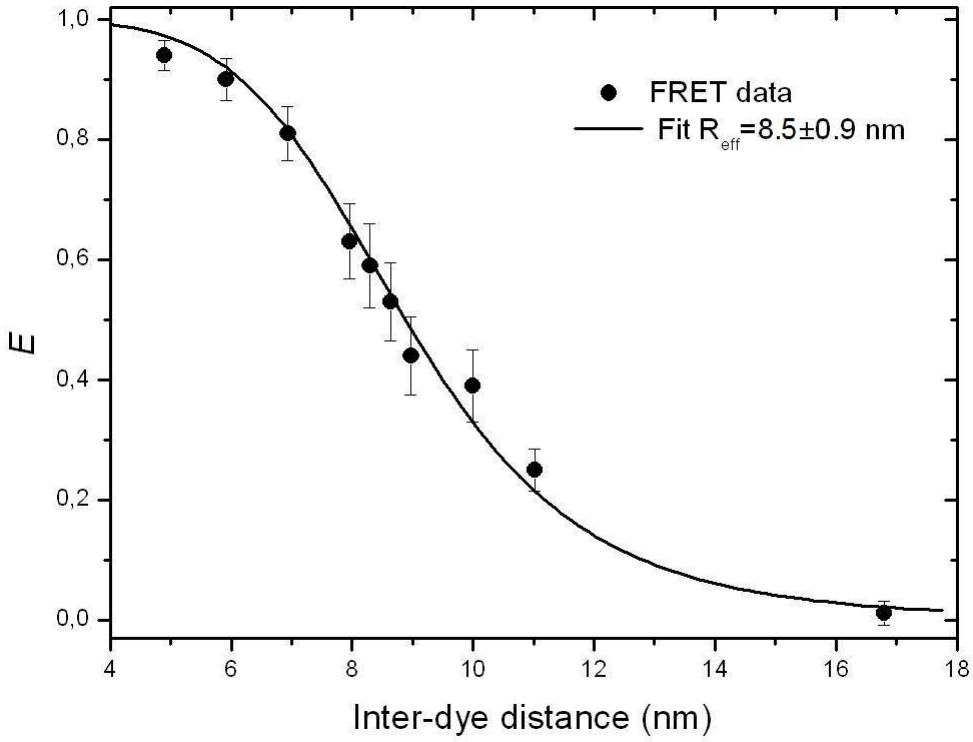

Figure 3.8.: Transfer efficiencies extracted from smFRET histograms as a function of the separation of the dyes. The solid line is a fit to equation (3.2) with $R_{eff} = 8.5 \pm 0.9$ nm.

efficiency curve can be used as a spectroscopic reference for single-molecule fluorescence experiments where nucleotides are involved since DNA and RNA have similar chemical behavior.

3.5. Width of DNA FRET distributions

The width of the histograms is slightly higher than the one limited by statistics Eq. (3.6). The extra broadening can be explained by the motion of the dyes through a spherical volume of radii dR with probability distribution $P(y) = \pi dy^2$, where $dy^2 = \pi(dR^2 - dz^2)$ is the radii of a disk at a variable height z (Fig. 3.9) The mean square displacement of each dye-linker is $(\int y^2 P(y) dy)^{1/2} = dR$. Thus the variation

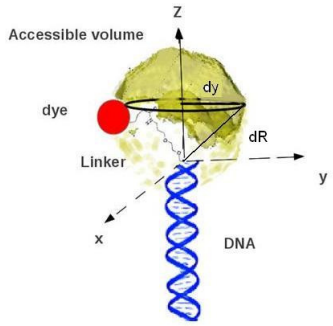

Figure 3.9.: Schematic representation of the accessible volume for a dye-linker attached to an end of DNA.

of FRET efficiency for dye motion is

$$dE = 6E^2 \left(\frac{R}{R_o}\right)^6 \left(\frac{1}{\sqrt{2}}\frac{dR}{R}\right) \tag{3.7}$$

This fluctuation in each dye-linker position contribute to the width of FRET histograms for DNA-alexas constructions. Therefore the total width of the FRET histograms coming from statistical contribution and dye-motion. The statistical width

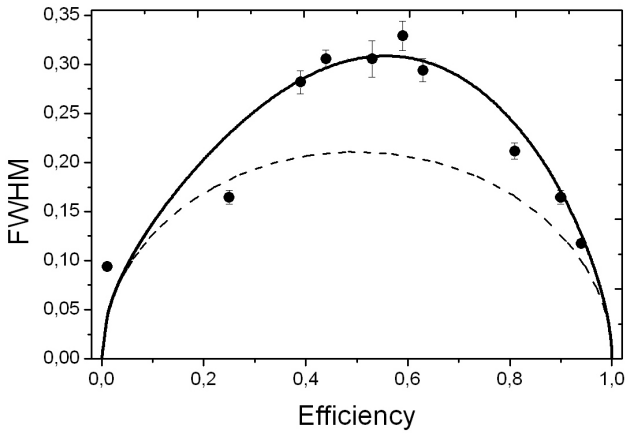

Figure 3.10.: Width of the FRET-DNA histograms (circles) compared to the one expected from statistics (dotted line) and that obtained once adding also the contribution from the motion of the fluorophores within a range of 0.8 nm (black line).

(Eq. 3.6) is symmetrical with respect to E as we can see in figure 3.10 (dot line), while for dE (3.7) it is not symmetrical due to the dR contribution (Fig. 3.10, black

line). The total width matches well the experimental data. Therefore according to S. Sindbert [64] the orientation effects of the fluorophores and other contributions such as the undefined dye environment or the correction to the calibration due to the fluorophore motion are well below the quoted (10%) uncertainty

Conclusion

Short DNA fragments labeled with donor (AF546) and acceptor (AF647) dyes placed in opposite ends were used as a spectroscopic ruler for single-molecule FRET distance measurements. Mean FRET efficiencies extracted from each DNA-alexa construction histograms were plotted vs the inter-dye distance. The resulting calibration curve gives a R_{eff} value of 8.5±0.9 nm, larger than the one reported by the manufacturer. The difference in value comes from the value of γ for our optical system and the dynamics associated with the linker. The width of the DNA FRET histograms can be explained well taking into a count the motion of the linkers as we could see in figure (3.10). The orientational effects of the dyes become negligible, and other contributions such as undefined dye environment or the correction of the calibration due to dye motion are well bellow the quoted (10%) [73].

4

How Closed are the Ends of Large RNA's?

Ribonucleic acid (RNA) is one of the principal biomolecules that are essential in all known forms of life. RNA is composed by a linear arrangement of four basic molecules called nucleotides (nt); adenine (A), cytosine (C), guanine (G), and uracil (U). Each nucleotide consists of a ribose sugar, a phosphate group, and a nitrogenous base (Fig. 4.1).

RNA plays a central role in many cellular processes such as splicing [77], translation tasks [78], signal recognition of proteins [79] and catalytic activity per se [80–84]. These variety of biological task are often intimately related to its native structure also know as tertiary structure. However, most functional RNA molecules exhibit secondary structure that is highly conserved in evolution [85–90] and moreover, stable secondary elements fold fast and determine the three-dimensional structure.

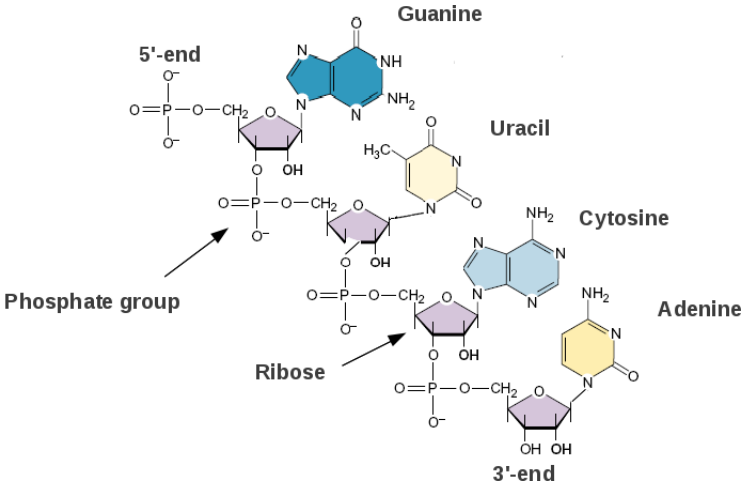

Figure 4.1.: Structural features of the RNA. The RNA is composed of alternating phosphate and ribose moietie

RNA secondary and tertiary structure is formed, in the Watson-Crick rule, by the base-pairing formation of A-U, G-C and G-A through hydrogen bond interactions.

Adjacent nucleotides that are linked through the 5',3'-phosphodiester bonds form the covalent structure of the molecule (Fig. 4.1). The labels 5' and 3' correspond to the carbon 3 and the carbon 5 in the ribose sugar. Nucleotides that do not form base-pairing generally leads to the formation of stem-loop structures in which the intervening RNA is looped out from the end of the double-helical segment as in a hairpin, a bulge, or a simple loop including the external loop. In RNA secondary structure the external loop is the loop that contains one or more nucleotides as well as the 5' and 3' terminals also known as the ends of the RNA molecule. A schematic representation of an RNA secondary structure is shown in figure (4.2).

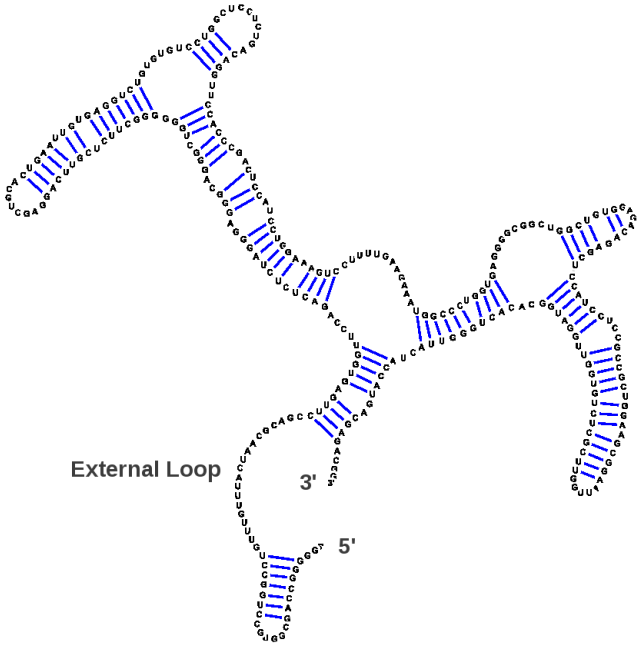

Figure 4.2.: Secondary structure predicted by mFOLD [95] of a 300 nt fragment of the messenger RNA *phr1* from *T. atroviride* fungus. External loop containing the 5'-3' ends.

In references [91–94] they found that the percentage of paired nucleotides (f), duplex length (k), and the average number of nucleotides in an internal loop (l) for a single stranded RNA (ssRNA) molecule, which form secondary structure, approach a constant value with increasing number of nucleotides (nt). The constancy of f, k and l has been confirmed for a wide range of viral and yeast ssRNA sequences

by application of the mFOLD [95] and RNA Vienna [96] algorithms. Based in these findings and using some theoretical tools and the mFOLD and the RNAsubopot programs, Yoffe *et al.* [97] showed that the ends of large RNA molecules ($\geq$ 1000 nt) are necessarily close independent of its base composition and length. The authors found that on average the 5'-3' ends for the viral RNA sequences tested were separated between 13 and 20 nt links, whereas for a random RNA sequence the separation was $\sim$ 12 nt. The biological relevance of this findings speculated by the authors reside in that many biological processes requires that the 5'-3' ends of the RNA molecules should be close to each other since that facilitates, for instance, the circularization of RNA molecule for effective translation. Examples include the translation control of specific messenger RNA (mRNA) by closing the 5'-cap and the 3' poly(A) tail through binding the eukaryote initiation factor protein eIF4E with the eukaryote initiation factor protein eIF4G and the poly(A) binding protein (PABP) [98–100] as is illustrated in figure (4.3). In the case of mRNAs that lack both a 5'-cap and

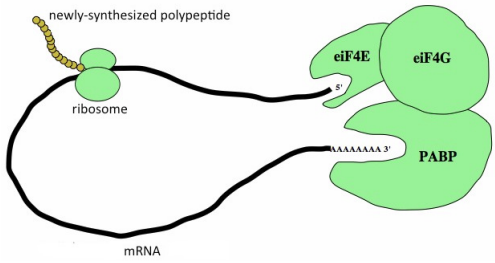

Figure 4.3.: Mechanisms of translation initiation. During eukaryotic cap-dependent translation, initiation factors mediated by the cap-binding protein, eIF4E, recognize an m7 GpppN-cap structure at the 5'-end of mRNA. Image adapted from http://en.wikipedia.org/wiki/Poly(A)-binding protein

3'-poly(A), complementary bases in the 5'- and 3'- UTR regions lead to RNA-RNA interaction [101, 102] allowing effective circularization of the RNA. This end-to-end communication has been observed in the modulation of translation in some RNA virus, for instance in the yellow fever virus [103], influenza A virus [104], dengue virus [105] and Tomato Bushy Stunt virus [106].

Peter Clote and co-workers [107] extend the work of Yoffe *et al.* from a rigorous mathematical framework. By defining the 5'-3' end distance as the shortest path

length from 5' node to 3' node length they conclude that the asymptotic expected 5'-3' distance of length n homo-polymers is ≈6.47 nt for RNAs with base-pairing fraction of $f = 0.6$. Li Tai [108], in order to explain the proximity of the 5'-3' ends, proposed a probabilistic model called randomly self-paired polymer (RSPP). He deduced an end-to-end distribution which for an RNA sequence of 1000 nt and pairing fraction $f = 0.6$, the mean end-to-end distance obtained was of 14.4 nt. Hill *et al.* [109] through combinatorial analysis and bivariate generating functions computed an end-to-end distribution for any length of RNA. Their results showed that the 5'-3' end distance for random RNA sequences are distinctively lower than those computed by Yoffe *et al.*

In this chapter we present briefly the outline followed by Yoffe *et al.* in order to conclude that the 5'-3' ends of large RNAs which form secondary structures are necessary close. We also mention the essential results of Peter Clote *et al.*, Li Tai, and Hill *et al.* about this topic. These results are presented in order to compared our experimental measurements of the 5'-3' end distance of some RNAs with the above theories.

4.1. Defining the End-to-end distance by graph theory

The 5'-3' end distance or the end-to-end distance from an RNA secondary structure molecule, may be determined by counting the total number of hydrogen bonds and covalent links that are in the exterior loop as [97]

$$D = l_{ext} + d_{ext}, \tag{4.1}$$

where l_{ext} is the number of covalent links and d_{ext} is the number of hydrogen bonds. Imagine each nucleotide in the external loop as a vertex, while each covalent link or hydrogen bond as an edge. Then the external loop can be represented as a graph as is illustrated in figure (4.4) for the RNA fragment of 300 nt *phr1* previously showed in figure (4.2). In that graph representation, the external loop is a tree graph. Therefore, it has $n - 1$ edges, where n is the number of vertices [110]. Thus the

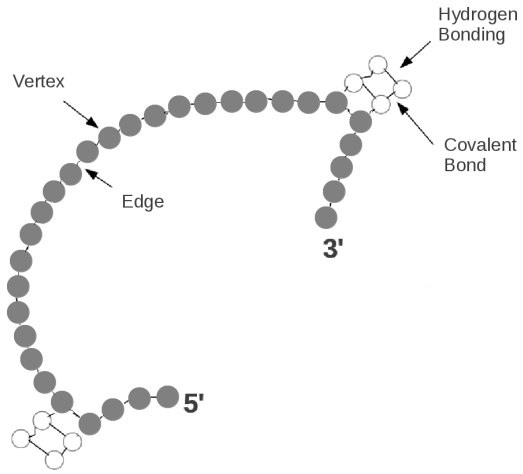

Figure 4.4.: External loop of a RNA secondary structure represented as a graph. The white nucleotides do not form part of the external loop.

end-to-end distance D can be expressed as

$$D = n_{ext} - 1 = (s_{ext} + 2d_{ext}) - 1, \tag{4.2}$$

where s_{ext} is the number of unpaired bases in the loop, while $2d_{ext}$ is the number of nucleotides linked by hydrogen bonds. The factor 2 arise because the hydrogen bonds are formed by two nucleotides. Eq. (4.2) gives an effective end-to-end distance D for a particular RNA sequence.

Using the above results and by observing that the free energy of a circular RNA obtained by attaching its 5'-3' ends (e.g. an internal loop in the secondary structure) is similar to the free energy associated with a linear RNA, (e.g. the external loop or the linear RNA obtained by cutting an internal loop in one of its $l+s$ covalent links), Yoffe *et al.* deduced a general result to determining the end-to-end distance for any RNA sequence. They proposed that the end-to-end distance of an RNA sequence of N nucleotides which form a secondary structure, can be determined as the average of all end-to-end distances D_α that arise from cutting all $L_\alpha(s,d)$-internal loops in the secondary structure in any of its $l+s$ covalent links as

$$D_\alpha = \frac{\sum_{s,d}(s+2d-1)(s+d)L_\alpha(s,d)}{(s+d)L_\alpha(s,d)}, \tag{4.3}$$

where $L_\alpha(s,d)$ is the α-th internal loop with s unpaired bases and d duplex, $(s+d)L_\alpha(s,d)$ is the statistical weight, $< s >_\alpha = \sum_{s,d} sP_\alpha(s,d) = \sum_s sP_\alpha(s,d)$, where $P_\alpha(s,d) = L_\alpha(s,d)/L_\alpha$ denoted the fraction of s,d-loops in the RNA structure. Computing the $P_\alpha(s,d)$ distribution by maximizing the functional entropy $\sum_i P_i \ln P_i$ through Lagrange multipliers the average end-to-end distance can be expressed as

$$D = \frac{< s >_\alpha^2 + 3 < sd >_\alpha + 2 < d >_\alpha^2 - < s >_\alpha - < d >_\alpha}{< s >_\alpha + < d >_\alpha} \tag{4.4}$$

where

$$\begin{aligned} < d^2 >_\alpha &= 2 < d >_\alpha^2 - < d >_\alpha, \\ < s^2 >_\alpha &= 2 < s >_\alpha^2 - < s >_\alpha . \end{aligned} \tag{4.5}$$

Here $< d >$ denote the mean number of hydrogen bonds that are in each loop, including the external loop of the RNA secondary structure. $< s >$ denote the mean number of covalent links in each loop.

4.2. RNA secondary structure as a Tree Graph

The average $< d >$ value of equation (4.5) can be determined by representing each loop, including the external loop, of the RNA secondary structure as a vertex and each duplex length as an edge, such as is illustrated in figure (4.5) for the mRNA *phr1* fragment of 300 nt showed in figure (4.2). Since this graph representation contains no cycles, the total number of vertices or loops that has the RNA secondary structure is $L = S + 1$, where S is the number of edges that in average has a length $k \approx 4$ base-pairs (bp).

It is know in graph theory that the grade of a vertex v, denoted by $deg(v)$, is the number of edges that incident with v while the sum of the grades of the vertices is equal to two times the number of edges [110]

$$S = \sum_{v \in V} deg(v) = 2|E|. \tag{4.6}$$

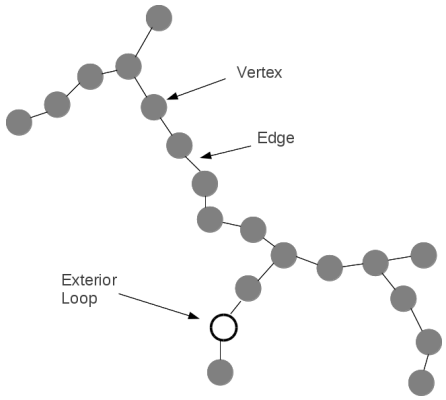

Figure 4.5.: Schematic representation of RNA secondary structure showed in figure (4.2), in a graph. Each internal loop as well as the external loop represent a vertex while each duplex length or doubled-stranded RNA represent an edge.

Therefore the total sum of the edges (or duplex fragments of RNA) is $2|E| = 2S = 2(L-1) = 2L-2$.

For a graph G is defined that the average of edges incident with a vertex v is given by

$$avgdeg(G) = \frac{\sum_{v \in V} deg(v)}{|V(G)|} \tag{4.7}$$

therefore, from equations (4.6) and (4.7), it follows that each internal loop or circular RNA in the RNA secondary structure is connected in average with duplex RNA

$$\begin{aligned} < d > &= \frac{\sum_{v \in V} deg(v)}{|V(G)|} = \frac{2|E|}{|V(G)|} \\ &= 2 - 2/L \end{aligned} \tag{4.8}$$

For large RNAs (N> 1000 nt), the number of loops is $L >> 2$, this imply that $< d >= 2$. This result indicates that on average each loop in the secondary structure of RNA (including the external loop) contains two hydrogen bonds.

