Real-time unfolding of DNA G-quadruplexes by helicases and polymerases

Samar Hodeib

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à L'école normale supérieure

Real-time unfolding of DNA G-quadruplexes by helicases and polymerases

Résolution des G-quadruplexes d'ADN en temps réel par les hélicases et les polymérases

Ecole doctorale n°564

PHYSIQUE EN ÎLE-DE-FRANCE

Spécialité PHYSIQUE

COMPOSITION DU JURY :

Mme CRIBIER Sophie
UPMC, présidente du jury
Mme SAINTOME Carole
UPMC, Rapporteur
M. MERGNY Jean-Louis
IECB, Rapporteur
Mme ALBERTI Patrizia
MNHN, Membre du jury
M. NICOLAS Alain
Institut Curie, Membre du jury
M. BOULE Jean-Baptiste
MNHN, invité
M. CROQUETTE Vincent
ENS, Directeur de thèse
M. TANNER Kyle
IBPC, Codirecteur de thèse

Soutenue par Samar HODEIB
le 28 juin 2017

Dirigée par **M. Vincent CROQUETTE**
M. Kyle TANNER

ENS ÉCOLE NORMALE SUPÉRIEURE | PSL RESEARCH UNIVERSITY PARIS

IBPC

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my principal advisor professor Vincent Croquette for the valuable opportunity to perform my ph.D study in his group. I am sincerely thankful to him for his continuous guidance, patience, availability, positive attitude and motivation during the last years. It is a pleasure to work with a person who has a huge scientific knowledge, a lot of inspiring ideas and precious "human qualities". And it is an honor for me to be one of his ph.D students.

Next, I would like to thank my second advisor professor Kyle Tanner and his wife Josette Banroques for their support, advices and kindness.

I would like to thank all my committee members: Mme Carole Saintomé, M. Jean-Louis Mergny, Mme Patrizia Alberti, M. Alain Nicolas, Mme Sophie Cribier et M. Jean-Baptiste Boulé.

I would like to acknowledge our collaborators: Michelle Spiering, Kevin Raney, Shubeena Chib and Eric Johansson.

I would like to thank professor David Bensimon, the second leader of the group, for his kindness and for the opportunity he provided me to join his team. Besides, I would like to thank the permanent members of the group. A special thanks to Jean-François Allemand for the musical atmosphere in the lab, and for all the Christmas lunches during the last four years, and to Nicolas Desprat for the valuable advices for my presentations.

I would also like to thank the current and former members of the team for their support and for this lovely atmosphere in the lab Saurabh, Fatima, François-Xavier, Thibault, Elena, Elise, Marie-Cécilia, Romain, Gaël, Hugo, Caroline and Bertrand.

I am also grateful for the members of Depixus especially to Jimmy Ouellet who taught me how to make the DNA constructions. A special thanks to Laurène and Gordon for their kindness.

I would like to thank Annie Ribaudeau, Nora Sadaoui, Fabienne Renia, Benoît Paulet and Marie Gefflot.

I would also like to acknowledge Thierry Bizebard who provided me with the necessary training during my graduate studies, and for his encouragement and kindness.

I would like to extend my gratitude to all my wonderful friends Elsa, Marilyne, Layal, Hala, Elise, Hadia, Mathilde, Racha, Serine, Kaoula, Fida et Rim.

My deepest thanks go to my family for your sacrifices and for standing beside of me through my tough times. For my loving parents and brother who supported me, for my loving, encouraging, and patient husband Mahmoud who didn't let me give up, and for my precious daughter Yara who also had to be very patient, thank you.

Table of contents

Acknowledgments ..... **i**

Table of contents **iii**

List of figures **viii**

List of key abreviations **xi**

Introduction **1**

CHAPTER I: The DNA, Structure and mechanics **5**

I.1 DNA History 5

I.2 DNA structure 6

I.2.1 DNA components 6

I.2.2 Primary structure or sequence 8

I.2.3 DNA secondary structure 8

I.2.3.1 B-DNA 9

I.2.3.2 Melting temperature T_m 11

I.2.3.3 Base pair Mismatch 11

I.2.3.4 Non-canonical secondary structures 11

a. A-DNA 11

b. Z-DNA 11

c. Cruciforms 11

d. Holliday junctions 12

e. Triple helices 13

f. H-DNA 13

g. G quadruplexes 13
I.2.4 DNA tertiary structure 15
I.2.4.1 Topological parameters 15
a. Contour length: 15
b. Persistence length 15
c. Linking number LK 15
d. Twist 16
e. Writhe 16
I.2.4.2 DNA storage in chromosomes 16
I.3 DNA mechanics 18
I.3.1 dsDNA extension model 18
I.3.1.1 Free jointed chain model 18
I.3.1.2 Worm like chain model 19
I.3.2 ssDNA elasticity 20
I.3.3 Unzipping a DNA double helix 20
I.3.4 DNA supercoiling 21
CHAPTER II: Overview of the DNA replication 23
II.1 Proteins and Enzymes 23
II.2 Proteins structure 24
II.2.1 The primary structure 24
II.2.2 The secondary structure 24
II.2.3 The tertiary structure 24
II.2.4 The quaternary structure 25
II.3 Gene expression 26
II.4 DNA replication 27
II.4.1 Helicase 27
II.4.2 Polymerase 28
II.4.3 Telomere replication 29
CHAPTER III: DNA micromanipulation using Magnetic tweezers 31
III.1 Introduction 31
III.1.1 Atomic Force Microscopy AFM 31
III.1.2 Optical microfiber 32
III.1.3 Optical tweezers 32

III.1.4 Magnetic tweezers setup 32
III.1.4.1 DNA molecules preparation 33
III.1.4.2 Fluid cell assembly 34
III.1.4.3 Bead tracking and molecule length 35
III.1.4.4 Force calibration 35
III.2 Enzymatic activity study using Magnetic tweezers 39

CHAPTER IV: Thesis project 45

IV.1 G-quadruplexes 45
IV.1.1 C-MYC G4 46
IV.1.2 Human telomeric G4 47
IV.2 This work 48
IV.3 G4 ligands 48
IV.3.1 Phen-DC3 ligand 49
IV.4 G4s kinetics and stability dependence on cation type 50
IV.5 G4 and helicases 51
IV.5.1 Pif1 helicase and G4s 51
IV.5.2 RecQ helicase and G4s 52
IV.6 G4 and Polymerases 52
IV.7 G4 and proteins 53
IV.7.1 RPA and G4s 53
IV.7.2 Sub1 and PC4 proteins 53

CHAPTER V: Results and discussion 55

V.1 Folding a G4 in a double-stranded DNA 55
V.1.1 Simple unzipping and re-zipping of DNA hairpins 55
V.1.2 Unzipping and re-zipping the DNA hairpin in presence of an oligonucleotide complementary to a segment on the DNA hairpin 56
V.1.3 Folding the c-MYC (14, 23) G4 embedded in the DNA hairpin in a K+ buffer 58
V.2 C-MYC (14, 23) G4 kinetics 62
V.2.1 C-MYC (14, 23) G4 folding rate 62
V.2.2 C-MYC (14, 23) G4 Stability 63
V.2.3 Unfolding the c-MYC (14, 23) G4 by applying an external force 64
V.2.4 C-MYC (14, 23) G4 encircling 66
V.2.5 The G4 stability depends on cation type 66

V.2.5.1 Effect of Na^+ on the c-MYC (14, 23) G4 folding 66
V.2.5.2 Effect of Na^+ on the c-MYC (14, 23) G4 stability 68
V.2.6 Discussion of the c-MYC (14, 23) G4 Kinetics and stability 69
V.3 Folding other G4 sequences in the hairpin 73
V.4 G4 stability 75
V.4.1 Stability of different G4s sequences with and without phen-DC3 in K+ buffer 75
V.4.2 Increasing the G quadruplex stability 77
V.4.3 Promoting the G-quadruplex folding 78
V.5 Does the c-MYC (14, 23) G quadruplex form a roadblock on the enzymes' path? 80
V.5.1 Bacteriophage T4 replisome 81
V.5.1.1 gp41 helicase 82
V.5.1.2 gp43 polymerase 86
V.5.1.3 gp41 and gp43 coupled 88
V.5.2 Pif1 helicase 91
V.5.2.1 The helicase is added on the hairpins that closing is blocked by the lagging strand G4 93
V.5.2.2 Pif1 heliacase is added to the hairpins having an encircled G4 that is located on the lagging strand in order to visualize its collision with the G4 95
V.5.2.3 G4 on the leading strand 98
V.5.2.4 Does the Pif1 helicase have an affinity for the G4 structure? 99
V.5.2.5 Does Pif1 resolve the c-MYC (14, 23) at a high concentration of K+? 100
V.5.2.6 Does Pif1 resolve the c-MYC (14, 23) in the presence of Phen-DC3 ligand? 100
V.5.2.7 Does Sub1 protein enhance the G4 stability against unwinding by Pif1? 100
V.5.3 RecQ helicase 101
V.5.3.1 Does RecQ unfold the c-MYC(14,23) G4? 101
V.5.3.2 Folding of G4s due to hairpin opening by helicases 102
V.5.4 Does RPA or SSB unfold the G4? 103
V.5.4.1 G4 folding in presence of single stranded proteins 103
V.5.4.2 Unfolding the G4 by the single-stranded proteins 103
V.6 Unfolding the G4 structure by polymerases 105
V.6.1 T7 Polymerase 106
V.6.2 Manta polymerase 106
V.6.3 Pol ε polymerase 107

Conclusion 111

CHAPTER VI: T4 replisome study using DNA rolling circle 117

VI.1 Rolling circle construction 117

VI.2 Rolling circle replication 118

VI.2.1 gp43 exo- polymerase 120

VI.2.2 Manta polymerase 120

VI.3 T4 coupled leading strand and lagging strand polymerases 121

VI.3.1 GP43 polymerase 122

VI.3.2 Light replisome (gp41 helicase and gp43 polymerase) 124

Résumé du travail de thèse en français……..……129

Annexe 1 ……...151

Annexe 2: ……...177

Annexe 3 ……...179

Annexe 4 ……...181

Annexe 5 ……...183

References ……...185

List of figures

Fig. I.1: Nucleic Acid building blocks (nucleotides). 7
Fig. I.2: Glycosic bonds and sugar conformation. 9
Fig. I.3: DNA double helix. 10
Fig. I.4: DNA major and minor groove.. 11
Fig. I.5: DNA secondary structures. 12
Fig. I.6: DNA triple helix. 13
Fig. I.7: G-quadruplexes. 14
Fig. I.8: DNA storage in chromosome. 16
Fig. I.9: A chromosome at different magnifications. 17
Fig. I.10: G-quadruplex structures at the 3' overhang of telomeres.. 18
Fig. I.11: FJC and WLC models.. 20
Fig. I.12: Unzipping a double stranded DNA helix. 21
Fig. I.13: DNA supercoiling. 22
Fig. II.1: Schematic representation of proteins.. 23
Fig. II.2: Protein structure. 25
Fig. II.3: Protein tertiary structure stabilizing interactions. 26
Fig. II.4: DNA replication or DNA synthesis in the process of copying a double-stranded DNA molecule. 28
Fig. III.1: The magnetic trap setup. 33
Fig. III.2: The DNA hairpin used in the current work. 34
Fig. III.3: Fluid chamber assembly.. 34
Fig. III.4: Calibration image. 35
Fig. III.5 : The forces acting on the magnetic bead in the magnetic tweezers setup. 36
Fig. III.6 : – Power spectrum density. 38
Fig. III.7: Schematics of the unzipping assay. 41
Fig. IV.1: G quadruplexes. 46

Fig. IV.2: c-Myc(14,23) G4 structure..... 47
Fig. IV.3: Ligand binding to G4..... 49
Fig. IV.4 : Phen-DC3 structure 49
Fig. V.1: Force-Extension curves..... 55
Fig. V.2: Unzipping and rezipping of the DNA hairpin in presence of an oligo. 57
Fig. V.3: A serie of opening – closing cycles of the hairpin having a G4 sequence on one strand before the folding of the G4 structure..... 58
Fig. V.4: The method used to fold the G4 in the hairpin. ..... **Erreur ! Signet non défini.**
Fig. V.5: Folding the G4 structure in the DNA hairpin using an oligonucleotide complementary to the hairpin loop.. 60
Fig. V.6: The hairpin extension curves before and after the G4 folding. 61
Fig. V.7: Folding the G4 in the DNA hairpin using a loop oligo..... 62
Fig. V.8: C-MYC (14, 23) G4 folding rates. 63
Fig. V.9: C-MYC (14, 23) G4 stability. 64
Fig. V.10: G4 unwinding by applying an external force..... 65
Fig. V.11: The histogram representing the number of G4 resolved at different forces..... 65
Fig. V.12: C-MYC (14, 23) G4 encircling rates. 66
Fig. V.13: Force - extension curves before and after the G4 folding cycles in 60 mM Na^+. 67
Fig. V.14: Unfolding a G4 structure by exchanging K+ by Na^+. 68
Fig. V.15: C-MYC (14, 23) folding and unfolding rates.. 72
Fig. V.16: hairpin having 2 consecutive telomeric G4s..... 74
Fig. V.17: Human telomeric G4 stability.. 76
Fig. V.18: G4C2 lifetimes in 150 mM K+ and in presence of 1 nM Phen-DC3.. 76
Fig. V.19 : G4 stabilizing agents. 77
Fig. V.20: G4 chaperone assay..... 78
Fig. V.21 : G4 folding at high force in the presence of Sub1 protein. 79
Fig. V.22: gp41 helicase (60 nM) unwinding a hairpin that does not have a G4 structure. 81
Fig. V.23: gp41 helicase encountering a G4.. 82
Fig. V.24: gp41 helicase encountering a G4 that is blocking the hairpin closure. 83
Fig. V.25: gp41 helicase opening a DNA hairpin having an embedded G4 structure..... 84
Fig. V.26: gp41 helicase opening a folded hairpin containing a G4 structure on its lagging strand that is encircled in the duplex..... 85
Fig. V.27: The gp43 polymerase encountering a G4 structure.. 86
Fig. V.28: gp43 WT polymerase replicating the G4. 87

Fig. V.29: T4 light replisome encountering a G4. The T4 light replisome formed by the helicase and the polymerase is tested on a hairpin having the G4 on the leading strand. 88
Fig. V.31: T4 replisome encountering a G4.. 89
Fig. V.31 : The T4 replisome replicating the G4 where the helicase and the polymerase are coupled. 90
Fig. V.32: Pif1 helicase unwinding a hairpin that does not have a G4. 92
Fig. V.33: Pif1 unwinding rate.. 93
Fig. V.34: Pif1 added to hairpins that are blocked by the lagging G4 structure. 93
Fig. V.35: Pif1 resolving the G4..... 94
Fig. V.36: Pif1 helicase jumps the G4. 95
Fig. V.37: Pif1 unwinding a hairpin that has an embedded G4 structure. 95
Fig. V.38: Pif1 resolving a G4 embedded in a hairpin. 96
Fig. V.39: Zoom on the Pif1 pause on the G4. 97
Fig. V.40: Pif1 pause duration on the c-MYC (14,23) G4. 98
Fig. V.41 : Pif1 added to hairpins having a G4 on their leading strand.. 99
Fig. V.42: G4 unwinding by Pif1 in the presence of Sub1..... 100
Fig. V.43: RecQ delta opening the hairpin and leading to the G4 folding. 102
Fig. V.44: RPA unzipping DNA lex in the absence of ATP. 104
Fig. V.45 : Pif1 resolving a G4 embedded in a hairpin in the presence of RPA. 105
Fig. V.46: T7 polymerase replicating c-MYC (14, 23) G4. 106
Fig. V.47 : The Pol ε exo- replicating a molecule that does not have a G4 structure..... 107
Fig. V.48: Polymerase Epsilon WT replicating a G4. 108
Fig. V.49: The Polymerase Epsilon replicating a G4 in a primer extension mode. 109
Fig. VI.1: Construction of a rolling circle molecule. in of the DNA molecule. 117
Fig. VI.2: Replication of a rolling circle DNA molecule by a polymerase.. 119
Fig. VI.3: gp43 exo- polymerase replicating the rolling circle DNA molecule. 120
Fig. VI.4: Manta polymerase replicating the rolling circle DNA molecule..... 121
Fig. VI.5 : A model of the T4 bacteriophage DNA replisome[148].. 122
Fig. VI.6 : Rolling circle replication with the coupled lagging and leading strands polymerases. a) ... 124
Fig. VI.7 : The rolling circle is replicated without coupling between the lagging and leading strands polymerases. 124
Fig. VI.8 : Rolling circle replication showing a coupling between the lagging and lesding strands polymerases.. 125
Fig. VI.9: Rolling circle replication with the coupled polymerases..... 125

List of key abreviations

ATP: Adenosine triphosphate

Bp: Base pair

BSA: Bovine Serum Albumin

dNTP: deoxynucleoside triphosphates

dsDNA: Double-stranded DNA

DTT: Dithiothreitol

EDTA: Ethylenediamine tetracetic acid

G4 DNA: G-quadruplex DNA

k_{off}: Unfolding rate

k_{on}: Folding rate

RPA: Replication protein A

SF: Superfamily

SSB: Single stranded DNA-binding protein

ssDNA: Single-stranded DNA

Introduction

Introduction

In an effort to understand what living organisms are made of, how parents transmit hereditable traits to offspring, and how do organisms age and fall sick, scientists have discovered an impressive microscopic world that is encapsulated in every micron-sized cell of our body. In the cell, the craft workers that intervene in the various processes dealing with the DNA masterpiece are the proteins and enzymes. Each one of them has a specific expertise and a function that is complementary to the others. The DNA that contains all the organism's genetic information is compacted in entities called chromosomes and stored in the nucleus of every cell of the organism.

Cell multiplication is a process that ensures the growth and development of a living organism, where a cell divides into two daughter cells having the same genetic patrimony. The genetic code conservation among the cells after so many divisions is due to the complementarity between the two strands of the DNA double helix where adenine binds to a thymine, and a guanine binds to a cytosine. So before a cell division occurs, the DNA must be replicated in order to generate from a double-stranded DNA helix two similar double-stranded DNA helix. The replication is a process achieved by the replisome that is an assembly of different proteins and enzymes. It consists mainly of a helicase that unzips the double helix and polymerases that replicate the two DNA strands among other proteins.

Moreover, when the organism signals the need to express a protein, the transcription machinery first intervenes to copy the recipe of the protein that is encoded in segments of DNA called genes. Then, it synthesizes the protein following the given instructions.

Furthermore, proteins intervene in the aging process that was shown to be linked to the erosion of the ends of chromosomes called telomeres. The overexpression of some proteins is linked to diseases. For instance, the telomerase is overexpressed in 80% of cancers. It elongates and replenishes the telomeres of malignant cells, the thing that continuously rejuvenates them and makes them immortal.

Most of the time, the DNA forms a double-stranded helix *in vivo*. Some biological processes, such as the replication and the transcription mentioned above, require the opening of the helix in order to access to the genetic code embedded in its vicinity. But when a guanine rich sequence is single-stranded, it can form a kind of DNA knot called G-quadruplex (G4). The G4 consists of three or more parallel quartets of four guanines each. The intra-quartet guanines are linked by hydrogen bonds and the quartets are coordinated by cations. These DNA knots are considered as roadblocks on the path of molecular motors. Their

occurrence can interrupt the replication and thus lead to genetic instabilities. If a G4 is present upstream of a gene, it can suppress the transcription of the gene and provoke diseases. The presence of putative G4 sequences in various biologically relevant regulatory regions in the human genome such as oncogenes and the telomeres, makes them interesting pharmacological targets for disease treatments. For instance, the folding of G4 in a proto-oncogene can silence gene expression, while the stabilization of those roadblocks on telomeres can inhibit the telomerase in cancerous cells. Until today, nearly 900 G4 stabilizing compounds have been developed, and some of them are used in medicine.

The G-quadruplexes have been extensively studied by different biochemical and biophysical methods, by both bulk and single molecule assays. Thermodynamics of some biologically relevant G4s as well as their folding and unfolding kinetics has been studied by different techniques. Furthermore, many papers have reported their stabilization or also their unfolding by some proteins and helicases. However in most of these studies authors do not have a direct insight on the folded G4 structure and thus the G4 might have been unfolded before the enzyme encountered it.

In this study we present a new method to study G-quadruplexes using magnetic tweezers. The designed DNA molecule mimics a G4 in a double-stranded region of the DNA such as those embedded in the promoters of the genes. The G4 folding gives a stable signal that ensures the G4 still folded during the entire assay. This method allows measuring the kinetics as well as the stability of a G4 structure using various buffer conditions. It also allows the detection of a G4 chaperone activity of some proteins. Moreover, it permits for the first time to get a real-time insight into the interaction between a molecular motor and a G4 roadblock encountered on its template, to visualize if it constitutes a real barricade and how long the enzyme will require to overcome such structures.

The manuscript is structured as follows:

The first part introduces the different players: DNA, G-quadruplexes, and proteins that are reported in two different chapters; the third chapter describes the single molecule magnetic tweezers technique used in this work.

In the second part, the thesis project is presented in the fourth chapter. It consists of studying G4 structures embedded in double-stranded DNA molecules and visualizing in real-time the effect of proteins and enzymes on the folding, the stability and the unfolding of such DNA structures. In the fifth chapter we explain, comment and discuss the results that we have obtained in our study. Finally, we present a general conclusion about the G4 study.

A sixth chapter presents another research topic about the lagging strand replication. We present the methods and the results that we have obtained. However, we did not get good statistics, which has not allowed leading the project to the end.

Part I
DNA & enzymes

CHAPTER I: The DNA, Structure and mechanics

I.1 DNA History

How do children inherit their parents' traits? The mystery of inheritance has persisted over centuries before being finally solved by the discovery of deoxyribonucleic acid, famously known as "DNA". This genetic material encodes the biological information for living organisms and permits its transfer from generation to generation. DNA studies have revolutionized medicine, it revealed mutations on the DNA that cause genetic diseases, and permitted to customize genetic drugs and therapies. For untreatable diseases, DNA analysis permitted an earlier diagnosis that could help avoiding some symptoms or reducing pain. Another important DNA application is crime scene investigations, where DNA analysis had become irrefutable evidence to identify the suspect as well as the victim and thus to resolve criminal cases. The DNA profiling or genetic fingerprinting is also used for parenting tests.

In the following paragraph, we present an overview about the most remarkable stages in the DNA discovery.

In 1865, the botanist Gregor Mendel, who was pursuing experiments on pea plants, concluded that the traits found in the offspring are determined as discrete factors inherited from the parents. Those factors will be later known by "genes", DNA segments that encode proteins that build the organism's tissues and organs among other functions. The genes carry the code that determines the traits of an organism. A gene could be involved in the determination of multiple traits, and a polygenic trait, such as eye color and skin color, could be affected by many genes. Mendel concluded that a feature is determined by two genes alternates, each inherited from a parent. A gene alternate could be dominant or recessive. The dominant is expressed in the new generation even if there is only one copy of it (i.e from only one parent). Contrarily, the recessive one, in order to be expressed, must be transmitted by both parents. The nature of genes, their localizations in the living organisms, and the way they

pass to the next generation was still unknown. In 1869, while the chemist Friedrich Miescher was trying to extract the proteins components in the nuclei of human white blood cells, he discovered a substance with a higher phosphorous content, that he called the nuclein. In 1919, the biochemist Phoebus Levene discovered that the nuclein is a deoxyribonucleic acid[1]. He found the components of DNA: a phosphate group, a deoxyribose sugar, and a nitrogenous base with basic qualities, all together form a DNA building block called nucleotide. Four types of nucleotides were discovered: Adenine (A), Thymine (T), Cytosine (C) and Guanine (G) that have different nitrogenous bases. Levene stated that the DNA is a polynucleotide chain that contains the same amount of each nucleotide. This latter finding will be later proven to be false. In 1944 Oswald Avery demonstrated that genes are composed of DNA[2]. But at the time, it had been thought that the DNA was the chromosome component but it was difficult to accept that it could hold biological information. In 1950 Erwin Chargaff found that the DNA of different species had the same properties but different nucleotides order[3], [4]. He also concluded that the amount of Thymine equals the amount of Adenine, while the amount of Cytosine equals the amount of Guanine.

In 1950, James Watson and Francis Crick, had been working on a 2D model of the DNA, and finally suggested that the DNA should have a 3D structure with the backbone located at its center and the bases pointing outward. Linus Pauling proposed another structure, that of a triple helix[5]. In 1951 Rosalind Franklin obtained the famous X-ray diffraction pattern of DNA, called the “Photograph 51”, that revealed the double helix structure of DNA[6]. It enabled her to know the helix diameter, the distance by turn, the distance between two consecutive nucleotides, and that the bases are located in the center of the DNA molecule. This finding paved the way for one of the most important discoveries of the twentieth century, the 3D structure of DNA in a right-handed double helix published by Watson and Crick in 1953. In fact, when Watson saw the photograph 51, which was not published at that time, he directly understood that it resulted from a helical structure. He pursued working with Crick on their model, improved it by putting the bases inside the molecule, and concluded without any DNA manipulation, that Adenine pairs with Thymine while Guanine pairs with Cytosine[7], [8]. The discovery of the correct DNA structure shed the light on different biological mechanisms in the cell, as well as the inheritance chemical basis.

While great progress has been made during the last decades, much more research is required to answer the remaining questions surrounding the DNA.

I.2 DNA structure

I.2.1 DNA components

A DNA strand is a long polymer chain composed of nucleotides. Each nucleotide itself consists of a nitrogenous base, a deoxyribose sugar (forming together a nucleoside) and a phosphate group negatively charged. The phosphate group attached to the 5’ sugar carbon of a nucleotide is linked to the hydroxyl group attached to the 3’ sugar carbon of the next

nucleotide by phosphodiester bonds (fig.I.3.b). Thus the nucleotides of a single strand of DNA are joined by the sugar phosphate backbone. There are four different nitrogenous bases and hence four different nucleotides. Purine bases are Adenine (A) and Guanine (G), they have two carbon cycles: a six membered and a five membered ring containing nitrogen. Pyrimidine bases, that have only one carbon cycle, are Thymine (T) and Cytosine (C) (fig.I.1).

Similarly, the ribonucleic acid RNA is composed of nucleotides that, instead of having deoxyribose sugar, include a ribose sugar. The RNA nucleobases are the same as DNA bases, the only difference is that RNA does not have a thymine but instead it has a uracil (U).

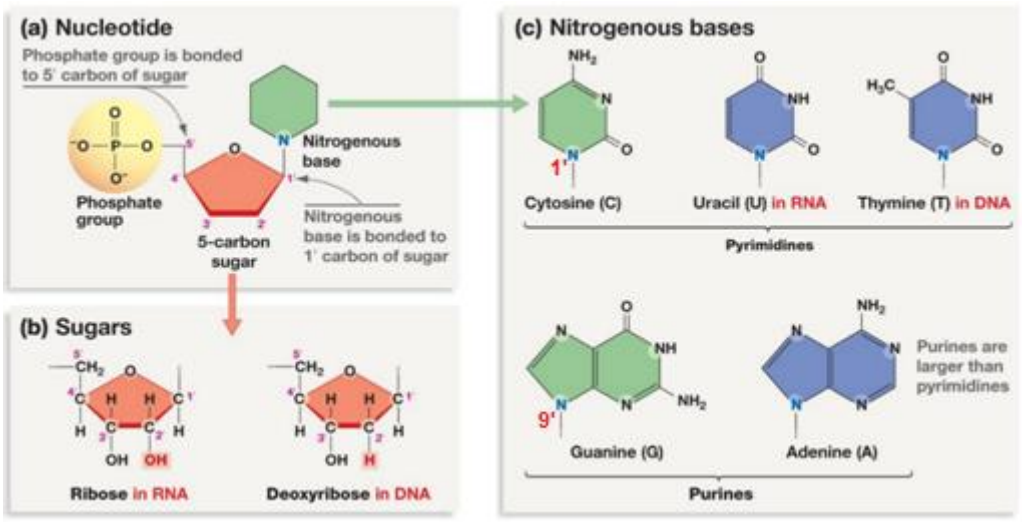


Fig. I.1: Nucleic Acid building blocks (nucleotides). ***a)*** *A nucleotide is the association of a nitrogenous base, a carbon sugar and a phosphate group. b) DNA nucleotides contain a deoxyribose sugar while RNA nucleotides contain a ribose sugar. The hydrogen on the 2'C of a deoxyribose is replaced by a hydroxyl group on the ribose. c) Four bases are possible for DNA: Purine bases (Adenine (A) and guanine (G)) and pyrimidine bases (Cytosine (C) and Thymine (T)). The first theree bases are also found in RNA, but instead of Thymine, RNA has a uracil (U). (www.sites.google.com)*

In a nucleotide (fig.I.1):

- The **ester bond** is the linkage between the triphosphate group oxygen and the 5' carbon of the sugar.

- The **glycosidic Bond** is the linkage between the 9' nitrogen of purine bases or 1' nitrogen of pyrimidine bases and the 1' carbon of the sugar. The rotation of the base around the glycosidic bond gives rise to an anti or a syn conformation. The nucleotide has an anti-conformation if the base is projecting away from the sugar.

I.2.2 Primary structure or sequence

The primary structure of the DNA, in other words the DNA sequence, is defined by the order of its nucleotides from 5' to 3' end; the same direction of DNA synthesis. The DNA sequence is usually written as the sequence of bases it contains. A change in the DNA sequence within a gene, even in one nucleotide could have a large effect on the organism. For example, different eyes colors come from different DNA sequences called allele i.e the possible sequence within a gene or the gene variant. In fact, the DNA permits building the organism tissues and organs as well as ensuring its biological functions by translating the genetic code contained in the genes into functional proteins. Along the DNA chain constituting the gene, every three nucleotides will finally contribute by one amino acid in the protein chain. Thus, a mutant base will lead to a different amino acid and thus a different protein that does not fulfill the required function and could cause diseases like cancer. For instance, the sickle cell anemia disease is caused by a genetic mutation where an adenine is replaced by a thymine. However, some amino acids could be encoded by different base sequences, thus a mutation that does not change the resulting amino acid leads to the synthesis of the normal protein and is called a silent mutation. Moreover, a mutation that occurs between genes, in other words in the noncoding region is also neutral. Finally, in some cases, the mutation could be beneficial and lead to the improvement of a biological function.

I.2.3 DNA secondary structure

In the cell, the DNA is double stranded most of the time. This is due to the binding between two antiparallel DNA strands by means of nitrogenous bases. This base pairing is not random, and only takes place between complementary bases i.e. an A binds to a T by means of two hydrogen bonds, and a G binds to a C by means of three hydrogen bonds. Although the DNA strands are negatively charged due to the phosphate groups, cations in physiologic buffer shield the electrostatic repulsion between them and allow their hybridization.

Due to the complementarity between the bases, the two strands hold the same information and allow its faithful transmission to the next generation. However, some non-Watson-Crick base pairing could occur. Those are called base pairing mismatches (see below).

The DNA secondary structure or the DNA folding pattern arises from the specific base pairing as well as from the base stacking inside of the DNA duplex. In fact, the DNA forms a 3D double helix where the two nucleotides chains are twisted around each other. The sugar phosphate backbone is on the outside of the helix and the bases point inward. Actually, the hydrophobic nature of the bases, as well as their flat form, makes them stack within the double stranded DNA. This structure is made possible due to the hydrophilic nature of the sugar phosphate backbone and its flexibility. Moreover the bases are tilted like the steps of spiral staircase preventinginsertion of water molecules.

Each base of the base pair is linked to the sugar of its phosphodiester backbone via a glycosidic bond. The two glycosidic bonds of the same base pair are not symmetric with respect to the short base pair axis, but the angle between them is around 120° from a side and

around 240° from the other side. This leads to the existence of two different sized grooves for the double helix, called major and minor grooves. The grooves twist around the DNA helix on opposite sides.

The sugar ring of nucleic acids is not planar, but some atoms are out of its plane due to steric reasons. Endo-pucker refers to the case where the C2' or C3' are out of the plane toward the O5'. Exo-pucker refers to a shift of C2' or C3' in the opposite direction. C2' endo and C3' exo are equivalent, likewise C2' exo and C3' endo are equivalent.

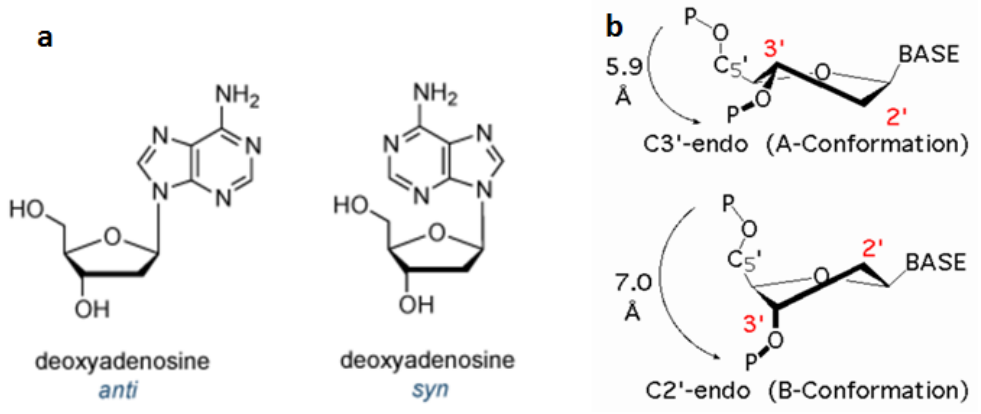


Fig. I.2: Glycosic bonds and sugar conformation. *(a) Glycosidic bond anti- and syn-conformations: The rotation of the base around the glycosidic bond determines its conformation. If the base is pointing toward to the sugar the glycosidic bond is in the syn-conformation. If it is pointing away from the sugar, the glycosidic bond is in the anti-conformation. (b) Sugar A- and B- conformation: The sugar of a nucleotide is not planar, the C2' and C3' are shifted in opposite directions, thus the one that is shifted toward the O5' is said endo-puckered. A-conformation refers to a C3' endo (also called C2' exo) and B-conformation to C2' endo (also called C3' exo).*

I.2.3.1 B-DNA

The predominant DNA structure found in nature is the canonical B-DNA form (fig.I.3.a) that is the preferred conformation for the DNA double helix at high humidity and low salinity. It consists of a right handed double helix of 2 nm in diameter (while the electrostatic diameter is 5nm), a distance of 0.34 nm between adjacent nucleotides and a rise of 3.4 nm per helix turn involving 10-10.5 bases per turn. The B-DNA helix has a narrow minor groove 12 nm wide and a wider major groove of 22 nm. The sugar pucker is C3' exo or C2' endo, and the bases are nearly perpendicular to the helix axis. However, DNA was discovered to be highly polymorphic and can adopt other non-canonical conformations.