Replacing $< d >= 2$ in the equation (4.5) it follows that $< d^2 >_\alpha = 6$. With these values the end-to-end distance for any RNA which form secondary structure is given now as

$$D = \frac{2\langle s \rangle^2 + 4\langle s \rangle + 10}{\langle s \rangle + 2}. \tag{4.9}$$

The percentage of base-pairs f in a RNA secondary structure is $f \sim 60\%$, then the number of unpaired nucleotides in the secondary structure is $< s >= N(1-f)/L \sim N(1-f)/S = 2k(1-f)/f$. For $f \sim 0.6$ and $k \sim 4$, $< s >= 5.33$. Substituting this value in equation (4.9), the end-to-end distance for any RNA length ($N \geq$1000 nt) which form secondary structure, is $D = 12$ nucleotides links. Taking the distance between nucleotide-nucleotide as the distance between phosphate-end-phosphate (0.59 nm [111]), the physical end-to-end distance would be $D_R \sim 12 \times 0.59 \sim 7.08$ nm.

4.3. End-to-end distance deduced by statistical mechanics

In order to prove the above results obtained for the end-to-end distance of any RNA molecule, Yoffe *et al.* used the mFOLD and RNAsubopt algorithms [95, 96]. These algorithms employ detailed energetic models to predict the most likely minimum RNA secondary structure. However, for a given RNA sequence with alphabet $\{G, C, A, U\}$ there are many suboptimal secondary structures with almost similar free energy than the minimum free energy structure, the above algorithms also provide an ensemble X_α of all sub-optimally folded RNA molecules within a desired energy range above the ground state. To estimated the end-to-end distance for an RNA sequence, Yoffe *et al.* used 1000 suboptimal RNA structures for each RNA sequence used in his study. This sample space was computed with RNAsubopt and the average end-to-end distance was determined with the mFOLD program as

$$D = \frac{\sum_{\alpha=1}^{1000} D_\alpha \exp(-\Delta \mathrm{G}_\alpha / \mathrm{k_B T})}{\sum_{\alpha=1}^{1000} \exp(-\Delta \mathrm{G}_\alpha / \mathrm{k_B T})} \tag{4.10}$$

where D_α and ΔG_α is the end-to-end distance and the free energy of the α-th structure relative to the minimum free energy for the RNA sequence, k_B is the Boltzmann's constant and T is the absolute temperature. The mean end-to-end ensemble averaged distance obtained are plotted in figure (4.6) for different RNA sequences. Surprisingly the mean end-to-end distance for randomly-permuted RNA sequence (500-8000 nt) of viral-like composition with equal proportions of A, U, G, and C,

was $D \sim 12$, value very close to the theoretically predicted by equation (4.9). For RNA viral sequence the values of the mean end-to-end distance and standard deviations are higher, but overlap with $D = 12$ for all taxa except for the Tymoviruses.

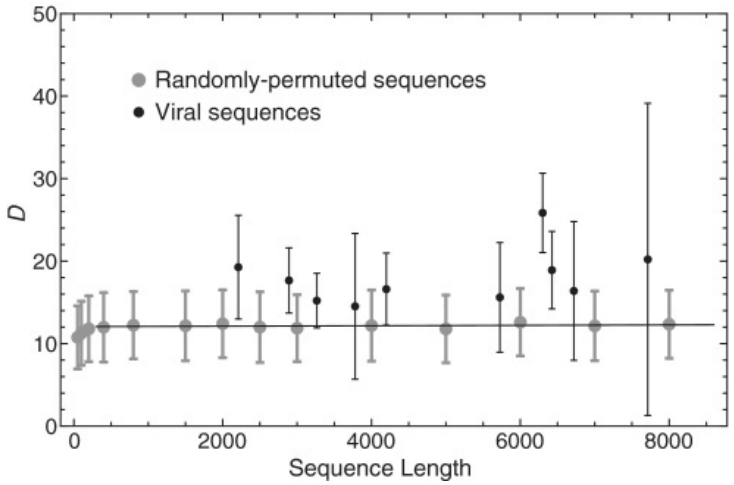

Figure 4.6.: Predicted end to end distance from a statistical mechanics approach for random RNA sequence with equal base composition and for RNA viral sequences. Figure taken from [97]

4.4. Expected distance between the ends of RNA secondary structures

Peter Clote and co-workers extend the work of Yoffe *et al.* from a rigorous mathematical framework [107]. By defining the 5'-3' end distance as the shortest path length from the 5' node to 3' node, they solved the recurrence relations of the end-to-end distance problem by the next mathematical sequence

$$\langle d_{1,n} \rangle = \sum_{S} \frac{\exp(-\mathrm{E(S)/RT})}{Z_{1,n}} d_S(1,n) \tag{4.11}$$

where n indicated the size of the RNA in number of nucleotides, $E(S)$ is the free energy of the suboptimal RNA secondary structure with sequence S, R is the constant

of gases, and $Z_{1,n}$ is the partition function defined as

$$Z_{1,n} = \sum_{S} \exp(-\mathrm{E(S)/RT}) \tag{4.12}$$

Note that the above problem is based in statistical mechanics approach. By solving the recurrence relations of $d_S(1,n)$ Pete Clote *et al.* conclude that the asymptotic expected 5'-3' distance of length n homo-polymer is $\approx \langle d_n \rangle \sim 6.47$ when the percentage of paired number of nucleotides in the secondary structure is $f \sim 0.6$. This value is much more lower than the one obtained by Yoffe *et al.*

4.5. The end-to-end distance of RNA as a randomly self-paired polymer

Li Tai in order to explain the proximity of the 5'-3' ends, proposed a probabilistic model called randomly self-paired polymer (RSPP) [108]. Whit this model he conclude that the probability, $\rho(X)$, of the end-to-end distance being exactly X unpaired monomers is

$$P(N_{T,eff}, N_{p,eff}, X) \;=\; 1 - \prod_{B=1}^{N_{p,eff}} (1-p)^{\Omega} \tag{4.13}$$

where P, is the probability that the end-to-end distance of a RSPP is less than or equal to X unpaired monomers, $N_{T,eff}$ is the effective number of monomers, $N_{p,eff}$ is the effective number of paired nucleotides and Ω is the number of possible structures that can be formed by a RNA sequence with fixed $f = 0.6$. Applying this theory to an RNA sequence of 1000 nt with paired fraction $f = 0.6$, the probability distribution of the end-to-end distance appears roughly Gaussian (Fig. 4.7). The mean end-to-end distance, X, from this distribution provides a value of 14.4 unpaired bases of separation, with the width of the distribution being approximately 5. Note that the above approach is based on statistical methods and do not involve thermodynamic parameters directly.

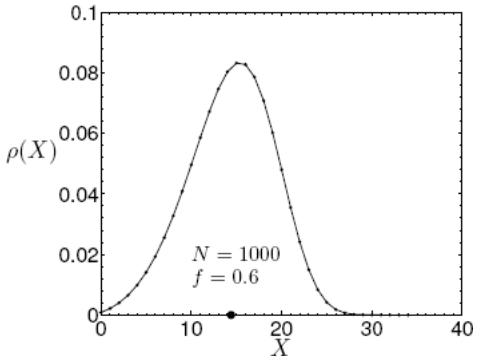

Figure 4.7.: End-to-end probability distributions for RNA sequence of $N = 1000$ nt of length and pairing fraction $f = 0.6$. Figure taken from reference [108]

4.6. The 5'-3' distance of RNA secondary structures

Hill S. W. and coworkers [109] define the end-to-end distance $d(\sigma_n)$ of RNA secondary structures as the minimum length of a path consisting of σ-arcs and backbone-edges from vertex 1 (the 5'-end) to vertex n (the 3'-end). From this definition and by using complicated bivariate generating function of the number of RNA secondary structures of length n with distance d, they computed end-to-end distributions from which we extracted an end-to-end distance for a random RNA sequence of 10 and 100 nt as ~ 2 nt and 3 nt respectively while for a random RNA sequence above 1000 nt of ~ 7 nt.

Conclusion

Aron Yoffe *et al.* predicted a common characteristic between different species of RNAs. This is that its terminals, also know as 5'-3' ends, are necessary close, between 10-20 nucleotide regardless of its composition sequence. If we assume that the distance between ssRNA nucleotides is 0.59 nm, the theory of Aron Yoffe tell us that the end-to-end distance corresponds to a physical distance of a few nm.
These results have originated a series of complementary theoretical works almost all of them based in advanced graph theory as well as thermodynamics approach [107–109]. The above predictions are very interesting, but an experimental confirmation was missing, that is the purpose of the present thesis.

5

Experimental Measurement of the 5'-3' End Distance of Eleven Large RNAs by smFRET

Although there are theoretical calculations, and even some biological requirements for the proximity of the two ends of RNA molecules, no experimental data on the RNA ends distance has been obtained to date. In this chapter we present the first systematical experimental determination of the 5'-3' end separation for RNA molecules of different lengths and sources. We used single-molecule fluorescence resonance energy transfer (smFRET) to measure the physical 5'-3' end distance of eleven RNAs of different length and base composition. The RNA molecules used were labeled at the 5'-end with the alexa fluor 647 and its 3'-end with alexa fluor 546. Mean FRET efficiencies were extracted from photon RNA-trajectories and compared with the FRET calibration curve presented on chapter 3, in order to determine the end-to-end distance. The results were compared with the theoretical predictions. The RNAs used in this work come from both a fungus (*Thrichoderma atroviride*) and two virus (Cowpea Chlorotic Mottle Virus (CCMV) and Brome mosaic virus (BMV)).

5.1. Total RNA extraction from *T. atroviride*

We select four DNA published sequences that encodes for different lengths of mRNAs of *T. atroviride* fungus. Two sequences are in the Genome Portal of the Department of Energy Joint Genome Institute [113], with code Protein ID 258498 and ID 146288. These DNA sequences encode for a 574 nucleotides and an 1012 nucleotides RNA lengths, respectively. For these genes we assigned here a short nomenclature as *fgen1* and *triat1*, respectively. Another two gene sequences were chosen from the GenBank

data base [114] with numbers DQ068751.1 and AJ009960.1, which correspond to the mRNA *chi18-4* (chitinase 18-4) of 1667 nucleotides of length [115] and mRNA *phr1* of 2012 nucleotides of length [116].

We design forward and reverse primers for each DNA sequences aforementioned on the 5'-3' untranslated regions (UTR's). DNA primer sequences are listed in table (5.1). They were provided by Integrated DNA technologies. The sequence of the forward and the reverse primers for *fgen1* containing Kpn1 site (GGTACC) on the 5'-end. The above design was realized for get sense (+, upstream transcription) and anti-sense (-, downstream transcription) RNA sequence of *fgen1* in order to determined weather the 5'-3' end distance of the sense and antisense of an RNA was similar. DNA forward and reverse primers for *chi18-4*, *phr1* and *triat1* were designed with Apa1 (GGGCCC) and BamHI (GGATCC) sites respectively on the 5'-end.

To obtain the above fungal mRNAs, 39 g of Potato Destroxe Agar (PDA, Potato

f-fgen1	5'-AAAGGTACCTCTCTCGCTATCTCGGAA-3'
r-fgen1	5'-AAAGGTACCGGCATAGCGAACATGTTA-3'
f-Triat1	5'-TATGGGCCCTTCTTTCTTTCTTTCTCT-3'
r-Triat1	5'-ATAGGATCCTAGTGCTGCAATGCTTTA-3'
f-chi18-4	5'-TATGGGCCCTACGCCAGACAATAGATTC-3'
r-chi18-4	5'-TAAGGATCCGCAGCCTGTATTCATTGCAC-3'
f-phr1	5'-AAAGGGCCCGCTATATCGTCAAGATTATC-3'
r-phr1	5'-TATGGATCCTTTTCAAGTACGAATTGACC-3'

Figure 5.1.: Forward (f) and reverse (r) primers used to get four DNA genes from *T. atroviride* fungus by Polymerase Chain Reaction.

Starch from infusion 4 g, Dextrosa 20 g, Agar 15 g, from Difco) was mixed thoroughly in 1 liter of nanopure water. PDA is a relatively rich medium for growing a wide range of fungi such as *T. atroviride* and *Rhizoctonia solani*. PDA medium was heated with frequent agitation and boiled for 1 minute to completely dissolve the powder. The medium was autoclave at 121 ^{o}C for 15 minutes and cooled at room temperature. Then it was deposited in Petri-dishes plates (9 cm diameter) each containing 20 ml of these media. When the deposited PDA medium in petri-dishes was solidified, a sterile cellophane disks of 8.9 cm of diameter was placed on the agar surface. The PDA plates were stored at 4^oC until use.

Two PDA-plates previously prepared were inoculated directly with a 5 mm diameter core of *T. atroviride* (IMI 206040 strain) mycelium placed at the center of the plate, sealed with parafilm and incubated in dark thermostat at 26^{o}C during 36 hours.

Others plates, with solid PDA, were separately inoculated with *T. atroviride* strain P1 (ATCC 74058) and *Rhizoctonia solani* fungus respectively and incubated at 26^{o}C in total darkness for 24 hours. Thereafter, in two new solid-PDA plates, *T. atroviride* strain P1 was set in confrontation with *Rhizoctonia solani* fungus. A plate confrontation experiment with fungus is commonly used to study the mechanism by which some species of fungi antagonize and parasite other fungi, for instance *T. atroviride* vs *R. solani*, during the pre-contact period through chitinases, glucanases, and proteases expression [123]. For plates confrontation assays, strips of 3 mm were cut from the growing front of *T. atroviride* strain P1 and *R. solani* wild-type. After that they were placed on PDA plates covered with cellophane at a distance of 4 cm from each other and the confrontation was incubated at 27^{o}C in total darkness for 28 hours. Figure (5.2) shows the growing of the above strains as well as the confrontation assay between *T. atroviride* P1 againts *R. solani* on the solidified PDA medium. The mycelium was collected from surface of the cellophane from each plate. In the

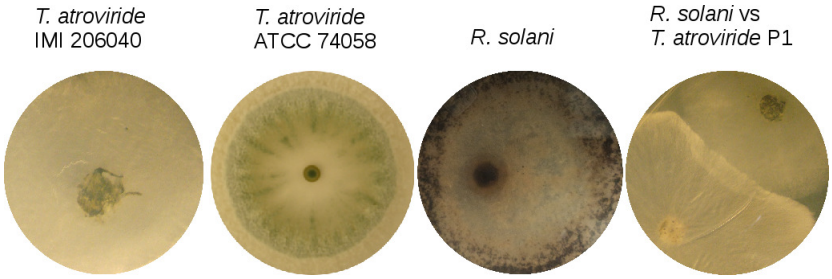

Figure 5.2.: Mycelia growing on solidified PDA medium of different strains of the fungi used in this work

case of confrontation assay, the mycelium were harvested when *T. atroviride* and *R. solani* was just touching.

Mycelia harvest from each plate was immediately deposited at mortar, previously frozen in liquid nitrogen, and grinned with a frozen pestle. The powder was deposited immediately into RNase free eppendorf tubes and incubated in liquid nitrogen. Total RNA was extracted from collected mycelia using Trizol reagent (Invitrogen life technologies) following the manufacturer's protocol. Figure (5.3) shows separated phases by Trizol reagent of the harvest mycelia obtained in confrontation assays. Lower red phase contains residues of phenol-chloroform. Inter-phase contains pro-

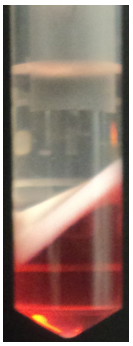

Figure 5.3.: Total RNA extraction from *T. atroviride vs R. solani* fungi through trizol reagent from Invitrogen. Upper phase contains the total RNA

teins, while the upper aqueous phase containing the total RNA of the mycelia. Total RNA was extracted from upper phase and placed into a new eppendorf tube, previously frozen in liquid nitrogen. Contaminating genomic DNA was removed by DNase treatment using TURBO DNase-free (Ambion). After that, TURBO-DNase was inactivated by adding 2 μl of 100 mM EDTA and heating the mixture at 65^{o}C for 5 minutes. Finally total RNA was precipitated by ethanol and re suspend in nuclease free water. The integrity of total RNA extracted was visualized in a denaturing formaldehyde/MOPS/1% agarose gel. Briefly, 1g of ultrapure agarose (Invitrogen Life Technologies) was deposited in 60 μl of DEPC treated water and heated in microwave until agarose were completely dissolved. We added 6 mL MOPS 10X pH=7.1 and 700 μl deionized formaldehyde. The mixture was poured in a pre-cleaned electrophoresis chamber with 10X benzene and used until solidified. One micro liter of total RNA extraction was mixed with RNA loading buffer (50% glycerol, 1 mM EDTA, 0.4% bromophenol blue, 0.5 mg/mL ethidium bromide) and heated at 65^{o}C for 3 minutes. The RNA reaction was deposited in the agarose gel running for 30 minutes at 6 V/cm. Agarose gel, was visualized in a transilluminator (Gel Doc XR, Bio Rad). Figure (5.4) shows the intact total RNA samples from each extraction. The bands exhibiting the 28S and 18S ribosomal RNA intensities. The 28S and 18S nomenclature refers to the large subunit and the small subunit of ribosomal RNA respectively in Sedverg units. All materials used in above process were free of proteases, DNases and RNases.

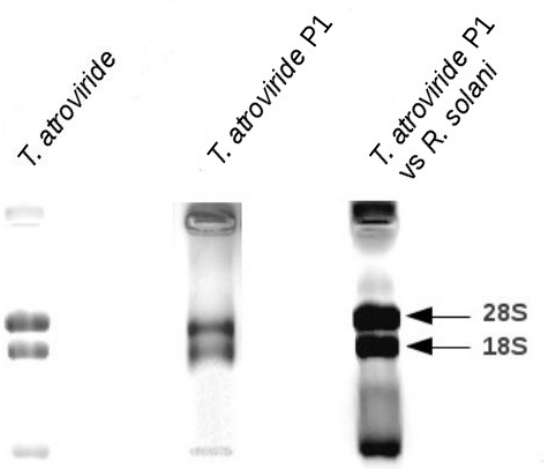

Figure 5.4.: Electrophoresis gel of total RNA from fungi. Total RNA was extracted as described in methods and subjected to electrophoresis through a native 1% MOPS agarose gel. The arrows indicate the 28S and 18S units of rRNA

5.2. Complementary DNA synthesis

Complementary DNA (cDNA) is a DNA copy synthesized from mRNA through a reverse transcriptase enzyme. In a mixture of mRNA molecules, as is the case of total RNA extraction from fungi, each mRNA can be a template and will produce a cDNA in the form of a single stranded molecule. In many biological process cDNA is a more convenient way to work rather than mRNA because RNA is very easily degraded by omnipresent RNases.

Complementary DNAs from each total RNA extraction from fungi were synthesized performing first a reaction of 12 μl containing 1μl oligo(dT)18 (500 μg/ml) primer, 5 μg of total RNA and 10 mM deoxynucleotides (dNTPs) mix. The mixture was heated at 65^{o}C for 5 min and quickly chilled on ice. We added 4 μl 5X First-Strand buffer (250 mM Tris-HCl, pH 8.3 at room temperature; 375 mM KCl; 15 mM $MgCl_2$) and 2 μl 0.1 M DTT. The mixture was incubated at 42^{o}C for 2 min. Thereafter 1 μl (200 U) SuperScript II Reverse Transcriptase (Invitrogen Life Technologies) was added to the reaction and nuclease free water up to a final volume of 20 μl. The mixture was incubated at 42^{o}C for 50 min followed by quantification with a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, USA) and stored at

-20^{o}C until use. Complementary DNA derived from total RNA extraction from *T. atriviride* wild-type contains the sense *fgen1* and *triat1* sequence, while cDNA derived from total RNA extraction from *T. atroviride* strain P1 contains the sense *phr1* gene sequence. The cDNA derived by confrontation assay contains the *chi18-4* gene sequence.

5.3. Synthesis of DNA by PCR technique

Complementary DNAs derived from total RNA extraction were used as template for polymerase chain reaction (PCR) amplification of the corresponding DNA coding sequences. PCR technique allows to produce millions of copies of a specific DNA sequence in approximately two hours from the starting material. First the template of DNA which contains the target sequence is heated to higher temperature. For the case of cDNA (ssDNA) the heating allows to disrupt the tertiary or secondary structure. In the case of dsDNA, the heating allow to separate the strands from each other. After that, the temperature is lowered to allow the DNA primer to match the target sequence. At this stage in the presence of the appropriate buffer and dNTPs a polymerase enzyme synthesize new DNA strands complementary to the target sequence. The first and most commonly used of these enzymes is the Taq DNA polymerase. The process is repeated again and again as is illustrated in figure (5.5).

PCR for each primer was carried out in a total volume of 25 μl containing 25 mM $MgCl_2$, 10 mM dNTPs mix, 0.5 μM of each primer, 2.5 μl 10X goTaq Polymerase Buffer, 0.25 μg cDNA, and 0.5 U GoTaq DNA polymerase (Promega). The amplification program for each PCR began with an initial temperature of 94 ^{o}C for 5 min, 35 cycles with 95 ^{o}C for 0.30 s, extension period at 72 ^{o}C and a final extension period of 10 min at 72 ^{o}C. Different conditions of annealing temperature and extension times were carried out for each primer. In the case for DNA-primers of *fgen1*, annealing temperature was at 53 ^{o}C for 0.30 s and an extension time for 1 min. The annealing temperature for primers of *triat1* was performed at 45 ^{o}C for 0.35 s and an extension time of 2 min. Annealing temperature for *phr1* and *chi18-4* were carried out at 55 ^{o}C for 0.30 s with extension time of 3 min.

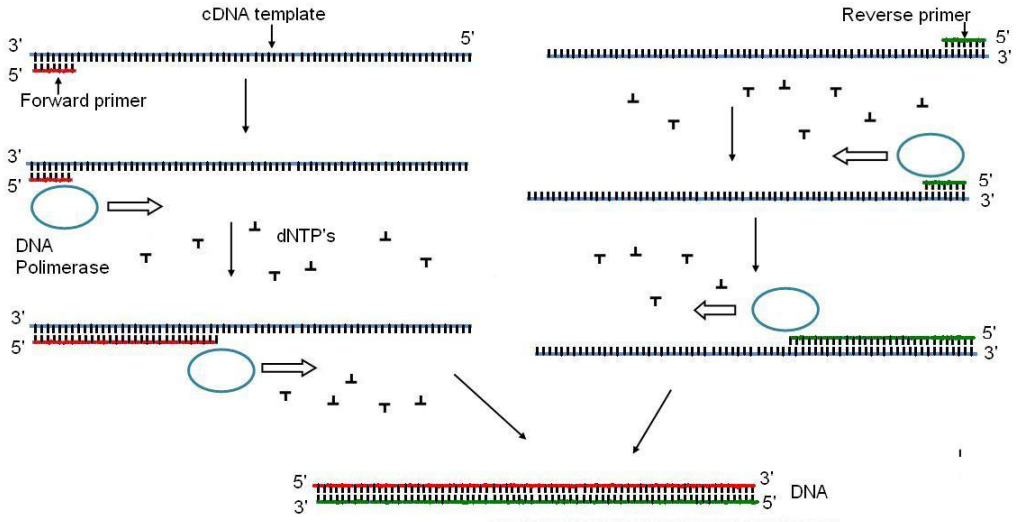

Figure 5.5.: The different steps in PCR

Analysis of the PCR products were examined by loading 4 μl of each PCR reaction onto 1% agarose gel. Electrophoresis of PCR products are shown in figure (5.6). From left to right, lane 1 is the molecular weight ladder (DNA ladder of 1 kbp (Promega)), lane 2-5 display the corresponding *fgen1*, *triat1*, *chi18-4* and *phr1* molecular weight respectively. The rest of PCR product of each DNA gen was purified by centrifugation using the Wizard DNA Clean-Up System following the manufacturer procedure. Quantification of each DNA was obtained with its 260 nm absorption in a nanodrop system.

5.4. Cloning DNA genes in the pBSK II(+) vector

To synthesize RNA by in-vitro transcription, it is necessary to have large quantities of the individual gene in pure form with a promoter sequence [117]. In our case, the genes previously obtained by PCR technique do not contain any RNA sequence promoter and moreover, the quantity of each of them is yet low for in-vitro transcription. To solve these problems we used the DNA cloning technique. There, the DNA fragment of interest is linked into a DNA vector containing one promoter sequenceThe inserted DNA is replicated along with the vector generating a large number of identical DNA molecules. DNA plasmids are the most common vectors used in the DNA

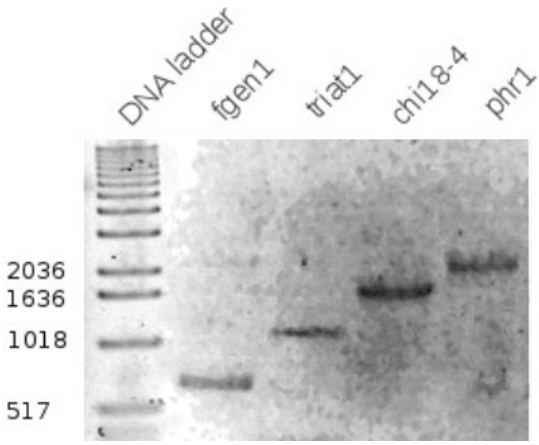

Figure 5.6.: Agarose gel electrophoresis for DNA fragments of *fgen1*, *triat1*, *chi18-4* and *phr1* obtained by PCR. In lanes 1 an 1 kb DNA ladder (Promega) was loaded to compared the lengths of genes aforementioned. The expected length of each gen gene corresponds to the reported one.

cloning technique and they are double-stranded circular DNA molecules that are separated from a cell's chromosomal DNA. To insert the foreign DNA fragment into the plasmid, the plasmid and the foreign DNA fragment are treated first with appropriated restriction enzymes. Thereafter the DNA fragment is ligated into the plasmid using a DNA ligase enzyme. The construction is introduced in a bacterial cell in a process called transformation. Later the vector can easily be separated of the host cell, purified and precipitated as a pellet upon centrifugation. The basic scheme of the process is summarized in figure (5.7).

For DNA cloning we used the plasmid vector pBluescript II(+) (pBSK II+). This vector contains multiple cloning sites sequence (MCS), antibiotic resistance sequence to ampicillin, T7 and T3 promoter, and an *E. coli* and f1 helper phage origin of replication. The multiple cloning site sequence is located within a *LacZ* controlled gene in a way that it does not interrupt the reading frame or the functionality of the resultant *lacZ* gene protein product. This protein product is a galactosidase. In recombinant vectors which have an foreign insert DNA molecule cloned into one of the restriction enzyme sites in the polylinker, this insert DNA results in an altered *lacZ* gene and a non-functional galactosidase. The presence or absence of this protein can easily be determined through the use of a simple chromogenic assay using

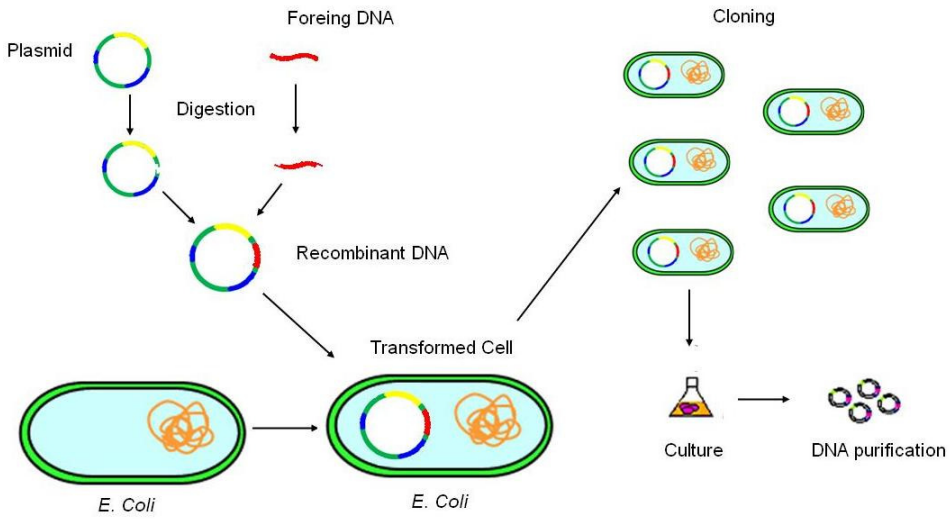

Figure 5.7.: Illustration of the procedure to form recombinant DNA. First the DNA and vector are treated with appropriated restriction enzymes for DNA ligation. After the recombinant DNA is introduced in the host cell and replicated. Finally the recombinant DNA is extracted from the host cell.

Isopropyl β-D-1-thiogalactopyranoside (IPTG) and (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) X-Gal. IPTG is the *lacZ* gene inducer and is necessary for the production of the galactosidase. X-Gal is a colorless, modified galactose sugar. When this molecule is metabolized by the galactosidase, the resultant products are a bright blue color. When IPTG and X-Gal are included in a plasmid DNA transformation, blue colonies represent bacteria harboring non-recombinant vector DNA since the *lacZ* gene region is intact. IPTG induces production of the functional galactosidase which cleaves X-Gal and results in a blue colored metabolite. It follows that colorless colonies contain recombinant DNA since a nonfunctional galactosidase is induced by IPTG which is unable to cleave the X-Gal.

Circular maps and lists of features for the pBluescript II phagemids are shown in figure (5.8).

For DNA cloning, pBSK SK II(+) plasmid and DNA sequences *triat1*, *chi18-4* and *phr1* were digested with restriction endonucleases ApaI and BamHI in an reaction of 40 μl that contained 4 μg of each DNA, and 4U of each enzyme. The reaction was incubated at 37^o all night. Similar procedure was realized with *fgen1*, but pBSK

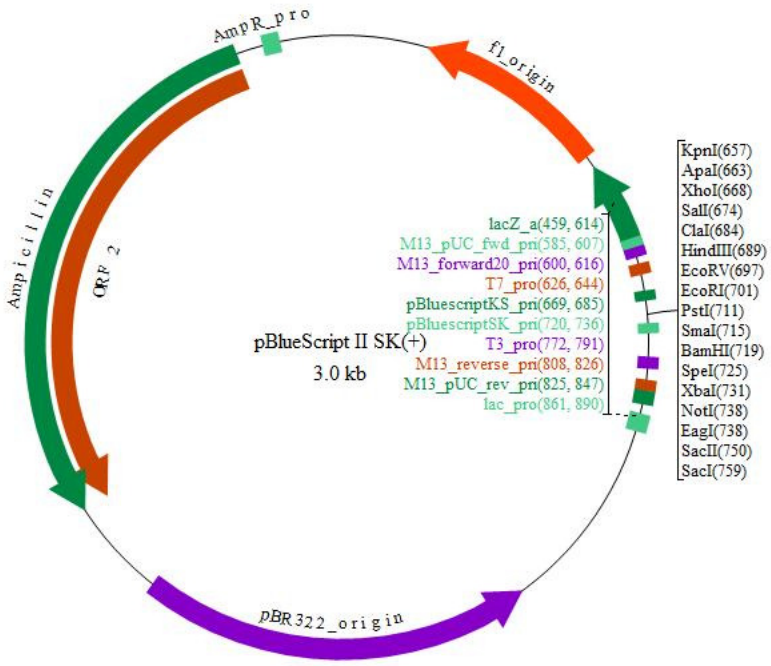

Figure 5.8.: The pBluescript II SK (+) phage-mid vector. The arrow in red indicate the multiple cloning site of the plasmid. T7 promoter transcription initiation site is at 643.

II(+) and *fgen1* where treated with KpnI enzyme. All DNA digestions were purified with Wizard SV Clean-up system (Promega).

DNA ligation was realized by following the next reaction: 5 μl 2X Rapid Ligation Buffer (60 mM Tris-HCl pH 7.8, 20 mM $MgCl_2$, 20 mM DTT, 2 mM ATP and 10% polyethylene glycol) were mixed separately with 30 ng of pBSK II(+), 90 ng of digest-gen PCR product and 1U of T4 DNA ligase (Promega) in a final volume of 10 μl. Thereafter, the reaction was incubated at room temperature (25^{o}C) for 3 hours. The constructs were called pfgen1, pTriat1, pChi18 and pPhr1 respectively where p stand as plasmid.

5.5. Competent cells

DNA transformation in cell bacteria is performed using the "so called" competent cells. These bacterial cells possess more easily altered walls by which foreign DNA can

be passed through easily. The standard method for making the bacteria permeable to DNA involves treatment with calcium ions. This helps to open the pores in the cell membrane so plasmid DNA from the surrounding environment can enter the cell.

To prepare competent cells of it E. coli strain TOP10F for plasmid DNA transformation, 1 L of liquid Lysogeny broth (LB) medium pH 7.0, (Difco) was prepared. Briefly, 10 g Bacto-tryptone, 5 g Bacto-yeast extract and 5 g NaCl were diluted in 1 L of dH_2O, autoclaves for 15 min and stored at 37^oC until use.
A single colony from a freshly grown plate TOP10F *E. Coli* cells were growing in 15 ml of Lysogeny broth (LB) medium, previously prepared, at 37^oC overnight with shake at 150 revolution per minute (rpm). 1 ml of previous growing were inoculated in 100 ml of LB medium and shaken vigorously at 37^oC until a cell optical density $OD_600 \sim 0.3$ at 600 nm was reached. The culture was chilled on ice for 15 min along with 0.1 M $CaCl_2$ solution and 0.1 M $CaCl_2$ plus 15% glycerol. The cells were centrifuged for 10 min at 3300 g at 4^oC. The medium was discard and cells were re-suspend in 40 ml of cold 0.1 M $CaCl_2$. Thereafter, the cells were kept on ice for 30 min followed by centrifugation just as above and the supernatant was removed for re-suspend the cells in 6 ml of 0.1 M $CaCl_2$ solution plus 15% glycerol. Finally, 0.2 ml of the cell suspension were deposited into sterile 1.5 ml micro-centrifuge tubes and incubated at -80^oC until use.

5.6. DNA transformation

LB agar plates were used for DNA transformation. Briefly 2.5g Bacto-tryptone, 1.25g Bacto-yeast extract, 2.5g NaCl and 3.75g agar were mixed well in 250 ml of dH_2O and autoclaved for 15 min. 0.5 mM of IPTG and 80 μg/ml of X-Gal were added to LB agar medium. 30 ml of LB agar medium then were poured in petri dishes and incubated at room temperature in sterile conditions, until agar solidified. After which, the petri dishes were stored at 4^oC until use.
For DNA transformation into *E. coli*, the competent cells were taken from -80^oC and thawed on ice (approximately 20-30 min). Later, 2 ng of each plasmid construction was added to competent cells and the mixture was incubated for 20 min on ice fol-

lowed by an incubation at 42^{o}C for 1 min and quickly returned to ice for 5 min. We added 0.5 ml of liquid LB medium to the above reaction and we incubated it at 37^{o}C for 60 minutes at 150 rpm. After which the transformed cells were obtained by centrifugation at 6000 rpm for 5 minutes and spread onto a LB agar petri dishes with an sterile handle to assure that cells were widely spaced on the plates. The plates were incubated at 37^{o}C overnight in an inverted position to avoid condensation.
Figure (5.9) shows the plates previously incubated at 37^{o}C overnight for DNA transformation of pfgen1, pTriat1, pChi18 and pPhr1 plasmids. *E. Coli* colonies with the plasmid-DNA insert are in white color while *E. Coli* bacterias with blue color only contains vector plasmid or neither.

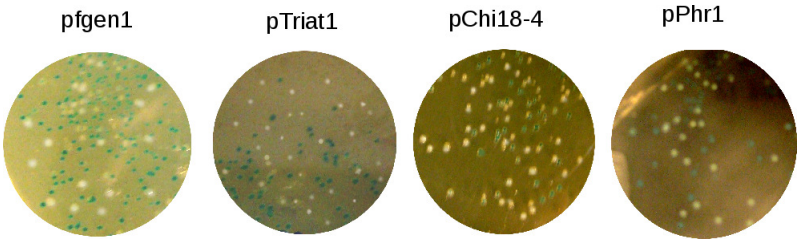

Figure 5.9.: Blue and white colonies of the DNA cloning probes

Several white colonies from each plate were taken and deposited in an eppendorf tube with 5 mL of liquid LB medium and 100 μg/ml of carbenycillin. Thereafter the tubes were incubated at 37^{o}C overnight with shaking at 150 rpm.
To extract and purify plasmid from *E. Coli* bacterias, eppendorf tubes with bacterias-plasmid were centrifuged for 3 min at 7000 rpm. The supernatant was removed and the cells were re-suspend in 200 μl of Birnboim and Doly solution I (Tris-HCl 25 mM pH 8.0, 10 mM EDTA and 100 mM of Glucose) followed by cells lysed in 400 μl of Birnboim and Doly solution II (0.2 M NaOH and 1% of SDS prepared at the time of the lysis). The plasmids were precipitated with 300 μl of Birnboim and Doly solution III (3M of potassium acetate pH 4.8 adjusted with acetic acid), incubated on ice for 10 min, and centrifuged for 10 minutes at 12000 rpm. The supernatant was recuperated and placed into new eppendorf tube with 5 μg/ml of RNaseH (Promega). The reaction was incubated at 37^{o}C for 1 hour. After which 0.8 mL of absolute ethanol and 0.08 mL of 3M sodium acetate, pH 5.2, were added and the reaction was

incubated at -20^{o}C for 3 hours. Finally the plasmid was recovered by centrifuging at 12000 rpm for 10 min in form of pellet. The pellet was washed with 70% ethanol and centrifuged at 12000 rpm for 5 min and re-suspend in 50 μl of nuclease free water.

5.7. DNA cloning probe

To obtain sense and anti-sense orientation of the gen *fgen1*, we cut several pfgen1 plasmids with Pst1 (NEB) enzyme. The sense plasmid (pfgen1+) contains Pst1 sites at positions 1142 bp and 1285 bp, therefore Pst1 digestion generates two fragments; one of 143 bp of length and another of 3431 bp of length. In contrast with pfgen1+, the anti-sense plasmid (pfgen1-) contains Pst1 sites at 718 bp and 1285 bp, then with Pst1 digestion the expected DNA fragments will be of 567 bp and 3007 bp. The figure (5.10) shows the expected base pair lengths for the two orientation of the construction pfgen1.

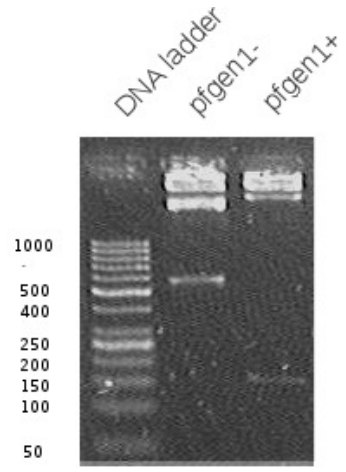

Figure 5.10.: The resulting sense and anti-sense bands after digestion in an 1.5% agarose gel running in TBE buffer. Digestion of pfgen1 with Pst1 (NEB) enzyme to verify both orientations. Given the different location of the Pst1 sequence, the sense orientation generates two fragments of 143 and 3431 bp, while the anti-sense orientation generates fragments of 567 and 3007 bp. DNA marker of 50 bp (Fermentas).

5.8. In-vitro Transcription

In-vitro transcription is a simple procedure that allows for template-directed synthesis of RNA molecules of any sequence. It is based on the engineering of a template that includes a bacteriophage promoter sequence (e.g. from the T7 coliphage) upstream of the sequence of interest followed by transcription using the corresponding RNA polymerase.

For in-vitro transcription, plasmids pfgen1+ and pfgen1- were linearized with restriction endonuclease XhoI, while pTriat1, pPhr1 and pChi18-4 with BamHI enzyme. Figures (5.11) and (5.12) shows all linearized plasmids including the plasmids which contain the DNA sequence for RNA1 of BMV (pT7B1), RNA2 of BMV (pT7B2), RNA3 of BMV (pT7B3Sn), RNA1-3 of BMV (pRDCT7B1B3), RNA2 of CCMV (pCC2), and RNA3 of CCMV (pCC3). These plasmids are reported in [112]. To linearized pT7B1, pT7B2 and pT7B3Sn we used BamHI (NEB) enzyme. Plasmid pRDCT7B1B3 was linearized with EcoRI (NEB). CCMV plasmids were linearized with XbaI (NEB).

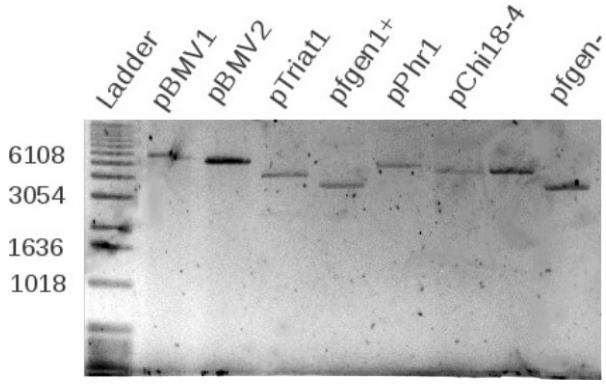

Figure 5.11.: Viral and fungi constructions linearized with appropriated endonuclease enzyme. The length of each digestion corresponding to the expected DNA length. As marker, DNA ladder 1 Kbp from Invitrogen.