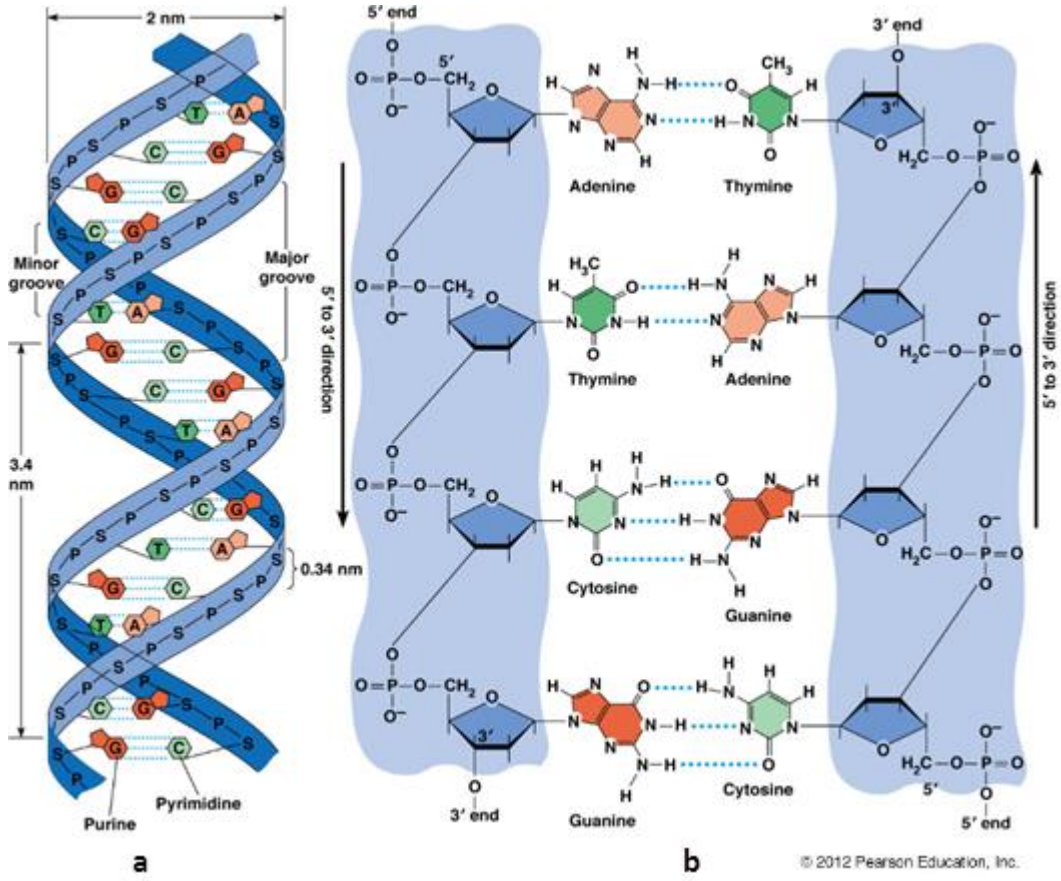


Fig. I.3: DNA double helix. ***a)*** *A B-form DNA that is a right-handed double helix formed by two anti-parallel strands. A DNA strand is a string of nucleotides that are joined via the sugar phosphate backbone. The bases are pointing to the interior of the double helix. The double helix has a diameter of 2 nm, a rise by Bp of 0.34 nm, and a major groove that is wider than the minor groove.* ***b)*** *An A is bound to a T via two hydrogen bonds while a G is bound to a C via three hydrogen bonds.*

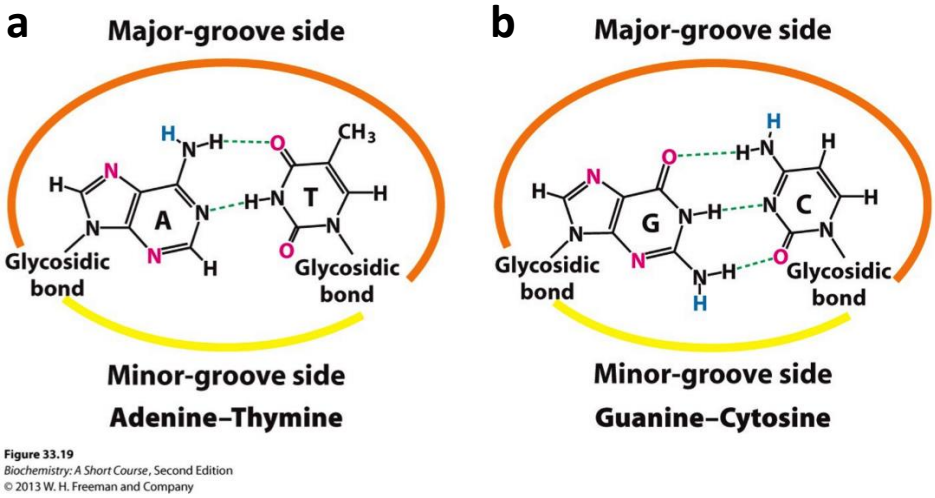


Figure 33.19
Biochemistry: A Short Course, Second Edition
© 2013 W. H. Freeman and Company

Fig. I.4: DNA major and minor groove. a) *An **A-T** base pair bound by two hydrogen bonds, and **b)** a **G-C** base pair bound by three hydrogen bonds. The glycosidic bonds linking the bases to the sugar phosphate cytoskeleton defines the major and the minor grooves.*

I.2.3.2 Melting temperature T_m

The DNA melting temperature Tm is the temperature needed to denature the DNA double helix. It is the temperature required to dissociate half of DNA duplexes into single strands by breaking the hydrogen bonds. Hence, the more stable the double stranded DNA, the higher the Tm (molecules having a big content of G-C base pairs, long DNA chains, less mismatch base pairing, high cations concentration in the surrounding buffer). The average DNA T_m is around 65°C.

I.2.3.3 Base pair Mismatch

During DNA replication, a strand is used as a template to synthesize the complementary strand. If a mistake is made and a wrong base is added, two types of mismatches could occur:

- Transition mismatch: When a purine pairs to the wrong pyrimidine (G-T and A-C baisepairs).
- Transversion mismatch: When a purine pairs to a purine or a pyrimidine to a pyrimidine.

I.2.3.4 Non-canonical secondary structures

a. A-DNA

A-DNA is also a widespread DNA conformation mostly adopted by DNA-RNA hybrids and purine DNA sequences. It consists of a right handed double helix that has a diameter of 2.3 nm, a distance of 0.26 nm between adjacent nucleotides, a rise of 2.8 nm per helix turn and 11 bases per turn. The minor groove of the A-DNA is wider than its major groove. The B-DNA is driven into an A-DNA form when proteins bind to it and shorten it, or under dehydrating conditions. The sugar pucker is C3’ endo, the bases are not perpendicular to the helix axis but inclined.

b. Z-DNA

The Z-DNA is a left-handed helix of 1.8 nm of diameter, an inter-base distance of 0.37 nm, and 12 bases per turn. This structure occurs in GC rich DNA stretches alternating pyrimidine-purine steps, and is promoted by DNA negative supercoiling.

c. Cruciforms

A DNA stretch can have direct repeated sequences; mirror repeated sequences, and inverted repeated sequences or palindromes. The first is a repeat of the sequence with the

same polarity, the second is a repeat with the opposite polarity (as if a mirror is placed between the two repeats) and the third is the complementary of the sequence (as if one has a mirror but this mirror returns for each base the image of its complementary base instead of his image). Inverted sequences are prone to form cruciforms under negative supercoiling. In order to form these non-canonical structures the DNA duplex, unzipping the DNA duplex is required. Then the sequence and its palindrom will bind to each other extruding a loop of unpaired bases, and this leads to a stem-loop structure.

d. Holliday junctions

It is a four way DNA junction that takes places as an intermediate structure during homologous genetic recombination. The homologous recombination happens when two homologous chromosomes each undergo a single strand break, and each incised strand invades the other DNA duplex leading to a four bridged Holliday junction. It was proposed by Robin Holliday in 1964. At high cation concentration permitting the phosphate repulsion to be overcome, the junction adopts a stacked conformation. While at small cation concentration, it adopts an open conformation, by moving every branch away from the others. At the end of the homologous recombination, the junction resolving enzymes separate the resultant duplexes as much as possible.

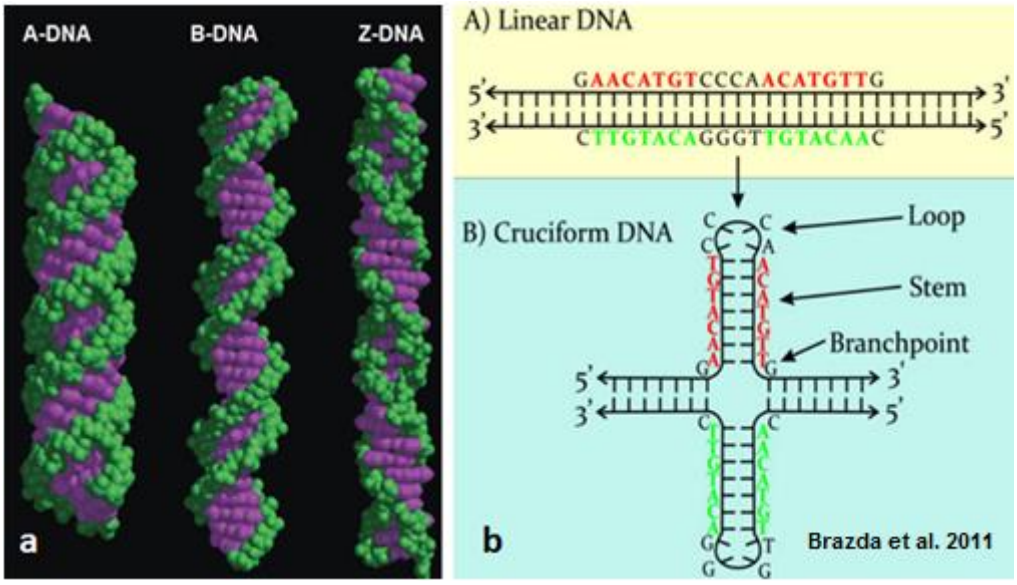


Fig. I.5: DNA secondary structures. ***a)*** *A-DNA is shorter and wider than B-DNA. Both are right handed double helices. Z-DNA is a left-handed double helix that is slightly longer and narrower than B-DNA.* ***b)*** *Palindromic sequences form cruciform structures under negative supercoiling: once unzipped, a linear DNA having inverted repeats can form a cruciform structure when palindromic sequences bind to each other.*

e. Triple helices

A Triple helix can form during recombination when a strand of a DNA duplex interacts with another DNA duplex via Hoogsteen (or reversed Hoogsteen) bonds at the major groove of the Watson-Crick duplex. A polypyrimidine invading strand binds to the polypurine strand of the canonical DNA duplex via Hoogsteen bonds in a parallel orientation and leads to C+:G-C and T:A-T triplets (where C+ represents a protonated cytosine on the N3 position). Similarly, a polypurine invading strand binds also to the polypurine strand of the duplex but this time via reverse Hoogsteen bonds and in an anti-parallel orientation and leads to G:G-C, A:A-T, and T:A-T triplets.

f. H-DNA

H-DNA is an intarmolecular triple helix structure adopted in a negatively supercoiled B-form DNA by polypurine-polypyrimidine regions having a mirror repeat symmetry. If the strand that is going to bind to the canonical duplex via Hoogsteen binds is a polypyrimidne (leading to T-A:T or C-G:C+ triplets) the triple helix requires an acidic pH to form. Now if the invading strand is a polypurine a triple helix can form at neutral pH but requires divalent cations for stability (that will lead to T-A: or C-G:G base triplets).

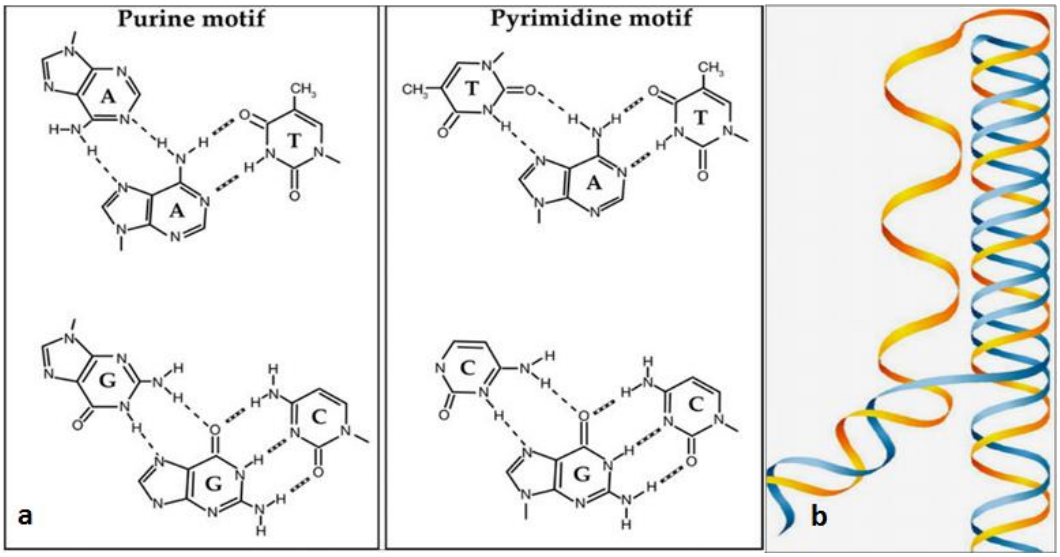


Fig. I.6: DNA triple helix. ***a)*** *Base triplets of a DNA triple helix: On the left panel, base triplets that occur when the invading strand is composed of purines only. On the right panel, when the third strand is composed of pyrimidine only.* ***b)*** *H-DNA or intramolecular triple helix.*

g. G quadruplexes

In 1910, Bang reported that higher order structures could be formed within G rich DNA sequences[9]. Fifty years later, in 1962, Gellert found that Guanine rich DNA sequences can form G-quadruplex structures[10], hereafter called G4. G4 is a non-canonical four stranded

DNA structure that arises from guanine quartets stacking (fig.I.7). A guanine quartet is a square planar arrangement formed by four guanine nucleotides where the neighboring guanines are linked by two hydrogen bonds. The minimal requirement to make an intramolecular G4 structure is four tracts of three or more Guanine ($G_{\geq 3}N_x\ G_{\geq 3}N_x\ G_{\geq 3}N_x\ G_{\geq 3}$), however some intramolecular G-quadruplexes are formed by repetitions of two guanines. The G4 structure is stabilized by a monovalent or a divalent cation such as k+, Na^+, Sr^{2+}. In fact, depending on their size, these cations occupy the intra-quartet, or inter-quartets cavities, neutralizing therefore the electrostatic repulsion between negatively charged guanines oxygens[11]. The bases chains between the guanine tracts are extruded from the tetrads and are called loops. Shorter loops and longer guanine tracts lead to more stable G4s[12]. Intramolecular G4s are those that engage Guanines of the same DNA strand while intermolecular G4s engage guanines of different strands and could be bimolecular, trimolecular or tetramolecular.

A G quadruplex is parallel if all of its G strands are parallel. On the contrary an antiparallel G4 has two parallel strands that are antiparallel to the two others. Besides, if one strand is antiparallel to the three others, the G4 is called hybrid. The big majority of parallel G4s have guanines in the anti-conformation while antiparallel and hybrid G4s have a mixture of syn- and anti-conformations of guanines. Moreover, the loops joining the G runs could be lateral, diagonal or double chain reversal. For all those reasons, the G4 is considered a DNA structure of high polymorphism.

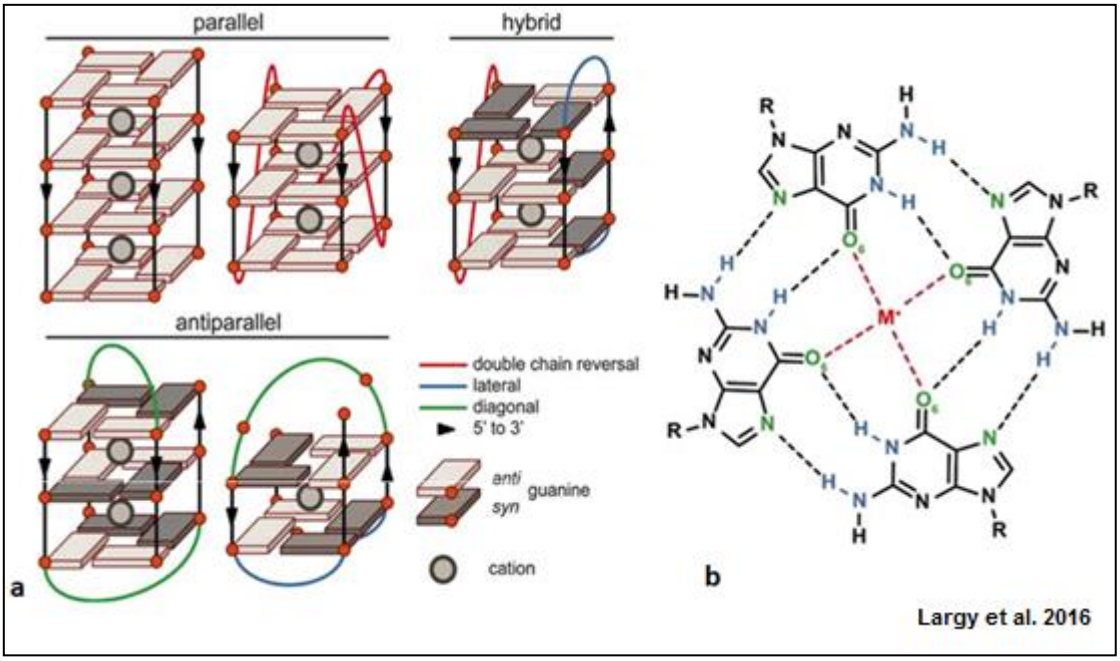


Fig. I.7: G-quadruplexes. ***a)*** *Tetrads can stack to form G-quadruplexes. G quadruplexes can fold into a variety of topology, which differ mainly by the relative orientation and number of strands (1 to 4), the number of tetrads (at least 2), and the geometry of the loops. b) Guanine tetrads contain four guanines linked by eight hydrogen*

bonds (donor and acceptor groups in blue and green, respectively). Guanine O6 selectively coordinate a metal cation (red). [13]

I.2.4 DNA tertiary structure

The DNA tertiary structure refers to the three dimensional arrangement in space of the DNA double helix and its confinement in chromosomes that is influenced by geometrical and steric constraints.

I.2.4.1 Topological parameters

a. Contour length:

The contour length of a polymer is its total length L_c=N.l.

Where N is the number of monomers composing the polymer, and l is the length of a single monomer. The contour length of DNA is thus the number of nucleotides times the length of a single nucleotide.

b. Persistence length

A polymer persistence length is the length scale beyond which the elastic cost of bending is smaller than thermal energy. The stiffer the polymer, the bigger the persistence length. The DNA persistence length is about 50nm[14] and is much higher than PVC or polyethylene polymers. The DNA is compared to a random coil in the absence of a stretching force, and its end to end length is $R = \sqrt{2Ll_p}$, where l_p is its persistence length. The DNA tertiary structure refers to the topology adopted by DNA at length scales larger than the persistence length L>>>lp.

c. Linking number LK

It is the number of times one strand is wrapped around its complementary one in a circular double stranded DNA. This number cannot be changed in a double stranded circular DNA because it does not have free ends. L_k could be also considered as the number of times a DNA strand has to be cut (breaking of a covalent bond) and turned around its complementary in order to entirely separate the two strands. For a given circular DNA of N base pairs, the more the number of base pairs per turn (n) the less the L_K. A relaxed DNA has $L_K=L_{K0}=N/n$, for B-form DNA n=10.5 base pairs. L_K is an integer and it is the sum of two other characteristic numbers the twist Tw and the writhe Wr.

d. Twist

It is the number of helical turns of one strand about the other. It is determined by the number of times a strand wraps completely around the other. Therefore, for a planar circular double stranded DNA, Tw=LK.

e. Writhe

It is the number of times the DNA double helix crosses around itself. A right handed writhe is negative while a left-handed writhe is positive. When the DNA is overwound the helix wrap around itself in a left-handed fashion, writhe is then positive and the linking number is higher than that of relaxed DNA. The DNA is said positively supercoiled. If the DNA is underwound, the double helix wraps around itself in a right handed fashion leading to a negative writhe and then a linking number smaller than that of relaxed DNA.

I.2.4.2 DNA storage in chromosomes

Living organisms are sorted into two groups: Eukaryotes (human, animals and plants) and Prokaryotes (archaea and bacteria). In Eukaryotes, the DNA is stored in every cell nucleus, which is surrounded by the cytoplasm and enveloped by the cell membrane. Thus, every cell contains the whole code that characterizes the organism except red blood cells and blood platelets that lack the nucleus. In order to fit into the cell nucleus of 6um of diameter, the DNA, long about 2m, is condensed into compact chromosomes. In a human cell, there are twenty three pairs of chromosomes.

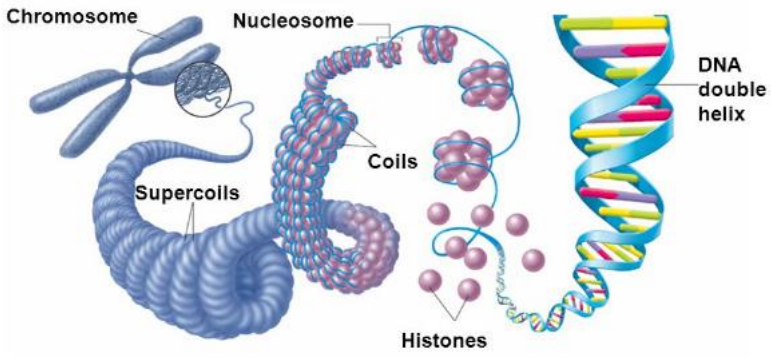


Fig. I.8: DNA storage in chromosome. *From PMG biology October 2015*

In Eukaryotes, the negatively charged double helix DNA is first wrapped around positively charged proteins called histones. Along the DNA molecule, every eight histone subunits form one nucleosome, around which 146 DNA base pairs are wrapped. The resulting complex DNA-protein that is three times shorter than the linear DNA is named chromatin and still accessible to enzymes and proteins. In this form, DNA strings link the beads-like nucleosomes, and hence the structure is called "beads on a string" and has a diameter of 10 nm. The DNA becomes more strongly compacted when every six nucleosomes are coiled together, and stack upon each other's to form a solenoid of 30nm of diameter. In this form,

the DNA is 50 times shorter than the linear form. The chromatin form does not only make possible fitting DNA in a tiny space but also control gene expression. For instance, when the cells are not dividing, some chromatin regions becomes tightly packed and thus could not be transcribed, those are called heterochromatin. Heterochromatin accumulates near the nucleus envelop. Contrarily, euchromatin are the loosely packed regions of chromatin that are not stainable but dispersed in the nucleus and are prone to transcription. At this stage of the cell cycle, called interphase, the DNA is replicated to two sister chromatids in order to prepare to the mitosis, or the cell division stage where a cell divides into two daughter cells that will inherit the same genome. The chromatid sisters are joined by cohesion proteins to form an X shape chromosome and are bound together at its centrosome. The chromosomes become more condensed during the first phase of the mitosis called prophase, and during the metaphase they become 10 000 fold shorter than the linear form. The chromosomes at this stage align along the cell equator in order to let the sister chromatids detachment and migration to the opposed cell poles during the anaphase. At the end, during telophase, every cell is divided into two new cells that contain chromosomes of one chromatid each.

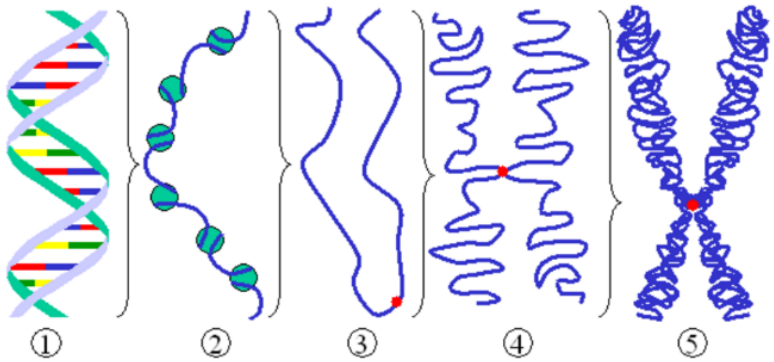


Fig. I.9: A chromosome at different magnifications. 1*) DNA double strand.* ***2)*** *Double-strand DNA wrapped around histone proteins (nucleosomes).* ***3)*** *Chromatine with a centromere during the interphase.* ***4)*** *Following synthesis, two chromatids are joined by a centromere; each chromatid consists of highly coiled nucleosomes.* ***5)*** *Two chromatids of a chromosome are tightly coiled in preparation for cell division. (From © 2012 T. F. Fletcher vanat.cvm.umn.edu)*

In Prokaryotes, the cells are nucleus free and the DNA is located in the cytoplasm in a region called nucleoid that is not membrane bound. The majority of prokaryotes have only one circular chromosome and their DNA is even more compact than the DNA of Eukaryotes.

Eukaryotic telomeres refer to the terminal segments of chromosomes that seal the genes and protect them from erosion. They consists of a double stranded DNA regions and a 3' single stranded DNA overhang, both are gene free and have a species-specific tandem repeats of G-rich sequence. Human telomeres contain repeats of TTAGGG (2-50 kilobase pairs) and a G-tail of 100–250 bases[15]. These single-stranded G-rich ends are believed to form G-quadruplex structures (fig.I.10) [16].

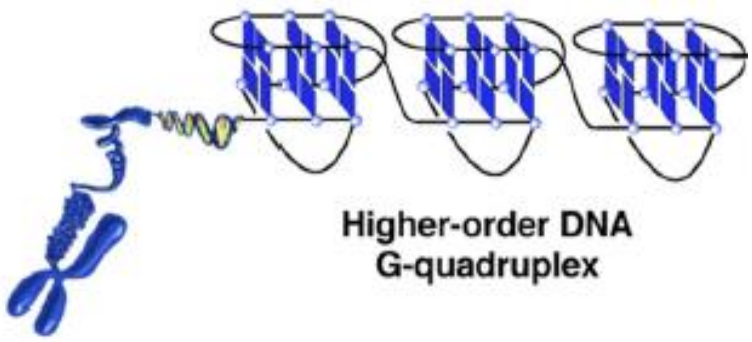


Fig. I.10: G-quadruplex structures at the 3' overhang of telomeres. *The repeated G rich sequence on the single stranded telomeric overhang forms successive G4 structures.*

I.3 DNA mechanics

I.3.1 dsDNA extension model

I.3.1.1 Free jointed chain model

In this model, the DNA is assimilated to a polymer chain of length **L** and having **N** rigid segments (fig.I.11.A). All the links have the same length **b** and an independent random direction $\vec{r}_i$. There is any self-avoiding condition and thus the chain can cross itself. The end to end length of the chain $\langle \vec{R}_0 \rangle$ is equal to 0 because the direction a link can adopt is random and thus two opposite directions are possible with equal chance:

$$\langle \vec{R}_0 \rangle = \langle \overrightarrow{r_N} - \overrightarrow{r_1} \rangle = \langle \sum_{i=1}^{N} b\vec{r_i} \rangle = b \langle \sum_{i=1}^{N} \vec{r_i} \rangle = 0$$

The root mean square end-to-end distance is:

$$\langle (\vec{R}_0)^2 \rangle^{1/2} = \langle (\overrightarrow{r_N} - \overrightarrow{r_1})^2 \rangle^{1/2} = b \langle \left(\sum_{i=1}^{N} \vec{r_i} \right)^2 \rangle^{1/2} = b \langle \sum_{i=j}^{N} \vec{r_i} . \vec{r_j} + \sum_{\substack{i,j=1 \\ i \neq j}}^{N} 2\vec{r_i} . \vec{r_j} \rangle^{1/2} = b\sqrt{N}$$

Although the FJC model is interesting since it explains the principle of entropic elasticity, it does not describe properly the elasticity of a DNA molecule.

I.3.1.2 Worm like chain model

The FJC model can be used for double stranded DNA at low forces. In fact, at intermediate and high forces the FJC model cannot describe the DNA behavior, because this latter is not a jointed chain but a double helix and its bending energy should be taken into account. The worm-like chain model (WLC) assimilates the DNA to a flexible rod that has a persistence length ξ (fig.I.11.B). In order to align the polymer segments in the range of the persistence length, one should apply a force of $\frac{K_B T}{\xi}$. If the DNA is subjected to a stretching, then the energy of the system is:

$E = E_{bending} + E_{streching}$

$$\frac{E}{K_B T} = \frac{1}{2}\xi \int_0^L \left(\frac{\partial \vec{r}}{\partial s}\right)^2 ds - \frac{F}{K_B T}\int_0^L \cos(\theta(s))ds$$

Where ξ is the persistence length, $\vec{t}(s)$ is the vector tangent to the chain at the curvilinear coordinate **s**, **F** is the stretching force modulus along the stretching direction **z**, and **θ** is the angle between $\vec{t}(s)$ and the **z** direction.

The force **F** required to induce an end to end extension of ⟨**z**⟩ can be approximated to[17]:

$$F = \frac{K_B T}{\xi}\left(\frac{\langle z\rangle}{L} - \frac{1}{4} + \frac{1}{4\left(1 - {\langle z\rangle}/{L}\right)^2} + \sum_{i=2}^{7} a_i \left(\frac{\langle z\rangle}{L}\right)^i\right)$$

At low forces $F < \frac{K_B T}{\xi}$, the double stranded DNA acts like an entropic spring, its extension is proportional to the force:

$$F = \frac{3K_B T}{2\xi}\left(\frac{\langle z\rangle}{L}\right)$$

At forces up to 10 pN, the dsDNA extension curve can be modelled by the WLC model. Above 10 pN, the dsDNA is stretched and its end-to-end length exceeds the contour length. At 65 pN the DNA is overstretched and undergoes a phase transition from B-form to S-form. Moreover, above 300 pN the double stranded DNA is unzipped and gives single strands DNA.

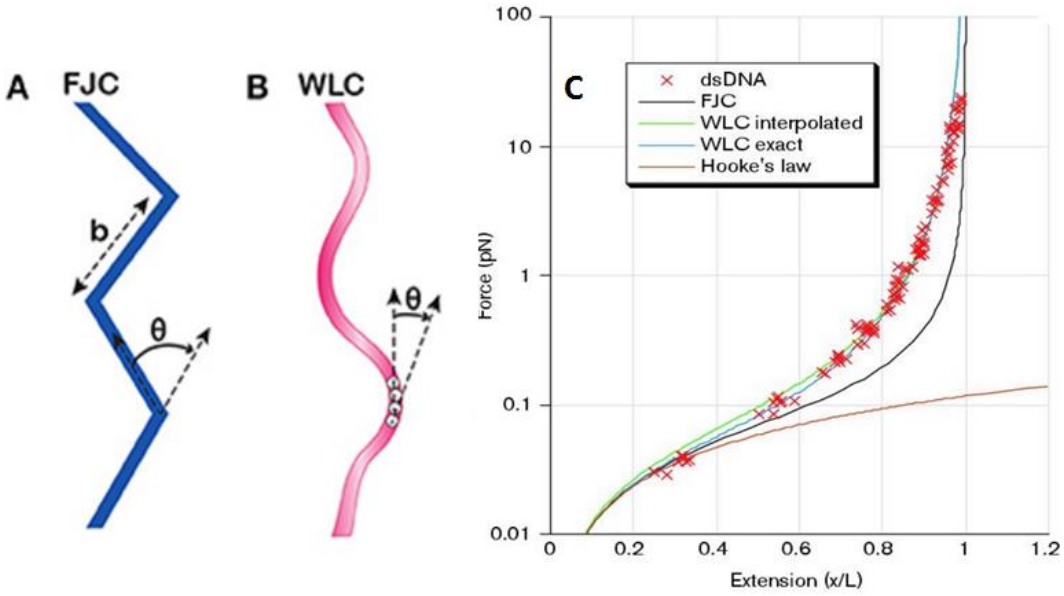


Fig. I.11: FJC and WLC models. A) *The FJC model assimilates the DNA to a chain of rigid segments having a length b and linked to each other. θ is the angle between two consecutive segments. B) The WLC model assimilates the DNA to a smoothly bending flexible rod that has a total length L and a persistence length ξ. Θ is the angle between the pulling force axis and the tangent vector to the curvilinear position s.*

I.3.2 ssDNA elasticity

Due to its high flexibility, the ssDNA is more contractile than dsDNA and needs a higher force to align it (around 5pN instead of 0.1pN $F=K_BT/\xi$). Up to 5 pN, the single stranded DNA is shorter than dsDNA, above 5 pN, however, the ssDNA length exceeds that of dsDNA. ssDNA can be modelized by the FJC model or the elastic FJC model.

I.3.3 Unzipping a DNA double helix

The force required to open a DNA double helix is around 15 pN[18]–[20]. The exact value, however, depends on the double helix **G-C** and **A-T** base pairs content. A **G-C** rich segment requires a higher force value.

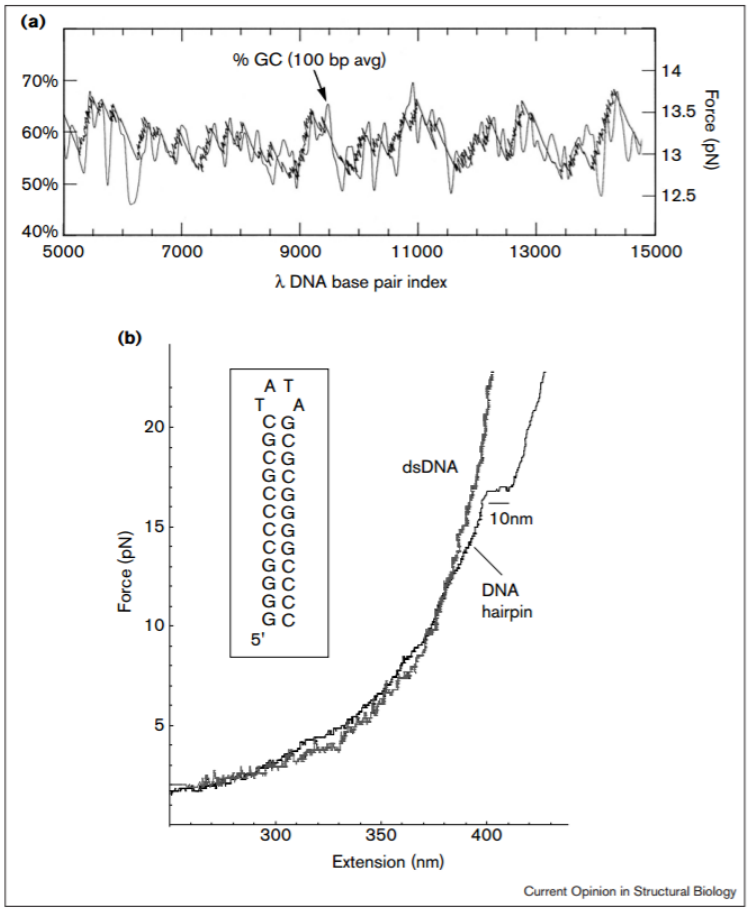


Fig. I.12: Unzipping a double stranded DNA helix. *On the top, the graph shows the force required to open 100 bases of dsDNA depending on the GC content.*

The fig. I.12 represents the curve of the Force versus the extension when opening a DNA hairpin. The hairpin opens around 15 pN and an abrupt extension of 10 nm is noticed.