Linear plasmids were purified using Wizard SV Clean-up system (Promega) or phenol-chloroform extraction. DNA was quantified using the UV-Vis Nanodrop System. Messenger RNAs were transcribed by assemble the next reaction at room temperature in the following order: 10 μl of nuclease-free water (Promega), 2 μl of

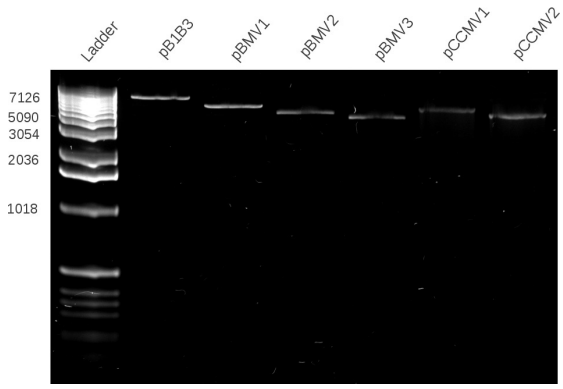

Figure 5.12.: Viral constructions linearized with appropriated endonuclease enzyme. The length of each digestion corresponding to the expected DNA length. As marker, DNA ladder 1 Kbp from Invitrogen. The agarose gel was stained with GelRed.

10X Reaction Buffer T7 (NEB), 2 μl of Ribonucleotide Triphosphates (rNTPs) mix (80 mM each), 1 μg of DNA template, 1U of Ribo-block (NEB), and 1U of T7 RNA polymerase (NEB). The reaction was incubated at 37^{o}C for 3.5 hours followed by 15 min with 1U of DNaseQ1 (promega) to remove the DNA template. Unincorporated rNTPs, were removed by phenol (pH 4–5):chloroform:isoamyl alcohol (125:24:1) (Invitogen) extraction followed by ethanol precipitation. The resulting RNAs were re-suspend in 40 μl of nuclease free water and quantified with Nanodrop system. To verify integrity and length of each RNA, 2 μl of previously reactions were loaded in denaturing formaldehyde/MOPS/1% agarose electrophoresis gel. Figures (5.13) shows the integrity and its corresponding length of each mRNA from *T. atroviride* as well as CCMV and BMV virus. No one of above mRNAs synthesized contained 5'-cap and 3'-Poly(A) tail.

The viral mRNAs obtained from the CCMV and BMV viruses have the next length; for CCMV, RNA2 2,774 nt and RNA3, 2,177 nt, for BMV, RNA1, 3234 nt, RNA2, 2865 nt and RNA3, 2117 nt, respectively.

5.9. RNA labeling

For RNA end-labeling the following protocol was applied: Purified mRNAs were fluorescent labeled on the 3' end OH using an ATP-dependent addition of a 3',5'(bis)phosphate

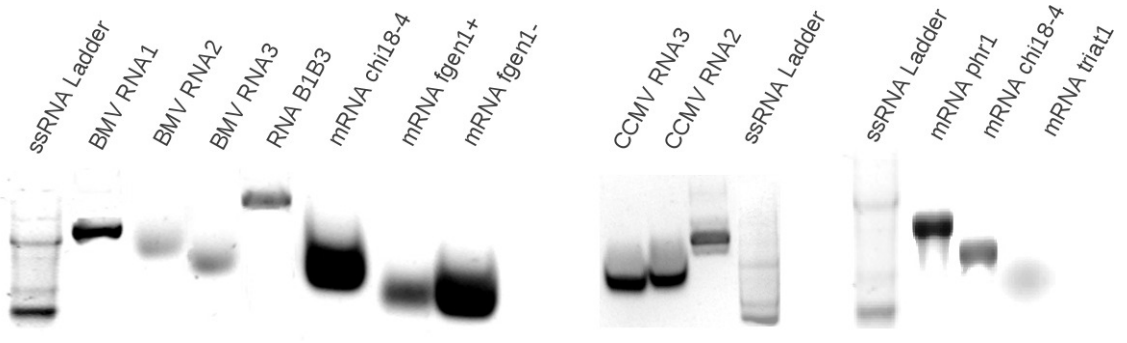

Figure 5.13.: RNAs were loaded on a 1% agarose gel in formaldehyde/MOPS buffer. RNA samples were heated at 65 ^{o}C for 10 min in denaturing agarose gel loading buffer before electrophoresis at 5 V/cm. The agarose gel was stained with GelRed. The marker was the single stranded RNA ladder from NEB

RNA-nucleotide as follows: in a eppendorf tube free of RNases were added 3 μl of nuclease free water, 3μl of 10X RNA ligase Reaction Buffer (0.5 M Tris-HCl, 0.1 M $MgCl_2$, 0.1M DTT, 10 mM ATP pH 7.8), 1.34U/μl of RNase Inhibitor (NEB), 1 nmol of mRNA, 1.5 nmol of custom made r-Adenosine-3',5'-bis-phosphate-8-[(6-Amino)hexyl]-amino-Alexa Fluor-546 (Jena Bioscience, Thuringia, Germany) diluted in DMSO, 1.34U/μl of T4 RNA ligase (NEB) and 15% of PEG 5000 previously heated at 37^{o}C for 8 minutes. The mixture was incubated at 4^{o}C overnight. Tagged 3'-mRNAs were purified two times using Amicon Ultra-0.5 mL Centrifugal Filters following manufacturer protocol and stored at -20^{o}C until use.

For 5' RNA end-labeling the following protocol was applied: purified mRNAs 3'-end labeled were fluorescent labeled on the 5' end in a three step strategy using T4 polynucleotide kinase (NEB). First, 1 nmole of each RNA were dephosphorylated on the 5'-end for 1 hour at 37^{o}C using Calf Alkaline phosphatase (NEB). Thereafter a thiophosphate group was added on the dephosphorylated 5' end using ATPγS and T4 polynucleotide kinase. Finally, the thiol-reactive C-2 Maleimide-Alexa Fluor 647 (6 nmoles in DMSO. Prepared at instant. Invitrogen) was allowed to react with the thiophosphate for 30 minutes at 65^{o}C. All 5'-3' end labeled mRNAs were phenol extracted, precipitated with Ethanol and re-suspend in nuclease free water. The labeling process was realized avoiding as much as possible light and RNA contamination.

5.10. 5'-3' end distance of mRNAs measured with sm-FRET

To measure the 5'-3' end physical distance of transcript RNA molecules, labeled-RNA samples were diluted to picomolar range in TE buffer or in TM buffer (5 mM EDTA, 10 mM of $MgCl_2$). Each sample was loaded into a sample chamber and smFRET measurements were carry out. FRET efficiency distribution of each RNA sample was computed and fitted with a Gaussian distribution as show in figures (5.14) and (5.15).

mRNA-FRET histograms contain two peaks; one centered at $E \approx 0$ which is caused probably by RNA molecules lacking an active acceptor dye and the other at $E > 0$ indicating certain RNA folding structure which keeps the two ends in close proximity. For instance, an RNA of 1000 nt has a fully stretched longitude of 590 nm but for mRNA *triat1* (1012 nt) we determine an $E = 0.68$ which corresponds to a dye-to-dye separation around 8 nm.

The inter-dye distances of RNA-labeled samples were computed by comparing mean FRET efficiencies extracted from FRET RNA histograms with the calibration curve (Fig. 3.8 Chap. 3). Figure (5.16) gives the separation distance of the dyes attached to RNA vs the lengths of RNA samples. It is remarkable that changing the RNA length by a factor of 10 changes the dyes separation by less than 50 %. The separation is not completely constant but we observe a small increasing trend of 0.7 ± 0.2 picometer/nt. The same result is obtained both in TE buffer and in the presence of 5 mM of $MgCl_2$.

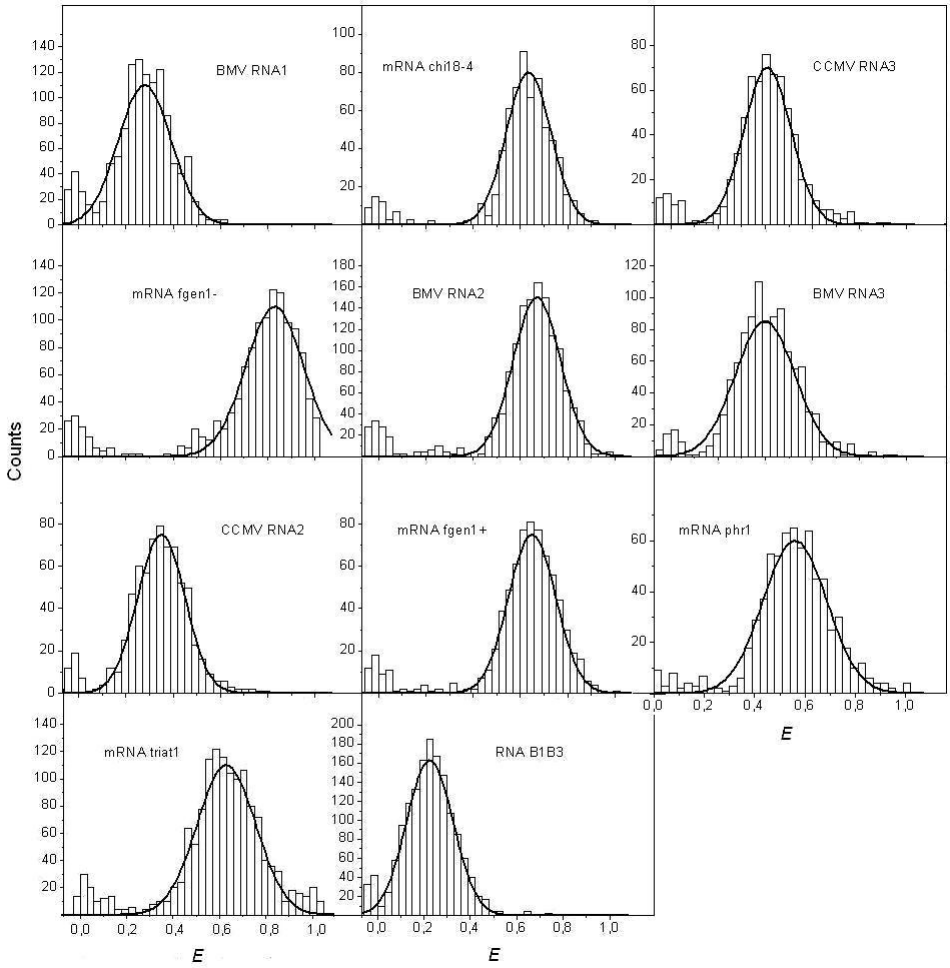

Figure 5.14.: smFRET efficiency histograms for mRNAs in 5 mM $MgCl_2$. The black lines are Gaussian fits

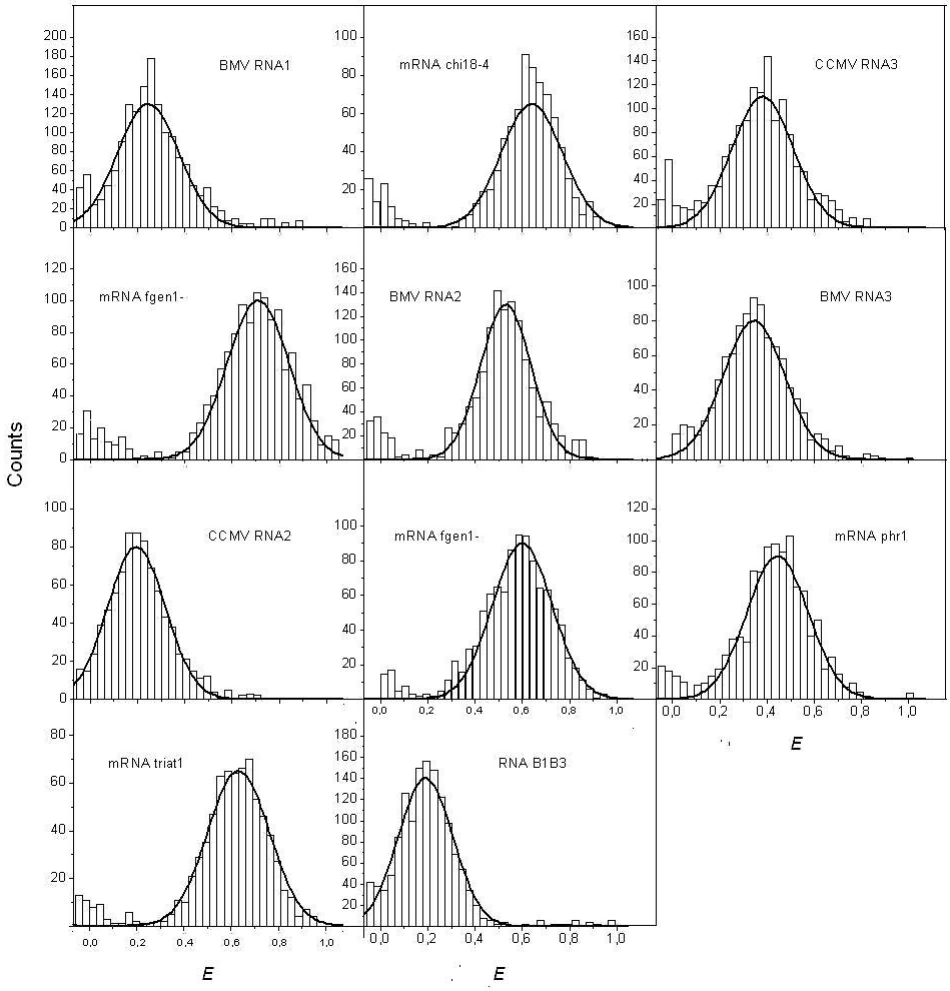

Figure 5.15.: smFRET efficiency histograms for RNAs in TE buffer solution. The black lines are Gaussian fits

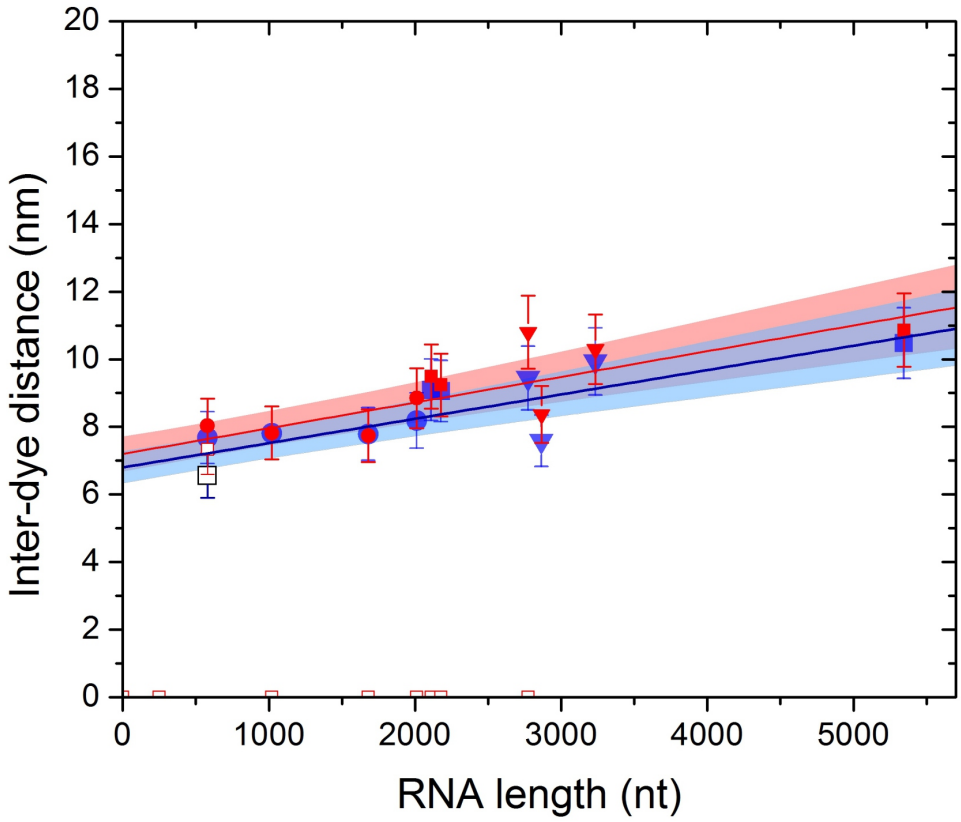

Figure 5.16.: Inter-dye distances for end-labeled mRNA molecules of different lengths. Circles and triangles represent monocistronic fungal and viral mRNAs, respectively. Filled squares represent dicistronic mRNA molecules whereas empty squares represent the antisense mRNA *fgen1*. Blue and red data are from TM and TE buffer solution smFRET experiments, respectively. Error bars correspond to $\pm 1\sigma$. The plot includes the linear fits ($y = a+bx$) with $a = 6.8\pm0.47$ nm and $b = 7.2\pm2x1\times10^{-4}$ nm/nt for TM buffer and $a = 7.2 \pm 0.5$ nm and $b = 7.6 \pm 2x1 \times 10^{-4}$ nm/nt for TE buffer with the 1σ band for each fit

The error bars in Fig. (5.16) are dominated by the 10% uncertainty of the calibration. The inter-dye separation fluctuates within a standard deviation of 1.2 nm, which corresponds to about 14% of the measured separation.

The separation between fluorophores we found is in the range between 6.5 to 10.5 nm (range 1. See Fig. 5.17). To extract the end-to-end distance of RNA molecules (range 2. Fig. 5.17) from data in Fig. (5.16), we need to subtract from range 1

the contribution coming from the fluorophore linker lengths. The dyes attached to both ends of the short DNA molecules used for calibration point towards opposite directions adding an extra 1.5 nm to the total length. However, secondary structures of our RNA molecules predicted by mFOLD showed that in the case of the RNA molecules, dyes point mainly in opposite directions as well (Fig. 5.17). Therefore subtracting 1.5 nm to inter-dye distances we obtain an RNA end-to-end distance between 5 to 9.5 nm (range 2).

smFRET histograms of the RNA molecules, are 50 % wider than the limits of

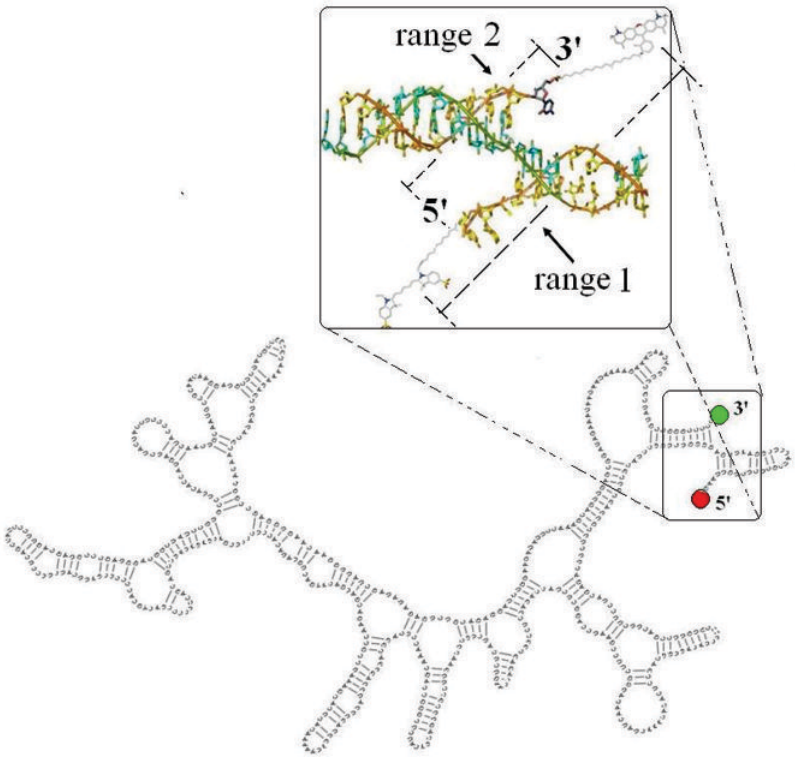

Figure 5.17.: Secondary structure of mRNA *fgen1+* predicted by mFOLD web server [95] and edited with jViz.Rna 2.0 [124], with zoom on the 5' and 3' region. The gray line indicated the contour length

statistics as we can see in figure (Fig. 5.18a). As in the case of DNA (3.10), dye linker motion on RNA molecules contributes to the width (Fig. 5.18b). However, since the persistence length of ssRNA (2.1 nm at 5 mM $MgCl_2$ [21]) is much smaller than that of dsDNA, there should be considerable motion of the terminals that intrinsically contributes to the end-to-end separation (considering only the exterior

loop as a ssRNA). The extra broadening can be explained by a motion of the fluo-

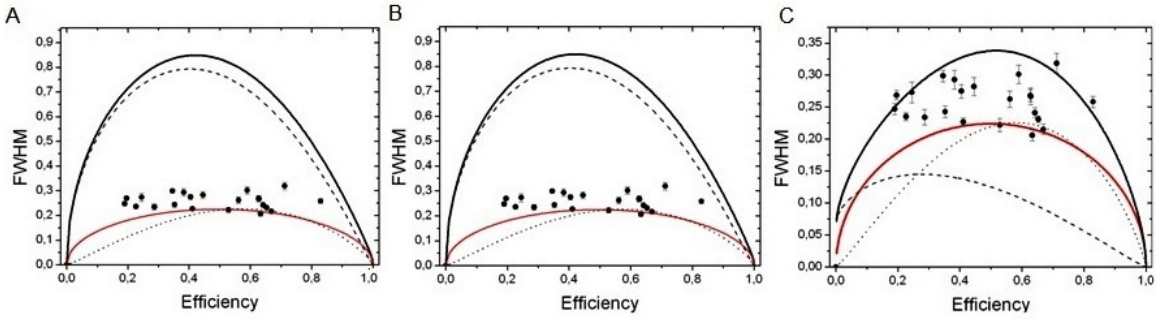

Figure 5.18.: Width of the RNA histograms (circles). It shows the contributions to the width coming from statistics (dotted line), from the motion of the fluorophores within a range of 0.8 nm (red line), the variations in the RNA terminals distance in an end to end distribution (dashed line) and the total width obtained by adding all the contributions (blue line); (a) persistence length of 2.1 nm

rophores through the end-to-end distribution deduced by Thirumalai and Ha [125] for polymers that obeys the worm-like chain model

$$P(r) = \frac{4\pi A r^2}{(1-r^2)^{9/2}} \exp\left(\frac{-4t}{3(1-r^2)}\right) \tag{5.1}$$

where A is a normalization constant, t is the contour length L in multiples of the persistence length L_p, $t = L/L_p$, and r is the end-to-end distance R normalized to the contour length, $r = R/L$. Therefore the averaged-FRET efficiency can be obtained by

$$< E >= \int_0^1 \frac{1}{1+(rL/R_o)^6} P(r)dr. \tag{5.2}$$

For a value of $R_o = 8.5$ nm, a persistence length of $L_p = 2.1$ nm, and a length of $\approx$0.59 nm between each nucleotide in the exterior loop, the solution for the integrable Eq. (5.2) for a value of $< E >= 0.68$ gives an histograms with a width greater than those we observed (Fig. 5.19). However, by increasing the persistence length L_p the experimental widths of the RNA-FRET histograms fits well with Eq. (5.2). Therefore the narrower widths measured indicate a higher rigidity of the ssRNA exterior loop (Fig. 5.17). The increased rigidity may come from the fact that in most cases the exterior loop, where the ends are located, is anchored by at least two

paired regions (stem-loops) on the mRNA [97] that restrict its movement.

In brief assuming a rigid ssRNA and 0.59 nm separation between nt, we obtain

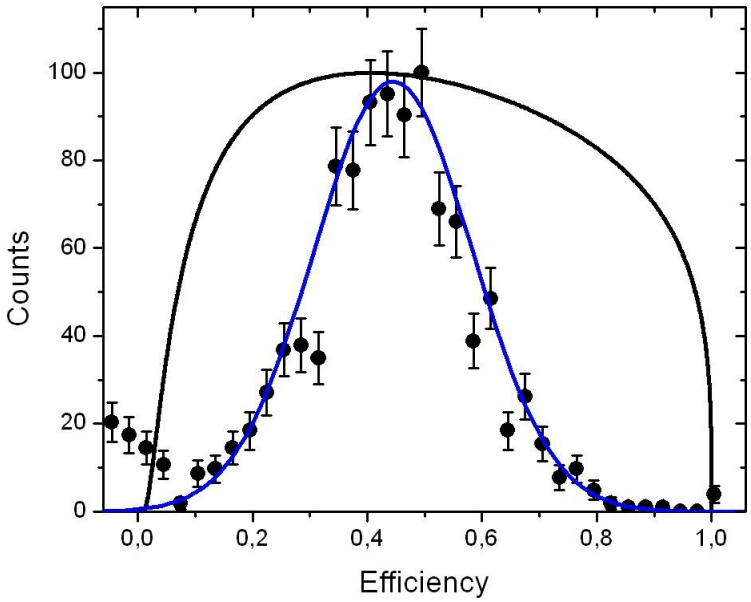

Figure 5.19.: Experimental and theoretical FRET histogram for the mRNA *triat1* obtained in TE buffer solution. The black line is the FRET distribution deduced by Eq. (5.2). The black circles are the experimental data fit with a Guassian distribution (Blue line).

an exterior loop contour length (L) equivalent to 11-19 nt for range 1 and 9-16 nt for range 2. If instead we use a persistence length of 2.1 nm and the end-to-end distribution (5.1) we obtain an L of 19-88 nt for range 1 and L of 12-46 nt for range 2. We propose that range 2 is the correct one for the comparison and that the persistence length of the exterior loop should be much higher than 2.1 nm to get results consistent with the width of the histograms (Fig. 5.18c). Under these assumptions, all the measured RNA molecules give an L between 9 to 16 nt, which is consistent with theoretical predictions [93, 97, 107, 109]. In addition, the end-to-end separation is also neither affected by the different structures that an RNA molecule can adopt [126], nor the difference in origin and length of all tested mRNAs.

The length of the majority of conserved proteins found in eukaryotes and prokaryotes species are between 70 and 1500 aminoacids (aa) in length [127–129]. Therefore, in this sense, the range in size of the mRNA molecules we used spans a range of what is

biologically relevant. Because of the short end-to-end separation we found, our results imply that the ssRNA molecules are "effectively circularized" and raise an intriguing question: is the end-to-end distance of RNAs conserved in all forms of life? If so, this structural feature must have played an important role in evolution, for example, in allowing RNA recognition to carry out their functions, a reminiscent that we can find nowadays in riboswitches, mRNA splicing [130] and transcription termination in prokaryotes [128]. However, there are mRNAs that code for exceptionally large proteins, e.g. Titin that is made of 27,000 aa [131]. Certainly, these large mRNAs are of biological significance in cell physiology; we speculate that even these large molecules also adopt similar end-to-end separation distance, since they use the same post-transcriptional and translational machineries.

Conclusion

We have measured experimentally the distance between the 5'-3' end terminals of ssRNA molecules of different length and obtained from different sources, a fungus and two viruses. The end-to-end separation of study RNA samples increases slowly in the range between 5.0 to 9.0 nm (9 to 21 links) as we vary the RNA length from 500 to 5500 nucleotides. With these measurements we finally provide the first experimental verification of the physical separation of the ends of long ssRNA molecules. We show that this separation is indeed small and remains fairly constant, independent of the RNA length, with the exception that we do observe a small linear growing trend as the molecular length increases. The result holds in the presence of TE buffer or in 5 mM of $MgCl_2$, it is independent of the RNA origin and it covers all the biological relevant range of RNA lengths. The fact that the ends of ssRNA molecules are close by indicates that most ssRNA molecules in nature could be effectively circularized, which can lead to an efficient translation as well as to ensure that only complete RNA molecules are translated.

6

Structural Changes of HIV-1 RNA targets and Cleavage Activity by the *Tt Ago*-RISC

An RNA molecule may coexist in different alternative structures in order to perform different functional roles [132, 133]. For instance, the cIIIm RNA of the bacteriophage lambda was found in two equilibrium structures called the structure A and the structure B. The structure B allows for efficient binding to the 30 S ribosomal subunit, while in the conformation A, the translation initiation region is occluded, thereby preventing the 30 S ribosomal subunit binding [134]. Biebricher *et al.* [135] found that the small RNA molecule SV11 (115 nt) that is replicated by Qβ replicase, exists in two major conformations; the metastable multicomponent structure and the native conformation in the form of a rod-like structure. The metastable conformation is a template for Qβ replicase while the native is not. Other interesting examples have been observed in some RNAs with catalytic activity (ribozymes) [136, 137] and gene-regulatory mRNA domains (riboswitches) [138–141]. Experimental techniques such as nuclear magnetic resonance and temperature jump kinetics have provided helpful insights into the details of RNA structure and its conformation [144, 145].
As an alternative experimental approach, single-molecule fluorescence resonance energy transfer (smFRET) has emerged to give us new insights on the RNA molecular arrangements [146]. Implementation of smFRET in RNA folding studies has revealed that RNA can fold into multiple conformations with different time rates [146–148], it has no-equilibrium period jumps and tertiary cooperative contact [149, 150]. These results have been generally observed by immobilizing the RNA molecule to a substrate and changing the experimental conditions, for example, magnesium ions concentration. However conformational changes of an RNA secondary structure might also be induced by temperature changing (RNA thermometer [159]) or nucleotide mutations along the native RNA sequence [151]. Nucleotide mutations are common

events in natural systems such as viral RNA, which use this mechanism to elude the host defense system, and therefore ensuring their survival [152]. For example, some RNA-target sequences into HIV-I RNA virus can escape from the RNA-induced silencing complex (RISC) cleavage by nucleotide substitutions or deletions. The RISC complex is a ribonucleoprotein particle composed of a single-stranded short interfering RNA (siRNA) and an endonucleolytically active Argonaute protein, capable of cleaving mRNAs complementary to the siRNA. It has been proposed that punctual mutations in some region of the RNA sequence leads to an alternative structure of the RNA that occludes the target sequence to the RISC system [153, 154].
In this chapter, we present the results obtained on four small conserved HIV-I RNA target-sequences called T1, T6, Ago-WT and Ago-Mut, in free diffusion single-molecule studies. RNA target T6 is derived from nucleotide substitution in the RNA T1 sequence in such a way that partially exposes its target sequence to a small interference RNA termed Pol [151]. RNA target Ago-Mut is obtained from a punctual mutation in the RNA Ago-WT sequence. RNA Ago-WT sequence is on the genomic HIV-I viral *Nef* gene. Here we observed that the RNA target T6 shows three peaks in the FRET signal at room temperature and the population of each peak changes with temperature. In contrast RNAs T1, Ago-Mut and Ago-WT show only one component. To analyze how the target RNA structures influence the cleavage activity by RISC, we incubated the target RNAs with a RISC system containing in its core an Argonaute protein (*Thermus thermophilus Argonaute, TtAgo*) and a short single-stranded interference RNA (siRNA) either Pol or Nef. Our results reveal that, depending on the point mutation, there is a different accessibility to the target sequence.

6.1. Material and methods

Ribonucleic acids RNA interference Pol (RNAi: 5'-CUGUAUCAUCUGCUCCUGU-3'), target T1 and target T6 were synthesized and purified in a RNase free HPLC (DNA Technologies Inc.). For smFRET experiments, T1 and T6 were also labeled by the vendor with Cy5 and Cy3 fluorophores on the 5' and 3' ends, respectively (T1: 5'-Cy5-ACAGGAGCAGAUGAUACAGUUCAAGAGACUGUAUCAUCUGCUCCUG UAAU-Cy3-3'; T6: 5'-Cy5-ACAGGAGCAGAUGAUACAGUUCAAGAGAAUGUA

UAAUAUGAUACUGUAAU-Cy3-3'). RNA targets Ago-WT (5'-Cy5-AGCAAUCAC AAGUAGCAAUACAGCAGCUACCA-CY3-AUGCUGCUUGUGCCUGGCUAGAA GCACAAGAGGAGG-3') and Ago-Mut (5'-Cy5-AGCAAUCACAAGUAGCAAUACA GCAGCUACCA-CY3-AUACUGCUUGUGCCUGGCUAGAAGCACAAGAGGAGG) were also labeled with Cy3 and Cy5 dyes. Interference RNA Nef (5'-UGUGCUUCUAG CCAGGCAC-3') was provided by T4 Oligo.

The RNA T6 is derived from nucleotide mutation of Cystidine's (C's) to Adeno-

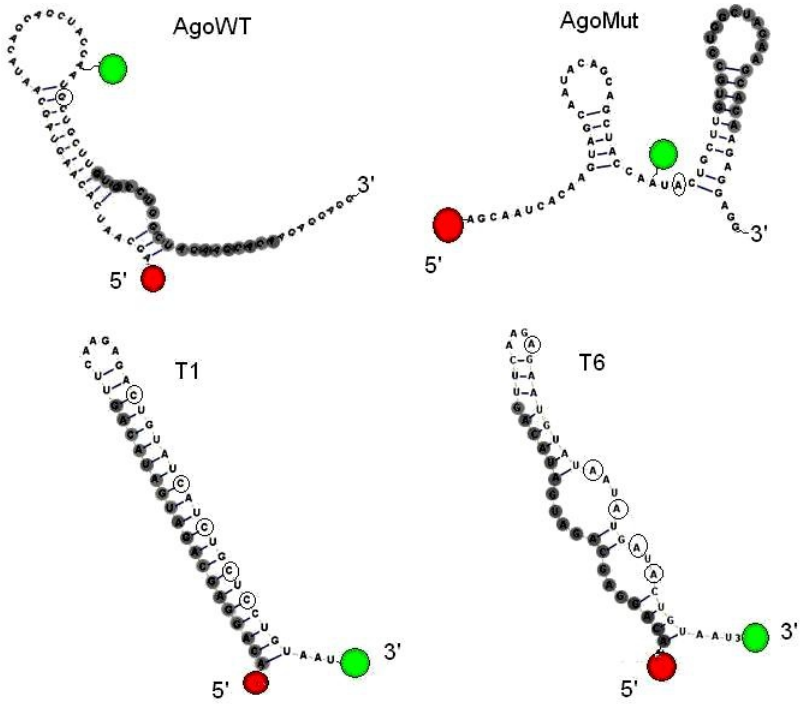

Figure 6.1.: Predicted secondary structures [96] of the RNA HIV-I targets used in this study. The nucleotides encircled denote the nucleotide substitution. The target sequence is highlighted by gray. Only in the RNA T1 is completely occluded the targeted sequence. The red and green circles in the RNAs represent the Cy5 and Cy3 dyes respectively.

sine's (A's) in the positions C29Δ, C35Δ, C38Δ, C41Δ, and C43Δ upstream of the RNA T1 sequence (Fig. 6.1). T1 and T6 targets were designed to minimize the interaction between the fluorophores; the AAU sequence was added to the 3'-end to separate the donor and acceptor and minimize their interaction. When the RNA is completely folded, and the 5'-3' ends are in the closest positions, the donor-acceptor

distance remains significant ($\sim$ 5 nt) as is shown in figure 6.1. The RNA Ago-Mut is derived from the RNA Ago-WT by an A-to-G substitution at nucleotide G35Δ upstream of the target sequence as is indicated in figure 6.1. Wild-type Argonaute from *T. thermophilus* (*TtAgo*) was obtained in the laboratory of Prof. Joel Stavans as described previously in reference [158], with appropriate modifications in the protocol. The full-length Argonaute was expressed in BL21 (DE3) cells cultured in LB medium during an overnight induction period at 20 ^{o}C with 0.4 mM IPTG at an OD600=0.6. Cells were harvested and frozen in liquid nitrogen. After thawing, cells were resuspended in 50 mM Tris pH 7.5, 0.5 M NaCl, 20 mM Imidazole, 1 mM EDTA, complete proteinase inhibitor (Roche), and lysed by sonication. After centrifugation (40,000 $\times$ g, 1 hr), the supernatant was filtered and loaded onto a Ni^{2+} affinity column (HiTrap chelating HP, GE Healthcare) that was equilibrated with 50 mM Tris pH 7.5, 0.5 M NaCl, 20 mM Imidazole. Argonaute was eluted with the same buffer containing 0.5 M Imidazole. The protein was then applied to a size exclusion column equilibrated with 25 mM Tris pH 7.5, 0.5 M NaCl and 10 mM DTT (HiLoad 16/60 Superdex 75, GE Healthcare).

6.1.1. Temperature controller device

To carry out smFRET experiments at different temperatures, the capillary tube sample container was placed on top of a copper holder connected to a Peltier element (Fig. 6.2). A control module (Mod. ITC133, Thorlabs, NJ, USA) fed the Peltier element. The temperature of the sample was monitored at all times using a thermocouple micro-probe (Model HH21, Omega, CT, USA). The temperature of the sample was controlled within $\pm$0.1 ^{o}C in the range between 4 and 70 ^{o}C. The average background signal increased from 2 to 4 photons/ms when going from room temperature to 45 ^{o}C.

6.2. Results and Discussions of the RNA targets T1 and T6

The results of smFRET performed with the RNA targets separately at room temperature (27 ^{o}C) are shown in Fig. (6.3). The Gaussian fit to the RNA T1 gives a mean

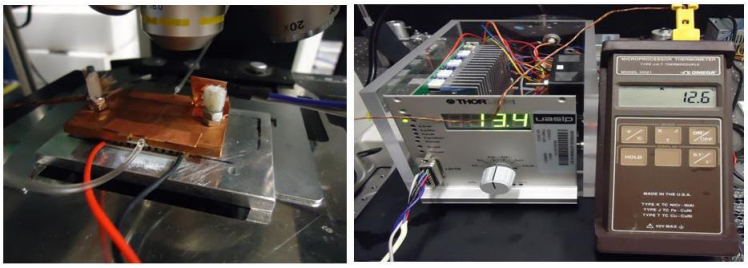

Figure 6.2.: Temperature controller device built for single-molecule FRET experiments at different temperatures. From left to right; the copper holder with the Peltier element, the controller module, and the digital thermometer.

FRET efficiency of $E = 0.96 \pm 0.05$ and a Full-Width at High Maximum (FWHM) $= 0.06 \pm 0.002$. The data could indicates the presence of only one RNA T1 sec-

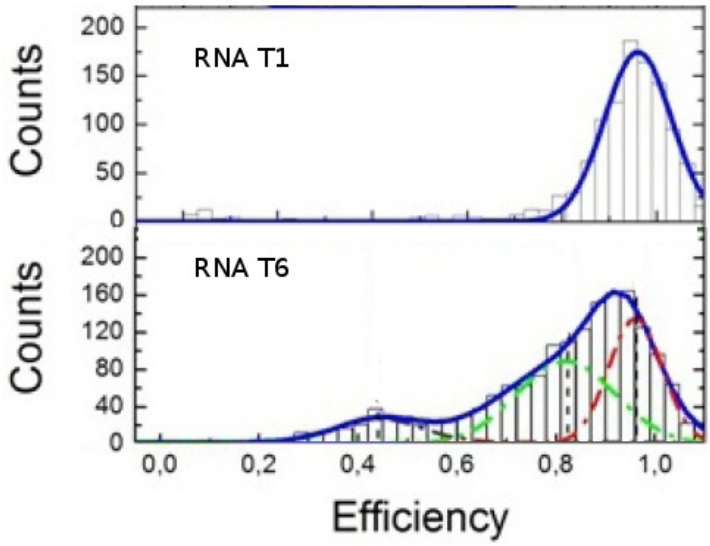

Figure 6.3.: FRET histograms of the RNA HIV-1 based targets T1 and T6 measured at room temperature (27 ^{o}C).

ondary structure and, according to equation (1.9), the structure has its terminals in close proximity, in good agreement with the secondary structure predicted by RNA Vienna (Fig. 6.1). The FRET histogram for the RNA T6 (Fig. 6.3), in contrast, is much wider and indicates the presence of several peaks. Deconvolution of the FRET histogram shows the presence of three peaks, indicating the presence of three or more coexisting RNA T6 secondary structures. The percentage of molecules in

each state was obtained by determining the area of each distribution. The first state assigned as P_1, has a mean energy transfer efficiency of $E = 0.955 \pm 0.03$ and covers $40.5 \pm 5\%$ of the total molecular population. The second state, called P_2, corresponds to a transfer energy efficiency $E = 0.82 \pm 0.07$ and covers a molecular population of $48.2 \pm 10\%$. The third state, called P_3, with a mean transfer energy efficiency $E = 0.44 \pm 0.01$, covers $11.2 \pm 2\%$ of total detected molecular population. The first state P_1 has similar mean FRET efficiency to that of RNA T1. It is probably, then, that this E value is originated from the minimum free energy secondary structure predicted by RNA Vienna (Fig. 6.1). The other two states, P_2 and P_3, correspond to different RNA T6 configurations that have the two fluorophores further apart. To explore the possible configurations of RNA T6 that explain the above observations, we ran the RNA T6 sequence in the RNAsubopt package of RNA Vienna [96]. Figure (6.4) shows the lowest energy suboptimal RNA configurations of RNA T6. From these predictions we can immediately observe that the first three states con-

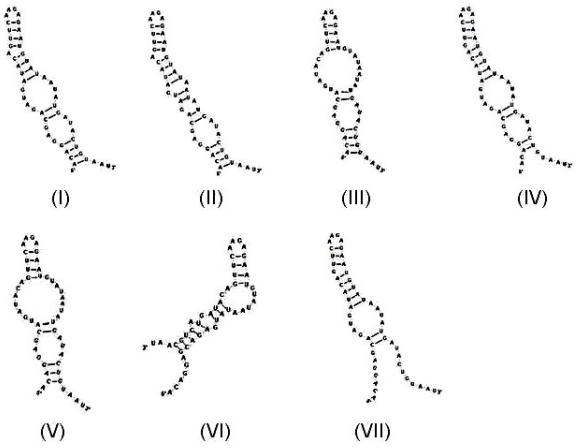

Figure 6.4.: Lowest energy suboptimal RNA T6 structures determined by RNAsubopt

tain the 5'-3' ends in similar close proximity to the lower energy secondary structure. Then the measured FRET efficiency $E = 0.955$ cannot be assigned directly to the minimum free energy secondary structure of RNA T6, rather to the set formed by the first three suboptimal RNA T6 structures. To obtain an estimate of the mean FRET efficiency originated by each suboptimal RNA T6 structure, the exterior loop of each structure was modeled as a semi-flexible ssRNA chain of contour length L

and persistence length $l_p = 2.1$ nm, but rigid and immobile within the fluorescence lifetime of the dyes burst. Then the FRET efficiency for each RNA T6 structure was determined through the end-to-end distribution [125],

$$E = \int_0^1 \frac{1}{1 + (rL/R_o)^6} P(r)dr. \tag{6.1}$$

For a value of $R_o = 5.4$ nm for the pair of dyes Cy5-Cy3 [157], and a length of $l_{nt} \approx 0.59$ nm between each nucleotide in the exterior loop, the mean FRET efficiencies obtained by equation (6.1) for the suboptimal RNA T6 structures (Fig. 6.4) are summarized in table 1.