I.3.4 DNA supercoiling

If the DNA molecule is tethered with multiple bounds at both extremities, twisting the DNA molecule becomes possible. If we increase the torsional constrain, the molecule extension remains constant until we reach a threshold above which plectonemes appear leading to the shortening of the molecule.

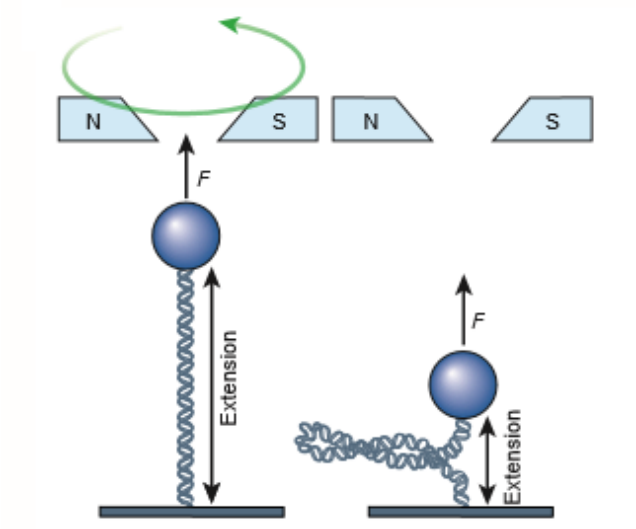


Fig. I.13: DNA supercoiling. ***a)*** *A dsDNA pulled at a force F and twisted by rotating the magnets.* ***b)*** *The same dsDNA that is shortened after plectonemes formation.*

CHAPTER II: Overview of the DNA replication

II.1 Proteins and Enzymes

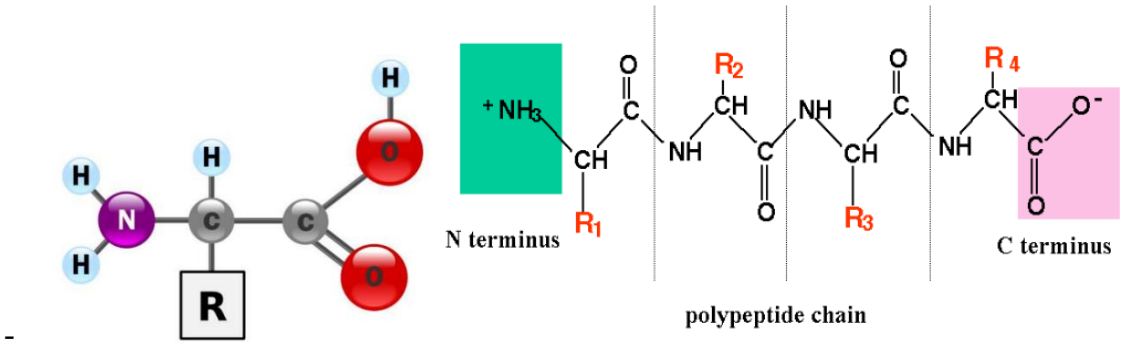


Fig. II.1: Schematic representation of proteins. ***a)*** *Amino acid: It is formed by a carboxyl group and a central atom to which is linked a lateral chain. The lateral chain determine the type of the amino acid. b) Polypeptide chain: It is a formed by the association of amino acids via peptidic bonds and is read from N-terminus to C-terminus domain. A protein is formed by one or several polypeptidic chain.*

Proteins and enzymes are biological macromolecules that fulfill most functions in the organism. Enzymes are proteins that lower the activation energy for a reaction through a process called catalysis. Proteins consist of long chains of amino acids linked together by covalent peptide bonds (fig.II.1). If this polymeric chain contains up to fifty amino acids it is called a peptide, if it contains more, it is referred to by polypetide or protein. A protein could be formed by one or more polypeptides. All of the twenty amino acids that serve to build proteins have an amino group, a carboxyl group and a central carbon atom to which is linked a lateral chain (R-group). This lateral chain is specific for every amino acid that could be hydrophilic, hydrophobic or negatively or positively charged. Proteins are defined by their

sequence, structure and function, and are therefore classified in different categories. Among proteins categories, one finds: Structural proteins that build the cells and the tissues, transport proteins that carry molecules and ions, gene regulatory proteins that trigger or silence gene expression and enzymes that catalyze reactions.

II.2 Proteins structure

II.2.1 The primary structure

The protein primary structure or the sequence of a protein is defined by the order of amino acids forming the molecule from the N-terminus to the C-terminus. The primary structure of a protein determines the secondary structure; therefore a change in the first generally leads to a change in the secondary structure and may affect the protein function.

II.2.2 The secondary structure

The secondary structure of a protein is its shape, more precisely the shape of every petide chain within the protein, which is also called domain). In fact, every peptide chain folds in a three dimension structure owing to the hydrogen bonding between the backbone atoms (and not the side chains atoms). At small distances, this interaction gives the polypeptide chain a shape of α-helix (a rode shape), β-pleated sheet, or unstable random coil. The α-helix is a right handed helix that arises from the stable hydrogen bonds between the oxygen of C=O of an amino acid (i) and the hydrogen of the NH group of the amino acid (i-4) in the helix. There is 3.6 amino acids in every turn of the α-helix and the amino acids side chains point to the exterior. β-sheet arises from the hydrogen bonds between strands. These bonds take place between the oxygen of the carbonyl group of an amino acid and the hydrogen of the amino group of another amino acid. The beta sheet is called parallel (anti-parallel) when the directions of the strands forming it are the same (opposite). The pleated form of the β-sheet is conferred by the coplanar hydrogen bonds.

II.2.3 The tertiary structure

The tertiary structure of a protein is the overall 3-D structure that results from the folding of all the protein domains and their positioning with respect to each other, which is fashioned by the amino acids R-groups (side chains) interactions. In fact, many stabilizing interactions (Hydrogen bonds, salt bridges, ionic interaction, stacking) lead to this structure in order to achieve the lowest energy state. The hydrophobic R-chains are buried within the protein core while the hydrophilic ones are exposed to the aqueous medium. The tertiary structure of a protein is also highly stabilized by covalent bonds formed between oxidized cysteine amino acids, or disulfide bridges, that connect amino acids even if they are distant in terms of sequence.

II.2.4 The quaternary structure

The quaternary structure of a protein is the way subunits assemble to make a functionalized protein complex (fig.II.2). Depending on the protein, the subunits can be similar and form a homodimer, or different and form a heterodimer. This structure is also stabilized by the same interactions cited above between amino acids side chains (fig.II.3).

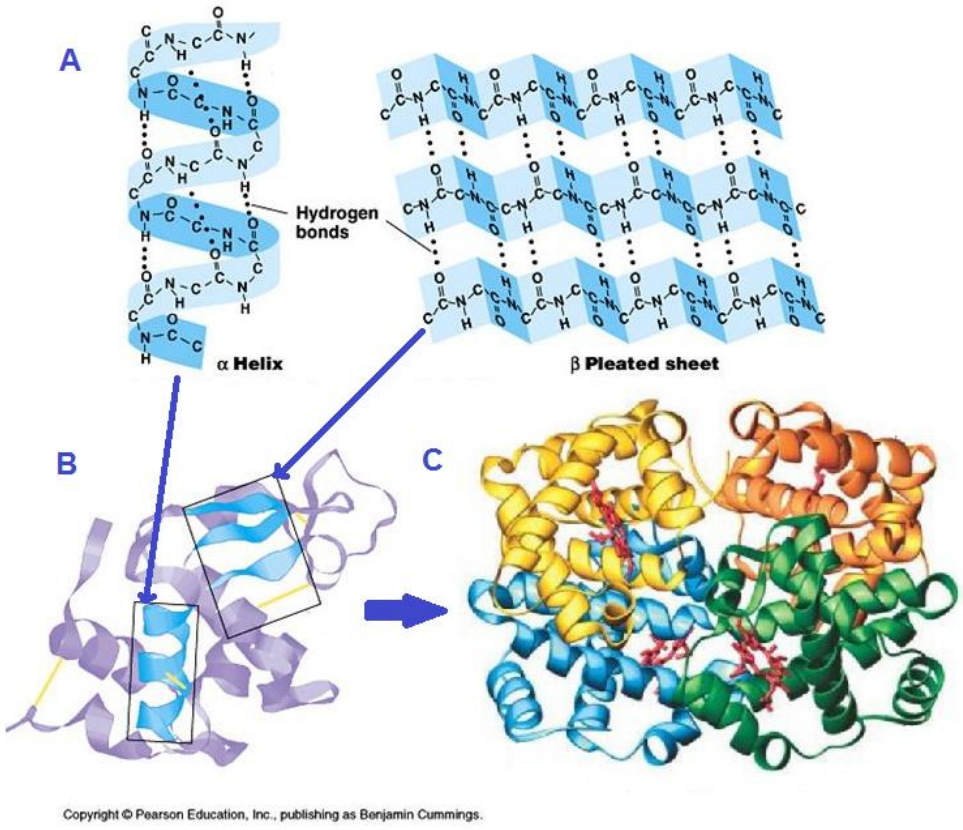


Fig. II.2: Protein structure. ***A)*** *Protein secondary structure: α-helix and β-sheet formed by hydrogen bonds between backbone atoms.* ***B)*** *Tertiary structure: The overall shape of the protein that is achieved by weak interactions between side chains and covalent bonds between oxidized cysteine amino acids.* ***C)*** *Quaternary structure: This is the fourth level of the protein structure to form a protein complex by assembling protein subunits.*

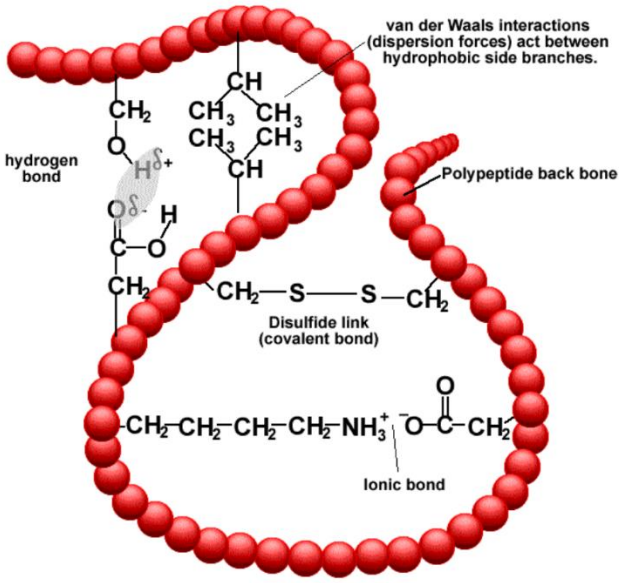


Fig. II.3: Protein tertiary structure stabilizing interactions. *Interactions between amino acids side chains: Ionic bonds, hydrogen bonds, van der Waals interactions, and covalent disulfide bridges.*

II.3 Gene expression

Gene expression refers to protein synthesis using the specific gene code in order to respond to the organism demand and provide a biological function. This is a two steps process: **Transcription** and **translation**. The transcription takes place in the nucleus and consists to copy the gene or the protein recipe encoded in the DNA. The transcript is an RNA called messenger mRNA. The mRNA quits the nucleus afterwards, and delivers the message to the DNA synthesis machinery, a nucleoprotein called ribosome located in the cytoplasm. The ribosome translates the code and synthesizes a polypeptide chain by adding one by one the amino acids according to the order provided by the messenger.

II.4 DNA replication

II.4.1 Helicase

Helicases constitute one of the largest class of enzymes [21], [22] , and are implicated in almost all cellular mechanisms that engage DNA and RNA such as DNA replication, transcription, translation, combination, DNA repair, and ribosome biogenesis [23]–[26] . A helicase is an enzyme that unzips the DNA double helix using ATP hydrolysis. In fact, the helicase opens an eye shaped region (also called replication fork fig. II.4) of two single strands in the nucleic acid duplex to initiate replication and allow the loading of different proteins in order to achieve the replication. The helicase unzips the double stranded DNA permitting the progression of the replisome that is formed by the helicase, the polymerases, the clamps and the clamps loaders (see next section).

The human genome codes for 64 RNA helicases and 31 DNA helicases [27]. Accordingly, defects in their function or deregulated expression lead to severe diseases [28], [29]. Helicases are classified into six super-families (SF) based on their sequences, structures, and mechanistic features [30], [31]. SF 3 to 6 include hexameric helicases that contain six RecA-like domains that form a ring shape complex, while SF 1 and 2 include non-ring-shaped helicases that contains two RecA-like domains [30]. The helicase core domains are referred to by RecA-like domain due to their resemblance to the ATP-binding core of the recombination protein RecA.

A Helicase of type alpha translocates on ssDNA while a helicase of type beta translocates on dsDNA. Another feature to distinguish helicases is the translocation polarity: A helicase A has a 3’-5’ directionality while a helicase B moves in the opposite direction. Conserved sequence motifs between helicases are those involved in ATP binding, ATP hydrolysis and translocation along nucleic acids. Seven of these conserved motifs had been identified by bioinformatics and led to the helicase discovery [32].

Helicases may be active or passive, in both cases they are molecular motors using ATP to translocate along ssDNA. The active or passive behavior characterizes the way they displace the opposite strand: a passive helicase cannot melt the dsDNA but wait for thermal fluctuations to open the dsDNA and then advances preventing the fork reclosing. Active helicases are able to translocate and melt dsDNA at the same rate. In fact an active helicase holds a special structure which enhances the melting fluctuations of dsDNA. The unwinding rate of an active helicase is nearly equal to translocation rate. Contrarily, the unwinding rate of a passive helicase is depends on the force applied to dsDNA and is lower than its translocational rate.

II.4.2 Polymerase

The polymerase is the enzyme that performs the DNA synthesis in the 5'-3' direction. When a replication eye is opened, the polymerase can access the single stranded DNA and duplicate it by adding nucleotides containing the complementary bases to a primer. The association of the polymerase with a ring shaped protein called clamp increase its processivity. The clamp loader as denoted by its name allows loading the clamp. Since the parent strands are anti parallel, the replication directions on these strands are also anti-parallel and generate two anti-parallel new strands. The new strand that is synthetized in the same direction as the helicase motion is continuously replicated and called the "leading strand". The other strand is called the "lagging strand" and is synthetized in segments called Okazaki fragments. As the helicase moves forward, it separates the two strands of the parent DNA. RNA primers are synthetized on the lagging strand by an enzyme called primase and are used to initiate new Okazaki fragments. The end of an RNA primer having a 3' OH group is extended by the polymerase till it encounters previous RNA primer. Okazaki fragments are ligated together after replacing RNA primers with DNA segments by means of an enzyme called ligase.

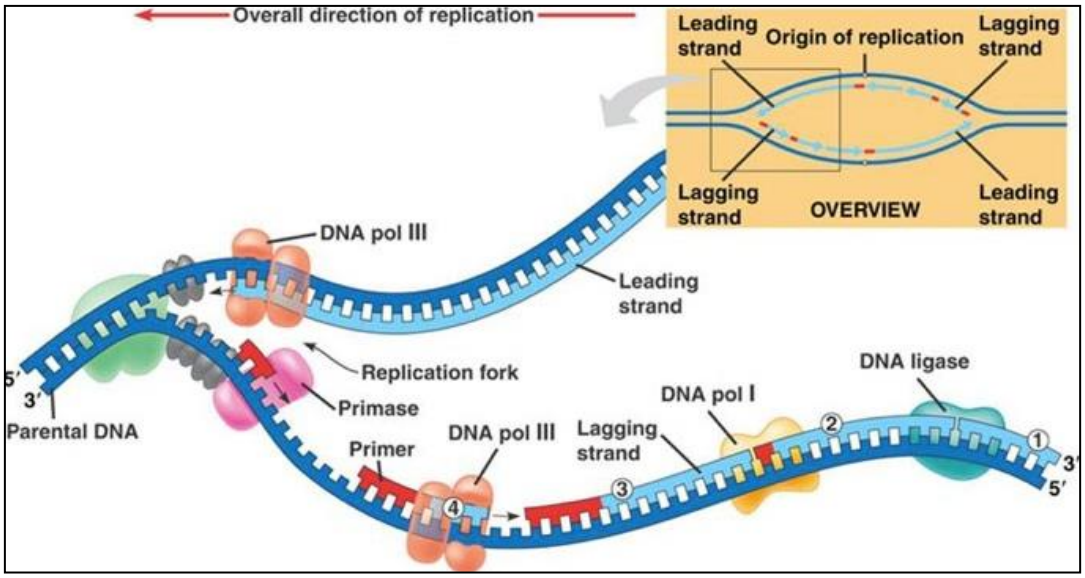


Fig. II.4: DNA replication or DNA synthesis in the process of copying a double-stranded DNA molecule. *A helicase unzips the DNA, and expose the leading and the lagging strands in order to be replicated by polymerases. The leading strand is replicated in the same direction of the helicase progress, thus in a continuous fashion. While, the lagging strand replicated in the opposite direction, thus it is replicated in segments, called Okazaki fragments. The primase adds an RNA primer to initiate these fragments, and a ligase ligates them together by replacing an RNA primer by a DNA primer* (*http://faculty.irsc.edu/FACULTY/TFischer/images/DNA%20replication.jpg*).

II.4.3 Telomere replication

During replication and when the helicase reaches the end of the chromosome, a segment of a tens of base pairs located on the 3’ extremity of the telomere remains unduplicated[33]. The polymerase uses the 3’ OH extremity of the newly synthetized Okazaki segments to prime DNA synthesis. But when the last added RNA primer is removed, the lack of an OH group on the 3’ extremity makes its replication unachievable. This is called the end replication problem[34]. Therefore, at the end of every replication, the 3’ end of the telomere is shortened. However, the genome integrity still protected as long as the chromosomes erosion is taking place at telomeres, which are non-coding segments.

Some cells respond by the DNA damage by activating the ribonucleoprotein called telomerase. The telomerase contains an RNA segment that partially hybridizes with the tip of the telomeres and serves as a template to elongate them by the catalytic telomerase reverse transcriptase (TERT) component[35]–[37]. If the telomeres’ shortens faster than the telomerase action, or if the telomerase is deactivated, the totality of telomeres will be lost after nearly 60 replications. This stage in the cell life is called the Hayflick limit and the cell becomes senescent and does not divide further. In other words, the telomeres shortening are directly linked to aging, and are described as a biological clock. Besides, the telomerase is overexpressed in 80% of cancers.

Therefore, one possibility to inhibit malignant cell proliferation is to deactivate the telomerase. Telomeres and telomerase constitute an interesting research topic that might allow fight cancer and prevent, or reverse aging.

CHAPTER III: DNA micromanipulation using Magnetic tweezers

III.1 Introduction

Biochemical measurements in bulk average the behavior of a large number of molecules behaviors and often lack the real time insight. In a biological sample, the majority of molecules are not synchronized, and their heterogeneous activity distribution is called the static disorder. Moreover, the activity of every molecule varies over time, and this is called the dynamic disorder. Therefore averaging static and dynamic disorders is known as a common disadvantage for bulk experiments.

In the 1990s, the real time observation and micromanipulation of individual biomolecules became possible thanks to the progress that microscopy techniques. The real-time temporal resolution made the observation of transient events possible. In addition of the conditions that are monitored in bulk experiments (temperature, pH, concentrations), micromanipulation techniques apply a physical constraint on the molecule to determine its characteristics such as elasticity, extension, and interaction force. The constraint is applied to the molecule of interest via a sensor bound to it. The smaller the sensor size, the smaller the crowding effect and the viscous loss, and the higher the signal to noise ratio.

III.1.1 Atomic Force Microscopy AFM

It allows imaging the surface of a sample with an atomic resolution. A cantilever with a nanometric tip scans the sample point by point. The deflection of the cantilever is used to determine the interaction between the sample and the tip of the cantilever. A laser beam reflected on the cantilever is sent to a detector, so that when the cantilever is deflected, the beam deviation produces a signal on the detector.

III.1.2 Optical microfiber

A DNA molecule that is attached to a bead is held to a micropipette by on end to an optical fiber by the other end. When a force is applied on the DNA by the micropipette, the bending of the optical fiber change and leads to a change of the laser beam position on the detector. The applied force can be calculated using the known microfiber bending rigidity and the deflection value.

III.1.3 Optical tweezers

A Laser beam is highly focalized using an objective with a high numerical aperture. The focalization point forms a trap for dielectric particles. In fact, a dielectric particle near the focalization point is polarized by the electric field of the light. The induced dipole interacts with the high gradient of the electric field and is subjected to a force that attracts it to the focalization point[38]. In order to study biomolecules by optical tweezers, the molecule of interest is bound to a microscope coverslip by one end and by the other end to a dielectric bead having a refraction index higher than that of the medium.

III.1.4 Magnetic tweezers setup

This is the technique used in this work, a magnetic force allows to stretch or rotate a biopolymer tethered to a surface, typically a DNA molecule, due to a super paramagnetic bead attached to its extremity' [39]–[41]. The setup consists of an inverted microscope permitting to observe the sample from below while a pair of antiparallel magnets is placed above. These magnets can be translated closer or furtherer away of the sample or rotated. The magnets generate a horizontal magnetic field $\vec{B}$ that has a magnitude of 1 Tesla. The magnetic field decreases exponentially while moving away along z axis. Above the magnets is placed a red led that illuminates the sample. The sample is a fluid chamber containing the DNA tethers bound to magnetic beads, which allow pulling and twisting the DNA by means of magnetic force. Pulling the DNA is done by approaching the magnet to the sample while twisting the DNA is done by rotating the magnets. A microscope air objective 40x positioned underneath the sample collects the direct light and the forward scattered light on the beads. The real time tracking is performed by a CCD camera connected to the computer.

The temperature of the sample is controlled but small thermal fluctuations lead to the expansion of the objective and the sample holder and this change the objective-sample distance of some nanometers. . In order to overcome this issue, the objective holder has been designed specially to compensate the expansion of the different materials.

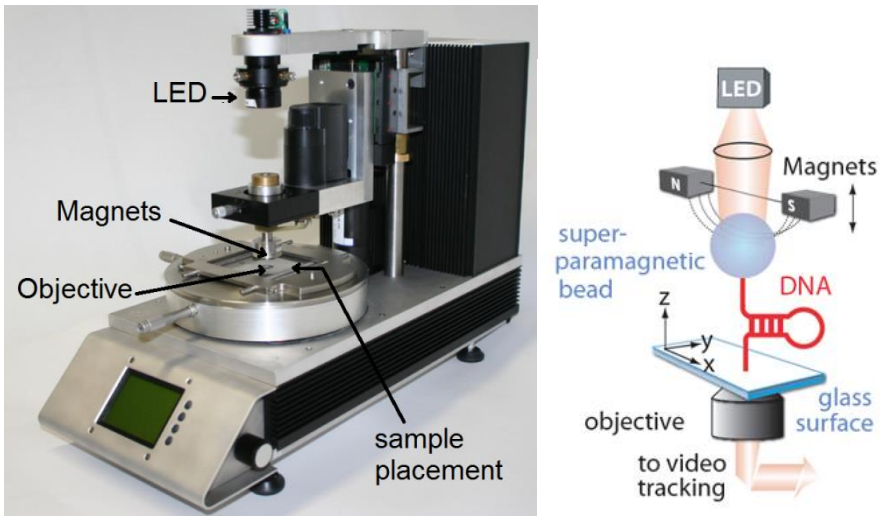


Fig. III.1: The magnetic trap setup. *(**On the left**) **PicoTwist** magnetic tweezers device. (On the right) A DNA hairpin, tethered to the glass surface and bound to a magnetic bead on the other end, is pulled by a vertical force generated by a pair of magnets. The sample is illuminated from above by a LED, and imaged from below by an objective and a CCD camera.*

III.1.4.1 DNA molecules preparation

The used DNA molecules have a digoxygenin modification on one extremity by which they are anchored to an anti-digoxygenin coated surface. A biotin modification at the other end of the molecule makes possible its binding to a streptavidin coated para-magnetic bead, and thus the DNA manipulation by magnetic tweezers. In a certain force range (up to ~ 20 pN), the mentioned bindings are strong enough to keep the DNA molecules in place for some hours, the time needed to perform the experiment. In fact the binding Streptavidin-Biotin is known to be the strongest of the weak bindings and need 160pN to be broken. The digoxygenin - anti-digoxygenin binding, however, could be broken by smaller forces.
In this work, we used DNA hairpins made by ligating:

1- A double stranded DNA of ~1Kb containing a sequence of interest.
2- A loop of 6 bases.
3- The hairpin arms:
   - A junction of a single stranded DNA – double stranded DNA containing some digoxygenin-dUTP on the double stranded region.
   - A 5’ biotin modified oligonucleotide.

   The hairpin arms are partially annealed together.

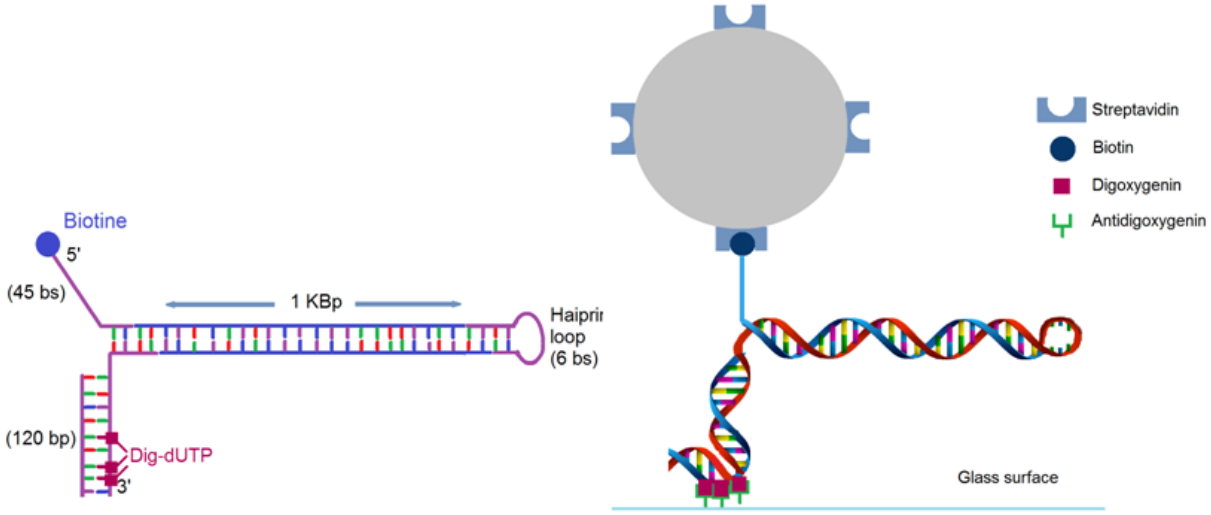


Fig. III.2: The DNA hairpin used in the current work. *Having a Biotin modification on its 5' end, the hairpin is bound to a streptavidin coated magnetic bead, while digoxygenin on its 3' end permits to tether it to an antidigoxygenin coated glass surface.*

(Annexe 2)

III.1.4.2 Fluid cell assembly

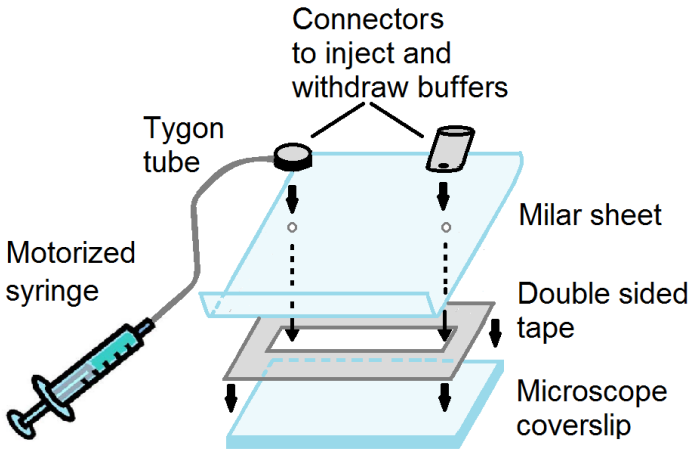


Fig. III.3: Fluid chamber assembly. A *double sided tape is sandwiched between a microscope coverslip and a mylar sheet. The tape contains a rectangular shaped channel that defines the fluid chamber size. Two connectors are glued to the top of the mylar sheet: The first is used to inject the buffers and the second is linked to a motorized syringe pump via a Tygon tube to withdraw the sample.*

The fluid cell is assembled by sticking a microscope slide to a mylar sheet using a double sided tape with an empty channel at its center. The fluid channel dimensions are 5x40 mm. An inlet and an outlet are glued to the mylar sheet using double sided tape. Injecting and

sucking buffers are achieved by means of a syringe pump. The smallest distance between the magnets and the beads is limited by the thickness of the Mylar sheet and the tape which corresponds to ~100 µm. (Annexe 3)

III.1.4.3 Bead tracking and molecule length

The tracking algorithm allows the real-time tracking of hundreds of beads in a field of view of 200 µm ^{2}. Every bead is visualized as a set of concentric rings corresponding to the bead diffraction pattern. This is due to the interference between the light coming directly from the LED and the light scattered by the bead. The rings radii become larger when the bead image is driven away from the objective focus plane. The position of the center of the diffraction rings is used to localize the bead in the horizontal plane (x, y). It is calculated by performing an autoconvolution of the bead image captured at the time t. The z coordinate of the bead is deduced from the size of the rings using a calibration image. In fact, a calibration image is done for every bead at the beginning of the experiment. It records the rings intensity profiles for different distances between the bead and the objective focal plane position by moving the objective while maintaining the DNA molecule stretched at a constant force. So the z position of the bead at a time t is obtained by comparing it to the calibration image. The force is then reduced to zero in order to bring the bead down on the glass surface in order to record a reference position.

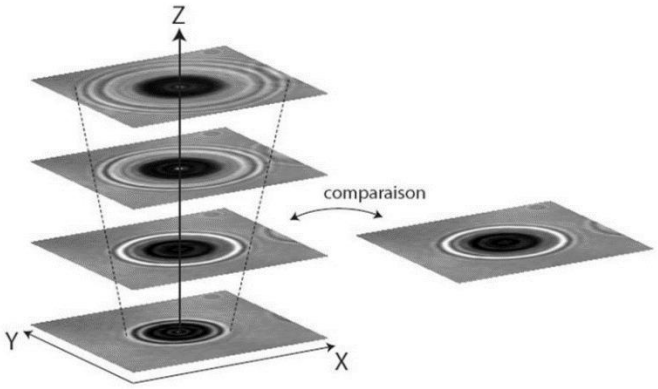


Fig. III.4: Calibration image. *The calibration image is achieved by moving the objective while maintaining the beads at the same position by assigning a constant force value for calibration. Note that the diffraction pattern is wider for bigger bead- focal plane distance. During the experiment, comparing the diffraction pattern of a bead with its calibration image, leads to determine its position along z.*

III.1.4.4 Force calibration

In this setup, the magnetic beads are exposed to an external magnetic field generated by the permanent magnets. Every bead becomes magnetized and acquires a magnetic moment proportional to its size and the magnetic field strength. Therefore the bead is subjected to a vertical magnetic force that attracts it towards the magnets.

$$\vec{m}(\vec{B}) = \frac{V\chi\vec{B}}{\mu_0}$$

$$\vec{F} = \frac{1}{2}\vec{\nabla}(\vec{m}(\vec{B}).\vec{B})$$

Where **m** is the bead magnetization, V the magnetic bead volume, χ the magnetic susceptibility, μ_0 the vacuum permeability, **B** the external magnetic field, and **F** the applied magnetic force.

Supercoiling DNA is achievable by rotating the magnetic field in order to apply a torque and the beads will rotate to align their magnetic moments with the magnetic field.

The applied force depends on the size of the magnets, the material that they are made of, the fixed gap between them, the distance separating them from the beads, and the bead size. The size of the used magnets is 1x1.7x3 mm, the gap between the two antiparallel magnets is 0.25mm, and the highest force that could be applied to the beads is limited by the thickness of the mylar and the double sided tape that is about 100 µm, and the beads used having a diameter of 1 µm. The magnetic field is assumed homogeneous in the field of view (~200 µm ^{2}) but the beads are slightly inhomogeneous thus the force applied is not exactly the same for all the beads (variation of ~2pN for a mean force of 10pN)

The beads surrounded by the buffer molecules are prone to Brownian motion due to thermal fluctuations. The beads are then fluctuating in the horizontal plane around their equilibrium position i.e when the DNA tether is aligned with z axis. The system is similar to an inverted pendulum. The higher the pulling force, the higher the restoring force applied on the bead and thus the smaller the Brownian motion amplitude. The restoring force could be assimilated to the force exerted by a virtual spring along x-axis having a spring constant of K_x.

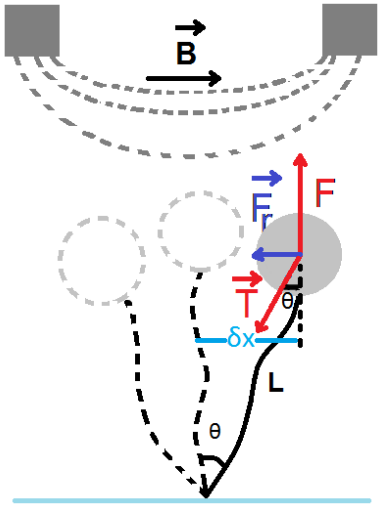


Fig. III.5 : The forces acting on the magnetic bead in the magnetic tweezers setup.

The restoring force along x is:

$$F_x = F\delta\theta = K_x\langle\delta x\rangle$$

For small angles, $\theta = \sin\theta$; Thus $F_x = F\sin\theta = \frac{F}{L}\langle\delta x\rangle$

$$\text{Then } K_x = \frac{F}{L}$$

Where θ is the angular bead deviation from the equilibrium position i.e. the angle between the axis of the DNA tether that is deviated from its equilibrium position and the z-axis.

Using the equipartition theorem stating that the energy is equal to $\frac{1}{2}K_BT$ for every degree of freedom, where K_B is the Boltzmann constant and T is the absolute temperature in K.

$$E_p = \frac{1}{2}K_x\langle\delta x\rangle^2 = \frac{F}{2L}\langle\delta x\rangle^2 = \frac{1}{2}K_BT$$

$$F = \frac{K_BTL}{\langle\delta x\rangle^2}$$

Therefore, for a single DNA tether of known length, and for different z magnet position, one can calculate the applied force F on the bead from its Brownian fluctuations in the horizontal plane. This measurement allows the calibration of the magnet and provides the curve of magnet position versus applied force. The zero position of the magnets i.e. the z-axis origin is the top surface of the coverslip where the DNA is tethered. In fact, in order to define the z-axis origin, the magnets are approached to touch the Mylar and a reference is set to 100 µm (thickness of the double sided tape and the Mylar sheet).

The fluctuations are not measured in real space it is more interesting to view them in frequency space. This representation gives the characteristic frequency of the system, the cutoff frequency f_c. In fact, the spectrum of the bead fluctuations along x is a Lorentzian it is flat until f_c, then it decays like $1/f^2$. The higher the force applied on the bead, the higher the cutoff frequency. The sampling frequency should be higher than at least $2*f_c$, in practice 4*fc is better. The sampling frequency is the camera video rate which is constrained by the camera and the tracking algorithm execution time.

The equation of motion of the system is:

$$m\frac{d^2x(t)}{dt^2} = -6\pi\eta r\frac{dx(t)}{dt} - K_xx(t) + F_L(t)$$

Where m is mass of the bead, x(t) the position of the bead at t, η the viscosity of water

($\eta = 10^{-3}$ Nsm^{-2}), r the radius of the bead, F_L the Langevin force that is a stochastic force.

The bead is subjected to collisions with water molecules from all directions, then the time average of the Langevin force is equal to zero:

$$< F_L(t) >= 0$$

For small $\tau_{(collision)}$, the correlation function could be written as delta-function:

$$< F_L(t).F_L(t') > = 24K_BT\pi\eta r\delta(t-t')$$

Then the autocorrelation function is:

$$< F_L(\omega) >^2 = 24K_BT\pi\eta r$$

During a time interval of $\Delta t = \frac{1}{\Delta f}$, the collisions of water molecules with the bead apply a force of:

$$\int_0^{\Delta f} F_L(f)df = \sqrt{24K_BT\pi\eta r\Delta f}$$

The Fourrier transform gives the amplitude of bead fluctuations:

$$X(f) = \frac{F_L(f)}{K_x(1+if/2\pi\tau)}$$

Where $\tau = 6\pi\eta r/K_x$

$$X^2(f) = \frac{|F_L(f)|^2}{{K_x}^2\left(1+\left(\frac{f}{f}\right)^2\right)} = \frac{24K_BT\pi\eta r}{{K_x}^2\left(1+\left(\frac{f}{f}\right)^2\right)}$$

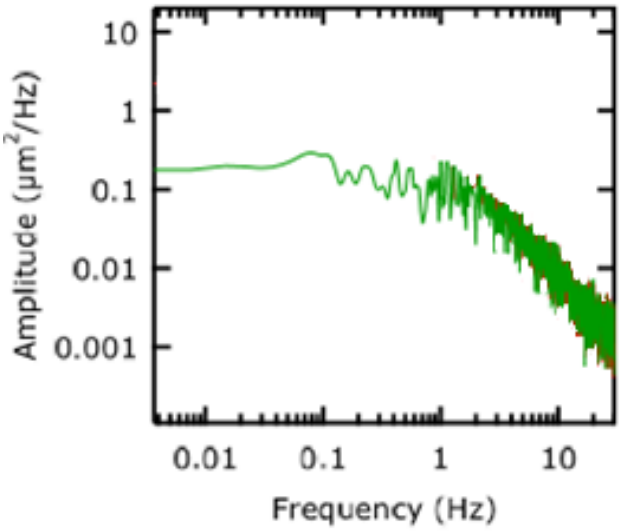


Fig. III.6 : – Power spectrum density.

By fitting the power spectrum with this equation of a Lorenzian, the cutoff frequency can be calculated.

Integrating the power spectrum density over all frequencies gives the bead fluctuations in the x direction. Therefore we can calculate the spring constant along x "K_x" as well as the applied force F.

The higher the applied force, the smaller the bead fluctuations and the higher the cutoff frequency.

III.2 Enzymatic activity study using Magnetic tweezers

This paragraph is extracted from the paper in annexe1 [42].

The general approach for single molecule mechanical assays of helicases is to stretch a DNA molecule and observe the change in its extension resulting from the transformation of double-stranded DNA (dsDNA) into single-stranded DNA (ssDNA). This can be implemented by two different configurations: the unpeeling configuration where a tension is applied between the two ends of the DNA molecule and the unzipping configuration where the force is applied between the two complementary strands at the same DNA extremity.

Force, the mechanical parameter introduced in these studies of helicases by micromanipulation techniques, has two roles:

1. It prevents re-hybridization of the strands in the wake of the helicase. By separating the two complementary strands, the force hinders their re-hybridization. As this effect increases with the force, the probability of re-hybridization is strongly dependent on the force, displaying an all or none behavior. Above a critical force F_r (~25 pN for the unpeeling assay, ~3 to 5 pN for the unzipping assay), re-hybridization is virtually impossible while it occurs readily below F_r (as reannealing in the wake of the enzyme becomes more frequent). This sets a lower limit to the force that can be used in these assays.

2. The force determines the molecule's extension and its fluctuations via the bead's Brownian motion, thus it sets the signal to noise ratio of the assay. The larger the difference between the ssDNA and dsDNA extensions the better is the signal. As the extension of both DNA forms is zero if no force is applied, the signal is null at zero force (and in the unpeeling situation at a force F ~ 5 pN where the extensions of both are equal, see below). Considering the elastic response of ssDNA and dsDNA the sensitivity of the measurements increases with the force, so it is tempting to apply strong forces to improve the signal to noise ratio of the experiment but this comes at the cost of applying a mechanical tension that might be out of the physiological range.

Indeed a large applied tension may destabilize the dsDNA. The mechanical unfolding of a hairpin construct in the absence of a helicase can be characterized, above a critical force of

F_u = 15±1 pN the hairpin spontaneously unfolds, while it remains otherwise stably folded at forces F_c < 12±1 pN. Thus at forces F < F_c, any unfolding observed in the presence of a helicase is a result of its activity. Indeed in its absence, the extension of the DNA molecule remains constant and equals to the folded hairpin. In the unpeeling configuration the destabilizing force is F_u ~ 60 pN. However, in both configuration, there exists a substantial range of forces (F_r < F < F_u) where the opening of the dsDNA can only be caused by the helicase and where the tension on the molecule prevents re-hybridization of the strands in the wake of the enzyme.

In the unpeeling configuration, the molecule is a stretched dsDNA (Fig.III.7.A). A nick or a gap in one of the strands of the DNA serves as a loading site for the helicase which proceeds by unpeeling one strand from the other. To prevent the reannealing in the wake of the helicase a large enough tension on the molecule has to be applied (F > 25 pN = 25 x 10^{-12} N). In these conditions the mismatch in the inter-nucleotide distance in the peeled (free) strand and in the strand under tension is large enough: the product of the force (25 pN) by the distance (about 1 nm) between the just separated complementary strands leads to an estimate of the energy cost that thermal fluctuations should overcome to re-hybridize the strands in the wake of the helicase. At F > 25 pN this energy exceeds several times the typical scale of thermal energies, k_BT (≈ 4.1 x 10^{-21} J, at room temperatures) resulting in a rare encircling of the helicase. As the extension of ssDNA is longer than that of dsDNA (for force > 5 pN) the result of helicase activity is an increase in the distance between the bead and the surface (the extension of the molecule) proportional to the amount of peeled ssDNA. Typically, the change of extension is a fraction of the extent of a base-pair (0.34 nm) depending on the force. Thus, in this configuration one has to apply quite a large force to get a good signal albeit with a poor sensitivity. The main advantage of this assay is its simplicity.

In the unzipping assay a tension is applied between the two ends of a hairpin structure. This configuration mimics a DNA replication fork structure, (Fig.III.7 B). The tension applied on the molecule acts on the two arms of the fork. The molecule's opening by the helicase increases the length of these arms while the tension prevents their re-hybridization in the wake of the enzyme. When the enzyme reaches the apex of the hairpin the molecule is completely unzipped. Since nothing prevents helicase translocation on the extended ssDNA strand, further translocation of the enzyme along that ssDNA allows the hairpin to refold in its wake, regenerating a fork that may push on the enzyme. In this configuration, the separation between the two arms of the fork is about twice the extension of ssDNA: 0.93 nm/bp at 10 pN. Consequently, the signal (i.e. the change in the bead/surface distance) is very sensitive to DNA unwinding. With a typical precision in the measurement of extensions of a few nanometers, this unzipping assay approaches single base resolution. In addition the critical force to prevent re-hybridization, F_r is low, typically 5 pN which allows an investigation of these enzymes in a more reasonable range of forces.

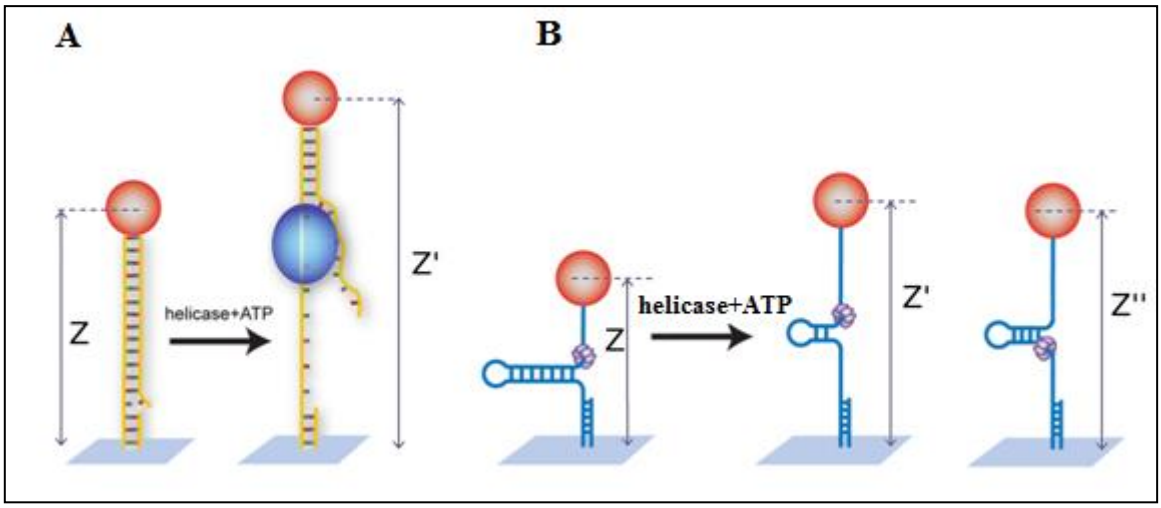


Fig. III.7: **Schematics of the unzipping assay.** *(a) Schematics of the unpeeling configuration used to study helicases: a helicase (blue blob) loads on a nick or gap in the dsDNA molecule under tension. Unwinding of the molecule results in an increase (Δz = z'-z) of the overall extension. The two ssDNA (one under tension and one free) are unable to match in the wake of the enzyme due to a mismatch in their extension. (b) Schematics of the unzipping assay: a helicase (the violet hexamer) loads at the fork of a DNA hairpin under tension. Unwinding of the hairpin results in an increase (Δz = z'-z) of the molecule's extension. The tension on the released ssDNA strands prevents their reannealing in the wake of the enzyme. As the helicase reaches the hairpin apex, its continuing translocation on a ssDNA template allows for reannealing of hairpin in its wake, monitored by the decreasing change in extension (Δz = z''-z).*

Part 2

Thesis project, Results & Discussion

CHAPTER IV: Thesis project

IV.1 G-quadruplexes

Nucleic acid sequences potentially forming G4 motifs are found in Human, Bacteria[43] and human DNA and RNA viruses[44]–[46]. *In vitro* biochemical and biophysical studies demonstrated the folding of these G-rich sequences into G-quadruplex structures[47], [48]. Since then, multiple studies were carried in order to check whether these non-canonical structures are present in-vivo. In fact, the first evidence providing the presence of G4 structures in-vivo was found in 2001 using specific fluorescent antibodies for the ciliate Stylonychia lemnae telomeric G4s[49]. Later on, a study reported that the disruption of a G4 structure in the proto-oncogene c-MYC done by a single base substitution, leads to an increase in gene expression, whereas the stabilization of the same G4 by TMPyP4 ligand leads to a decrease in transcription. This constitutes an evidence for the existence of G4 structures in-vivo[50].

Bioinformatic studies reported the presence of 716 310 putative G4 sequences in the human genome[51]. However, it is still unknown how many of those sequences do fold in a stable G4 *in vivo*. In the human genome, the pG4s are not randomly distributed, but mostly located in the single-stranded overhangs of telomeres, and in the oncogenes and proto-oncogenes that are located in double stranded DNA regions[51]–[54]. Putative G4 forming sequences are also present in the 5'UTR of mRNA[55] and in the RNA transcripts of telomeres (TERRA) [56].

Due to their presence in biologically relevant positions G4 structures have a biological role in controlling telomeres length and contributing to the gene regulation by their folding and unfolding mechanisms. They are considered as anti-cancer and anti-aging targets.

A G rich sequence embedded in a DNA double helix could potentially form a G4 structure when the duplex is transiently opened, like what happens during transcription, replication and repair. At this stage, there is a dynamic equilibrium between single stranded DNA, duplex DNA and G4 structure. Once folded, the G4 structure forms a roadblock

perturbing the progress of the replication fork and can thus lead to genomic instabilities[57]. Moreover, the folding of a G4 structure in the gene promoter can suppress the gene transcription and thus lead to genetic deletions[58]. It has been reported that, a G4 structure survives multiple cell divisions while keeping the same conformation[59].

The folding of telomeric G4 on the 3' single-stranded overhang of telomeres can inhibits the telomerase. This could be beneficial in cancerous cells by stopping the cell proliferation. However, in benign cells, telomeres shortening without being replenished by telomerase leads to cell death.

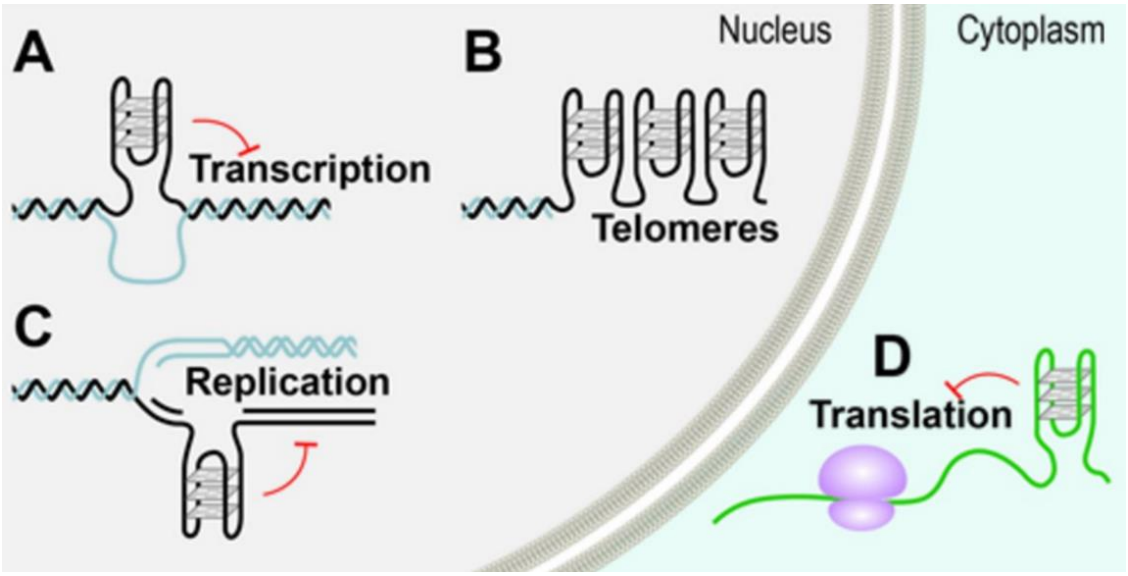


Fig. IV.1: G quadruplexes. (A) *G4 structure in a promoter.* ***(B)*** *G4 structure at telomeres. (C) G4 structure in a replication origin; the replication is interrupted at the G4 structure. (D) G4 structure in an RNA messenger.*

IV.1.1 C-MYC G4

The overexpression of c-MYC is associated to a large number of malignancies and diseases such as breast cancer, colon cancer, small cell lung cancer and Myeloid leukemia[60]. The guanine rich nuclease hypersensivity element NHEIII1, which is located upstream of the P1 promoter of c-MYC and required for 80-95% of c-MYC transcription[61], [62, p.], was shown to fold into a G quadruplex that repress the c-MYC expression *in vivo*[50], [62]. For instance, a study has reported the downregulation of c-MYC expression after the addition of the G4 stabilizing ligand TMPyP4[63]. Another study demonstrated that a single nucleotide mutation in this G4 (A instead of G) destabilizes its structure and lead to a 3 fold increase in the c-MYC expression[50]. Due to this fact, the c-MYC G4 is considered as a target for anti-cancer therapies.

The 27 nucleotide c-MYC G4 5'T**GGGG**A**GGG**T**GGGG**A**GGG**T**GGGG**AAGG3' has been reported to form a very stable parallel G4 structure of three quartets that does not

involve the first G tract, and to have a melting temperature above 95°C[62]. It has been also reported that this G4 has four loop isomers that are all parallel stranded G4s, and can be separately formed by dual G-to-T substitutions at (14,23), (11,23), (14,20), and (11,20) positions, respectively. The major loop-isomer is the (14,23) 5'TGA**GGG**T**GGG**TA**GGG**T **GGG**TAA3' having a melting temperature of 80°C in 50 mM of k+[64] has the same structure of the WT c-MYC[65]. *In vivo,* all the four isomers contribute to the c-MYC silencing.

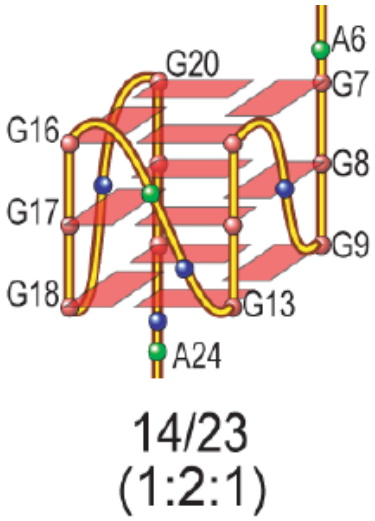


Fig. IV.2: c-Myc(14,23) G4 structure. *This parallel stranded G4 is formed by three G-quartets and have (1:2:1) loop. Guanines are represented by red beads, thymine by blue beads and adenine by green beads.*

IV.1.2 IV.1.2 Human telomeric G4

The human telomeric G4s formed in tandem by repetitive TTA**GGG** sequences on the 3' telomeric overhangs were observed in benign and cancerous human cells due to targeting by a specific G4 ligand 3H-360A[66]. However, the stabilization of telomeric G4s in malignant cells inhibits the telomeres lengthening by the telomerase, and leads to the senescence of cancerous cells. Due to this, the telomeric G4 stabilization by small molecule ligands is considered as an anti-cancer treatment.
The telomeric G4 (5'ATTA**GGG**TTA**GGG**TTA**GGG**3'), also called tel22, was shown to adopt various structures: an antiparallel basket G4 in Na^+[67], and a parallel propeller G4 structure with double chain reversal loops in k+[68]. While in k+ AAA(**GGG**TTA)$_3$**GGG**AA forms hybrid 1[69], and TTA(**GGG**TTA)$_3$**GGG**TT form hybrid 2[70].
The melting temperature of the telomeric G4 is ~60°C[71]. Many helicases have been tested on the telomeric G4 and were found to unwind it such as Pif1, WRN, BLM and FANCJ. Furthermore POT1 protein[72] and the replication protein A[73] were also found to resolve the human telomeric G4.

IV.2 This work

Due to their biological significance, the human telomeric and the c-MYC G4s and their variants have been intensively studied by different biologic and biophysical techniques. However, to our knowledge, in all the reported single molecule studies, the putative G4 sequence was single stranded and ligated to one or more DNA duplexes. Therefore, the folding of the G4 in a double stranded DNA that mimics the *in vivo* situation has not been assessed yet, nor has been investigated the real-time observation of the collision between a molecular motor and the G4 structure.

In this work, we simulate a G4 structure in the gene promoter by studying a c-MYC variant, the c-MYC(14,23), as well as the telomeric G4, and other G4 sequences embedded in the double stranded region of a DNA hairpin.

Using magnetic tweezers, we investigate the G4 folding and unfolding kinetics, the G4 stability, as well as the real-time unfolding of G4 by some helicases and polymerases that are implicated in G4 kinetics regulation.

In the next paragraphs we present an overview that might be important to the comparison to our results.

IV.3 G4 ligands

Since the finding of high thermodynamically stable G4 structures[68], [74], [75], a new research field emerged: that of G4 specific small molecular weight ligands. Nowadays, there are approximately 900 G4 ligands such as Phen-DC3, Phen DC6, TMPyP4, PIPER, QQ58, RHSP4….. A G4 ligand can be either a universal or a specific G4 ligand. Specific ligands can be used to stabilize a specific G4 in order to control gene expression or telomere lengthening. For instance, the ToxAPy ligand was found to be a G4 ligand specific for telomeric G4 coordinated by Na^+ cation and not k+[76].

The stabilization of the G4 structure by ligands was assessed in a large number of studies[77], it was shown to inhibit telomerase and G4 resolvases such as BLM and VRN[78], and regulate the expression of genes[79]. A study has reported the downregulation of c-MYC expression after adding the TMPyP4 (meso-tetra(N-methyl-4-pyridyl)porphine) G4 ligand[63]. A single molecule study done by optical tweezers has assessed the mechanical stability of a single stranded telomeric G4 that is linked to two DNA duplexes, one on each side. The unfolding force was twice higher in the presence of the PDS ligand[80].

However, the binding of the ligand to the single stranded potential G4 forming sequences might also drive the dynamic equilibrium between folded and unfolded G4 structures toward the G4 folding and act as a chaperone. Therefore G4 ligands can either enhance the G4 stability or/ induce the G4 folding .For instance, Balasubramanian et al. have reported a 4.8

folds increase of the G4 folding upon ligand addition. Wang reported that TMPyP4 ligand can induce formation of quadruplex from single-stranded DNA[81].

The G4 ligand can interact with the G4 loop(s), or by stacking on an external quartet, or also with the G4 groove. One can imagine that the ligand intercalates between the G4 quartets. But because of the high energetic cost, this interaction is improbable. The majority of G4 ligands are positively charged in order to binds the negatively charged DNA. Ligands have an aromatic cycle that facilitates the stacking interaction with the G4 quartets and have side chains that interact with the G4 loop[82].

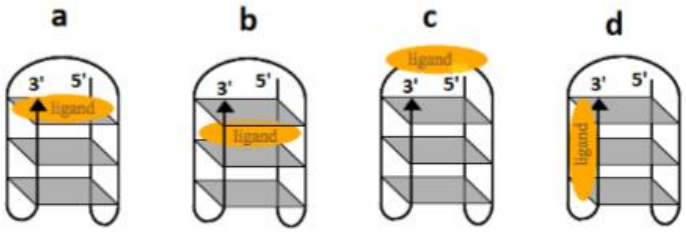


Fig. IV.3: Ligand binding to G4. *The figure shows the different ligand binding fashions to a G4 structure: a) stacking on an external quartet, b) intercalating between the quartets, c) binding to the loop or d) binding in the G4 groove.*

IV.3.1 Phen-DC3 ligand

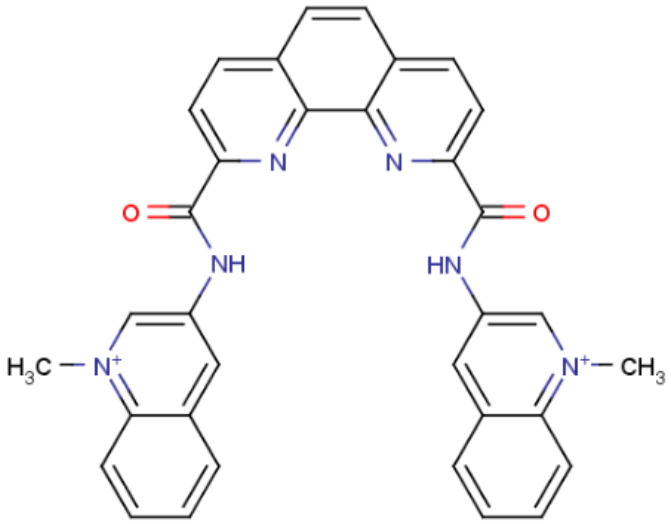


Fig. IV.4 : Phen-DC3 structure

Phen-DC3 is a bisquinolinium-dicarboxamide derivative that binds G4 structures and inhibits the fork progression downstream of the G4[83], and that exhibits a high affinity for G-quadruplexes and a high selectivity for G-quadruplexes over DNA duplexes[84], [85]. Phen-DC3 is a universal G4 ligand[84]–[86]**,** that was shown to enhance the G4 thermal

stability by 29.7°C. Structural information from NMR studies revealed that with 1:1 ratio of c-MYC G4: Phen-DC3, the ligand interacts with the G4 through extensive π-stacking on the 5' guanine tetrad, while with 1:2 ratio one ligand molecule stack on the 5' quartet and another on the 3' quartet. Phen-DC3 increases the c-MYC thermal stability by 12°C[87]. The binding of Phen-DC3 to the human telomeric G4 in k+ buffer removes one k+ ion bound to the G4 and induces a conformational change. Phen-DC3 changes the htel G4 structure from hybrid 1 or hybrid 2 to hybrid 3 having anti-patrallel characteristics ($\Delta G_{transition}$ >20kcal/mol) [88]. Phen-DC3 inhibits G4 unwinding by RHAU helicase[89], as well as the parallel CEB1 G4 unwinding by the Pif1 helicase in Yeast[86].

IV.4 G4s kinetics and stability dependence on cation type

The G4 stability is largely affected by the cation coordination. Most of the G4 studies were done in K+ and/or Na^+ buffers due the *in vivo* relevance of these cations. It is known now that the G4s are more stable in K+ than in Na^+ due to the difference between the ionic radii and the hydration energies of these cations. However, the effect of other monovalent and divalent cations has been also assessed. A classification of these cations from the most to the least stabilizing the G4 has been reported to be Sr^{2+} > k+ > Ca^{2+} > $NH4^+$, Na^+, Rb^+ > Mg^{2+} > $Li^+ \geq Cs^+$[90].

C-MYC WT and has been studied by Surface Plasmon Resonance (SPR) biosensor and has been found to be more stable in K+ than Na^+[91]. However, another study reported that c-MYC does not fold in Na^+[92]. The human telomeric G4 also has been reported to be more stable in K+ than Na^+, while the inverse is true for two successive human telomeric G4s[93].

Mg^{2+}, in the presence of K+, was shown to increase the stability of c-MYC among other G4s that form in the promoters regions where the stability of the G4 is competing with the hybridization of the c-rich complementary strand. However, the same study showed that the human telomeric G4, which usually forms in the 3' single-stranded overhang of chromosomes, is not affected by the presence of Mg^{2+} ions[94].

The stability of a G4 is the lifetime of the folded structure. The studies usually report the unfolding rate of a G4 structure (k_{off}) as the inverse of the folded lifetime. The folding rate (k_{on}), however, is the inverse of the time needed to fold the G4. Depending on the used method, the folded and unfolded G4 lifetimes can be directly assigned by the used method or indirectly deduced from other parameters. For instance, in the assays using single molecule magnetic and optical tweezers, as well as those using FRET microscopy, the lifetime can be directly calculated. However other techniques allows the calculation of the ratio between the unfolding and folding rates from the dissociation constant value ($k_d=k_{off}/k_{on}$). This ratio does not give the precise value of each rate. You and coworkers have reported the folding and unfolding rates of the c-MYC and the telomeric G4 using single molecule magnetic tweezers method in two different papers [95], [96].

IV.5 G4 and helicases

Many proteins have been reported to be implicated in G4 kinetic regulation such as helicases that resolve G4s. These helicases are important to resolve G4 structures and provide a faithful genome replication. Some of these helicases have 5'-3' directionality such that Pif1 helicase [97], while others have 3'-5' directionality such that RecQ, WRN and BLM [98]. It is not known if, *in vivo,* the helicases have a complementary role in untangling the G-quadruplexes. Some G4 unwinding helicases resolve the G4 in an ATP dependent manner such that Pif1 and WRN, while others unfold the telomeric G4 without ATP such that BLM and RHAU [99]. Pif1 and BLM helicase unwind either parallel and antiparallel G4s[100], while there exist other helicases that resolve the G4s in a conformational dependent manner such that RHAU helicase that has a specific affinity to parallel G4s [89], [101]. Some helicases are inhibited by the G4-ligand compounds while others are not. For instance, Pif1 and RHAU helicases unwind the G4s that are bound to Braco 19 ligand, while the RecQ family helicase is inhibited by this compound [78]. Similarily, Phen-DC3 inhibits the G4 unwinding by Pif1 helicase [86].

IV.5.1 Pif1 helicase and G4s

The Pif1 helicase family belongs to the SF1B superfamily and contains 5'-3' helicases that are evolutionarily conserved from bacteria to humans. These helicases have an ATP dependant DNA and RNA-DNA duplexes unwinding activity. In Yeast, Pif1 helicase is active in nucleus at telomeres, at the rDNA, and at mitochondria. Boulé et al. reported that Pif1 unwinds the RNA-DNA hybrid duplex that is formed by the telomerase and the 3' single stranded telomeric overhang leading to the inhibition of telomerase activity[102]. Pif1 is not involved in general replication fork progression but the fork progression is slowed near G4 forming sites in absence of Pif1. Furthermore, when expressed in yeast, human Pif1 suppressed both G-quadruplex associated deletions and telomere lengthening and it shows the highest specificity in G4 unwinding [103]. Besides, a biophysical study has shown that the Pif1 unfold the G4 and when acting as a monomer, it unfolds repetitively the G4 in order to keep the G-rich sequence single stranded without unfolding the dsDNA [104, p. 1]. Mendoza and coworkers demonstrate that Pif1 processes parallel c-MYC G4 and antiparallel telomeric G4 with the same rate and thus Pif1 does not have a preference for a specific G4 conformation[100]. A previous study has reported that the G4 structure stimulate the DNA duplex unwinding by Pif1[105].

Zhou et al. have also shown that G4 unwinding occurs in 3 steps in order to unravel the G4 one strand at a time. They also found that in k+ Pif1 requires ATP and needs a ssDNA overhang to unwind G4 while in Na+ the unfolding occurs even without ATP. The molecular mechanism of the G4 untangling by Yeast Pif1 helicase, one of the Pif1 family helicases, was found to occur as follow: The Pif1 translocates as it is unwinding DNA duplex, strips the first G column having the 5' end and then strips the last G column having the 3' end[89].

IV.5.2 RecQ helicase and G4s

RecQ helicases belongs to the SF2 superfamily and are highly conserved from bacteria to man. RecQ family includes Human Bloom helicase (BLM), human WRN and Escherichia coli RecQ among others. RecQ helicases are defective in Bloom and Werner syndromes[106], [107]. RecQ helicases have been reported to unwind G4. Single molecule studies by FRET revealed that RecQ unwinds G4, even without ATP, if there is a single stranded region downstream of the G4 knot. Actually, the RQC domain of RecQ binds to the G4 and its HRDC domain binds to the single-stranded segment in order to achieve the G4 unfolding. It has been also reported that in absence of this single-stranded DNA flap, RecQ cannot unfold the G4 even in presence of ATP [108].

IV.6 G4 and Polymerases

Multiple studies have reported that stable G4 structures block prokaryotic and eukaryotic polymerases [57], [109]–[111] and therefore it is believed that most DNA polymerases cannot process the G4 structures that arrest the replication fork. A biochemical study has reported that eukaryotic replicative polymerase Pol δ resolves DNA G4s when it is helped by the WRN helicase, but it is not clear if the G4 unfolding precedes the synthesis by Pol δ or occurs simultaneously; however this is not true for Pol α and for a mutant of Pol δ having only 2 subunits. The electrophoresis shift assay of the substrates replicated by Pol Epsilon indicates the resolution of G4 on some molecules[111].

PrimPol polymerase was shown to bypass G-quadruplex by repriming downstream of these structures only during leading strand replication without directly replicating the G4s [112].

The translesion (TLS) DNA polymerases are Pol η, Pol ι and Pol κ and belong to the Y-family polymerases. TLS polymerases resolve G-quadruplexes and reactivate stalled replication forks. The catalytic site of the TLS polymerases is much more opened than the error-free replicative DNA polymerases Pol δ and Pol ε [113]. This characteristic makes possible the accommodation of the non B-DNA bases in their active site. A biochemical study, however, assessed the effect of replicative and translesional polymerases on intramolecular G4 and showed that all the individually tested polymerases are inhibited by this secondary structure [114].

Further Eddy S. et al. (2015) investigated the effect of Pol ε from B family and Pol η from Y family on the c-MYC G4. They have shown that, contrarily to Pol η, Pol ε does not have a preference for G4 binding. When copying the G4 Pol ε retains only 4% of activity and hPol η shows an increase by 15 fold in fidelity, while the fidelity of hPol ε decreases by 33-fold[115].

Bacteriophage T4 and T7 are viruses that infect E.Coli. In an unpublished work of Bochman et al., the T4 replicative helicase gp41, that is one of the three helicases encoded in

the T4 genome, has been found to unwind G4. To our Knowledge, the effect of T4 and T7 polymerase on G-quadruplexes has not been tested yet by single molecule assays.

IV.7 G4 and proteins

IV.7.1 RPA and G4s

The eukaryote Replication protein A (RPA) is a single strand binding protein that binds to single stranded DNA in a non-sequence specific manner and prevents secondary structures formation. The RPA intervenes in different processes in the cell such as replication, transcription, recombination and telomere maintenance[116]. The RPA protein was shown to resolve the G4 structures[117], maintain the telomeres and activate the DNA reparation when the replication is arrested[118]. It has been reported that for cancerous human cells, the mutation of RPA subunit 1 leads to telomeres shortening[119]. It has been also shown that the human RPA regulates the telomerase activity[120], and resolves the G4 knots[73]. RPA has a higher affinity for G4 in Na+ buffer than in k+ buffer. Salas and coworkers proposed that RPA binds to the loop region and destabilizes actively the telomeric G4 [73]. It has been reported that the RPA have a 5'-3' directionality in unfolding the telomeric G4 [115]. It has been reported that the unfolding of the Gq23 (5 -TAGGGGAAGGGTTGGAGTGGGTT-3) by RPA was achieved in Na+, however in k+ it was not complete [121]. Another study reported that RPA unwinds other G4 structures with different stability than the telomeric G4. However, they have shown that the most thermodynamically stable structure (forming four quartets and having a melting temperature of Tm=86°C) is not the most stable against RPA[122]. Ray et al. reported that the stability of a G4 against RPA unwinding is enhanced by shorter loops and higher number of quartets [123]. For both studies the unfolding time by RPA was about 0.3 s.