Length L (nt)	Structure	$< E >$
5	I, II, III	0.98
7	V	0.95
9	IV	0.87
11	VI	0.75
19	VII	0.44

Table 1. Mean FRET efficiencies for the lowest energy RNA T6 suboptimal structures from Fig. (6.4) determined by the end-to-end distribution

From these data, the measured FRET efficiency $E = 0.82$ could correspond to the set of suboptimal secondary structures IV and VI, while structure VII could be the contributing to the FRET efficiency peak at $E = 0.44$. These estimated FRET efficiencies do not change much for values of R_o ranging from 5.2 nm to 5.8 nm, nucleotide distance from 0.59 to 0.63 nm and persistence length between 1.9 and 2.6 nm. The results indicate that both RNAsubopt and the worm-like chain model may adequately describe our measured FRET efficiencies. We only consider the seven lowest energy configurations, but many more are possible. In order to determine the influence of temperature on the observed RNA T6 and T1 structures at room temperature, we additionally performed single molecule FRET of RNA T6 and T1 at 13.1, 36.7 and 44.3 ^{o}C. FRET histograms are showed in Fig. 6.5. The table 2 summarized the mean FRET efficiencies and FRET areas derived from the above FRET histograms (Fig. 6.5) only for RNA T6.

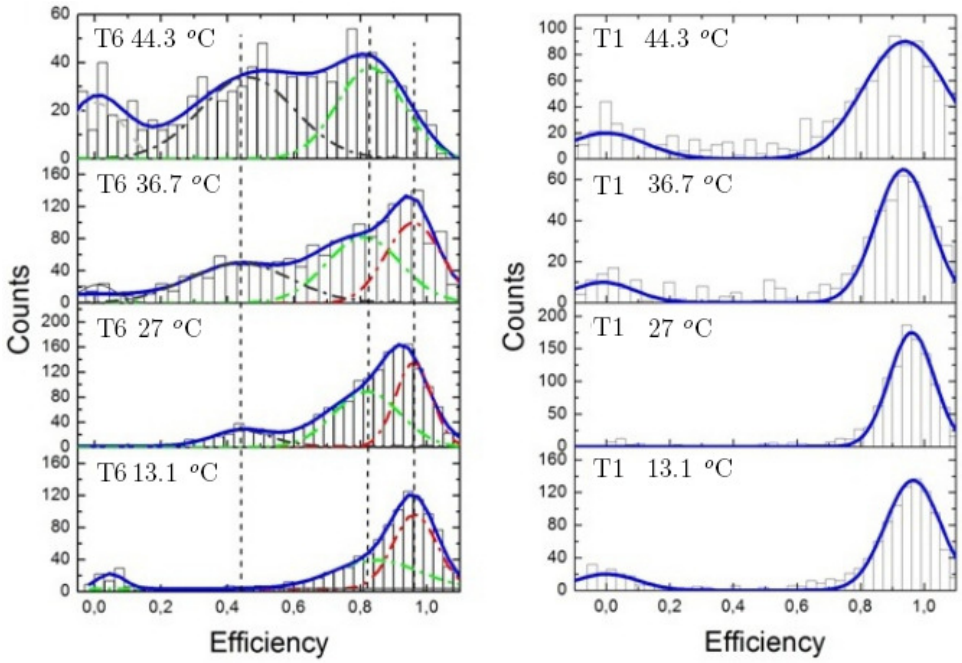

Figure 6.5.: FRET histogram of the RNA T6 and the RNA T1 obtained at different temperatures. Blue line is the sum of three or two Gaussian distributions. Red, black an green lines is the contribution of each individual Gaussian distribution.

Temp °C	E_1	E_2	E_3	$A_1\%$	$A_2\%$	$A_3\%$
44.3	0.80 ± 0.001	0.47 ± 0.002	0.07 ± 0.002	32.8 ± 8	38.1 ± 13	29 ± 11
36.7	0.96 ± 0.009	0.81 ± 0.007	0.42 ± 0.005	31.3 ± 10	39.2 ± 12	22.9 ± 8
13.1	0.96 ± 0.001	0.84 ± 0.002	0.05 ± 0.002	55.87 ± 12	34.2 ± 11	10.87 ± 2

Table 2. Average FRET efficiency, E, and FRET areas from a fit to a sum of three Gaussian distributions for the measured FRET efficiency histograms of RNA T6 (Fig. 6.5).

From data in table 2 we can observe that mean FRET efficiencies assigned to states P_1, P_2 and P_3 of RNA T6 always appear but its concentration changes with temperature. A more clear evidence of this behavior is showed in figure (6.6). P_1 decreases significantly by increasing the temperature, from about 57% at 13.1 °C to about 1.1% at 44.3 °C. In contrast, the fraction of RNA T6 molecules in the P_3 state is favored when the temperature increases; at the lowest temperature it has a

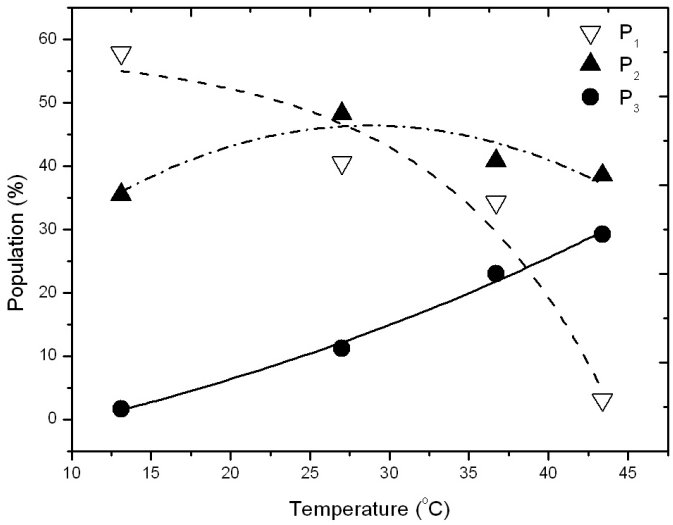

Figure 6.6.: Populations of RNA T6 observed as a function of temperature. P_1 decreased by increasing the temperature while P_3 increased. P_2 changes slightly in the temperature range. The curves are only a guide of the data.

low molecular population close to 0%, while at the highest temperature its molecular population grew significantly to about 30%. The behavior of P_2 as function of temperature remains fairly constant. Furthermore, at the highest temperature one new RNA T6 conformations appear with about 12% of the total molecular population. From the above analysis the behavior of RNA T6 is similar to an RNA thermometer. RNA thermo-sensors has been under investigation for many years in order to understand how they modulate translation efficiency of an mRNA according to the ambient temperature [160].

On the other hand, FRET histograms of RNA T1 performed at the different temperatures do not show great change in mean FRET efficiency (Fig. 6.5). This implies that the distance of separation between the dyes remains in close proximity such as is depicted in the predicted secondary structure of RNA T1.

6.3. Structural Stability of the RNA Targets Ago-WT and Ago-Mut at Different Temperatures

Figure 6.7 shows the RNA FRET histograms obtained for AgoMut and AgoWT at different temperatures. Two Gaussian distributions give a good fit to the data. Table 3 summarized the fit parameters. The transfer efficiency obtained does not change much from 12 to 37^{o}C for Ago-Mut (Fig. 6.7). However, at 44 ^{o}C, the FRET area associated with nonzero mean FRET efficiency decreases together with an increase of the FRET area with FRET efficiency near zero (Fig. 6.7). Assuming that the small hairpin (5 bp) of AgoMut structure at 44 ^{o}C opens (Fig. 6.1), the distance between Cy5-Cy3 dyes will be around 32 nt. Then, for a distance of ≈ 0.59 nm between each nt, and a ssRNA persistence length ≈ 2.1 nm, the end-to-end distance (Eq. 6.1) for a contour length $L = 32$ nt, gives an averaged FRET efficiency of $< E > \approx 0.09$. Therefore, we propose that the increased population at $E = 0$ for AgoMut at 44 ^{o}C, is due to the disruption of the 5 bp small hairpin.

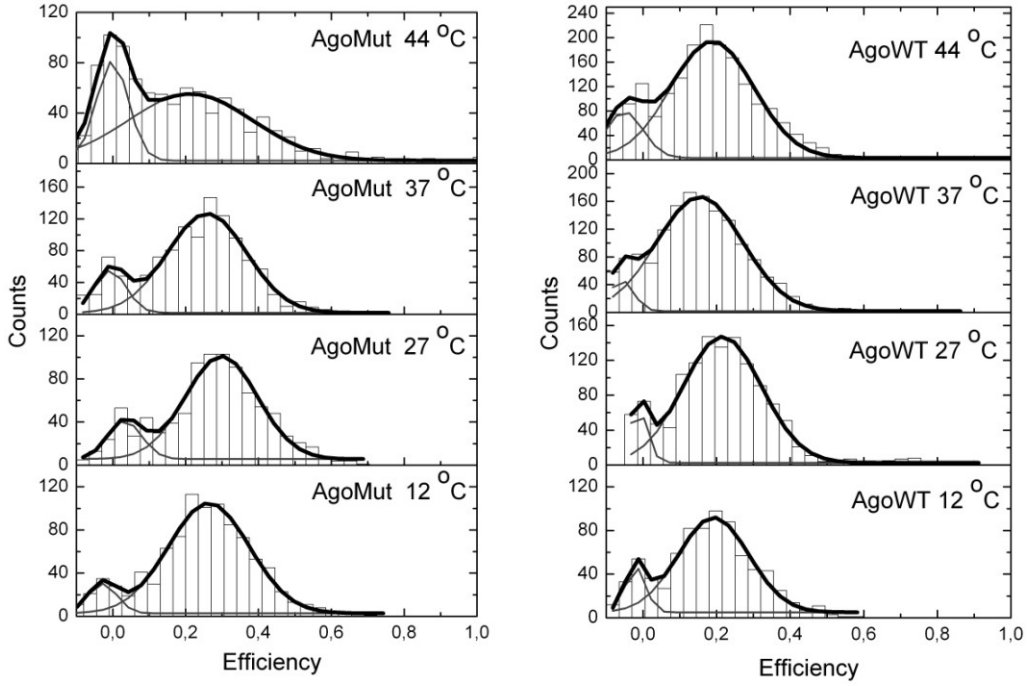

Figure 6.7.: FRET histograms obtained from single-molecule counts of RNAs Ago-Mut and Ago-WT at different temperatures.

AgoMut					AgoWT				
Tem	E1	FWHM1	E2	FWHM1	Tem	E1	FWHM1	E2	FWHM1
44	0.05±0.001	0.05±0.01	0.21±0.014	0.15±0.001	44	0.05±0.009	0.06±0.001	0.18±0.005	0.12±0.007
37	0.006±0.001	0.04±0.01	0.26±0.006	0.11±0.02	37	.05±0.006	0.05±0.001	0.155±0.01	0.11±0.008
27	0.03±0.001	0.05±0.001	0.29±0.006	0.09±0.005	27	0.01±0.003	0.04±0.001	0.21±0.003	0.1±0.008
12	0.001±0.0009	0.04±0.01	0.27±0.004	0.11±0.01	12	0.02±0.001	0.05±0.005	0.19±0.005	0.095±0.006

Table 3. Average FRET efficiency, E, full width at half maximum (FWHM) parameters from a fit to a sum of three Gaussian distributions for the measured FRET efficiency histograms of RNAs AgoWT and AgoMut (Fig. 6.7).

In the case of AgoWT (Fig. 6.7), note that the histograms are very similar at all temperatures. However the FWHM increase as a function of temperature. It is probable that a higher temperature the stem of 3 bp in AgoWT structure opens (Fig. 6.1) causing variations in distance between dyes. But this effect will not cause great variations in the mean FRET efficiency due to its small distance contribution to the separation of the Cy5-Cy3 dyes. Nevertheless, the disruption of the stem of 11 bp will provides a distance of separation between Cy5-Cy3 of length $L = 32$ nt. If this is the case, then according to Eq. 6.1, the population close to zero efficiency would increase but the FRET histogram data does not show that behavior.

6.4. Target RNA-interaction with mini-RISC System

For FRET dicing experiments, we prepare several mini-RISC systems following the next reaction: 500 nM of Ago were incubated for 30 min at 55 ^{o}C with 500 nM either siRNAs Pol or Nef in reaction buffer (10 mM HEPES-KOH, pH 7.5, 100 mM NaCl and 5 mM of $MgCl_2$). After that 400 nM of RNA target (WT, WT6, Ago-WT or Ago-Mut) were added to the above mixture (mini-RISC with siRNA Pol was used to cleavage RNA targets WT and WT6 and mini-RISC containing siRNA Nef was used for RNA targets Ago-WT and Ago-Mut) in addition to 4U of RNase Inhibitor, Murine (NEB). The incubation was continued for 60 min at 37 ^{o}C or 44 ^{o}C followed by phenol extracted. For FRET measurements, the mini-RISC RNA targets reactions were diluted to the picomolar range in reaction buffer. Figure (6.8) shows the FRET histograms derived from the mini-RISC RNA target interactions. Under the experimental conditions, the smFRET histograms show that for AgoWT (Fig. 6.8) at 37 ^{o}C and 44 ^{o}C the energy transfer efficiency peak decreases significantly

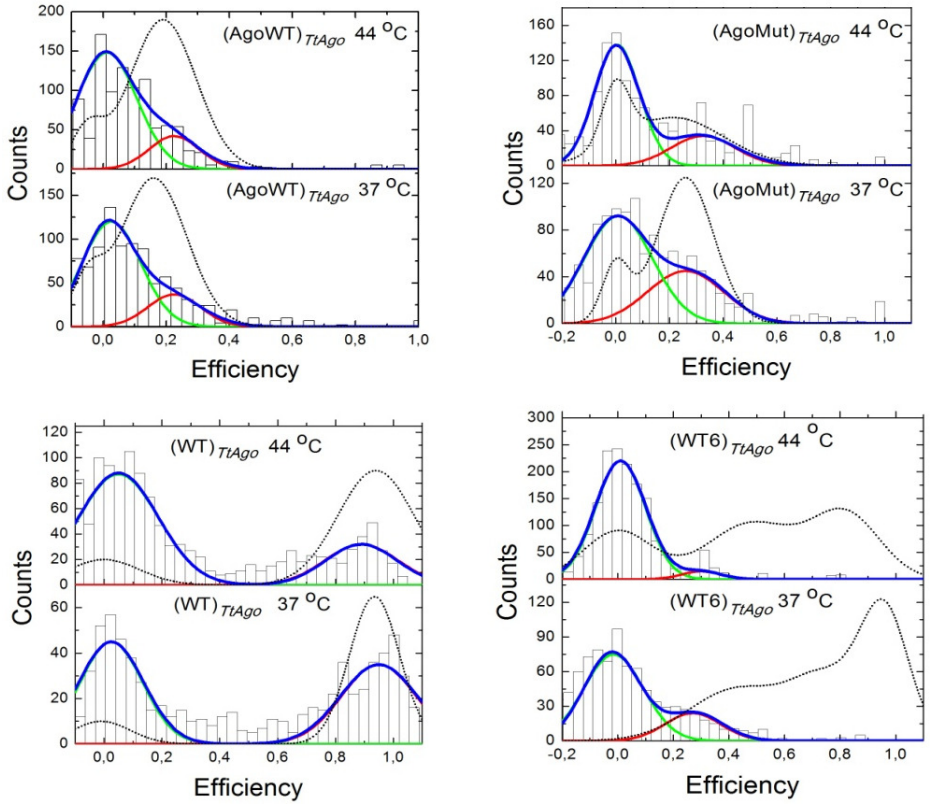

Figure 6.8.: FRET histograms from the experiments with the mini-RISC-RNA targets. The dashed line is the FRET fit of RNA targets without RISC complex.

compared to the intact AgoWT (see histogram of Figs. 6.7), while the population at about zero efficiency increases. If these effects are associated with RISC cleaved activity, then according to Eq. 6.1 the distance between Cy5 and Cy3 dyes in RNA fragments derived from cleaved activity of RISC, should be greater than $L = 28$ nt. Figure 6.9 shows the most probably RNA cleaved structures derived from the cut of mini-RISC on targets AgoMut and AgoWT, respectively. In both cases, the dye molecules remain on one of the diced pieces, so we still have FRET signal. However, the energy transfer efficiency should be different for AgoWT, especially if the two stems S1 and S2 can open-up; in this case, the distance between Cy5 and Cy3 is greater than 30 nt, therefore the energy transfer efficiency must be close to zero such as is showed in Fig. 6.8. For Ago-Mut, the transfer efficiency could be similar to that of the uncleaved molecule, but as the case of AgoWT if the two stems S1 and

S2 can open-up the energy transfer efficiency should be close to zero.
For the RNA T1 and RNA T6, if the mini-RISC cleaves them, there should be practically no energy transfer because the dyes will be located in different pieces of the diced RNA. However for RNA T6 the FRET histograms show a small peak located at about 0.3 efficiency which indicate that there are small amount of unknown RNA T6 structures that keeps the distance of separation of Cy5 and Cy3 around $R = R_o((1 - E)/E)^{1/6} \approx 6$ nm. In the case of the RNA T1 the FRET histograms also show a great amount of cleavage by the mini-RISC. At 44 ^{o}C almost all RNA molecules are cleavage while at 37 ^{o}C there is significant amount of intact RNA T1 molecules since there is energy transfer efficiency at 0.94.

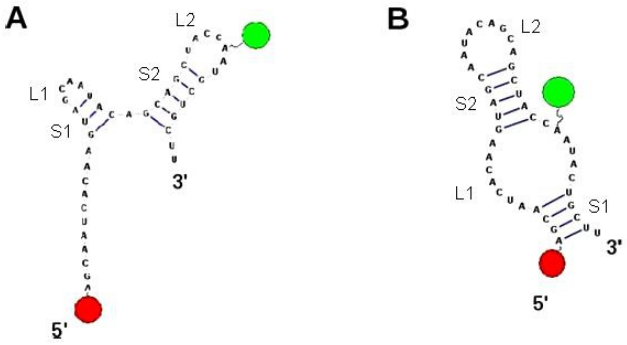

Figure 6.9.: Minimum free energy secondary structures of the cleaved RNAs pieces by the mini-RISC system from (A) Ago-WT and (B) Ago-Mut. Donor (Aceptor) position is labeled by a green (red) circle

6.5. Gel Electrophoresis Shift Assays

To perform tests in gel shift assays, the mini-RISC 5'-3'-labeled target reactions were stopped in loading buffer (25% glycerol, 0.5% SDS, 50 mM EDTA) and 10 μg Proteinase K; incubated for 30 min at 45 ^{o}C; after heat denaturation (65 ^{o}C, 2 min); phenol extracted; and then, were loaded directly onto an 8 % nondenaturing polyacrylamide TBE gel run at 4 ^{o}C using 70 V for 1 hour, then scanned at 546 nm. Figure 6.10 shows the cleavage activity assay of the mini-RISC system on denaturing polyacrylamide gels, where lanes 1, 4, 7 and 12 correspond to the labeled RNAs Ago-MUT, Ago-WT, T6 and T1 respectively, as controls. The activity of mini-RISC on

RNA Ago-MUT both 44 ^{o}C (lane 2) and 37 ^{o}C (lane 3) shows some amount of cleaved activity but a small amount of the intact RNAs target still remain, also in good agreement with the smFRET experiments. The cleaved activity on RNA Ago-WT at 37 ^{o}C (lane 5) and 44 ^{o}C (lane 6) shows a great amount of RNA cleaved such as in the case of RNA target T6 at 37 ^{o}C (lane 8) and 44 ^{o}C (lane 9). These results agree with the FRET data. However RNA target T1 do not shows a great amount of RNA cleaved both 37 ^{o}C (lane 11) and 44 ^{o}C (lane 10) of incubation. This last result does not match well with smFRET data (Fig. 6.8).