IV.7.2 Sub1 and PC4 proteins

Sub1 is a single stranded DNA binding protein and transcription co activator that interacts with RNA polymerase. Lopez and coworkers suggests that Sub1 is a suppressor of genome instability associated with unresolved G4 DNA structure and that it interacts with Pif1 by recruiting it at the co-transcriptionally formed G4 DNA structures. Gao et al. showed that Sub1 and its human homolog PC4 preferentially bind to c-Myc (14,23) G-quadruplex DNA over ssDNA and dsDNA[124]. They reported that Sub1 preferentially binds to G4-DNA *in vitro*, with a dissociation constant (K_d) value of 0.38 ± 0.02 nM for the winged G4-DNA, and 0.48 ± 0.05 nM for the tailed G4-DNA. The authors have also shown that Sub1 does not unwind the G4.

The multifunctional transcription positive co-activator 4 (PC4) shares 34% identity and 45% homology with Sub1, it is overexpressed in cancer and is known to bind to the promoter of c-MYC to regulate its expression. PC4 was also shown to bind preferentially to the winged

G4-DNA with an affinity constant K_d of 2.1 ± 0.4 nM, and to the tailed G4-DNA with an affinity constant of K_d = 1.4 ± 0.3 nM.

CHAPTER V: Results and discussion

V.1 Folding a G4 in a double-stranded DNA

V.1.1 Simple unzipping and re-zipping of DNA hairpins

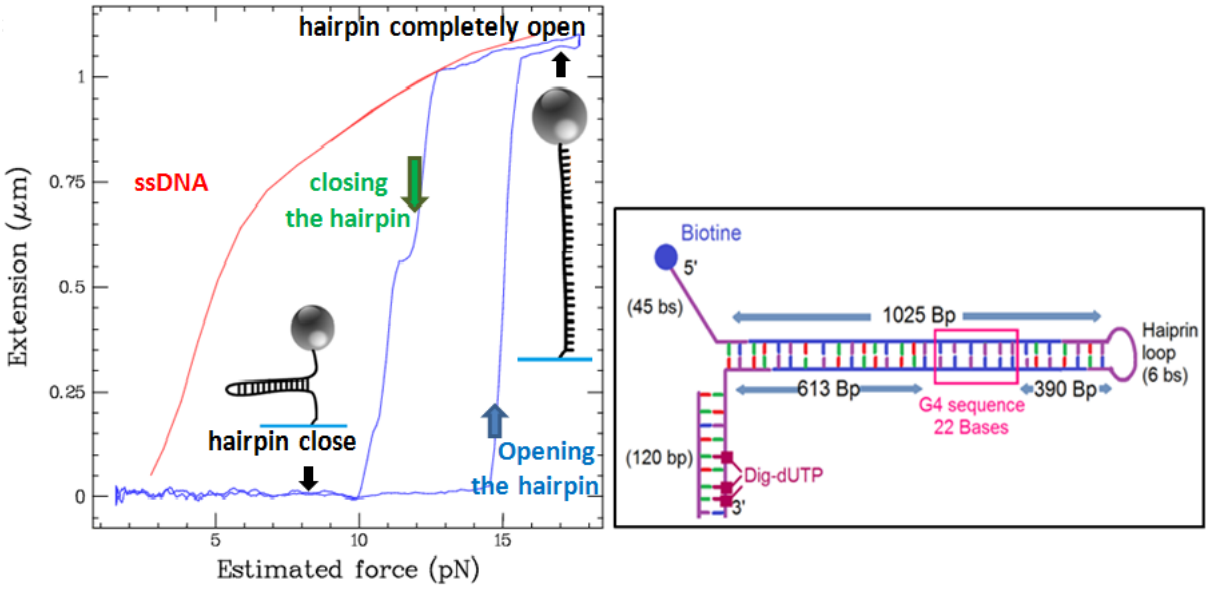


Fig. V.1: Force-Extension curves. ***a)*** *The extension of a single stranded DNA having ~ 2000 bases (red curve). The opening of a DNA hairpin of 1000 Base pairs (blue curve) occurs at 15 pN. The closing of the hairpin occurs below 15pN around 10 pN.* ***b)*** *Structure of the DNA hairpin that we have designed and that has a G4 sequence on one strand. If the G4 sequence is on the strand having the 5' end, it represents a G4 on the lagging strand. If it is on the strand having the 3' end, it represents a G4 on the leading strand.*

In order to study the G4 structure by single molecule experiments using magnetic tweezers setup, we designed DNA hairpin pairs, each having a G4 sequence of about 20 bases (fig.V.1.b). The first has the G4 sequence on its leading strand (on the strand having the 3' end) and the second has the G4 sequence on its lagging strand (on the strand having the 5' end). The hairpins are made of ~1000 Base pairs and have a 5' flanking oligonucleotide ending by a biotin in order to bind a streptavidin coated magnetic bead. On the other end, the hairpins have a double-stranded stem, and a 3' single stranded overhang with multiple digoxygenin in order to be tethered to an anti-digoxygenin coated surface. Once, the hairpins are tethered to the surface, if no vertical magnetic force is applied to pull the beads upwards, the hairpins remain closed and the micron size beads lay on the surface. Now if the magnets are brought within a fraction of millimeter to the sample, a magnetic force in the pN range is applied, attracting the beads upwards. If the force is lower than 15 pN, the hairpins remain closed. However, if the magnetic force exceeds 15 pN, the hairpins of 1000 base pairs are unzipped and become single-stranded DNA of ~2000 bases. Beyond 15 pN, the single-stranded DNA is stretched further. In order to close the hairpins, the force should be lowered to approximately 12 to 10pN. Actually, the re-zipping force is lower than the opening force. This is a signature of a hysteresis (fig.V.1.a). In fact, in order to refold, a hairpin needs to have a nucleation center at the apex involving a few base pairs that will lead to a complete refolding. Such a nucleation requires a transient shortening of the molecule allowing a small dsDNA formation, this center does not form at 15 pN, but appears at lower forces when the fluctuations become strong enough to bring the complementary strands close together.

V.1.2 Unzipping and re-zipping the DNA hairpin in presence of an oligonucleotide complementary to a segment on the DNA hairpin

Detecting if a G4 is folded or not on a DNA molecule is not completely straightforward. Previous studies have used the minute change of molecule extension (10 nm) occurring when the G4 folds or unfolds[95]. This strategy is not very reliable for long molecules: if the step size generated on folding or unfolding is within the magnetic tweezers sensitivity, it is only accessible when a high force is applied to the molecule but what is really worse is that the detection can only be done during a transition! In principle, the measure of the absolute molecule extension should reveal the state of the G4 but slow drifts in the experiment blur this measure. We have conceived an assay that circumvents this issue, using a fork closing assay, we are able to amplify the G4 signal which has been fundamental in our study. In the following paragraphs, we explain the principle of the used method applied to oligonucleotides and then to G4. Using oligonucleotides to transiently block the hairpin refolding also provides a means to maintain a hairpin in its open metastable state that will be most useful (fig.V.2).

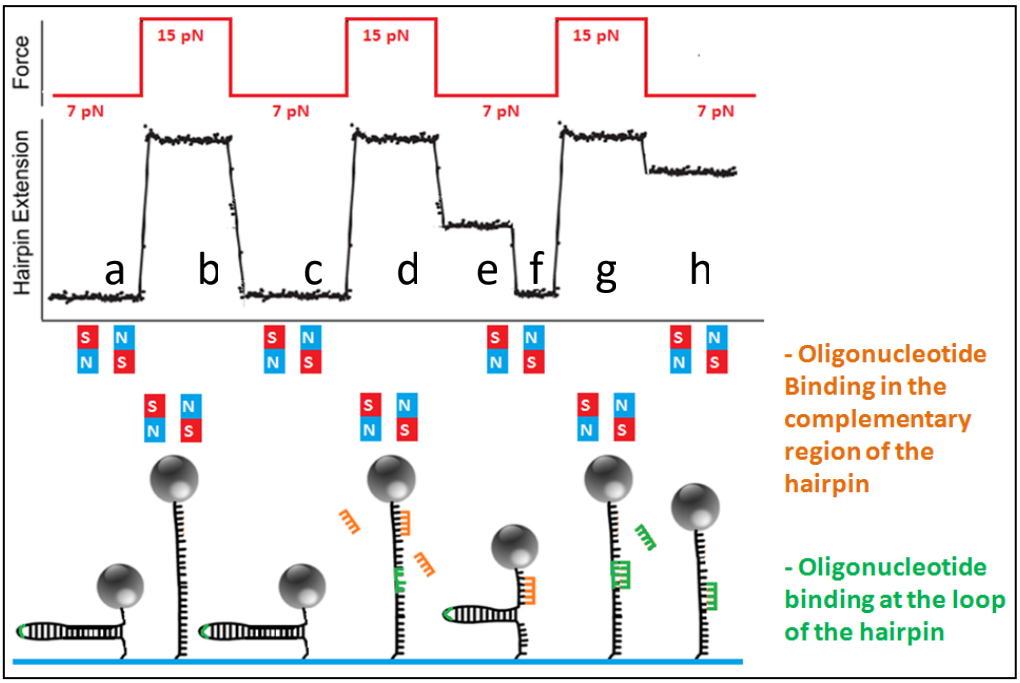


Fig. V.2: Unzipping and rezipping of the DNA hairpin in presence of an oligo. *The force cycles and the measured hairpin extension are shown on the top of this figure. At the bottom, the cartoon represents the corresponding hairpin configuration and magnetic tweezers positions.* ***(a)*** *At 7 pN, the hairpin is close,* ***(b)*** *we open the hairpin at 15 pN by approaching the magnets close to the beads,* ***(c)*** *reducing the force to 7 pN closes the hairpin completely.* ***(d)*** *When we add an oligonucleotide complementary to a segment of the duplex, it binds to the open hairpin at 15 pN,* ***(e)*** *now when the force is reduced to 7 pN, the hairpin closes partially and remains blocked by the oligo for a given time,* ***(f)*** *then when it dissociates (not showed in the cartoon but shown in the hairpin extension curve), the hairpin completely re-folds.* ***(g)*** *Now, if we add an oligonucleotide that covers the hairpin apex, when the hairpin is open, this oligo occupies the nucleation center of the hairpin and* ***(h)*** *inhibits the refolding when the force is reduced to 7 pN, keeping the hairpin completely open.*

a- Oligo complementary to a segment of the duplex

If we add an oligonucleotide that is complementary to a segment on the hairpin strand, the oligonucleotide can only hybridize to this segment when the hairpin is open, when we lower the force below 10 pN, this oligonucleotide will block the hairpin closing. Therefore the hairpin will close partially and remains blocked by the oligonucleotide for a while, before its dissociation (fig.V.2 d,e,f). When this blockage occurs, the extension of the hairpin depends on the number of bases that are open, and this is determined by the oligonucleotide position. The blockage duration depends on the length of the oligonucleotide and on the applied force. The longer the oligonucleotide, the longer the blockage and the lower the force, the shorter the blockage.

b- Oligo complementary to the apex

Now if we add in the solution an oligonucleotide that hybridizes to the hairpin apex when it is open, we strongly hinder the hairpin nucleation mechanism and this inhibits its complete re-zipping (fig.V.2.g,h). The blocking duration varies exponentially with the oligonucleotide length and also moderately with the applied force. A 8 bases oligonucleotides lead to transient blockages of a few seconds, 9 bases are stable for tens of seconds. Interestingly, when the oligonucleotide is hybridized to the loop, the hairpin can be kept open in a metastable state even if the force is decreased to a few piconewton. Adjusting the length of the oligonucleotide allows to adjust the blocking time at a given force.

V.1.3 Folding the c-MYC (14, 23) G4 embedded in the DNA hairpin in a k+ buffer

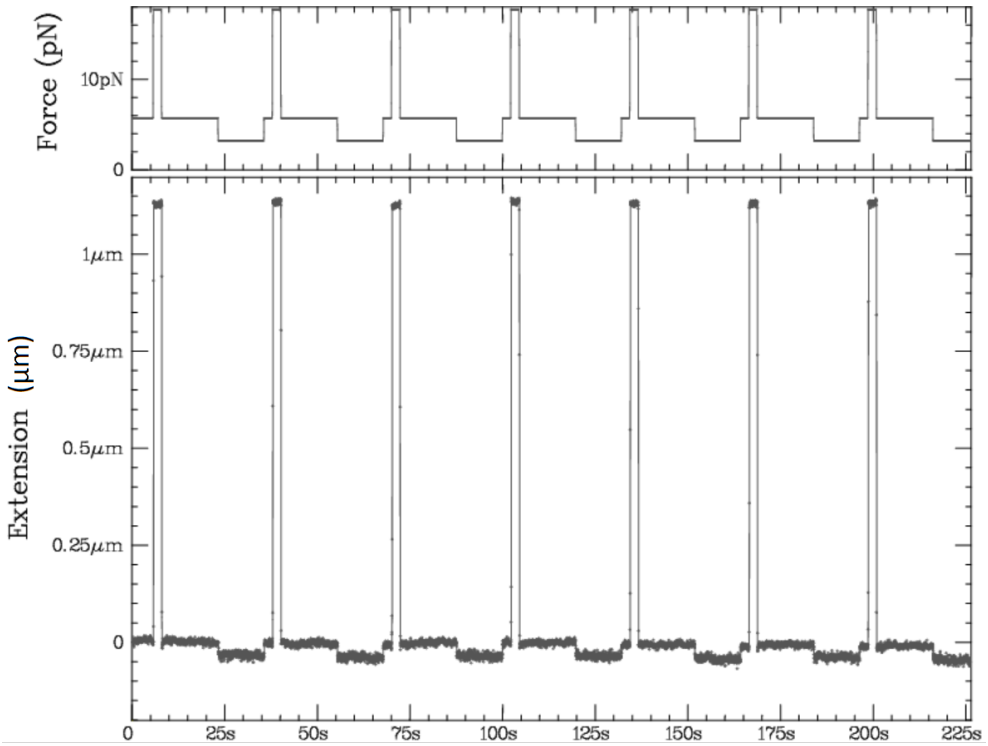


Fig. V.3: A serie of opening – closing cycles of the hairpin having a G4 sequence on one strand before the folding of the G4 structure. *On the top, the force cycles done in 60 mM k+ buffer, consisting in high force phase followed by low force one to open and close the hairpin. At the bottom, the hairpin extension shows that the molecule opens at high force and closes completely when the force is reduced below 10 PN. Clearly, no G4 structure is folded.*

If a DNA secondary structure takes place on a strand of the hairpin when it is open, such as a G quadruplexe structure, it will constitute a roadblock to the hairpin re-zipping fork. The blockage duration reflects the stability of the structure and will also depend on the applied force. The signature of this blockage is comparable to that of a hybridized oligonucleotide. It

corresponds to a partially opened hairpin when the force is lowered below 10 pN. The position of the roadblock on the hairpin defines the number of open base pairs and then the molecule extension.

In order to fold the G quadruplexes, we need the sequence to be in the ssDNA form and thus one needs to open the hairpins to separate the G-rich sequence from the C-rich sequence. In fact the G4 folding occurs after a finite time which strongly depends on the applied force. At high force this formation time becomes very large. Therefore, when we achieved a series of simple opening - closing cycles, the hairpins were opening and closing completely and no G4 structure was observed (fig.V.3).

In the simple hairpin opening and closing process, when the force is decreased below 10 pN, the hairpin immediately refolds in dsDNA which prevents G4 folding. In this case, the G4 can only fold at the high force when the hairpin is open. Because we weren't getting any G4 folding, we thought that the opening force was too high hindering the folding of the guanine quartet structure in a reasonable time. Therefore, we tested a novel strategy: We add a 7-mer oligonucleotide in the solution complementary to the hairpin with 6 bases covering the loop. Now when we open the hairpin, this oligonucleotide can bind to the loop, and when we reduce the force, the hybridized oligonucleotide will block the hairpin closing and the hairpin will remain single-stranded even at low force at least for a transient time. This situation is convenient to fold the G4 structure (fig.V.4).

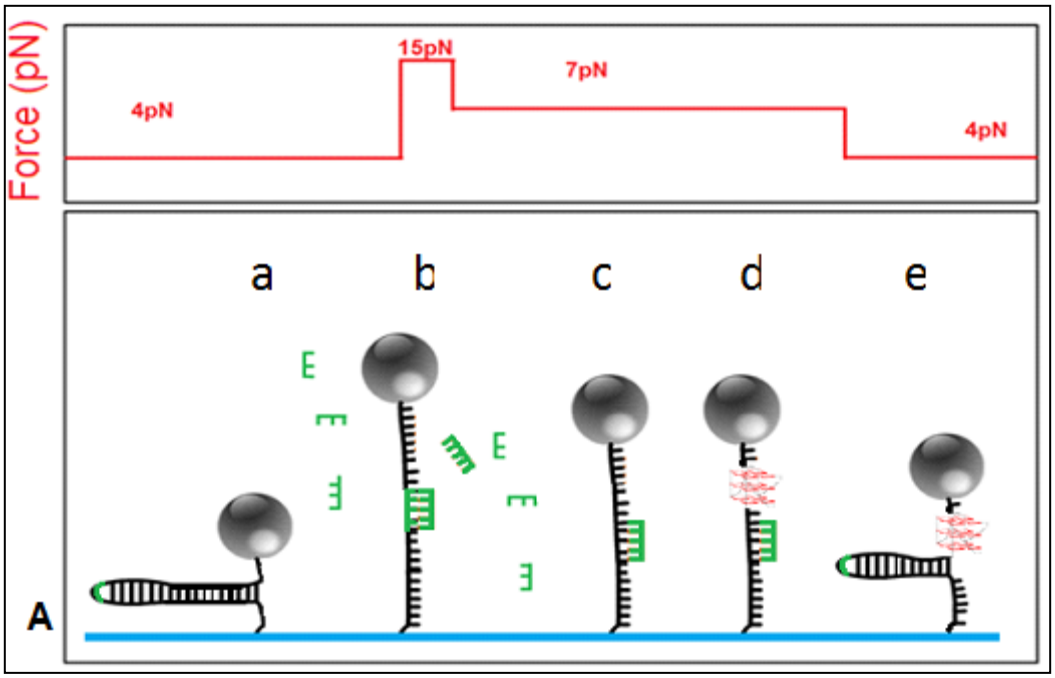


Fig.V.4: *G4 folding in a DNA hairpin substrate using an oligonucleotide complementary to the apex.*

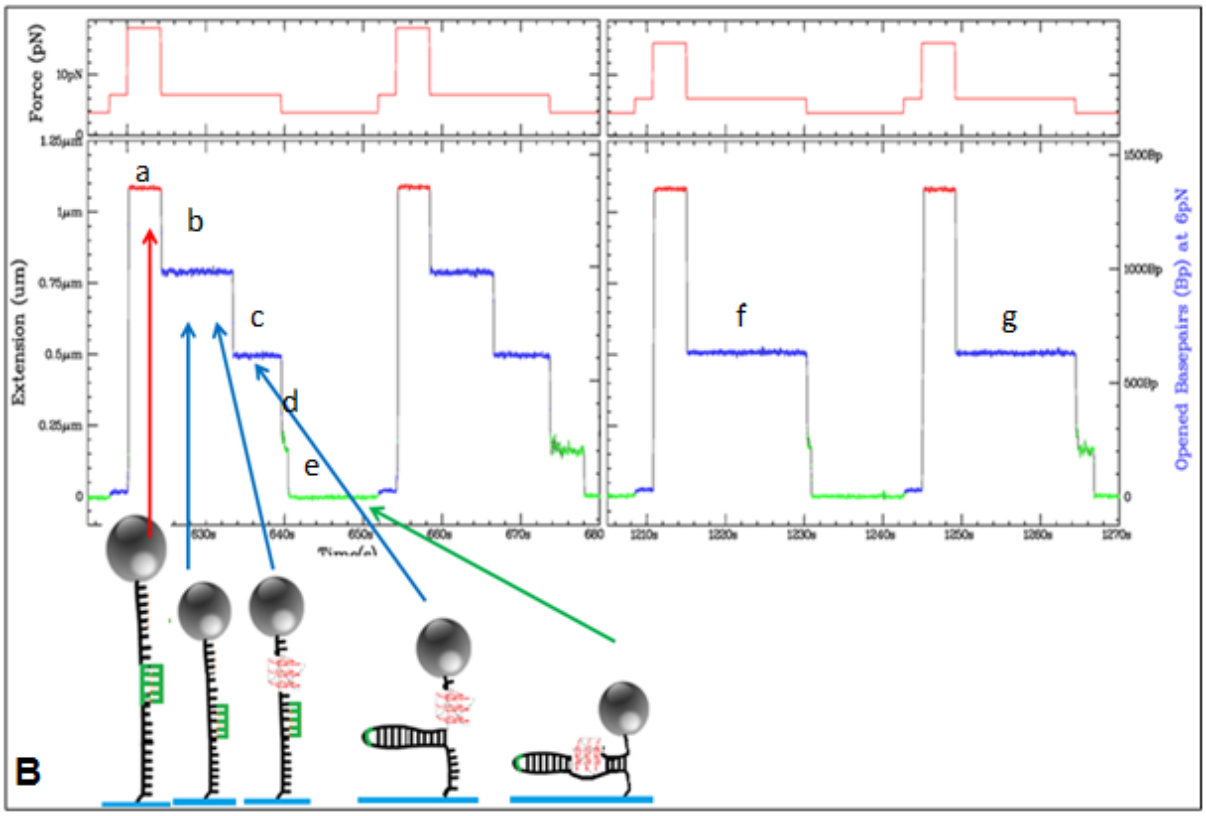


Fig. V.4: Folding the G4 structure in the DNA hairpin using an oligonucleotide complementary to the hairpin loop. ***A)*** *The method used to fold the G4.* ***B)*** *Experimental output. The panels on the top represent the force cycles. On the bottom* ***(a)*** *the red plateau represents the opened hairpin at 15 pN.* ***(b)*** *When the force is reduced to 7pN the hairpin remains open due to the hybridized oligonucleotide in the loop, but the DNA extension decreases (blue plateau at 0.75 µm).* ***(c)*** *When the oligonucleotide dissociates the hairpin refolds partially until it is blocked by the G4 structure (blue plateau at 0.5 µm). (d) When the force is reduced to 4 pN, the hairpin remains blocked by the G4 structure for a while (green plateau at 0.2 µm)* ***(e)*** *and then closes completely encircling the G4 in the duplex and the total extension is null.* ***(f,g)*** *The oligonucleotide was rinsed away but the G4 structure remains folded. The cartoon shows the time course of changes in the hairpin conformation.*

Actually, this was a successful method to fold the G4 structure in a k+ rich medium (60 mM). The graph above (fig.V.5) shows the time course extension of a hairpin on which the G4 has been folded. The hairpin opened at high force is visualized by a plateau in the extension curve (around 1 µm). At this stage, the oligonucleotide can bind or not to the loop. If it does not bind, the hairpin will close when the force is reduced below 10 pN. However if it binds, upon lowering the force, the molecule will remain open, but its extension is reduced because the force is lowered (see fig.V.1.a ssDNA force-extension curve). At this stage the G4 has a finite probability to fold. If it does not fold, upon dissociation of the oligonucleotide

from the loop, the hairpin will completely refold. But if the G4 is folded, the dissociation of the oligonucleotide from the loop will lead the fork to close only partially. As discussed before, the bases near the loop will close until the fork bumps in the G4 roadblock while the bases after the G4 will remain unpaired. Reducing the force further shows an interesting feature: The complementary bases after the G4 actually pair to each other so that the hairpin closes completely without unfolding the G4, but encircling it in the duplex. The assay is somewhat complex because both the oligonucleotide hybridization and the G4 folding lead to a fork blockage. Thus occurrence of a fork blockage needs some attention in order to distinguish its cause. Moreover, when the oligonucleotide is hybridized in the hairpin apex, it blocks the hairpin open screening transiently the visualization of the blockage of the G4. However, the positions of the two blockages are easily distinguishable and the G4 folding presents a very interesting feature that helps us to characterize it: once fold it is extremely stable while the oligonucleotide is stable for a few seconds. After rinsing the loop oligonucleotide, the force cycles then only show the blockage by the G4 structure while re-zipping the hairpin (fig.V.6). This c-MYC (14, 23) G4 structure was stable for hours; this allows us to study its kinetics and the effect of different proteins and enzymes on its stability.

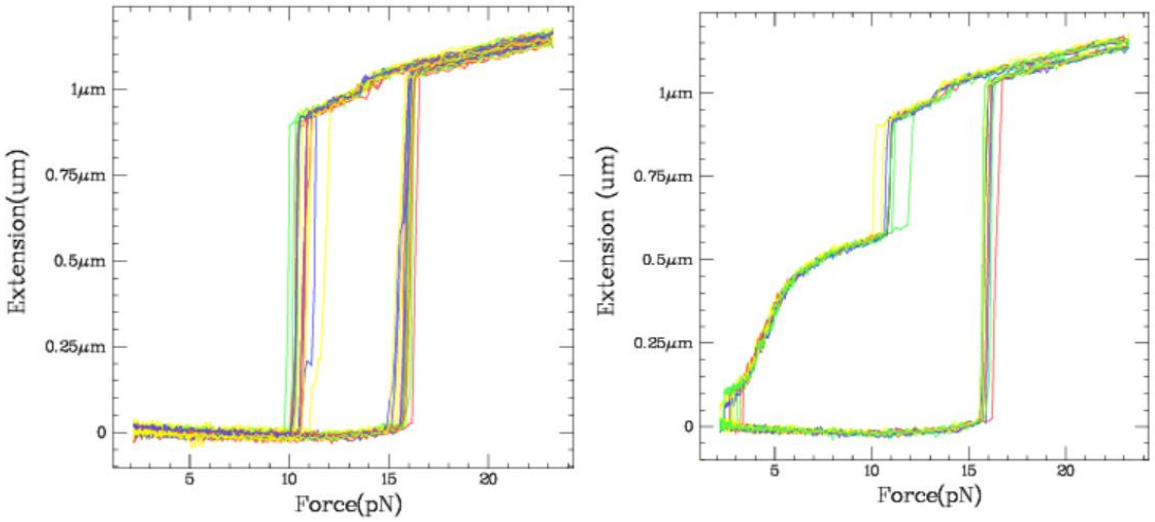


Fig. V.5: The hairpin extension curves before and after the G4 folding. *Multiple cycles of opening and closing the hairpin were achieved by slowly increasing and decreasing the force, the cycle corresponding to the same hairpin were overlapped: On the left, before the G4 folding, the hairpin opens at 15 pN and the extension increases by nearly one step from 0 to ~1 µm. Increasing the force further stretch the single stranded DNA a bit more. Reducing the force makes the hairpin closing at 10 pN, and the extension goes back to zero presenting a hysteresis curve. On the right, after having folded the G4 and rinse the loop oligonucleotide, the hairpin opening still occurs at 15 pN, but the closing is stopped by the G4 structure. The extension of the hairpin in this configuration, which is given by the number of open bases, depends on the applied force. Decreasing the force below 4 pN leads to encircle the G4 structure in the DNA duplex. The presence of the G4 in next cycles confirms that the G4 has not been unfolded.*

V.2 C-MYC (14, 23) G4 kinetics

V.2.1 C-MYC (14, 23) G4 folding rate

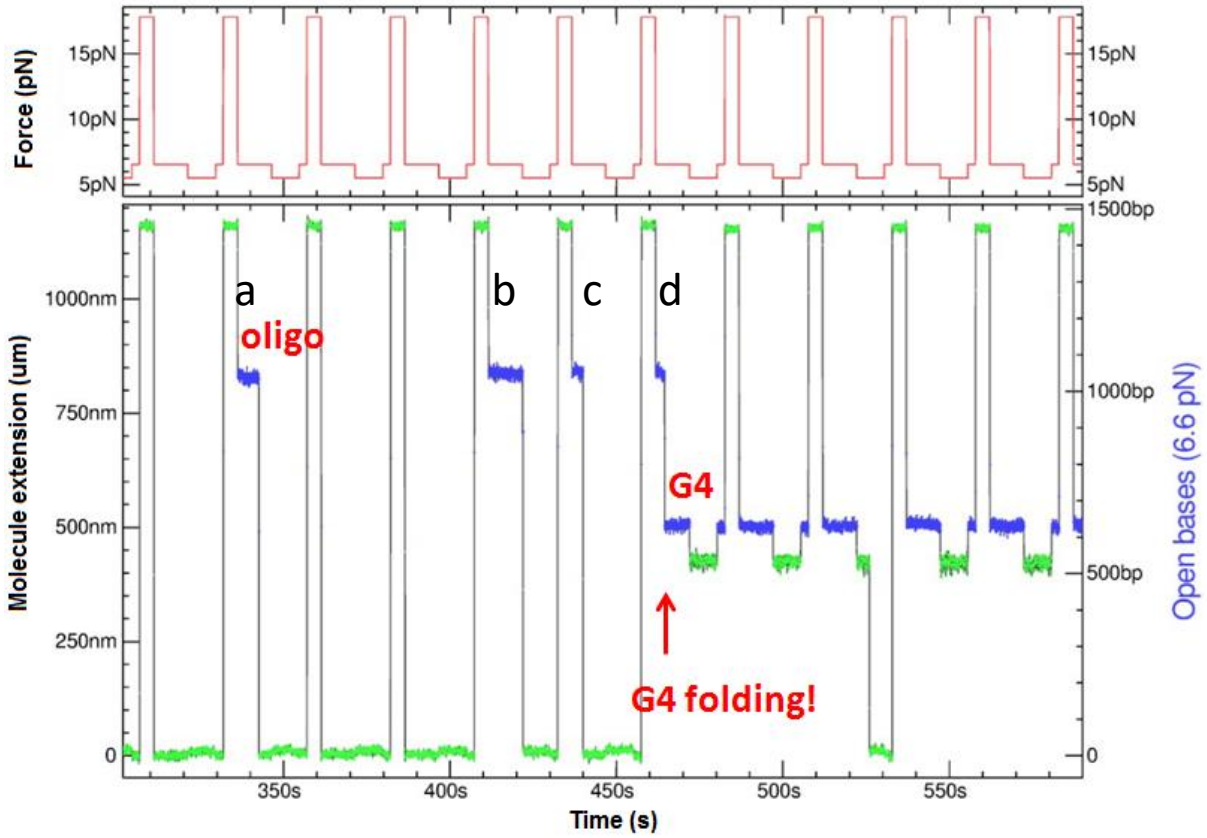


Fig. V.6: Folding the G4 in the DNA hairpin using a loop oligo. *The panel on the top shows the force cycle, while the panel on the bottom shows the molecule extension and the G4 folding. We have highlighted in blue the extension of the hairpin when it is blocked by the loop oligonucleotide (around 770 nm) as well as its extension when it is blocked by the G4 (around 500 nm) both at 6.6 pN. The nm-base pair conversion shown on the right axis was calculated at 6.6 pN (on only matches blue data points). On the graph one can see the cycles where the loop oligonucleotide did not hybridize during the opening (like in the frst cycle), and other cycles where it did hybridize but did not lead to any G4 folding* ***(a, b and c).*** *And finally,* ***(d)*** *the loop oligonucleotide blockage that leads to the G4 folding.*

Measuring the exact instant of the G4 folding on the opened single stranded DNA is not easy to observe directly when the molecule is open, the G4 folding is easy to detect when it blocks the hairpin refolding. In fact, at ~ 7 pN, folding 22 bases into a G4 shortens the molecule but by a very small amount that is hard to recognize on our setup because of the measurement noise. Therefore, in order to study the G4 folding kinetics, the folding duration

had to be calculated from the instant one sees an oligonucleotide binding at the loop till the time when a clear signature of the G4 appears (fig.V.7). However, the oligonucleotide at the loop can stay bound for a random time, thus hiding the G4 folding. In this case the G4 folding duration might be overestimated. Therefore, it is more precise to make force cycles with small transient states that consists of opening the hairpin at 15pN, reducing the force to ~ 7 pN to bind the oligonucleotide at the loop and reducing the force further to ~ 4 pN to expel the oligonucleotide if it still bound and reveals the presence or the absence of the G4 structure. By this method, the k_{on} can be calculated using the duration of all the phases that were convenient to G4 folding i.e where the oligonucleotide at the loop is hybridized until the appearance of the G4. The oligonucleotide can bind and dissociate many times before the G4 folds. In fact, the G4 folding is a rare event and thus the folding rate is well described using the Poisson distribution.

Different tests were done to fold the G4, each was done at a different folding force. This force value should be high enough to let the hairpin opened by the hybridization of the oligonucleotide in the loop region i.e. for smaller forces the fluctuations are high enough to let the hairpin close and expel this oligonucleotide, but not too high so that the interactions between the G-tracts take place and are not hindered by the stretching force. Therefore, we found that the smaller the force used to fold the G4 the more probable the folding (fig.V.8).

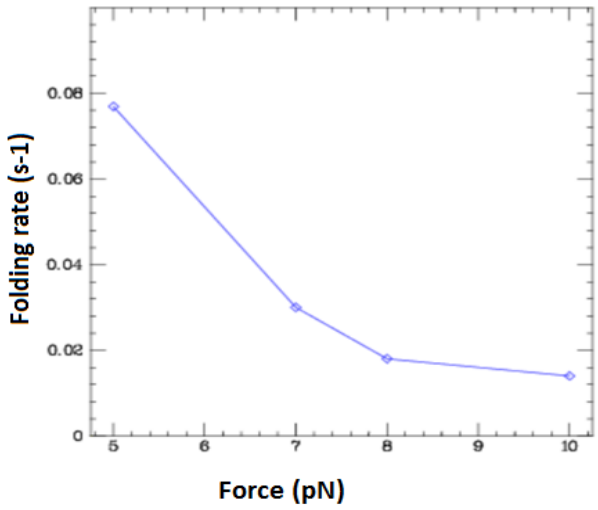


Fig. V.7: C-MYC (14, 23) G4 folding rates. *The folding rate is plotted as a function of the force at which the oligonucleotide is hybridized at the loop.*

V.2.2 C-MYC (14, 23) G4 Stability

As we have said, the G4 is extremely stable, this can be characterized by the folded time which is the duration over which the G4 remains folded. It is the duration between the moment when the first G4 blockage is seen and the moment when the G4 is unfolded.

For the c-MYC (14, 23) G4, the unfolding even has a very low probability, this means that the folded time of a G4 follows a Poisson distribution. Recording the state of 32 G4s all folded simultaneously, we can count the number of G4 that remain folded after a certain time, this leads to the exponential decrease shown in fig.V.9. The stability of this G4 may be characterized by the means of the folded time which appears to be around 6000s. That corresponds to an unfolding rate of $k_{off} = 1/6000 = 1.6x10^{-4}\ s^{-1}$ at 10 pN (fig.V.9). Owing to its high stability, the c-MYC (14, 23) G4 can be studied with enzymes.