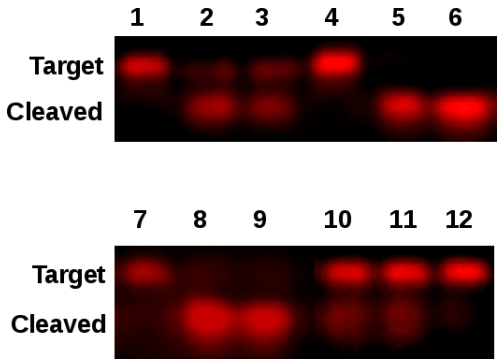

Figure 6.10.: Electrophoresis gel for mini-RISC-RNA labeled targets reaction products. See text for details.

Conclusion

Ribonucleic acids are some of the most important biomolecules of life. Its dynamic mechanisms by which they acquire different structural conformations are extremely important, not only in our understanding of their biological function but also for expanding the field of possible RNA therapy applications. Experimental and simulation results suggest that the RNA folding process have multiple conformations that can even have catalytic activity. These conformational changes may arise from changes in the thermodynamical conditions, such as temperature or ionic strength, as well as from mutations. Here, we show that by introducing mutations into otherwise perfect, small HIV-1 based RNA targets can results in the formation of various conformational coexisting secondary structures. The population on different conformations changes with temperature. Furthermore, we conclusively show that some

of these secondary structures, even at physiological temperature, expose their target sequence allowing the binding of the RISC complex and subsequently cleavage.

A

Labview algorithm

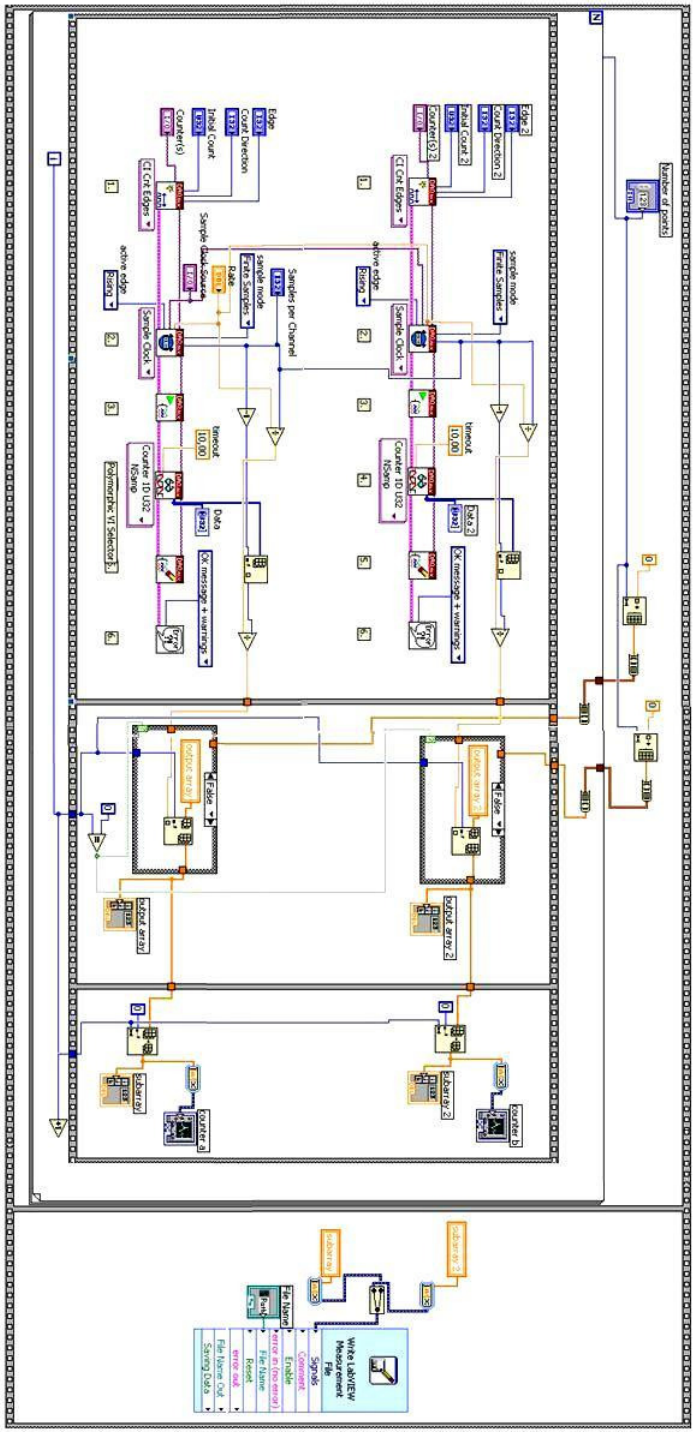
Counter(s)
Initial Count
Count Direction
Edge
CI Ctrl Edges
Sample Clock Source
Rate
Sample mode
Samples per Channel
active edge
Sample Clock
timeout
10,00
Counter 1D U32 NSamp
Polymorphic VI Selector
Data
I32
DBL
Ok, message + warnings
Counter(s) 2
Initial Count 2
Count Direction 2
Edge 2
CI Ctrl Edges
active edge
Sample Clock
sample mode
Finite Samples
timeout
10,00
Counter 1D U32 NSamp
Data 2
Ok, message + warnings
number of points
output array
False
output array 2
False
output array
output array 2
Subarray
Subarray 2
counter 0
counter 1
Write LabVIEW Measurement File
File Name
Data
Comment
Enable
Signals
error out
File Name Out
Saving Data

Figure A.1.:

B

Matlab algorithm for analysis of FRET-data

Step 1

```
clear
data=dlmread (Folder name\file name.txt );    Load FRET data file in format .txt

    sdata=max(size(data));    return the largest value of data
     bin0=data(2:sdata,1);    To data reading from line 2 (Time)
    det10=data(2:sdata,2);    To data reading from line 2 (donor counts)
    det20=data(2:sdata,3);    To data reading from line 2 (acceptor counts)
     sbin=max(size(bin0));    Maximum number of reading data
```

Find bins with counts too high electric noise and erase

```
  level0=1e7;    Fix upper limit
for ja=sbin:-1:1    Reading all data
if det10(ja)>level0 | det20(ja)>level0    Conditional
 det10(ja)=[];    deleting time bin
 det20(ja)=[];    deleting donor bin      end
  bin0(ja)=[];    deleting acceptor bin
end
```

Defining detectors as number of photons

```
unfoton=1000;
det1=det10/unfoton;
det2=det20/unfoton;
suma1=det1+det2;    Sum of total photon counts from both channels (This sum is compared with threshold)
```

Considers only the donor and acceptor sum counts above some threshold

```
    leveld=;                                                    To define a threshold
    sbin2=max(size(suma1));                                     Read all data
    suma2=suma1                                                 defining new variable
    det1b=det1                                                  defining new variable
    det2b=det2                                                  defining new variable
    binb=bin0;                                                  defining new variable
    for ja=sbin2:-1:1                                           reading all data
    if suma2(ja)<leveld          conditional to eliminate counts-sum below of the threshold
    suma2(ja)=[];                                               sum reading
    binb(ja)=[];                                                time bin deleting
    det1b(ja)=[];                                               donor bin deleting
    det2b(ja)=[];                                               acceptor bin deleting
end
end
suma2=det1b+det2b;    new sum from bins which where above the threshold
```

Subtracting the background of bins above the threshold

```
   leveld= ;    To define a background in D-channel
sbin3=max(size(det1));
det1c=det1;
for ja=sbin3:-1:1
if det1c(ja)>leveld
det1c(ja)=[];
end
end
ruido1=mean(det1c);
det1b=det1b-ruido1;
leveld=;    To define a background in A-channel
sbin4=max(size(det2));
det2c=det2;
for ja=sbin3:-1:1
```

if det2c(ja)¿leveld
det2c(ja)=[];
end
end
ruido2=mean(det2c);
det2b=det2b-ruido2;
sumat=det1b+det2b;

Computing FRET efficiency

Effi=det2b./sumat;
dlmwrite('Unit:\directory\file name',Effi); Write Effi in a file

Bibliography

[1] Armen R Kherlopian, Ting Song, Qi Duan, Mathew A Neimark, Ming J Po, John K Gohagan, and Andrew F Laine. *A review of imaging techniques for systems biology.* BMC Systems Biology **2**, 74, (2008).

[2] Daniel Figeys and Devanand Pinto. *Lab-on-a-Chip: A Revolution in Biological and Medical Sciences.* Analytical Chemestry, **11**, 330-335, (2000).

[3] Keith Wilson and John Walker, *Principles and Techniques of Biochemistry and Molecular Biology.* Seventh Edition, (2010). Cambridge University Press.

[4] Eli Barkai. YounJoon Jung, and Robert Silbey, *Theory of Single-Molecule Spectroscopy: Beyond the Ensemble Average.* Annual Review of Physical Chemistry, **55**: 457-507, (2004).

[5] Peter V. Cornish and Taekjip Ha, *A Survey of Single-Molecule Techniques in Chemical Biology*, ACS Chem. Biol. 2007, **2**, 53-61.

[6] Chirlmin Joo, Hamza Balci, Yuji Ishitsuka, Chittanon Buranachai, and Taekjip Ha, *Advances in Single-Molecule Fluorescence Methods for Molecular Biology*, Annual Review of Biochemistry, **77**: 51-76, (2008).

[7] Paul J. B., Michael S., ans Andres F. O., *Single-Molecule Methods, Molecular Biophysics for the Life Sciences.* Biophysics for the Life Sciences, Edition 6, pp 257-288, (2013).

[8] F Ritort, *Single-molecule experiments in biological physics: methods and applications*, J. Phys.: Condens. Matter **18**: 32, (2006).

[9] J. P. S. Farinha and J. M. G. Martinho, *Resonance Energy Transfer in Polymer Nanodomains.* J. Phys. Chem. C, **112**, 10591-10601, (2008).

[10] Selvin P. R., *The renaissance of fluorescence resonance energy transfer*, Nat Struct Biol. 2000, **7** (9):730-734.

[11] Robert M. Clegg, *Fluorescence resonance energy transfer*, Current Opinion in Biotechnology, **6**, 103–110, 1995.

[12] T. Förster. *Intermolecular energy Transfer and Fluorescence.* Ann. Physics, **2**:55-75, (1948). [Translated by R. S. Knox]

[13] M. Margittai, J. Widengren, E. Schweinberger, G. F. Schröder, S. Felekyan, E. Haustein, M. Konig, D. Fasshauer, H. Grubmüller, R. Jahn, and C. A. M. Seide, *Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1*, PNAS, **100**, 26, (2003)

[14] Yusdi Santosoa, Catherine M. Joyceb, Olga Potapovab, Ludovic Le Restea, Johannes Hohlbeina, Joseph P. Torellaa, Nigel D. F. Grindleyb, and Achillefs N. Kapanidisa, *Conformational transitions in DNA polymerase I revealed by single-molecule FRET.* PNAS, **107**, 2, 715-720, (2009).

[15] Chung H. S., Louis J. M., Eaton W. A., *Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories.* PNAS, **106**:11837-44, (2009).

[16] Yiwei Jiaa, David S. Talagaa, Wai Leung Laub, Helen S.M. Lub, William F. DeGradob, Robin M. Hochstrassera, *Folding dynamics of single GCN-4 peptides by fluorescence resonant energy transfer confocal microscopy.* Chemical Physics, **247**, 69-83, (1999).

[17] A.A. Deniz, T.A. Laurence, G.S. Beligere, M. Dahan, A.B. Martin, D.S. Chemla, P.E. Dawson, P.G. Schultz, S. Weiss *Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2.* PNAS, **97**, 5179-5184, (2000).

[18] Elza V. Kuzmenkina, Colin D. Heyes, and G. Ulrich Nienhaus, *Single-molecule Förster resonance energy transfer study of protein dynamics under denaturing conditions.* PNAS, **102**, 15471-15476, (2006).

[19] Benjamin Schuler, and William A Eaton, Protein folding studied by single-molecule FRET. Current Opinion in Structural Biology, **18**, 16-26, (2008).

[20] Nancy Beerens1, Mette D.E. Jepsen1, Volodymyr Nechyporuk-Zloy, Asger C. Krüger, Jean-Luc Darlix, Jorgen Kjems1, and Victoria Birkedal, *Role of the primer activation signal in tRNA annealing onto the HIV-1 genome studied by single-molecule FRET microscopy*, RNA, **19**, 517-526, (2013).

[21] Huimin Chen, Steve P. Meisburger, Suzette A. Pabit, Julie L. Sutton, Watt W. Webb, and Lois Pollack, *Ionic strength-dependent persistence lengths of single-stranded RNA and DNA*, PNAS, **109**, 799-804, (2012).

[22] Sergey V. Solomatin, Max Greenfeld, Steven Chu, and Daniel Herschlag. *Multiple native states reveal persistent ruggedness of an RNA folding landscape*. Nature **463**(4), 681-686, (2010).

[23] Xiaowei Zhuang, Harold Kim, Miguel J. B. Pereira, Hazen P. Babcock, Nils G. Walter, and Steven Chu. *Correlating Structural Dynamics and Function in Single Ribozyme Molecules*. Science 296, 1473, (2002).

[24] Wilson TJ, Nahas M, Araki L, Harusawa S, Ha T, Lilley DM. *RNA folding and the origins of catalytic activity in the hairpin ribozyme*. Blood Cells Mol. Dis. 38(1), 8-14 (2007).

[25] Xie Z, Srividya N, Sosnick TR, Pan T, Scherer NF. 2004. *Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme*. Proc. Natl. Acad. Sci. USA 101:534-39

[26] Elizabeth A Jares-Erijman, and Thomas M Jovin. FRET Imagin, Nature Biotechnology, **11**:1387, (2003).

[27] Jablonski, Aleksander, *Efficiency of Anti-Stokes Fluorescence in Dyes*. Nature, **131**, 839-840, (1933).

[28] P. A. M. Dirac, *The Quantum Theory of the Emission and Absorption of Radiation*, Proc. Roy. Soc., **A114**, 243 (1927)

[29] J. J. Sakurai, *Modern Quantum Mechanics Revised Edition*, Capters V, pp 285-345. (1994). Addison Wesley Publishing Company, Inc.

[30] David L. Andrews and Gediminas J., *Intermolecular Energy Transfer: Retardation Effects*. J. Chem. Phys. 96 (9), 6606-6612. (1992).

[31] Gareth J. Daniels, Robert D. Jenkins, David S. Bradshaw, and David L. Andrews. *Resonance energy transfer: The unified theory revisited*. J. Chem. Phys., **119**(4), 2264-2274, (2003).

[32] Roi Baer and Eran Rabani, *Theory of resonance energy transfer involving nanocrystals: The role of high multipoles*, J. Chem. Phys. **128**, 184710, (2008).

[33] John David Jackson, *Classical Electrodynamics*, Chapter IX, pp 407-449, Third Edition, (1999). John Wiley Sons Inc

[34] http://www.lifetechnologies.com/mx/es/home/life-science/cell-analysis/labeling-chemistry/fluorescence-spectraviewer.html.

[35] http://www.lifetechnologies.com/mx/es/home/references/molecular-probes-the-handbook/tables/r0-values-for-some-alexa-fluor-dyes.html

[36] Theodorus W. J. Gadella, *FRET and FLIM Techniques*, Chapter 1, Laboratory Techniques in Biochemistry and Molecular Biology, Vol 33, Elservier.

[37] Gutierrez-Merino C. *Quantitation of the Förster energy transfer for two-dimensional systems, II. Protein distribution and aggregation state in biological membranes.* Biophys Chem **14**:259-266, (1981).

[38] E. Lang, J. Baier, and J. Köler,*Epifluorescence, confocal and total internal reflection microscopy for single-molecule experiments: a quantitative comparison.* Jounal of Microscopy, **222**, 118-123, (2006).

[39] R. Roy, S. Hohng, and T. Ha, *A practical guide to single-molecule FRET*, Nature Methods, **5**, 507, (2008).

[40] E. Haustein, P. Schwille, *Single-molecule spectroscopic methods*, Current Opinion in Structural Biology **14**, 531-540, (2004).

[41] S. J. Lord , Hsiao-lu D. Lee and W. E. Moerner, *Single-Molecule Spectroscopy and Imaging of Biomolecules in Living Cells* Anal. Chem., **82**, 2192–2203, (2010).

[42] W. Patrick Ambrose, et al. *Single Molecule Fluorescence Spectroscopy at Ambient Temperature* Chem. Rev., **99**, 2929–2956 (1999).

[43] W. E. Moerner and David P. Fromm, *Methods of single-molecule fluorescence spectroscopy and microscopy*, Rev. Sci. Instrum. 74, 3597 (2003).

[44] Rasband, W.S., ImageJ, *U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/* (2012).

[45] *Introduction to Gaussian Beam Optics*, http://mccombe.physics.buffalo.edu/lab-manuals/MG-GaussianBeams.pdf

[46] Pawley J B (ed) 1995. *Handbook of Biological Confocal Microscopy* (New York: Plenum)

[47] Elisabeth K. Hill and Andrew J. de Mello. *Single-molecule detection using confocal fluorescence detection: Assessment of optical probe volumes*. Analyst, 2000, **125**, 1033-1036.

[48] Musundi B. Wabuyele, Hannah Farquar, Wieslaw Stryjewski, Robert P. Hammer, Steven A. Soper, Yu-Wei Cheng, and Francis Barany. *Approaching Real-Time Molecular Diagnostics: Single-Pair Fluorescence Resonance Energy Transfer (spFRET) Detection for the Analysis of Low Abundant Point Mutations in K-ras Oncogenes*. J. Am. Chem. Soc. 2003, **125**, 6937-6945.

[49] T. Ha, *Single-Molecule Fluorescence Resonance Energy Transfer*. METHODS 25, 78-86 (2001).

[50] Maxime Dahan, Ashok A. Deniz, Taekjip Ha, Daniel S. Chemla, Peter G. Schultz and Shimon Weiss. *Ratiometric measurement and identification of single diffusing molecules*. Chemical Physics 247 (1999). 85-106

[51] https://www.life technologies.com/mx/es/home/references/molecular-probes-the-handbook/tables/fluorescence-quantum-yields-and-lifetimes-for-alexa-fluor-dyes.html

[52] Lee NK, Kapanidis AN, Wang Y, Michalet X, Mukhopadhyay J, Ebright RH, Weiss S., *Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation*. Biophys J. **88**(4):2939-53, (2005).

[53] James J. McCann, Ucheor B. Choi, Liqiang Zheng, Keith Weninger, and Mark E. Bowen. *Optimizing Methods to Recover Absolute FRET Efficiency from Immobilized Single Molecules*. Biophys. J. **99**, 961-970, (2010).

[54] Hinterdorfer, Peter; Oijen, Antoine, *Handbook of Single-Molecule Biophysics* pag 40-55, Springer (2009).

[55] Joachim R. Fries, Leif Brand, Christian Eggeling, Malte Köllner, and Claus A. M. Seidel, *Quantitative identification of differentsingle molecules by selective time-resolved confocalfluorescence Spectroscopy*. J. Phys. Chem. A, (1998), **102**, 6601-6613.

[56] Irina V. Gopich, and Attila Szabo. *Single-Molecule FRET with Diffusion and Conformational Dynamics*. J. Phys. Chem. B **111**, 12925-12932, (2007).

[57] Clegg, R. M.; Murchie, A. I. H.; Zechel, A.; Carlberg, C.; Diekmann, S.; Lilley, D. M. J. Biochemistry 1992, 31, 4846-4856.

[58] Robert B. Best, Kusai A. Merchant, Irina V. Gopich, Benjamin Schuler, Ad Bax and William A. Eaton. *Effect of flexibility and cis residues in single-molecule FRET studies of polyproline*. PNAS 104(48), 18964-18969, (2007).

[59] Eyal Nir, Xavier Michalet, Kambiz M. Hamadani, Ted A. Laurence, Daniel Neuhauser, Yevgeniy Kovchegov, and Shimon Weiss. *Shot-Noise Limited Single-Molecule FRET Histograms: Comparison between Theory and Experiment*. J Phys Chem B. Nov 9, 2006; 110(44): 22103-22124.

[60] Ivan Rasnik, Sean A. McKinney and Taekjip Ha. (2005). *Surfaces and Orientations: Much to FRET about?*. Acc. Chem. Res., 38 (7), 542–548.