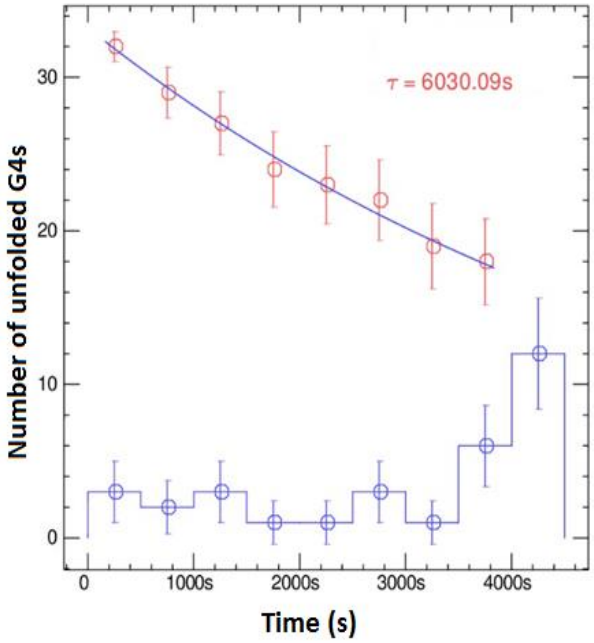


Fig. V.8: C-MYC (14, 23) G4 stability. *The histogram represents the time course (red points) of the number of folded G4s (once the loop oligo has been rinsed away). Al the histogram bins in blue represent the number of G4s that has been unfolded during a time interval. Except the last 'bin' that contains the G4s that were stable for more than 4000s, therefore the backward integrated histogram (red points) gives the number of G4s that remain folded for a given time, and the characteristic time deduced from its exponential fit, gives the c-MYC (14, 23) G4 stability that is around 2 hours.*

V.2.3 Unfolding the c-MYC (14, 23) G4 by applying an external force

Once the G4 is folded, its unfolding duration (T_{off}) depends on the applied force that mimics pulling on the G4 extremities. The higher the force, the shorter the T_{off}. We have performed a series of hairpin opening – closing, with an increasing opening force at each opening. We start with a force of 15 pN to unzip the hairpin, and increase the opening force until 65 pN. The mean c-MYC G4 untangling force was around 65 pN (fig.V.10-11).

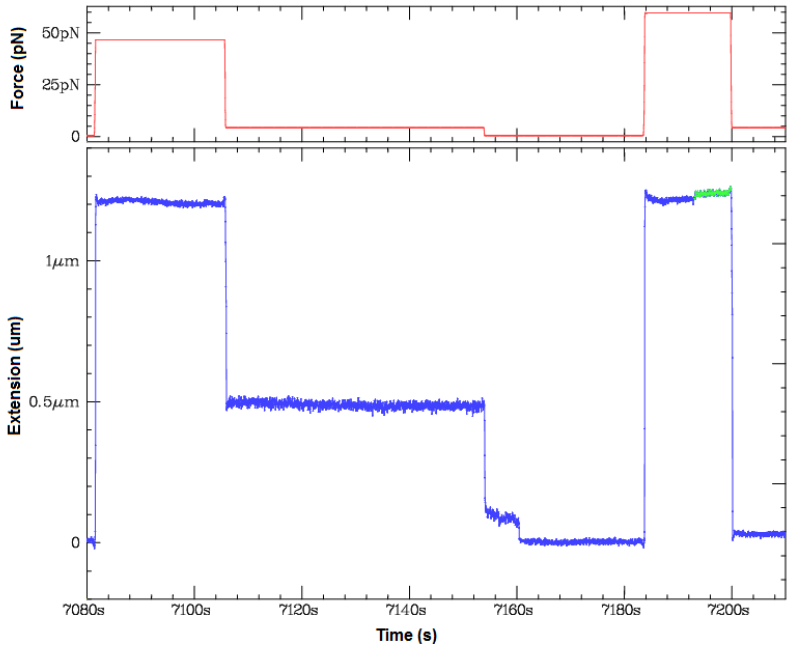


Fig. V.9: G4 unwinding by applying an external force. *Two cycles of hairpin opening – closing are shown in this graph with an increasing opening force. After the first hairpin opening, the fork closing blocks on the G4. While after the second hairpin opening, the fork closes completely indicating that the G4 has been resolved during the 58 pN surge.*

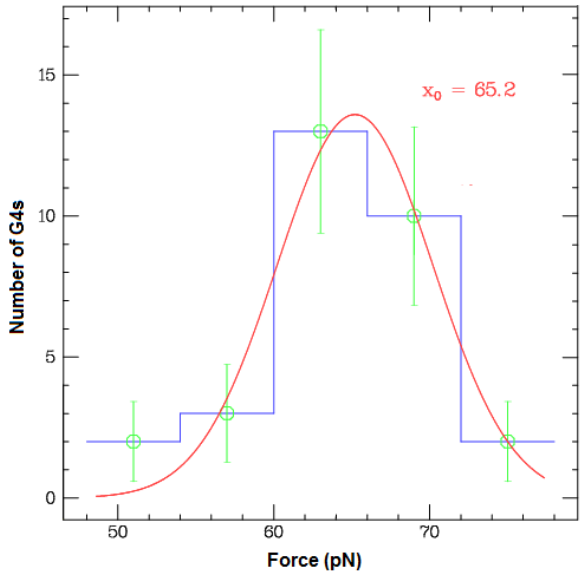


Fig. V.10: The histogram representing the number of G4 resolved at different forces. *Applying a Gaussian fit to the histogram gives the most probable G4 opening force at 65 pN.*

V.2.4 C-MYC (14, 23) G4 encircling

As we have discussed, lowering significantly the force led to a new feature: the G4 encircling by the duplex. We have measured the rate of this phenomenon which depends on the force applied on the hairpin. We first fold the G4 using the loop oligonucleotide, then we rinse this oligonucleotide, and we partially re-zip the hairpin at a medium force such that the re-zipping is stopped by the G4 roadblock. If we now decrease the force applied, the hairpin fluctuations becomes sufficient to bring closer the G4 flanking strands, the hairpin then close completely encircling the G4 leaving its complementary sequence unpaired. The encircling time depends on the applied force, the smaller the force the more probable the encircling (fig.V.12). The hairpin closing at this stage could also be the result of the G4 unfolding. To demonstrate that this is not the case we open the hairpin and reduce the force to a medium value (around 7 pN), and we observe the presence of a blockage characteristic of the G4 structure.

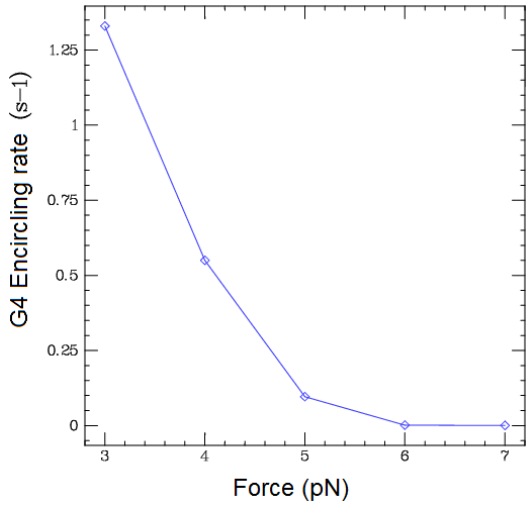


Fig. V.11: C-MYC (14, 23) G4 encircling rates. *The encircling rates are plotted as a function of the force at which the hairpin closing is blocked by the G4 structure.*

V.2.5 The G4 stability depends on cation type

V.2.5.1 Effect of Na^+ on the c-MYC (14, 23) G4 folding

In the literature we find authors explaining that the G4 stability and structure actually depends on the cation in the buffer. Since k+ and Na+ are the most relevant monovalent cations in the cell, the majority of the studies was done in the presence of at least one of them, or compared their influence on the stability of the G4s. Potassium leads in most cases to the higher stability. Telomeric G4 was shown to adopt different structures depending on the coordinating cation. Vairamani et al (Vairamani 2003) demonstrated that the ion radius is not the only factor contributing to the G4 stabilization. In fact, cations stabilize the G4 due to their ionic radius, hydration energy and coordination to the O6 of guanines. The hydration

energy of monovalent cations is proportional to their ionic radii. Kankia et al. (Kankia 2001) reported that two types of dehydrations occur in order to fold the G4: The dehydration of the cation and the O6 of the guanines O6, and the water uptake upon folding of a single strand into a G4 structure. The k+ cation facilitates dehydration because its size is slightly larger than the cavity of the G4 contrarily to the Na^+ cation that fits better in the G4 cavity. The ionic radii of k+ and Na^+ are 1.33 and 0.95 A°, respectively.

The method described above to fold the c-MYC (14, 23) G4 in a buffer of k+, was used to fold the same G4 in a buffer containing 60 mM Na^+ without k+. The Na+, like k+, is reputed to stabilize the G4 structures. However, in our experiment we couldn't fold any G4 in Na+ (fig.V.13). This indicates that this G4 could not be folded in a Na^+ buffer under moderate force. This means that in sodium buffer the folding time is considerably increased, in our assay, we apply a force that hinders the G4 folding, presumably the force combined with the lower probability of G4 folding in sodium prevent the G4 folding in our assay in a reasonable time.

However, in a buffer containing the physiologic concentrations of Na^+ and k+ (15 mM Na^+ and 150 mM k+) the c-MYC (14, 23) folding was achievable and has a folding rate of $1.6x10^{-2}$ s^{-1}.

Halder et al. (2005) reported that, in the presence of 150 mM Na^+, the c-MYC G4 folding was unfavorable. The equilibrium constant for quadruplex formation k_{eq} = kf/ku was reported equal to 0.54. This number indicates that the G4 is half of the time open. Another fluorescence study has shown that the c-MYC fold in k+ but does not fold in Na^+[92].

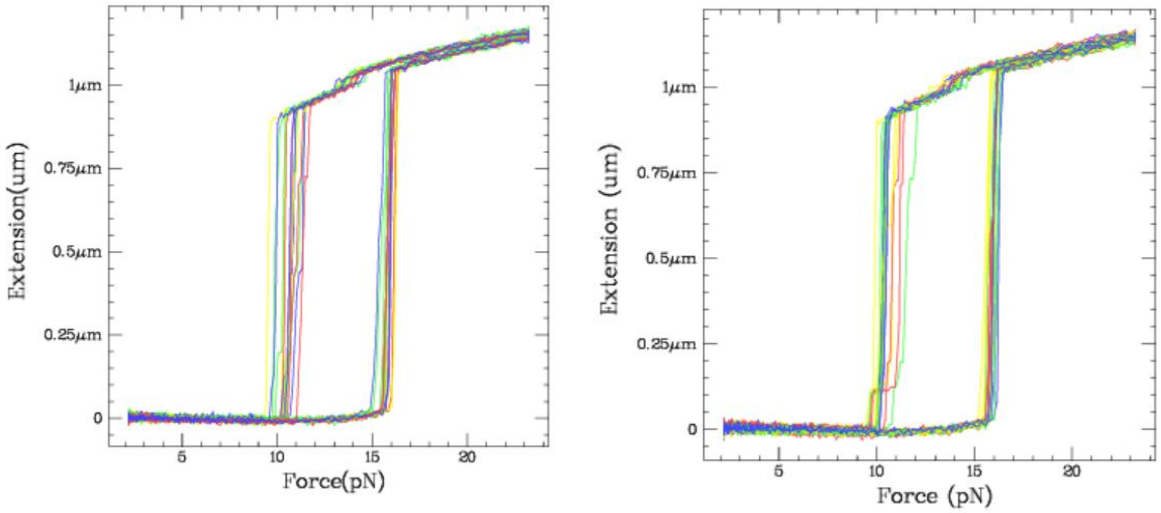


Fig. V.12: Force - extension curves before and after the G4 folding cycles in 60 mM Na^+. *On the left (right): superimposed force ramps in a buffer of Na^+ before (after) applying the folding G4 method using an oligonucleotide. Clearly the G4 did not fold in Na^+.*

The better stabilization of G-quadruplexes by potassium cations than by sodium cations is due to the greater energetic cost of Na^+ dehydration compared to that of k+in the G4

structure. The k+ cation facilitates dehydration because its size is slightly larger than the cavity of the G4 contrarily to the Na+ cation that fits better in the G4 cavity. The hydration energy of monovalent ions is inversely proportional to the ion radii[125].

V.2.5.2 Effect of Na^+ on the c-MYC (14, 23) G4 stability

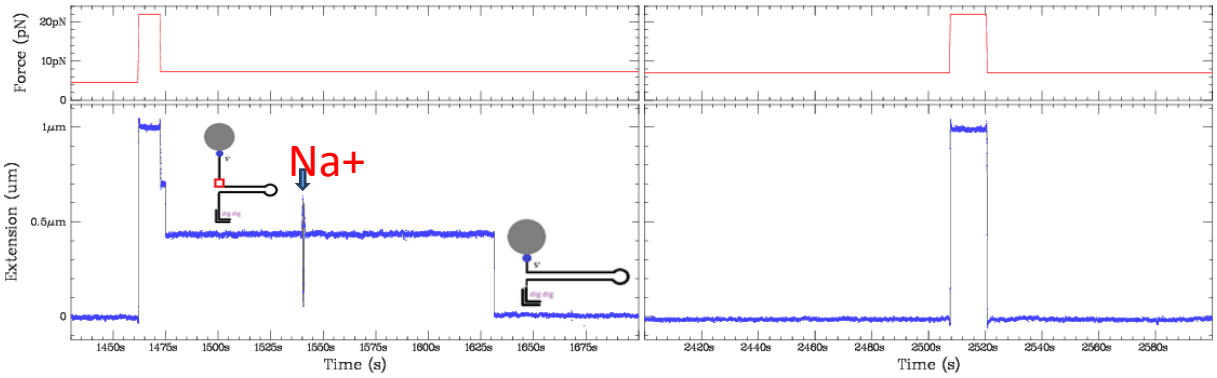


Fig. V.13: Unfolding a G4 structure by exchanging k+ by Na^+. *Na^+ is added to a hairpin that is blocked by a G4 structure on the lagging strand, the corresponding hairpin extension curve shows that the G4 is unfolded 200s after the injection of the Na^+ buffer. The corresponding hairpin extension curve shows that in the opening – closing cycle that follows the Na^+ addition, the hairpin refolds completely and is not blocked by any G4 structure. This means that the G4 has been indeed unfolded when the Na^+ has been replaced by k+.*

We have not been able to fold G4 in a sodium buffer but we still can measure the G4 stability in sodium buffer after having folded G4 in potassium. In order to check the effect of Na^+ on a folded G4 structure, we first fold the c-MYC (14, 23) G4 in the presence of k+ cations, then we partially close the molecules by reducing the force to a medium value, blocking the re-zipping by the G4 structure, and then we rinse the sample with the Na^+ buffer. The G4s remained folded for some time but rapidly unfold (fig.V.14) **with a characteristic mean time of 50s.** The same test was done with molecules having the G4 structure encircled in the DNA duplex. After rinsing the sample with the Na^+ buffer, all the G4s were unfolded.

A previous study has reported a decrease in the stability of the k+ coordinated c-MYC (14, 23) G4 upon the addition of Cs+, leading to the G4 unfolding at high concentrations of this ion. However, the same cation did not exert any influence while added to the telomeric G4[126].

However, rinsing the DNA hairpins having a folded G4 by a buffer containing 30 mM Na^+ and 30 mM k+ can keep the G4 folded for more than 3000s. This means that the k+ cation is required for the stability of this G4 structure.

V.2.6 Discussion of the c-MYC (14, 23) G4 Kinetics and stability

Various methods are used in G4 studies. X-ray crystallography, nuclear magnetic resonance spectroscopy and circular dichroïsm are used to determine the G4 folding topology. However, separative techniques, such as native gel electrophoresis and size exclusion HPLC, are used to determine the strand stochiometry of G-quadruplexes. These latter are also used to study the thermodynamics and the kinetics of these DNA secondary structures.
Single molecule FRET has been also used to study the kinetics and the effect of enzymes on G4 DNA knots. However, this technique requires labeling the G4 extremities by fluorophores, the thing that has been shown to affect the G4 folding and stability[127].

Single molecule magnetic tweezers technique has also been used to study G4 kinetics and stability as well as the effects of some enzymes on theses DNA structures. In our work, we used single molecule magnetic tweezers technique to study a DNA hairpin molecule into which we insert a G4 sequence. This configuration permits to detect the folding of the G4 by a clearly distinguishable and stable signal, and it simulates a replication fork that permits to visualize how enzymes act when encountering a G4 structure.
Therefore, contrarily to other studies that assessed the kinetics of folding and unfolding of a G4 structure in a single-stranded conformation we are simulating a G4 structure in its *in vivo* context or a G4 embedded in a double stranded DNA fragments (excepted telomeric and RNA G4s).

In most biochemical studies, G4 thermodynamics are calculated using Circular Dichroïsm (CD) from Van't Hoff analysis in order to get the melting temperature. The T_m is deduced from Arrhenius relationship that consists to plot the equilibrium constant logarithm as a function of the inverse of the temperature.

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

Also, $$\Delta G = -RT\ln(K_{eq})$$

Therefore, $$\ln(K_{eq}) = -\frac{\Delta H}{R} x \frac{1}{T} + \frac{\Delta S}{R}$$

$-\frac{\Delta H}{R}$ is deduced from the slope and $\frac{\Delta S}{R}$ is deduced from the y-intercept of the Arrhenius relationship.

ΔG: Gibbs free-energy change ($kJ.mol^{-1}$)

ΔH: Enthalpy change ($kJ.mol^{-1}$)

ΔS: Entropy change ($kJ.K.mol^{-1}$)

T: Absolute temperature (K)

R: Gas constant ($J.K^{-1}.mol^{-1}$)

k_{eq}: Equilibrium constant

a is the folded ratio and $\ln(K_{eq}) = \ln(\frac{a}{1-a})$

To calculate the melting temperature T_m, ΔG is taken as equal to zero

$$\Delta H - T_m \Delta S = 0$$

$$T_m = \frac{\Delta H}{\Delta S}$$

The G4 stability is usually expressed in term of T_m.

The equilibrium constant is also equal to the ratio between the folded and the unfolded G4s. Therefore these studies measure the ration between G4 unfolding and folding.

The stability of the G4 is equivalent to the lifetime of the folded structure. From our study, this duration can be directly measured as well as the folding duration at various force values. Therefore, we can deduce the free energy cost for G4 folding using the measured unfolding and folding rates, k_{off} and k_{on}.

For instance, at 10 pN and at 28°C, the free energy cost for the c-MYC (14,23) G4 folding is:

$\Delta G = RT\ln (k_{off}/k_{on})$

$= RT\ln (k_{unfolding\ G4}/k_{folding\ G4})$

$= 8.314 \times 10^{-3} \times 301 \times \ln (0.00016/0.04) = -13.817$ kJ.mol^{-1} $= -5.57\ k_BT$.

k_{off}: G4 unfolding rate (s^{-1})

k_{on}: G4 folding rate (s^{-1})

In previous biochemical studies, the energy cost of c-MYC (14,23) and c-MYC WT unfolding were reported of 5.48 kcal.mol^{-1} at 37°C [128], and 7.6kcal.mol^{-1} at 20°C [129] respectively.

A single molecule FRET assay reported a c-MYC2345 unfolding rate of $3x10^{-2}$ s^{-1} in 100 mM k+ and in the presence of 4000 fold of the complementary oligonucleotide [130]. Another single molecule study using magnetic tweezers has reported a c-MYC2345 unfolding cost at zero force of 6.2 kcal.mol^{-1}[95]. The study has also shown that the c-MYC WT fold in two species. The unfolding rate at a given force was deduced from Bell's model describing the force dependent unfolding rates.

$$k_u(f) = k_u(0)\exp(\Delta x\, F / k_B T)$$

k_B: Bolzman constant

T: Absolute temperature (K)

k_u (0): zero force unfolding rate (s^{-1})

k_u (f): zero force unfolding rate (s^{-1})

Δx_u:transition distance to the transition state

k_u (0) and Δx_u are determined by best-fitting the equation below, knowing the peak of the unfolding distribution $p_{unfold}^{Bell}(f)$:

$$p_{\text{unfold}}^{\text{Bell}}(f) = \frac{k_u^0}{r} \exp\left\{\frac{\Delta x_u f}{k_B T} + \frac{k_B T k_u^0}{\Delta x_u r}\left[1 - \exp\left(\frac{\Delta x_u f}{k_B T}\right)\right]\right\}$$

The study reported a zero force c-MYC WT unfolding rate in 100mM k+ and at 25°C, for the major and the minor species of $1.4\text{x}10^{-6}$ s^{-1} and $1.2\text{x}10^{-3}$ s^{-1}, respectively.
In the same study, the reported zero force unfolding rate of MYC2345 G4 (5' TGAGGGTGGGGAGGGTGGGGAA 3') is of $2.6\text{x}10^{-6}$ s^{-1}, thus ~2 times faster than that of the c-MYC WT.

In the mentioned study, we can deduce the unfolding rate of MYC2345 at 10 pN from Bell's model fit: $K_{u,10pN} = 0.66\text{x}10^{-4}$ s^{-1}. In our study, however, we have obtained a 2.5-fold faster unfolding rate of the c-MYC (14, 23) G4 at 10 pN: $K_{u,\ 10pN} = 1.6\text{x}10^{-4}$ s^{-1}. However, at 65 pN, we get an unfolding rate of 0.1 s^{-1}, this is slightly slower than the value they have gotten (fig.V.15).

On the other hand, the authors found that the folding rates of MYC2345 at 5 and 7 pN are of $3.5\text{x}10^{-2}$ s^{-1}, and $1.2\text{x}10^{-2}$ s^{-1}, respectively, while we have obtained $7.6\text{x}10^{-2}$ s^{-1} and $2.8\text{x}10^{-2}$ s^{-1} for the c-MYC (14,23). A bulk assay using FRET reported a folding rate of $5\text{x}10^{-2}s^{-1}$ of the c-MYC (14,23) upon mixing with KCl [131]. The G4 c-Myc (14,23) was also studied by Eddy et al. using fluorescence quenching assay, the obtained folding rate was of $2.29\text{x}10^{-2}$ s^{-1} at 40 mM k+[132].

The measurement discrepancies can be due to the difference in the stability between the c-MYC2345 and the c-MYC (14,23). Furthermore, the DNA constructs containing the G4 sequences were not the same in the two studies. In the previous study, the G4 forming sequence was single-stranded linked from both sides to a DNA duplex, while we have studied a G4 embedded into a double stranded region.

The values that we obtained in the thermodynamic and kinetic study of the c-MYC (14,23) are slightly different but in the same range of those obtained by other techniques.

In order to calculate the c-MYC (14,23) G4 unfolding cost in Na^+, we suggest to use the folding rate in k+ because the c-MYC (14,23) does not fold in Na^+.
From our measurements at 7 pN, $k_{off\ Na+}/k_{on\ k+} = 0.02/0.04 = 0.5$

$\Delta G_{unfolding\ Na+} = -RT\ln (k_{off\ Na+}/k_{on\ k+})$

$=-RT\ln (k_{unfolding\ Na+}/k_{folding\ k+})$

$=-8.314 \times 10^{-3} \times 301 \times \ln (0.02/0.04) = 1.734$ kJ.mol^{-1} $=0.7k_BT$

For the c-MYC (14, 23) G4, we can calculate the ratio from the previous calculated folding energies in Na+ and k+:

$$\frac{\Delta Gunfoldingin\ Na+}{\Delta GunfoldingK+} = \frac{\Delta Gunfoldingin\ Na+}{-\Delta GfoldingK+} = 0.125$$

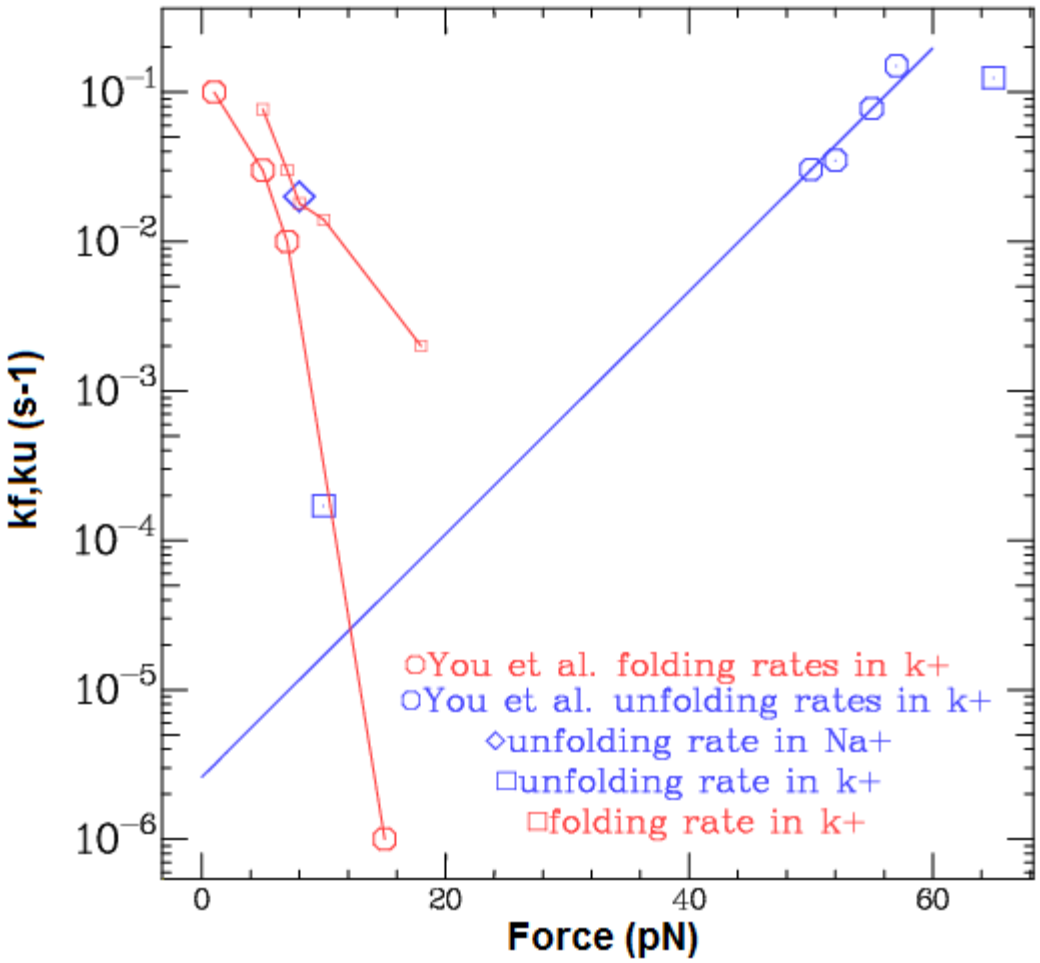


Fig. V.14: C-MYC (14, 23) folding and unfolding rates. *The folding and unfolding rates are plotted as a function of the force. The folding and unfolding rates obtained by a previous study [95] are represented by red and blue circles, respectively. These values are fitted by Bell's Model. The folding and unfolding rates that we obtained in 60 mM k+ are represented by red and blue squares. The folding rate value at 20 pN is in fact a superior limit to the real folding value, because at this force and during the assay duration, no G4 has been folded, and thus the folding duration is longer and the rate is smaller than the represented value. The unfolding rate that we obtained in Na+ is represented by blue diamond.*

V.3 Folding other G4 sequences in the hairpin

As G4 stability varies a lot with their sequences, other hairpins containing other G4 sequences have been constructed and tested in our setup (sequences are listed in table V.1).

G4	G4 sequence
Modified-c-MYC	5'TGAGGGTGGGTAGGGTGGGTAA3'
Modified-c-MYC triplex	5'TGAGGGTGGGTAGGGTAA3'
Telomeric-G4	5'GGGTTAGGGTTAGGGTTAGGG3'
2-Telomeric-G4s	5'GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG3'
G4C2	5'GGGGCCGGGGCCGGGGCCGGGG 3'
2-quartets-G4	5'GGATGGTGGTGGTT 3'

Table V.1: sequences of the tested G4s.

a. Modified c-MYC triplex

When we performed the cycles to fold the c-MYC (14, 23) quadruplex in 60 mM k+ buffer, we noticed on multiple molecules that a brief blockage of few seconds occurred before the folding of the stable G4 structure. This suggests a folding of an intermediate G4 structure probably a G triplex. In order to check if it is true, we designed a hairpin having a G triplex structure derived from the c-MYC (14, 23) sequence. But we have not succeeded in getting any folded structure. This might indicate that the intermediate structure results from another sequence and that it is unstable. This intermediate might involve non-consecutive G tracts.

b. Telomeric G4

The T_m of telomeric G4 in 70 mM k+ was reported to be 63°C (Balagurumoorthy 1994). The free energy cost of unfolding of k+ induced telomeric G4, G_0, has been reported in a wide range of 3.4–14.8 kcal mol^{-1} [88].

Folding the telomeric G4 in the double-stranded DNA was successful in a 60 mM k+ buffer, but this structure was not very stable. We find that the folding rate is $k_{on} = 1.3x10^{-3}$ s^{-1} at 8 pN, and that the folded structure lived, on average, for 22 s (fig.V.17.a) thus its $k_{off} = 4.5x10^{-2}$ s^{-1}. Adding 500 pM of the G4 ligand Phen-DC3 increased its average stability to 78 s (fig.V.17.b). In a 60 mM Na^+ buffer, however, the folding of this G4 in the hairpin was not achievable. A previous single molecule studies using magnetic tweezers (Huijuan You 2014) that studied the human telomeric G4 in 100 mM k+, has reported the folding durations of 15.8 s at 5pN, 53 s at 6 pN and 101 s at 7 pN, corresponding to a k_{on} in the range of 10^{-2} - $6x10^{-3}$.

They have also reported two unfolding distributions, with a short and a long lifetime, and both depend on the applied force. For instance they found that the mean G4 lifetimes for the two distributions are:
- at 5 pN, 3 and 45 s

- at 6 pN, 2.1 and 25 s
- at 7 pN, 3.2 and 16.6 s

Thus the mentioned study has reported an unfolding rate in the range of 10^{-2}-1 s^{-1}, and a zero force unfolding rate of $9x10^{-3}$ s^{-1}.

Therefore our results are in the same range of the results in the mentioned study.

c. Two consecutive telomeric G4s

A hairpin containing two consecutive telomeric G4s has been designed. On the cycles, we see the G4 folding as seen before with the hairpin having only one G4. However, after some seconds the hairpin does not open completely anymore (fig.V.16). We think that at this stage, both of the G4s have been folded and have locked the hairpin opening at the G4 level.

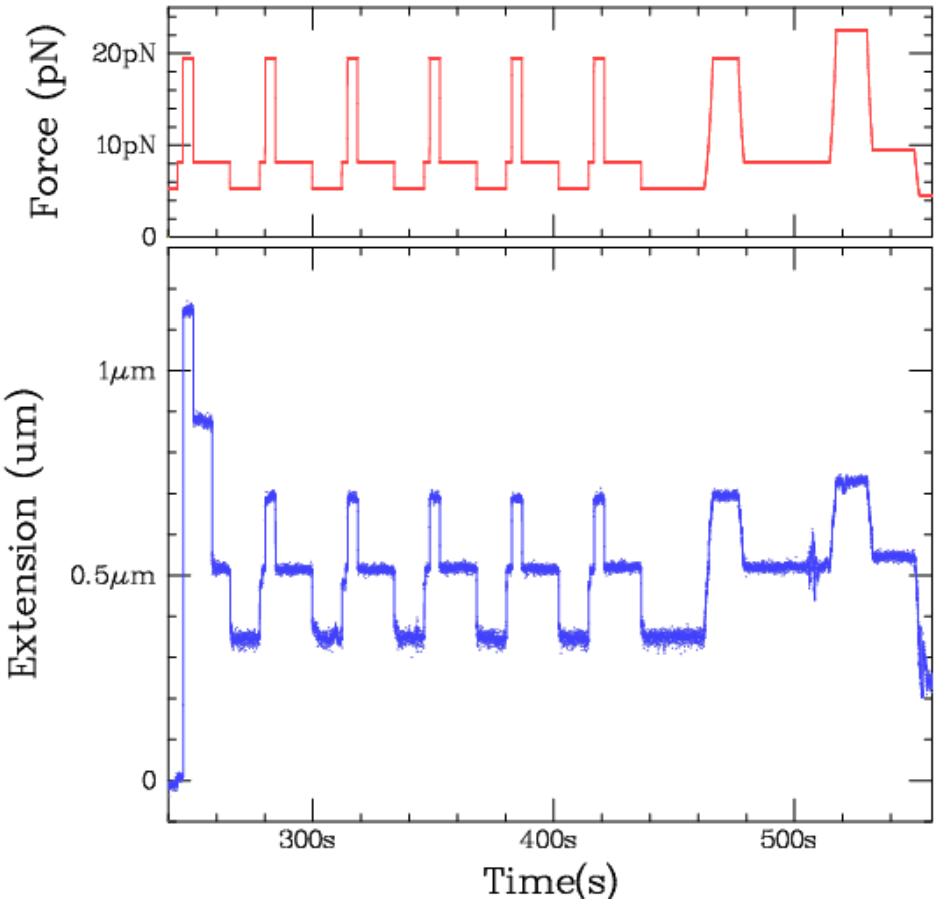


Fig. V.15: hairpin having 2 consecutive telomeric G4s. *The force cycles are shown in the top panel. In the panel on the bottom, in the first cycle, the hairpin is completely open, the G4 has been already folded before those cycles. However, from the second cycle, the hairpin does not open completely anymore. The test was done in 60 mM Na+ and 1 nM Phen-DC3.*

In a previous study, the authors demonstrated that the G4 structure formed by two consecutive telomeric G4s is more stable in sodium than potassium. And this is due to the interactions between G4 structural domains that can make a higher-order structure more stable in sodium than in potassium, even though its G4 structural domains are individually more stable in potassium than in sodium[93].

d. G4C2 quadruplex

Expansion the GGGGCC sequence repeat within the first intron of the C9orf72 gene causes neurodegenerative disorders such that Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Previous NMR studies shows the coexistence of two antiparalleL predominant G-quadruplex species[133]. With this sequence we got few structures folded at 150 mM k+. Adding Phen-DC3 ligand increases the number of folded G4s. However we found two distributions of the G4s lifetime. The mean lifetimes of the two distributions are 38 s and 1836 s, it corresponds to the lifetimes of two different G4 topologies or a G4 structure with and without the ligand.

e. 2-quartets G4

This G4 of two quartets did not fold without a ligand.