[61] Van Der Meer, B. W., Coker, G., III, and Chen, S. Y. S. (1994). *Resonance Energy Transfer: Theory and Data* (VCH, New York).

[62] Lee, N. K.; Kapanidis, A. N.; Wang, Y.; Michalet, X.; Mukhopadhyay, J.; Ebright, R. H.; Weiss, S. *Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation* Biophys. J. 2005, 88, 2939.

[63] Sabanayagam, C.R., J.S. Eid and A. Meller (2005) *Using fluorescence resonance energy transfer to measure distances along individual DNA molecules: Corrections due to nonideal transfer*. J. Chem. Phys. 122, 061103

[64] Simon Sindbert, Stanislav Kalinin, Hien Nguyen, Andrea Kienzler, Lilia Clima, Willi Bannwarth, Bettina Appel, Sabine Müller, and Claus A. M. Seidel. *Accurate Distance Determination of Nucleic Acids via Forster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity*. J. Am. Chem. Soc. 2011, 133, 2463-2480.

[65] Benjamin Schuler, Everett A. Lipman, Peter J. Steinbach, Michael Kumke, and William A. Eaton. *Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence*. PNAS 102(8), 2754-2759. (2005).

[66] Watson J.D. and Crick F.H.C. (1953). *A Structure for Deoxyribose Nucleic Acid*. Nature 171 (4356): 737-738.

[67] Hagerman P J., (1988) Annu Rev Biophys Biophys Chem 17:265-286.

[68] Brinkers S, Dietrich HR, de Groote FH, Young IT, Rieger B. *The persistence length of double stranded DNA determined using dark field tethered particle motion*. J Chem Phys. 2009, 130(21):215105.

[69] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. (2nd ed.), 5.40-5.43. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

[70] Irina V. Gopich, and Attila Szabo. *Single-Molecule FRET with Diffusion and Conformational Dynamics*. J. Phys. Chem. B **111**, 12925-12932, (2007).

[71] Stellwagen, N. C., Magnusdottir, S., Gelfi, C. and Righetti, P. G. (2001). *Measuring the translational diffusion coefficients of small DNA molecules by capillary electrophoresis*. Biopolymers, 58: 390-397.

[72] Dovichi, N. J., and Chen, D. D. In *Single Molecule Optical Detection, Imaging and Spectroscopy*, Basche, T., Moerner, W. E., Orrit, M., Wild, U. P., Eds.; Wiley: New York, (1997); P223.

[73] Eyal Nir, Xavier Michalet, Kambiz M. Hamadani, Ted A. Laurence, Daniel Neuhauser, Yevgeniy Kovchegov, and Shimon Weiss. *Shot-Noise Limited Single-Molecule FRET Histograms: Comparison between Theory and Experiments*. J Phys Chem B. Nov 9, 2006; 110(44): 22103-22124.

[74] Rahul Roy, Sungchul Hohng and Taekjip Ha. *A practical guide to single-molecule FRET*. Nature Methods 5, 507-516 (2008).

[75] Stanislav Kalinin, Evangelos Sisamakis, Steven W. Magennis, Suren Felekyan and Claus A. M. Seidel. *On the Origin of Broadening of Single-Molecule FRET Efficiency Distributions beyond Shot Noise Limits*. J. Phys. Chem. B, 2010, 114 (18), 6197-6206

[76] Adam Muschielok, Joanna Andrecka, Anass Jawhari, Florian Brückner Patrick Cramer and Jens Michaelis. *A nano-positioning system for macromolecular structural analysis*. Nature Methods 5, 965-971 (2008)

[77] R. F. Gesteland, T. R. Cech, J. F. Atkins, *The RNA World* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 2, 1999).

[78] Peter A. T., Joseph L. D., James E. W., and Ronald D. V., *Plasma Membrane Compartmentalization in Yeast by Messengers RNA Transport and a Septin Diffusion Barrier*. Science 290, 5490-341.

[79] Moore PB, Steitz TA. 2010. *The roles of RNA in the synthesis of protein*. Cold Spring Harb Perspect Biol 2: a003780.

[80] Alexander Serganov and Dinshaw J. Patel, Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. Nature Reviews Genetics **8**, 776-790, (2007).

[81] Elizabeth A. Doherty and Jennifer A. Doudna, *Ribozyme Structures and Mechanism*. Annu. Rev. Biophys. Biomol. Struct. 2001. **30**:457-75

[82] Hashim M Al-Hashimi and Nils G Walter, *RNA dynamics: it is about time*. Current Opinion in Structural Biology 2008, **18**:321-329.

[83] Lin He and Gregory J. Hannon. *MicroRNAs: small RNAs with a big role in gene regulation*. Nature Reviews Genetics 5, 522-531 (July 2004)

[84] Miranda-Rios, J., Navarror, M. and Soberon, M. *A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria*. (2001) Proc. Natl. Acad. Sci. USA 98(17), 9736-9741

[85] M. Sprinzl, C. Horn, M. Brown, A. Ioudovitch, S. Steinberg. *Compilation of tRNA sequences and sequences of tRNA genes*. Nucl. Acids Res., 26 (1998), 148-153.

[86] R.R. Gutell, J.J. Cannone, Z. Shang, Y. Du, M.J. Serra. *A story: unpaired adenosine bases in ribosomal RNA*. J. Mol. Biol., 304 (2000), 335-354.

[87] B.L. Maidak, J.R. Cole, T.G. Lilburn, C.T. Parker Jr, P.R. Saxman, R.J. Farris *et al. The RDP-II (ribosomal database project)*. Nucl. Acids Res., 29 (2001), 173-174.

[88] Y. Van de Peer, J. De Rijk, J. Wuyts, T. Winkelmans, R. De Wachter. *The European small subunit ribosomal RNA database*. Nucl. Acids Res., 28 (2000), 175-176.

[89] M. Szymanski, M.Z. Barciszewska, J. Barciszewski, V.A. Erdmann. *5S ribosomal RNA database Y2K*. Nucl. Acids Res., 28 (2000), 166-167.

[90] J.P. Wuyts, P. De Rijk, Y. Van de Peer, T. Winkelmans, R. De Wachter *The European large subunit ribosomal RNA database*. Nucl. Acids Res., 29 (2001), 175-177.

[91] Fontana W., Konings D. A. M., Stadler P. F., and Schuster P. (1993) *Statistics of RNA secondary structures*. Biopolymers, 33 1389-1404.

[92] Hofacker IL, Schuster P, Stadler PF. *Combinatorics of RNA secondary structures*. Discr. Appl. Math. 1998;88:207-237.

[93] Clote P, Kranakis E, Krizanc D, Stacho L. *Asymptotic expected number of base pairs in optimal secondary structure for random RNA using the Nussinov-Jacobson energy model*. Discr. Appl. Math. 2007;155:759-787.

[94] de Gennes PG. *Statistics of branching and hairpin helices for the dAT copolymer*. Biopolymers. 1968;6:715-729.

[95] M. Zuker. *Mfold web server for nucleic acid folding and hybridization prediction*. Nucleic Acids Res. 31 (13), 3406-3415, 2003.

[96] Ivo L. Hofacker. *Vienna RNA secondary structure server*. Nucl. Acids Res. 31: 3429-3431. (2003).

[97] Aron M. Yoffe, Peter Prinsen, William M. Gelbart and Avinoam Ben-Shaul. *The ends of a large RNA molecule are necessary close*. Nucleic Acid Res. 2011, 39(1): 292-299.

[98] Joel D. Richter and Nahum Sonenberg. *Regulation of cap-dependent translation by eIF4E inhibitory proteins*. Nature 433, 477-479. (2005).

[99] von der Haar, T., Gross, J.D., Wagner, G., McCarthy, J.E., 2004. *The mRNA cap-binding protein eIF4E in post-transcriptional gene expression*. Nat. Struct. Mol. Biol. 11, 503-511.

[100] Gallie, D.R. *The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency*. Genes Dev. 5 (11), 210-2116 (1991).

[101] Elizabeth L. Pettit Kneller, Aurélie M. Rakotondrafara, W. Allen Miller. *Cap-independent translation of plant viral RNAs Virus*. Research 119 (2006) 63-75.

[102] Dianna Edgil, and Eva Harris. *End-to-end communication in the modulation of translation by mammalian RNA viruses*. Virus Research 119 (2006) 43-51.

[103] Corver, J., Lenches, E., Smith, K., Robison, R.A., Sando, T., Strauss, E.G., Strauss, J.H.: *Fine mapping of a cis-acting sequence element in yellow fever virus RNA that is required for RNA replication and cyclization.* J. Virol. 77 (3), 2265-2270 (2003).

[104] Hsu, M.T., Parvin, J.D., Gupta, S., Krystal, M., Palese, P. *Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle.* Proc. Natl. Acad. Sci. U.S.A. 84 (22), 8140-8144 (1987)

[105] Diego E. Alvarez, María F. Lodeiro, Silvio J. Ludueña, Lía I. Pietrasanta, and Andrea V. Gamarnik. *Long-Range RNA-RNA Interactions Circularize the Dengue Virus Genome* J. Virol. 2005, 79(11):6631-6643.

[106] Marc R. Fabian and K. Andrew White.*5'-3' RNA-RNA Interaction Facilitates Cap- and Poly(A) Tail-independent Translation of Tomato Bushy Stunt Virus mRNA: A POTENTIAL COMMON MECHANISM FOR TOMBUSVIRIDAE.* J. Biol. Chem. 2004, 279:28862-28872.

[107] Peter Clote, Yann Ponty, and Jean-Marc Steyaert. *Expected distance between terminal nucleotides of RNA secondary structures.* J. Math. Biol. (2012) **65**:581-599

[108] Li Tai Fang. *The end-to-end distance of RNA as a randomly self-paired polymer.* Journal of Theo. Biology. 280(1); 101-107. (2011).

[109] Hillary S. W. Han, and Christian M. Reidys. *The 5'-3' Distance of RNA Secondary Structures.* Journal of Computational Biology. July 2012, 19(7): 867-878.

[110] John M. Harris, Jeffry L. Hirst, and Michael J. Mossinghoff. *Combinatorics and Graph Theory* Sringer-Verlag New York. (2000)

[111] Mills J. B., Vacano E, Hagerman P. J. *Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA).* J. Mol. Biol. 1999 Jan 8; 285(1):245-57.

[112] Cadena, R. D., Comas-Garcia, M., Rees, F. G., Rao, A. L. N., Knobler, C. M. and Gelbart, W. M. *Self-Assembly of Viral Capsid Protein and RNA Molecules of Different Sizes: Requirement for a Specific High Protein/RNA Mass Ratio*, J. Virol., **86**, 3318-3326, (2012).

[113] The Genome Portal of the Department of Energy Joint Genome Institute: Nucleic Acids Res 2011 0: gkr947v1-gkr947. or http://genome.jgi-psf.org

[114] http://www.ncbi.nlm.nih.gov/pubmed.

[115] Seidl,V., Huemer,B., Seiboth,B. and Kubicek,C.P., *A complete survey of Trichoderma chitinases reveals three distinct subgroups of family 18 chitinases.* FEBS J. **272**(22), 5923-5939 (2005).

[116] Berrocal-Tito,G., Sametz-Baron,L., Eichenberg,K., Horwitz,B.A. and Herrera-Estrella, A, *Rapid blue light regulation of a Trichoderma harzianum photolyase gene.* J. Biol. Chem. **274**(20), 14288-14294 (1999).

[117] Farahnaz Movahedzadeh, Susana González Rico, and Robert A. Cox. *In Vitro Transcription and Translation. E. coli* Plasmid Vectors. Methods in Molecular Biology, Volume 235, 2003, pp 247-255

[118] R. F. Gesteland, T. R. Cech, J. F. Atkins, The RNA World (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 2, 1999).

[119] Alexander Serganov and Dinshaw J. Patel, Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. Nature Reviews Genetics **8**, 776-790, (2007).

[120] Elizabeth A. Doherty and Jennifer A. Doudna, Ribozyme Structures and Mechanism. Annu. Rev. Biophys. Biomol. Struct. 2001. **30**:457–75

[121] Hashim M Al-Hashimi and Nils G Walter, *RNA dynamics: it is about time.* Current Opinion in Structural Biology 2008, **18**:321-329.

[122] Peter Clote, Yann Ponty, and Jean-Marc Steyaert. *Expected distance between terminal nucleotides of RNA secondary structures.* J. Math. Biol. (2012) **65**:581-599 I. V. Grigoriev, H. Nordberg, I. Shabalov, A. Aerts, M. Cantor, D. Goodstein, A. Kuo, S. Minovitsky, R. Nikitin, R. A. Ohm, R. Otillar, A. Poliakov, I. Ratnere, R. Riley, T. Smirnova, D. Rokhsar, and I. Dubchak. Nucleic Acids Res (2011)

[123] Dennis, C. and Webster J. 1971. Antagonistic properties of species group of Trichoderma. 1. Production of non-volatile antibiotics. Transactions of the British Mycological Society, **57**: 25-39.

[124] K. C. Wiese, E. Glen, and A. Vasudevan. *jViz.Rna 2.0 java tool for RNA secondary structure visualization.* IEEE Transactions on NanoBioscience, **4**(3), 212-218, (2005).

[125] Thirumalai D., and Ha B. Y. *Theoretical and Mathematical Models in Polymer Research*, Grosberg, A. (Academia, New York), pp. 1-35, (1998).

[126] Ding Y., Lawrence C. E. *A statistical sampling algorithm for RNA secondary structure prediction.* Nucleic Acid Res., **31**, 7280-7301, (2003).

[127] Brocchieri L. and Karlin S. *Protein length in eukaryotic and prokariotic proteomes.* Nucleic Acid Res., **33**, 3390-3400, (2005).

[128] Yang X., Tschaplinski T. J., Hurst B. G., Jawdy S., Abraham P. E., Lankford P.K., Adams M. R., Shah M. B., Hettich R. L., Lindquist E., *et al. Discovery and annotation of small proteins using genomics, proteomics, and computational approaches.* Genome Research, **21**, 634-641, (2011).

[129] Lee F. and Yanofsky C. *Transcription termination at the trp operon attenuators of Escherichia coli and Salmonella typhimurium: RNA secondary structure and regulation of termination.* Proc. Natl. Acad. Sci. U.S.A., **74**, 4365-4369, (1977).

[130] Solnick D. *Alternative Splicing Caused by RNA Secondary Structure.* Cell, **43**, 167-676, (1985).

[131] Fulton A. and Williams B. *Titin, a huge, elastic sarcomeric protein with a probable role in morphogenesis.* BioEssays, **13**, 157-161, (1991).

[132] Yang Lu, Robert J. Turner, and Robert L. Switzer. *Function of RNA secondary structures in transcriptional attenuation of the Bacillus subtilis pyr operon.* PNAS, **93**(25), 14462-14467, (1996).

[133] Astrid R. W. Schröder, Tilman Baumstark and Detlev Riesner. *Chemical mapping of co-existing RNA structures.* Nucl. Acids Res. **26**(14) 3449-3450, (1998).

[134] Altuvia S., Kornitzer D., Teff D., and Oppenheim A. B. *Alternative mRNA structures of the cIII gene of bacteriophage lambda determine the rate of its translation initiation.* J. Mol. Biol. **210**(2), 265-280, (1989).

[135] Biebricher, C. K., and R. Luce. *In vitro recombination and terminal elongation of RNA by Qβ replicase.* EMBO J. **11**, 5129-5135, (1992).

[136] Schultes, E. A. and Bartel, D. P. *One sequence, two ribozymes: implications for the emergence of new ribozyme folds*. Science, **289**, 448-452, (2000).

[137] Harry F. Noller. *RNA Structure: Reading the Ribosome*. Science, **309**, 1508-1514, (2005).

[138] Truus E. M. Abbink, Marcel Ooms, P. C. Joost H., and Ben Berkhout, *The HIV-1 Leader RNA Conformational Switch Regulates RNA Dimerization but Does Not Regulate mRNA Translation*. Biochemistry, **44**, 9058-9066, (2005).

[139] Evgeny Nudler and Alexander S. Mironov. *The riboswitch control of bacterial metabolism*. TRENDS in Biochemical Sciences, **29**, 11-17, (2004).

[140] Tina M. Henkin. *Riboswitch RNAs: using RNA to sense cellular metabolism*. Genes and Dev. **22**, 3383-3390, (2008).

[141] PLoss, M. Schmitz, G. Steger and D. Riesner. *Formation of a thermodynamically metastable structure containing hairpin II is critical for infectivity of potato spindle tuber viroid RNA*. The EMBO Journal **10**, 719-727, (1991).

[142] Mathews D. H., Turner D. H. *Prediction of RNA secondary structure by free energy minimization*. Curr. Opin. Struct. Biol. **16**, 270-278, (2006).

[143] Paul P. Gardner and Robert Giegerich. *A comprehensive comparison of comparative RNA structure prediction approaches*. BMC Bioinformatics, **5**, 140-151, (2005).

[144] Michael F. Bardaro Jr. and Gabriele Varani. *Examining the relationship between RNA function and motion using nuclear magnetic resonance*. Wiley Interdisciplinary Reviews, RNA. **3**(1), 122-132, (2012).

[145] Boris Furtig, Philipp Wenter, Luc Reymond, Christian Richter, Stefan Pitsch, and Harald Schwalbe. *Conformational Dynamics of Bistable RNAs Studied by Time-Resolved NMR Spectroscopy*. JACS, **129**, 16222-16229, (2007).

[146] Gregory B., and X. Zhuang. *Single-Molecule RNA Folding*. Acc. Chem. Res., **38**, 566-573, (2005).

[147] Xiaowei Zhuang, Harold Kim, Miguel J. B. Pereira, Hazen P. Babcock, Nils G. Walter, and Steven Chu. *Correlating Structural Dynamics and Function in Single Ribozyme Molecules*. Science **296**, 1473-1478, (2002).

[148] Wilson T. J., Nahas M., Araki L., Harusawa S., Ha T., Lilley D. M. *RNA folding and the origins of catalytic activity in the hairpin ribozyme*. Blood Cells Mol. Dis. **38**(1), 8-14 (2007).

[149] Andrei Yu. Kobitski, Alexander Nierth Mark Helm Andres Jaschke and G. Ulrich Nienhaus. *Mg^{2+}-dependent folding of a Diels-Alderase ribozyme probed by single-molecule FRET analysis*. Nucleic Acids Research, **35**, 2047-2059, (2007).

[150] Bernie D. Sattin, Wei Zhao, Kevin Travers, Steven Chu, and Daniel Herschlag. *Direct Measurement of Tertiary Contact Cooperativity in RNA Folding*. J. Am. Chem. Soc. **130**, 6085-6087, (2008).

[151] Mark Helm. *Post-transcriptional nucleotide modification and alternative folding of RNA*. Nucleic Acids Research, **34**(2), 721-733, (2006).

[152] Domingo E., and Holland J. J. *RNA Virus Mutations and Fitness For Survival*. Annu. Rev. Microbiol., **51**, 15178-15185, (1997).

[153] Ellen M. Westerhout, Marcel Ooms, Monique Vink, Atze T. Das and Ben Berkhout. *HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome*. Nucleic Acids Research, **33**, 796-804, (2005).

[154] Daniel Boden, Oliver Pusch, Frederick Lee, Lynne Tucker, and Bharat Ramratnam. *Human Immunodeficiency Virus Type 1 Escape from RNA Interference*. J. Virol., **77** 11531-11535, (2003).

[155] Ellen M. Westerhout and Ben Berkhout. *A systematic analysis of the effect of target RNA structure on RNA interference*. Nucleic Acids Research, **35**, 4322-4330, (2007).

[156] Ana Maria Contreras, Yoichi Hiasa, Wenping He, Adam Terella, Emmett V. Schmidt and Raymond T. Chung. *Viral RNA Mutations Are Region Specific and Increased by Ribavirin in a Full-Length Hepatitis C Virus Replication System*. J. Virol. **76**(17), 8505-8517, (2002).

[157] Sanghwa Lee, Jinwoo Lee, and Sungchul Hohng. *Single-Molecule Three-Color FRET with Both Negligible Spectral Overlap and Long Observation Time*. PloS ONE **5**(8), e12270 (2012).

[158] Wang Y, Juranek S, Li H, Sheng G, Tuschl T, Patel DJ. *Structure of an Argonaut Silencing Complex with a Seed-containing Guide DNA and Target RNA Duplex*. Nature **456**, 921-926, (2008).

[159] Gregor G. Webera, Jens Kortmannb, Franz Narberhausb, and Karl E. Klosea. *RNA thermometer controls temperature-dependent virulence factor expression in Vibrio cholerae*. PNAS, **111**(39), 14241-14246, (2014).

[160] Francesco Righetti and Franz Narberhaus. *How to Find RNA Thermometers*. Front Cell Infect Microbiol. **4** 132 (2014).