V.4 G4 stability

V.4.1 Stability of different G4s sequences with and without phen-DC3 in k+ buffer

In order to study the stability of the different G4s sequences mentioned above, the G4s were folded using “the oligonucleotides complementary to the loop” method, as explained previously. Then the hairpin was set partially open while blocked by the G4 structure. The G4 lifetime was measured as the duration between the first G4 blockage and the moment when the hairpin refolded after the G4 is unfolded. For the c-MYC (14, 23) G4, the stability was around 6000s. For the telomeric G4, the lifetime was typically 20 seconds. Adding 500 pM of Phen-DC3 ligand, increases the stability to ~80s (fig.V.17). The telomeric G4 stability in a double-stranded DNA configuration, even with Phen-DC3 ligand, is not long enough for enzyme studies. In fact, it is highly probable that the unfolding of G4 occurs before being reached by enzymes. For the G4C2 G-quadruplex stability studied in 150 mM k+ and in the presence of 1 nM Phen-DC3 shows two distributions that might correspond to two different possible structures of this G4 bound to the ligand. It might also correspond to G4 structures with and without a bound ligand (fig.V.18).

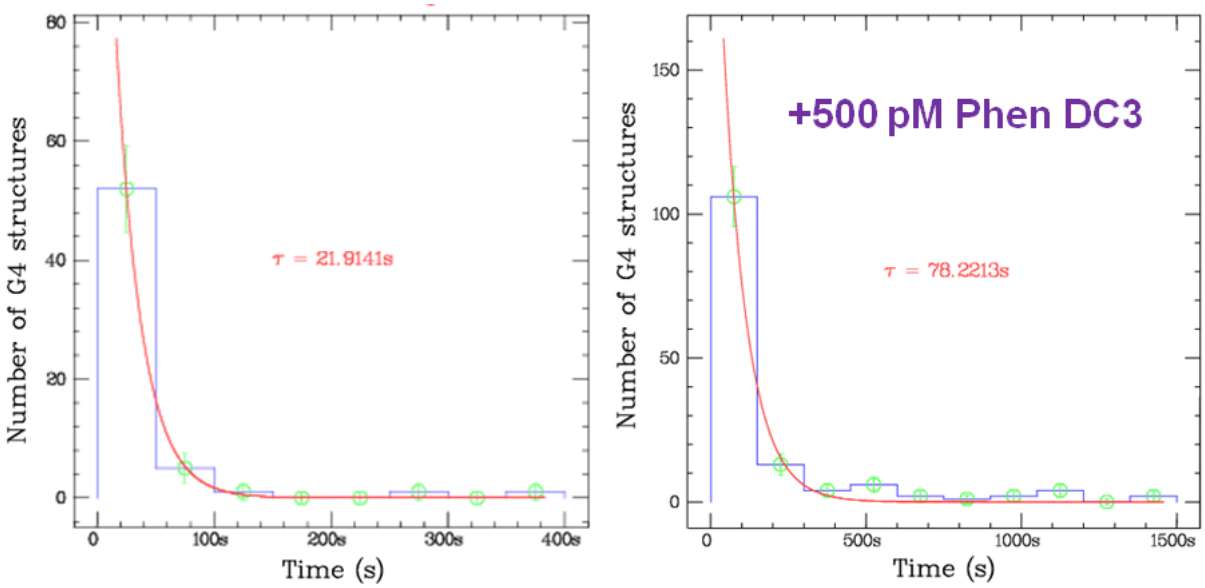


Fig. V.16: Human telomeric G4 stability. *On the left, the histogram representing the number of folded G4s during the time is fitted with an exponential decay having a mean lifetime of 22 s. On the right, in the presence of 500 pM of Phen-DC3 ligand, the mean lifetime is increased to 78 s.*

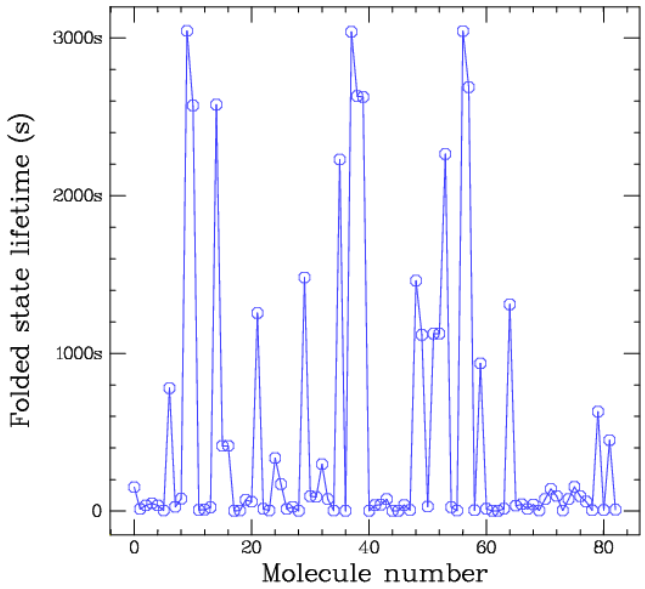


Fig. V.17: G4C2 lifetimes in 150 mM k+ and in presence of 1 nM Phen-DC3. *The graph shows two lifetime distributions.*

Phen-DC3 compound has been shown to enhance the thermal stability of telomeric G-quadruplexes by 29.7°C[85]. It has been reported as universal ligand that binds to different G4 topologies. Upon testing Phen-DC3 ligand on the Pu24 G4, one of the parallel c-MYC G4

mutants, it was shown to stack on the external G tetrad[87], confirming that two Phen-DC3 molecules can bind to a Pu24 G4 structure. The thermal stability of the Pu24 G4 was increased by 12°C.

V.4.2 Increasing the G quadruplex stability

Measuring folding time or unfolding time is interesting but we are facing an issue that is not simple to circumvent: As an experimentalist we can detect lifetimes that fold in a finite range bound for a small time by our resolution (a few seconds) and on a long time by our patience and also the lifetime of a single molecule studied. This last time is a few thousands seconds. But lifetime varies exponentially with energy and adding a ligand that stabilizes G4 leads to a strong increase in the G4 lifetime that can easily escape our measuring window. For instance, if we add a ligand to the c-MYC in k+ buffer the lifetime is so large that we cannot measure it. Thus we have combined destabilizing effect with stabilizing one in order to measure the stability increase by a ratio of time (fig.V.19). To be specific, to evaluate the effect of Phen-DC3 ligand on c-MYC stability, we have done the experiment in Na^+, which in absence of G4 ligand, decreases the G4 lifetime to 50 s as mentioned before.

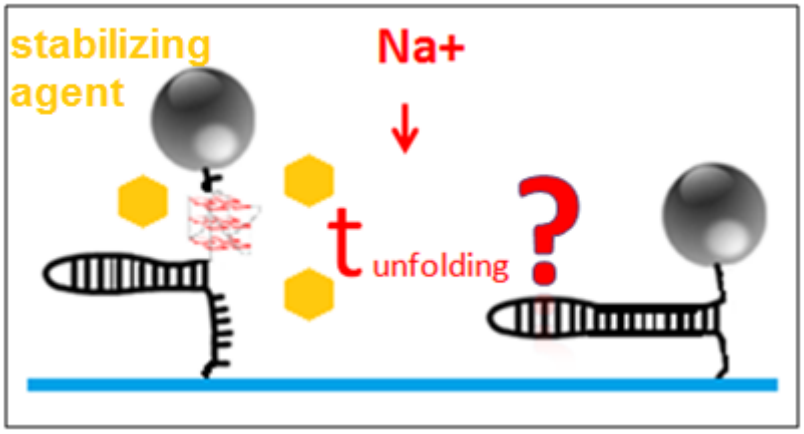


Fig. V.18 : G4 stabilizing agents. *In order to measure the stability of the c-MYC G4 in the presence of a stabilizing agent, first the G4 is folded, then the stabilizing agent is added to DNA molecules, and the chamber is rinsed by Na^+ buffer containing the stabilizing agent.*

The results show that Phen-DC3 increases the c-MYC lifetime in Na+ for hours (table V.2). The same assay was done by replacing Phen-DC3 by the Yeast Sub1 protein and then by the human homologue of Sub1 protein, the PC4 protein. First, we fold the c-MYC (14, 23) G4 in 60 mM k+ buffer, then we add the tested stability agent (20 nM of Sub1, or 20 nM of PC4 or 2 nM of Phen-DC3) and we inject the buffer containing 60 mM Na+ as well as the stability agent. We measure then the time needed to unfold the G4. As a result, the G4 stability was increased 20 folds with the Sub1 protein, 50 folds with the PC4 protein, and 60 folds by the Phen-DC3 ligand.

Stabilizing agent	G4 unfolding duration in Na+ (s)	Ratio
-	50	
Sub1 protein	1000	20
PC4 protein	~2500	50
Phen DC3 G4 ligand	3000	60

Table V.2: G4 unfolding duration in a buffer containing 60 mM of Na+ in absence and in presence of a stabilizing agent. *The ratio between the unfolding duration in presence of the cofactor and that in the absence of the co-factor is represented in the last column.*

Many proteins have been reported to stabilize G4 structures, such that the ribosome-associated protein Stm1p that bind to telomeric G4 and protect the telomeres [134], and the cellular protein nucleolin that stabilizes G-quadruplex structures folded by the LTR promoter [135]. Here we have shown that Sub1 and PC4 proteins, which have a high affinity for the c-MYC G4 structure over single - and double - stranded DNA, also stabilize the c-MYC (14, 23) G4.

V.4.3 Promoting the G-quadruplex folding

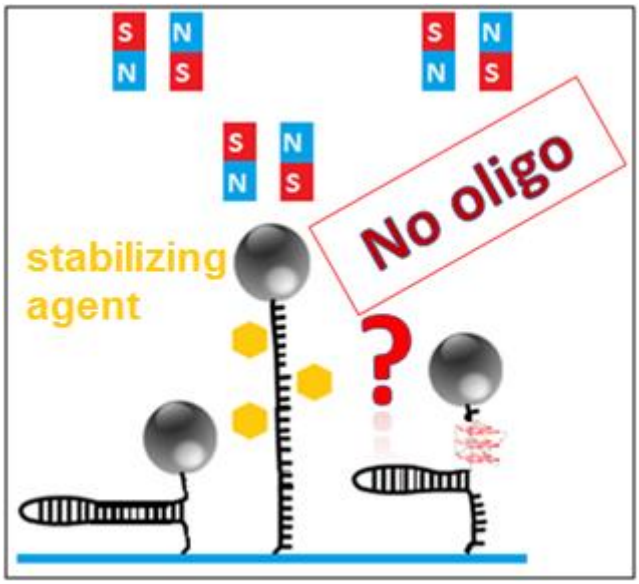


Fig. V.19: G4 chaperone assay. *To test if a protein or a ligand acts as a chaperone, a series of opening – closing cycles has been done in the presence of the cofactor and in the absence of the loop oligo. The cofactor is then rinsed away out in order to check the G4 folding status.*

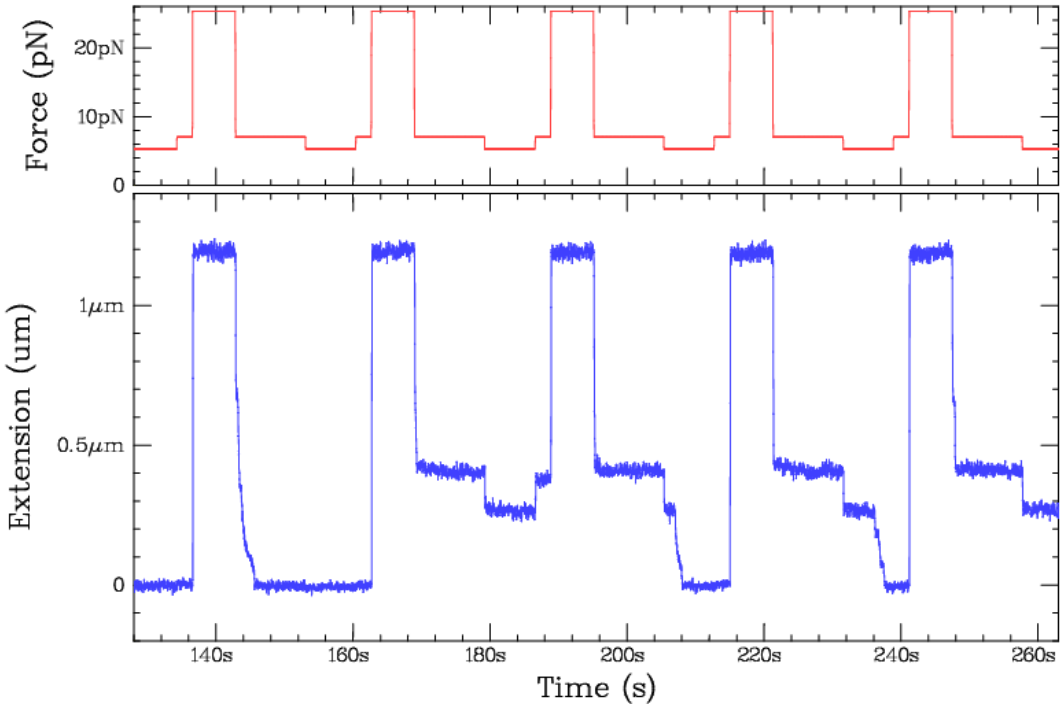


Fig. V.20 : G4 folding at high force in the presence of Sub1 protein. *Sub1 protein acts as a chaperone. It makes possible the G4 folding at high force.*

Co-factors can stabilize G4 by increasing their folded lifetime, but they can also decrease their folding time. We have done this with Sub1 and PC4 proteins (2 different test using 20 nM of the protein) as well as Phen-DC3 ligand (2nM) in order to check if they can act as chaperones for G4, in other words promoting the G4 folding. In this test we do not add the oligonucleotide complementary to the loop but we add the tested stabilizing agents in a 60 mM k+ buffer and we tried to fold the G4 by doing a series of opening closing cycles (fig.V.20). These cofactors allowed the folding of the G4 at high force (around 20 pN) when the hairpin was open (fig.V.21). This result confirms that they do not only enhance the G4 stability but they also promote the G4 folding (table V.3).

Stabilizing agent	folded G4s (%)	Folding time (s)
-	0	-
Sub1 protein	30	412
PC4 protein	12	892
Phen DC3 G4 ligand	12	700

Table V.3: The number of folded G4s in the presence of different proteins and of the Phen-DC3 ligand that acts as a G4 chaperone.

Some ligands, such that TMPyP4[81] and PIPER[136], have been shown to promote G4 folding[137]. In this work, we have found that the universal G4 ligand, PhenDC3 also induce the G4 folding. Some proteins have been also reported to promote G4 folding such that the telomere-binding protein in Oxytricha, and Rap1 protein[138]. In our assay, we have found that the G4 binding proteins Sub1 and PC4 promote the G4 folding, while other proteins such that SSB and RPA do not (The assays with SSB and RPA are reported in the next paragraphs).

V.5 Does the c-MYC (14, 23) G quadruplex form a roadblock in an enzyme's path?

The mechanistic view of DNA replication and transcription involves molecular motors (helicases and polymerases) moving along DNA. The existence of roadblocks on the DNA is now an accepted picture. Some are made directly by the cell, for instance, in *E.coli*, Tus protein binds Ter DNA sequences forming a one-way barrier that will arrest the first replication fork and prevent a second fork from entangling the DNA[139]. This roadblock has a direct interaction with the replicative helicase, DnaB. Other roadblocks appear when a protein remains stuck on the DNA after encountering damage or a mis-functioning, this is the case for RNA-pol which is the major causes of roadblocks in the cell [140]. The folded G4 forming a knot on ssDNA is likely to disrupt the normal functioning of helicases and polymerases and also to alter the action of RNA-pol. In the following, we address the question of whether G4 is a roadblock by directly testing the interaction of helicases and polymerases with G4 structures.

The high stability of the c-MYC G4 is ideal for these tests using single-molecule manipulation. We have arried out our experiment by preparing the DNA molecule with a single G4 structure and ensuring that this structure was folded before we injected the test enzyme, guaranteeing that the enzyme will encounter the G4 on its course. Using our G4 folding preparation method, after folding the G4, we were able to test various conditions:

1) We could keep the hairpin blocked at the G4 by keeping the force around 7 pN, and observe if the G4 is removed.
2) We could encircle the G4 in the dsDNA by reclosing the hairpin at low force $F < 4$ pN. In this situation, the enzyme must first open the hairpin and then encounter the G4.

Finally, we have prepared hairpins with a G4 on the lagging or leading strands. The collision between the helicase and the G4 may occur while the helicase is either opening the hairpin (if the G4 is encircled in the hairpin) or when it is translocating on the single-stranded DNA (if the hairpin is blocked by the G4 structure).

V.5.1 Bacteriophage T4 replisome

The gp41 is the replicative helicase from the T4 bacteriophage. It is a hexameric helicase that has a 5'-3' directionality and belongs to the SF4 superfamily of helicases. The gp41 helicase, along with the holoenzyme that is formed by the polymerase gp43, the clamp gp45 and the clamp loader gp44/62, constitute a minimal T4 replisome. A functional coupling takes place between the helicase and the polymerase and leads to efficient replication on the leading strand. For instance, in the presence of the whole replisome, the gp41 becomes more processive and unwinds the DNA with a faster rate than when it is acting alone.

The gp41 unzips the dsDNA and exposes the strands to be replicated by the gp43. In the absence of helicase, the gp43 cannot achieve strand-displacement synthesis. However, if we apply a pulling force on a tethered DNA hairpin molecule, we mimic the helicase activity and allow the DNA strand displacement by the gp43 polymerase. Now if we apply a force in the presence of the helicase and polymerase together, we facilitate the DNA unzipping by the helicase. If this force exceeds 4 pN, the coupling between the helicase and the polymerase cannot take place, due to the helicase moving more rapidly than the polymerase. Therefore, in order to couple the helicase and- polymerase, the force needs to be kept lower than 4 pN.

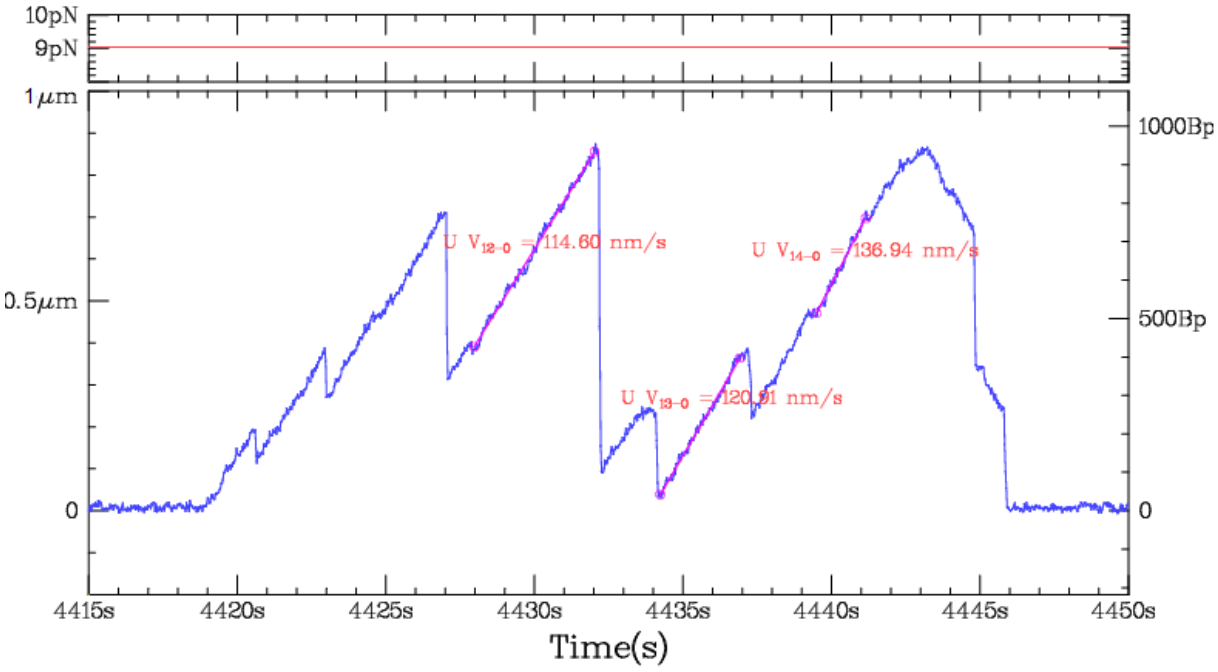


Fig. V.21: gp41 helicase (60 nM) unwinding a hairpin that does not have a G4 structure. *The helicase opens the hairpin and slips backward allowing the hairpin to close until another helicase opens it again. At t = 4444s, the helicase opens the entire hairpin, passes the apex and reaches the other strand. When it translocates on this strand, the hairpin closes after the helicase. On some occasions the gp41 dissociates and the hairpin closes completely.*

We have measured the unwinding velocity of the gp41 in nm/s on hairpins that do not have a G4 structure from hundreds of unwinding events (fig.V.22). After converting the velocity to bp/s (see annexe 5) we obtained a mean unwinding rate of 130 bp/s at 9 pN. We have also measured the processivity of gp41, or the number of bases opened by the helicase before it dissociates from the DNA as about 70 Bp at 9 pN. In measuring the processivity, we have considered that when the helicase makes a brief pause and then continue opening the DNA, it is still in tight contact to the DNA, but when the helicase slips, it probably means that the helicase looses its tight contact with DNA but has not dissociated into solution.

In this section, we evaluate whether the c-MYC (14, 23) G4 is a roadblock for the gp41 helicase, gp43 polymerase, or minimal replisome separately.

V.5.1.1 gp41 helicase

To simulate the effect of gp41 helicase on an ongoing replication fork stalled by the c-MYC (14, 23) G4 structure, we add 60 nM gp41 helicase to the hairpins having a G4 on the lagging strand (fig.V.23.b).

a. Hairpins blocked by the G4 structure

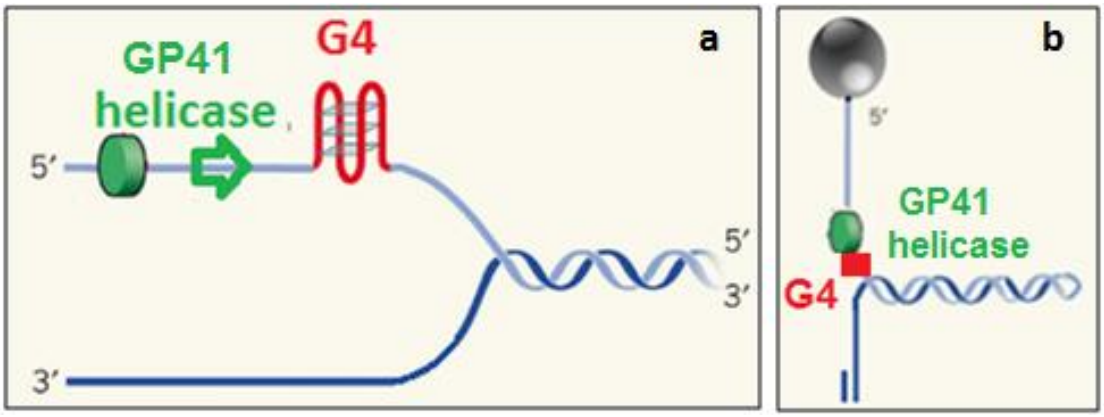


Fig. V.22: gp41 helicase encountering a G4. ***a)*** *A replication fork stalled by a G4 structure on the lagging strand.* ***b)*** *The substrate used to study the effect of 5'-3' helicases such as gp41 on the c-MYC (14, 23) G4 structure: a hairpin that is blocked at the G4 structure that mimics an ongoing replication fork stalled by a G4.*

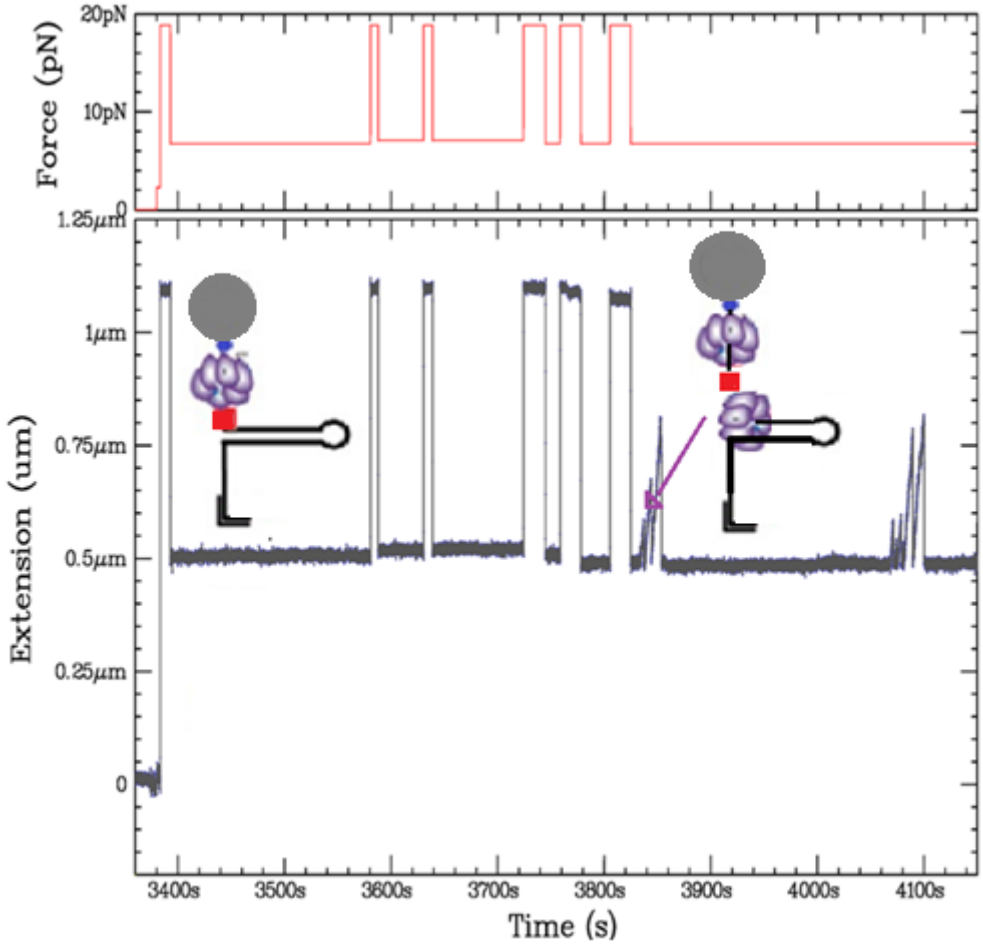


Fig. V.23: gp41 helicase encountering a G4 that is blocking the hairpin closure. *The panel on the top represents the applied force modulation. The panel on the bottom represents the time course of the hairpin extension blocked at the G4 structure located on the lagging strand; then the gp41 helicase unwinds the DNA hairpin, but the hairpin closes and is blocked again on the position of the G4. Thus the gp41 opens the duplex by loading after the G4 without resolving it.*

In the first test, including 60 mM gp41 and 1mM ATP, the hairpins were partially opened, blocked at the G4 structure, and maintained at 7 pN. The gp41 helicase will load on the available ssDNA and translocate from the 5' end towards the G4 structures. We expected the gp41 to resolve the G4 structure in order to unwind remaining fork; however we observed (fig.V.24) gp41 sometime opening the fork, dissociating from the DNA, and the hairpin closing back to the position of the G4 (fig.V.24). This indicates that the gp41 succeeded to load after the G4 and unwind the fork without resolving the G4. This could be explained by gp41 translocating towards the folded G4, bumps in the G4 structure, and is stopped by this roadblock. The first few base pairs of the duplex downstream of the G4 are fraying, so opening and closing. Most probably another gp41 helicase can then load on the single-stranded DNA provided by the opening of the mentioned bases and it can open the duplex.

But once the helicase dissociates from the DNA, the hairpin closes back until encountering the G4 that blocks it refolding. We thus propose that gp41 is able to jump the roadblock.

In some previous studies checking the G4 unfolding by helicases, a DNA molecule having a G4 structure linked to a dsDNA is used. The authors assume that when the double-stranded DNA that is located downstream of the G4 is unwound; it means that the G4 is unfolded. However, our result shows that the gp41 can bind between the G4 and the dsDNA, or can jump the G4 structure, without resolving it.

The gp41 has been reported as capable of protein displacement on its DNA path [141], however, we have shown that gp41 is not able to unfold the G4 structure. Therefore, we can deduce from this test, that the G4 unfolding mechanism is different from the protein displacement mechanism.

b. G4 encircled in the hairpin

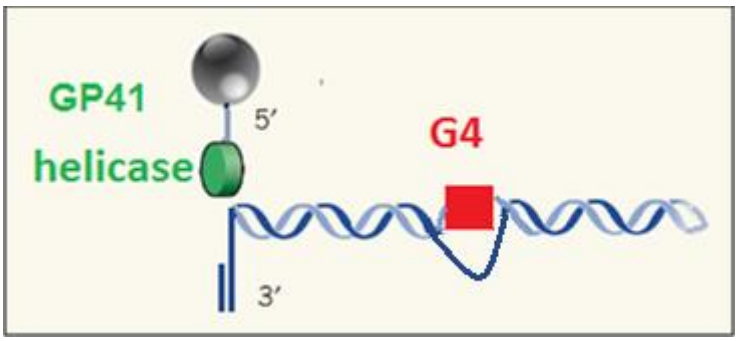


Fig. V.24: gp41 helicase opening a DNA hairpin having an embedded G4 structure. *The helicase is injected when the G4 is encircled in the hairpin. In order to visualize the collision between the helicase and the G4, the helicase has to open the hairpin first.*

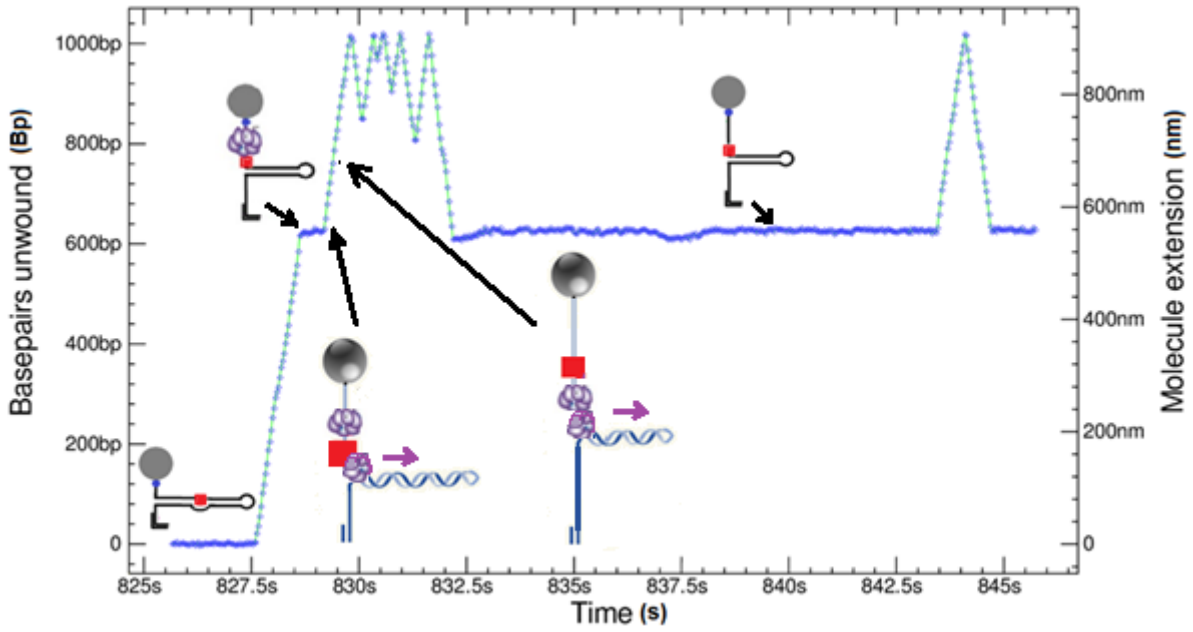


Fig. V.25: gp41 helicase opening a folded hairpin containing a G4 structure on its lagging strand that is encircled in the duplex. *The gp41 opens the duplex, bumps in the G4 structure and pauses for a while, then presumably another gp41 opens the bases downstream of the G4 and when it arrives at the loop, it continues its path on the other strand and the fork refolds after it. But the arrival of another gp41 helicase opens the fork again. When the fork closes, it encounters the G4 structure that has not been resolved and that blocks the hairpin closing.*

In the second test (fig.V.26), after folding the G4 we encircle it in the duplex by lowering transiently the force to 4pN. Then we add the gp41 helicase and we increase the force to ~ 8-10 pN while the G4 remains encircled. The gp41 opens the bases that are located upstream of the G4, then when it arrives to the G4 the system becomes similar to that described in the first test i.e partially opened hairpin that is blocked by the G4. On these molecules, we have observed an increase in the hairpin extension that is consistent with the hairpin opening until bumping into the G4. At this stage the gp41 can either be stopped by the G4 and then the extension remains constant or makes a pause before another helicase loads after the G4 and pursues opening the fork. As in the first test, when the helicase passes the apex and travels on the other strand, the hairpin closes and the molecule extension decreases until the helicase bumps again in the G4 that has not been unfolded by the gp41.

V.5.1.2 gp43 polymerase

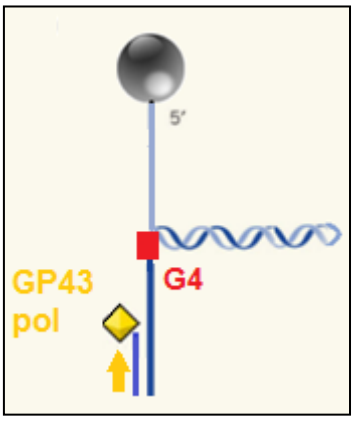


Fig. V.26: The gp43 polymerase encountering a G4 structure. *The gp43 polymerase is tested with the hairpin blocked by a G4 on its leading strand. The polymerase elongates the 3' end of the primer and moves towards the G4.*

The T4 replicative polymerase gp43 belongs to Pol B family. The gp43 polymerase replicates the DNA by adding dNTP to the 3' extremity of a primer. Therefore, in order to test if the polymerase is stopped by the G4, we add the 20 nM gp43 with 200 µm dNTP on the hairpins having the G4 on the leading strand (fig.V.27). The polymerase is added to the hairpins blocked by the G4. As described before, the hairpins have a self primer on their 3' end that can be used by the polymerase to start the replication. When the hairpin is blocked by the G4, the polymerase has to replicate around ~600 bases in order to reach the G4 5' extremity.

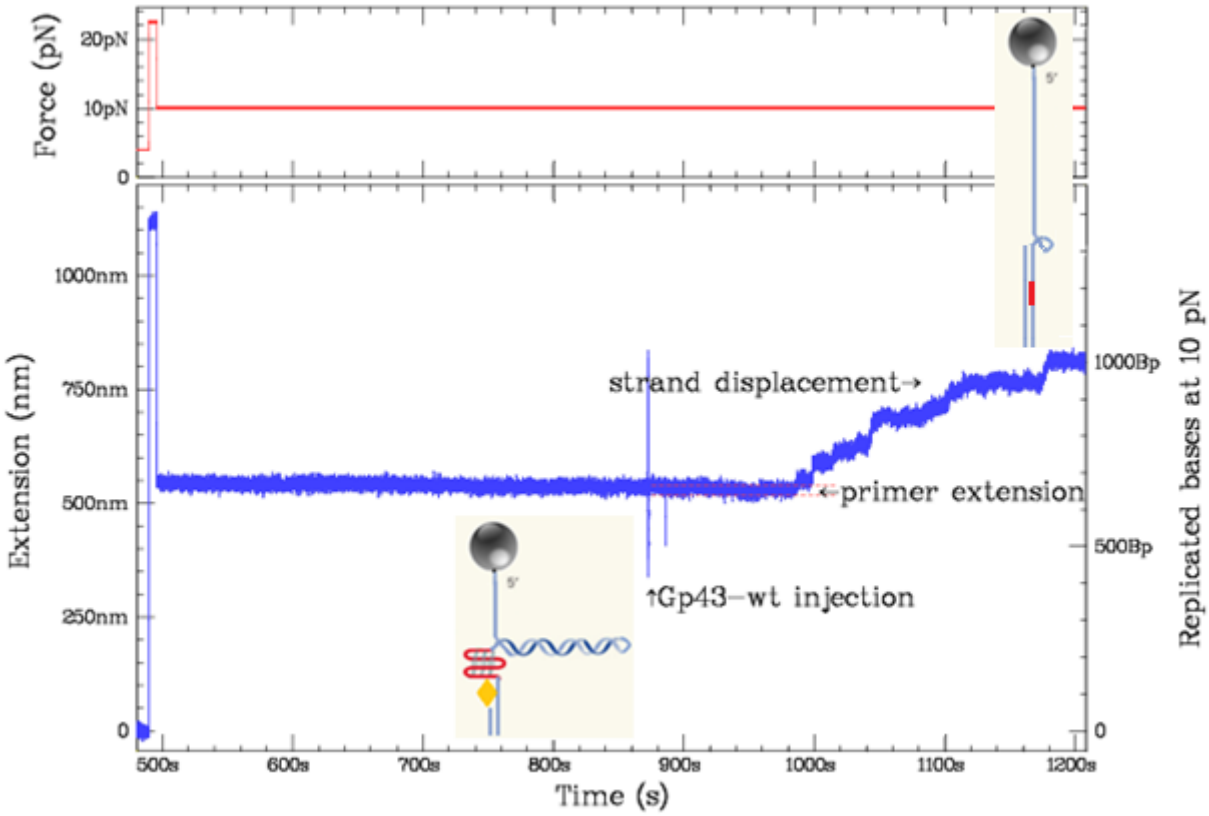


Fig. V.27: gp43 WT polymerase replicating the G4. *The GP43-WT is added to a DNA hairpin blocked by the G4 structure. The polymerase first works in primer extension: the length of the hairpin decreases when the polymerase replicates the single-stranded DNA segment that separates the primer from the G4 structure, because the extension of a base is larger than that of a base pair at 10 pN. The polymerase reaches the G4 structure, and now we see an increase in length that is due to the polymerase starting to work in strand displacement, opening and replicating the duplex. At this stage, the extension increases because the polymerase replicates the base that is on its template and releases the correspondent complementary base. The polymerase succeeds to replicate the G4 and ~360 bases downstream of the G4. The polymerase motion is transiently altered by pauses.*

When the polymerase is added to the solution with the hairpins having this conformation, it will extend the primer, transforming the single-stranded DNA in to double-stranded DNA until reaching the G4. At this force (10 pN) the extension of a base pair is smaller than that of a single base. Thus the molecule shortens when transforming the single-stranded DNA in to double-stranded DNA, by an amount that depends on the elasticity difference between single- and double-stranded DNA at the applied force (see annexe 5).

When the polymerase bumps into the G4, two cases can occur:

- The polymerase is stopped by the G4 structure and cannot go farther.

- Or, the polymerase can unfold the G4 structure (fig.V.28), replicates the G rich segment and pursues the replication process downstream of the unfolded G4. While replicating the duplex, the polymerase replicates the base on its template and releases thus the complementary base of the other strand. Therefore, the trace that we see is an overall increase in the molecule length that is the result of 1) the molecule lengthening due to the duplex opening and 2) the molecule shortening by replicating the bases that are on the polymerase path.

The gp43 succeeds in replicating past the G4 on the majority of the hairpins (88%). Thus the G4 cannot be considered as a strong of a roadblock in front of this polymerase. This result is surprising since the T4 polymerase is usually unable to synthetize under strand displacement conditions very well meaning that it is blocked by a simple oligonucleotide hybridized on the template. Therefore, observing the unfolding of the very strong c-MYC by the T4 polymerase is a surprise: How a molecular motor that is stopped by a relatively weak oligo, unfold the G4 without difficulty? Presumably a special configuration of the polymerase is involved in this process.

Contrary to the helicase the gp43 polymerase is unable to jump the G4 since this would require generating a new primer after the G4. This might be possible for the PrimPol polymerase [112] that is not able to replicate the G4 but restarts replication downstream of the G4.

V.5.1.3 gp41 and gp43 coupled

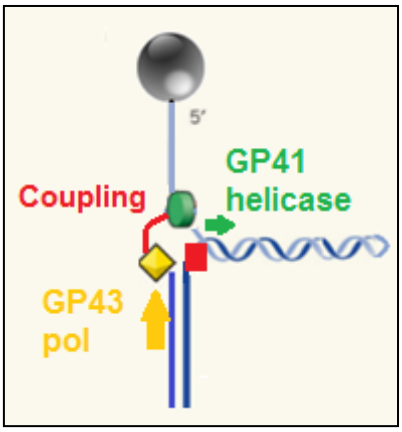


Fig. V.28: T4 minimal replisome encountering a G4. *The T4 minimal replisome formed by the helicase and the polymerase is tested on a hairpin having the G4 on the leading strand.*

Our laboratory has previously studied the T4 replisome and in particular the leading strand holoenzyme coupled with the replicative helicase gp41. This complex was shown to move in a tight collaboration with a very fast and regular motion that is not seen for the helicase or the polymerase alone. The helicase alone is seen unwinding dsDNA at a significant rate only if we apply a strong force (~10 pN) on the hairpin. Similarly, the

polymerase is seen replicating only if we apply a strong force (~10 pN) on the hairpin. The coupling between the two enzymes is seen only at low force (F < 4 pN) where they both become very active. The helicase with the polymerase works fast but its rate decreases if it detaches from the polymerase ensuring that the two enzymes remain in close contact.

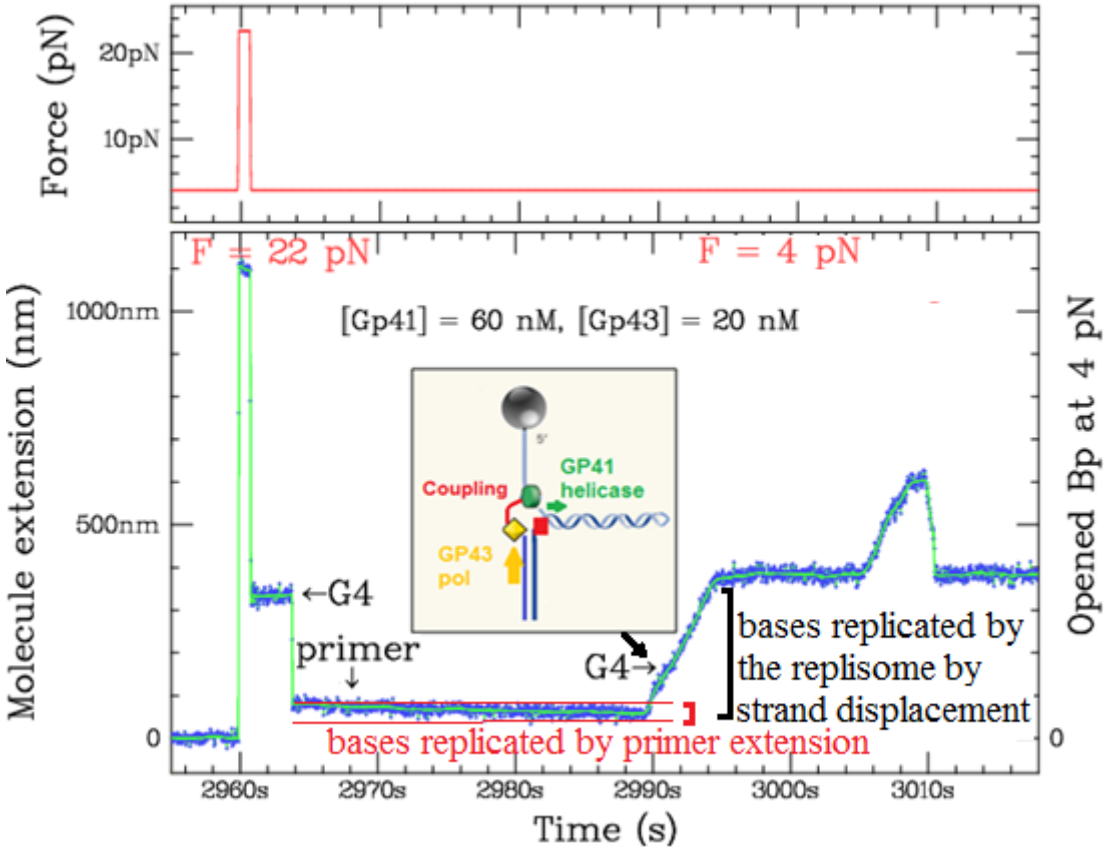


Fig. V.29: T4 replisome encountering a G4. *A 50-mer oligo is added to a DNA hairpin having a G4 on its leading strand. When the hairpin is opened, the oligo binds on the bottom of the hairpin. When the force is reduced to 4 pN, the hairpin refolding is stopped by the G4 structure, and then the hairpin encircles the G4 in the duplex but becomes blocked by the 50-mer oligo. After adding gp41 helicase and gp43 polymerase in the chamber, a decrease in the molecule length is due to the replication of the bases separating the molecule self primer and the 50-mer oligo. Then, an increase in the molecule length is observed. The first increasing length curve is due to the coupled action between gp41 and gp43, where the helicase is opening the duplex and the polymerase is replicating the template strand in an orchestrated manner. The replisome replicated the G4 and hundreds of bases downstream of it and then the helicase dissociates from the polymerase and opens the duplex, but then the helicase falls off and the fork closes in its behind until bumping in the double-stranded DNA that has been synthesized by the replisome.*

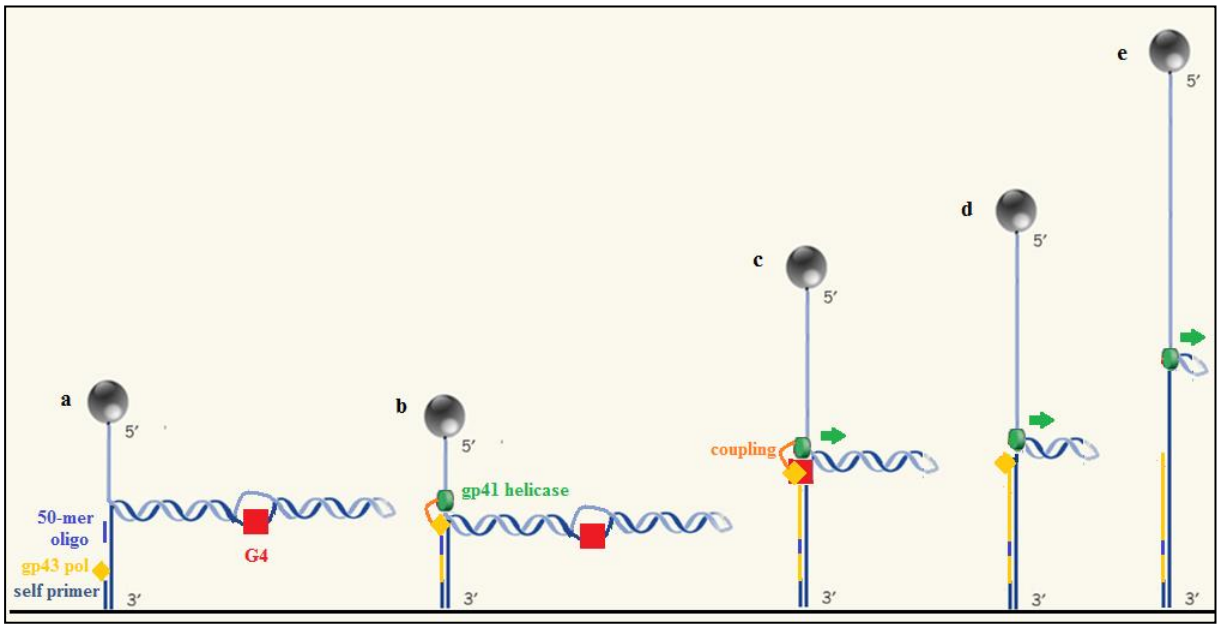


Fig. V.30 : The T4 replisome replicating the G4 where the helicase and the polymerase are coupled. *a) The gp43 replicates the single-stranded gap between the two primers. This results in a length decrease. b) The polymerase load after the second primer and is coupled to the helicase in order to open the duplex and replicate it. This result in an extension because every basepair opened will give a duplicated base on the polymerase template and a single-stranded base on the other strand. c) The coupled replisome arrives to the G4 level and replicates it. d) The helicase detaches from the polymerase and continue unwinding the rest of the duplex. e) The length of the molecule increases when the helicase pursue to open the duplex.*

The G4s have been pointed as blocking replication forks, this idea is logical but the direct interaction between a replisome and a G4 roadblock has not been seen in real-time yet. This is precisely what we wanted to assay here. In the current test (fig.V.29-30-31), we have added a 50-mer primer that is closer to the G4 structure than the molecule self primer where ~200 bases separate the primer from the G4. So the assay is as follow: First, the G4 is folded using the loop oligo. Second, the loop oligo is rinsed to avoid folding new G4s and also to avoide being a primer for the polymerase at the loop region. The 50-mer primer is added and hybridized to the hairpin. The 50-mer hybridization is checked in a similar manner than for the G4 folding i.e the blockage of the hairpin closing at the specific position. Then the hairpin is closed to encircle the G4 without expelling the 50-mer oligo by maintained the force around 4 pN. Then 60mM gp41 and 20 nM gp43 are added into the chamber along with 1 mM ATP and 200 µm dNTPs.

Some coupling events (4%) were observed, where the helicase and the polymerase open the duplex more efficiently (without pauses) and succeed in unfolding and replicating the G4. Actually in the fig.V.30 we first observe a decrease in the molecule length when it was blocked on the 50-mer primer. This length change corresponds to the replication of the single-stranded segment that separate the molecule self primer from the 50-mer primer given that the extension of a base pair is smaller than the extension of a base at 4 pN. Then an increase of the molecule length is seen when the replisome is opening the fork. The replisome opens the fork, replicates the bases on the polymerase path and releases the complementary bases. When it bumps in the G4, the molecule extension is not the same as the extension of the hairpin blocked on the G4 before the replication, but smaller. The replisome replicates 650 bases as well as the G4 structure and then occurs an uncoupling between the helicase and the polymerase. The uncoupling is seen later by the gp41 helicase acting alone to open the rest of the duplex, but when the hairpin closes back, it blocks on the double strand that has been synthesized by the coupled helicase and polymerase.

Therefore, we have demonstrated that the c-MYC G4, which is one of the most stable G4s, does not constitute a roadblock for the T4 replisome.

V.5.2 Pif1 helicase

Pif1 helicase is a 5'-3' helicase that belongs to the SF1 helicases family and that is conserved from bacteria to Human. Pif1 has been reported to unwind dsDNA with a speed around 75 Bp/s [142] and to have a processivity of 10 bases [142]. Pif1 have been also reported as a force regulated helicase. Under force, its unwinding rate and its total unwinding length have been shown to increase significantly [143].
We found that the *Saccharomyces cerevisiae* Pif1 at a concentration of 2nM and a force of 6 pN has a mean unwinding rate of 120 Bp/s (fig.V.32-33). The Pif1 has a processivity of 10 Bp at 6 pN and of 40 Bp at a force of 7 and 8 pN. Increasing Pif1 concentration to 10 nM enhances its processivity to 100 Bp.

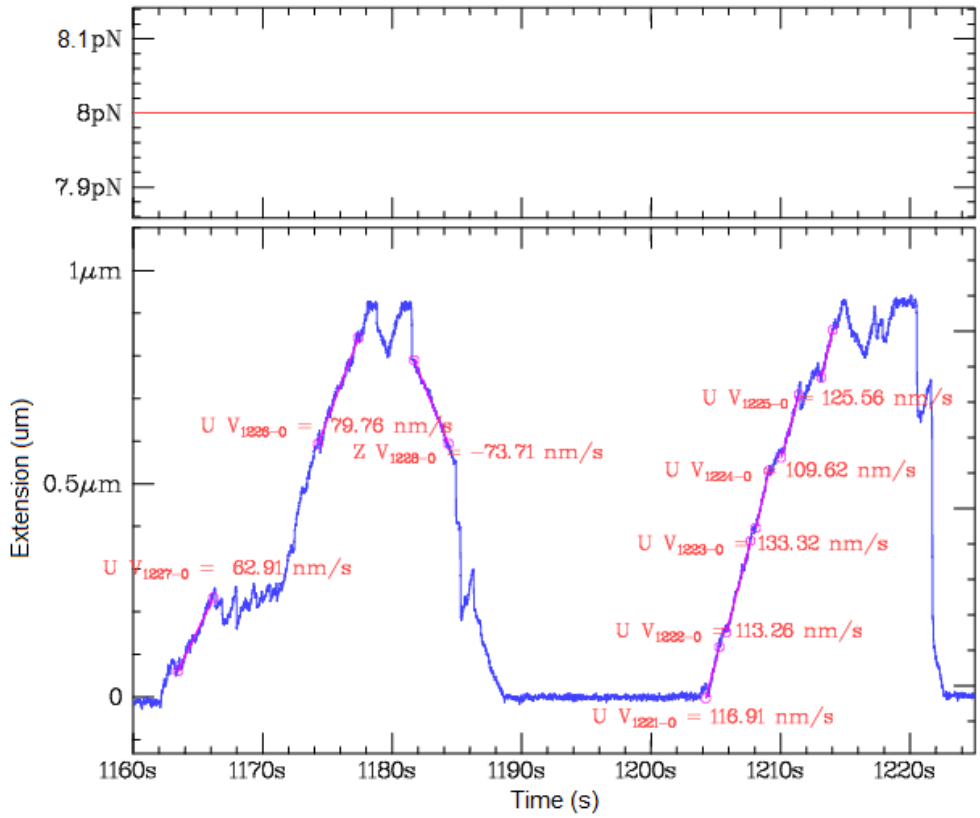


Fig. V.31: Pif1 helicase unwinding a hairpin that does not have a G4. *We can observe the first unwinding events where the Pif1 (10 nM) unzips few bases and then slips back, then at t = 1178 s we observe Pif1 unzipping all the base pairs of the hairpin, it slips back a little as seen by the abrupt decrease of extension then translocates towards the exit of the hairpin. Then it translocates again towards the hairpin apex and then it switches to the other strand and the hairpin closes progressively behind it. We can notice that Pif1 has nearly the same velocity when opening the fork of pushed by the fork.*

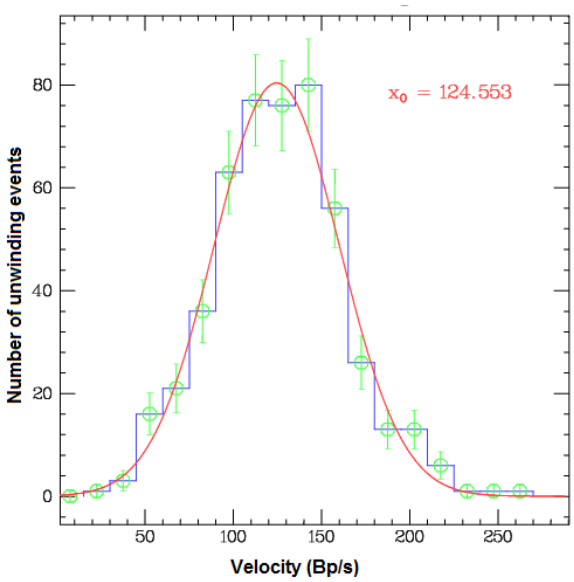


Fig. V.32: Pif1 unwinding rate. *The average Pif1 velocity was calculated at a force of 8 pN by fitting the histogram of the number of events by a Gaussian fit. The mean velocity is about 120 Bp/s.*

In order to check if Pif1 resolves the c-MYC (14, 23), we add 10 nM of Pif1 with 1mM ATP in 60 mM of k+ to the hairpins having a G4 structure.

V.5.2.1 The helicase is added on the hairpins that closing is blocked by the lagging strand G4

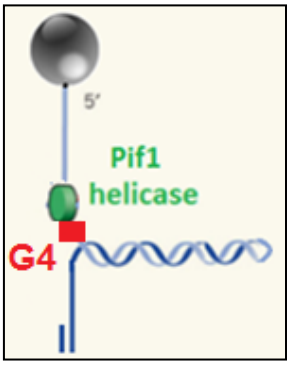


Fig. V.33: Pif1 added to hairpins that are blocked by the lagging G4 structure.

Two kinds of traces were found:

a- After a relatively short time, all the hairpins closed completely (fig.V.35) indicating that Pif1 has unfolded the G4 structures that were blocking the closing. The G4 removal was confirmed by a checking cycle.

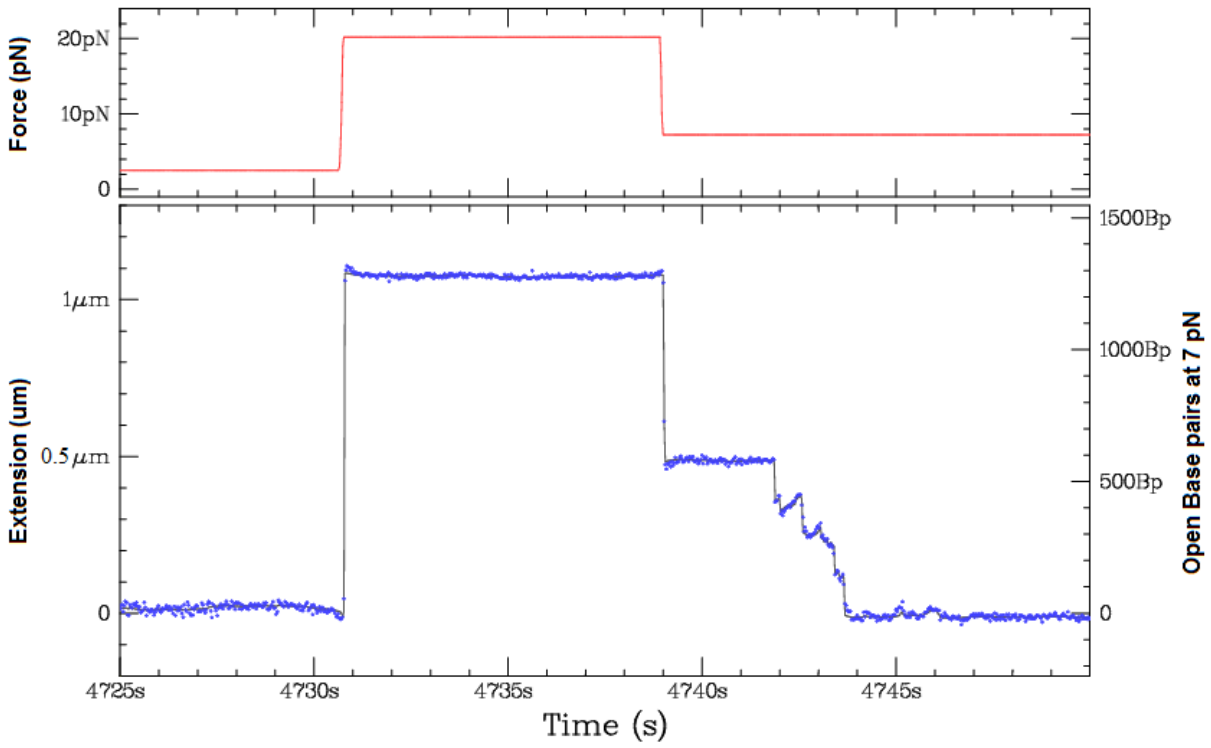


Fig. V.34: Pif1 resolving the G4. *The hairpin is blocked by the G4. Pif1 helicase unwinds the G4 and the hairpin closes completely.*

b- On some hairpins, a small increase in the hairpin extension is seen before the complete closing is observed (fig.V.36), meaning that the Pif1 opens a few base pairs of the duplex by loading after the G4 on the single strand that is made accessible due to the thermal fluctuations of the first-base pairs of the duplex. Pif1 thus displays the same behavior as gp41 that is the ability to jump the G4. The duplex blocks back on the G4. After some time the Pif1 helicase unfolds the G4 and the hairpin closes completely.

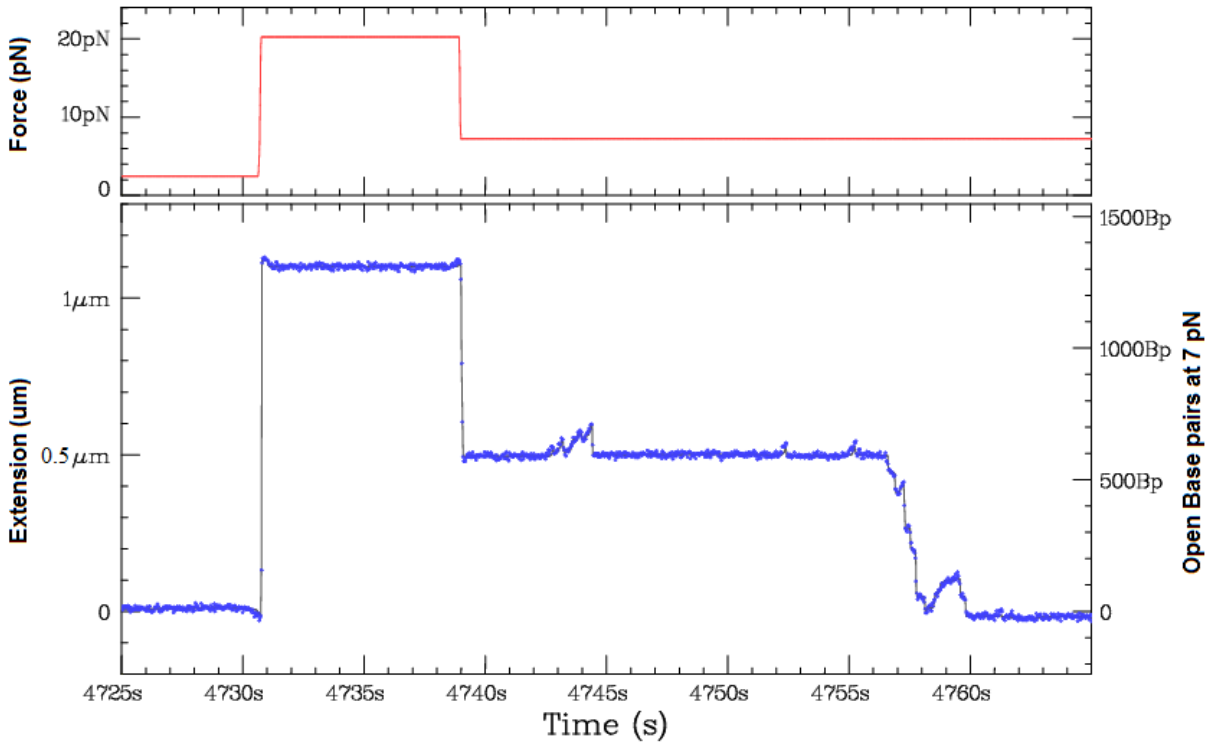


Fig. V.35: Pif1 helicase jumps the G4. *Pif1 helicase jumps the G4 structure without resolving it, unwinds a few base pairs of the duplex, and dissociates letting the hairpin closing partially and re-blocking on the G4. Another Pif1 helicase unwinds the G4 and after that the hairpin closes completely.*

V.5.2.2 Pif1 heliacase is added to the hairpins having an encircled G4 that is located on the lagging strand in order to visualize its collision with the G4

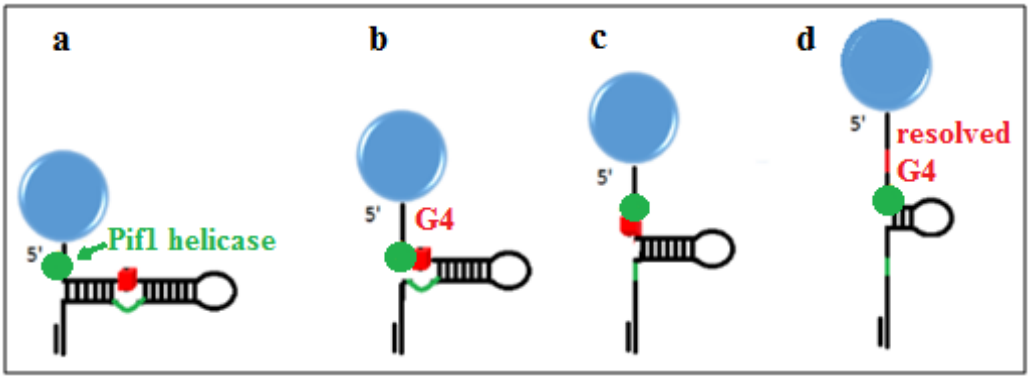


Fig. V.36: Pif1 unwinding a hairpin that has an embedded G4 structure. *This cartoon shows Pif1 unzipping a hairpin having an encircled G4 on its lagging strand.* ***a)*** *Pif1 unzipps the base pairs upstream of the G4,* ***b)*** *when the Pif1 bumps into the G4,* ***c)*** *a small abrupt jump in the molecule extension is seen due to the release of the complementary c-rich sequence. The Pif1 makes, or not, a pause on the G4 before* ***d)*** *pursuing the duplex unzipping.*

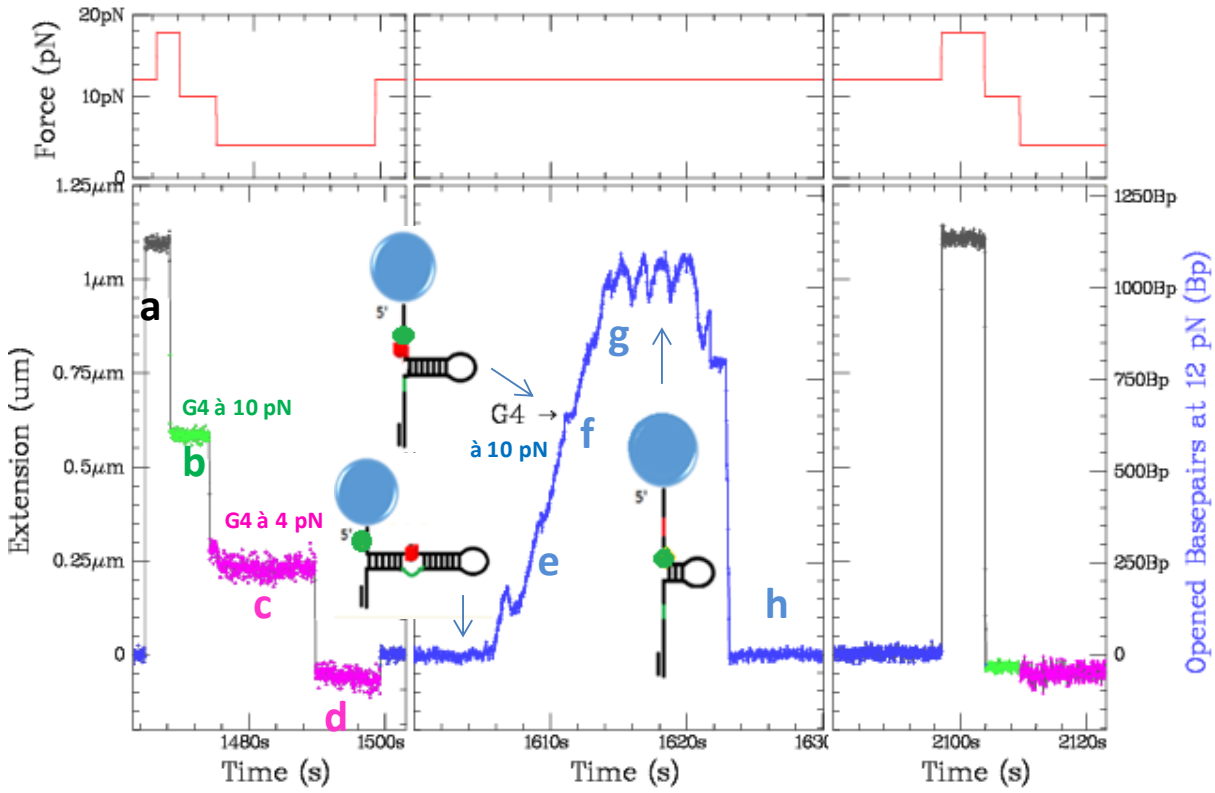


Fig. V.37: Pif1 resolving a G4 embedded in a hairpin. *On the first panel,* ***a)*** *the hairpin is open at a high force,* ***b)*** *then the force is reduced to 10 pN but the hairpin remains blocked by the G4 structure,* ***c)*** *the force is reduced further to 4 pN, the hairpin remains in the same configuration but its extension value decreases upon the change in ssDNA elasticity.* ***d)*** *The G4 structure is then encircled and the hairpin closes completely. In the middle panel, Pif1 is added, and one can see* ***e)*** *the Pif1 helicase unfolding the DNA duplex.* ***f)*** *A jump in the extension is seen before reaching the G4 level indicating the release of the c-rich segment, and a short pause is observed at the G4 level.* ***g)*** *The Pif1 unwinds the rest of the hairpin, reaches the apex, and then moves back and forth at the loop level leading to several closing and opening. Finally, the Pif1 helicase switches to the other strand and is finally dissociated from the DNA.* ***h)*** *The hairpin closes without being blocked by a G4 secondary structure, indicating that the G4 was unfolded by Pif1. The third panel confirms the G4 unfolding: if we open the hairpin and reclose it, the hairpin closes immediately.*

In this test (fig.V.37-38), after folding the G4, we encircle it and we add the 10 nM Pif1 helicase with 1 mM ATP in 60 mM k+. We then increase the force so that Pif1 can unfold the duplex. This time in order to reach the G4 structure, the Pif1 has to unzip the base pairs that are located upstream of the G4. Then when it bumps in the G4, the c-rich complementary

bases are released and this leads to a jump of extension of the hairpin, while the Pif1 makes a pause on the G4, and pursues opening the base pairs that are downstream of the G4. After opening all the hairpins, the Pif1 passes on the other strand leading to the gradual fork closing. The Pif1 helicase can also dissociate from the DNA when it reaches the loop and in this case the hairpin closes abruptly.

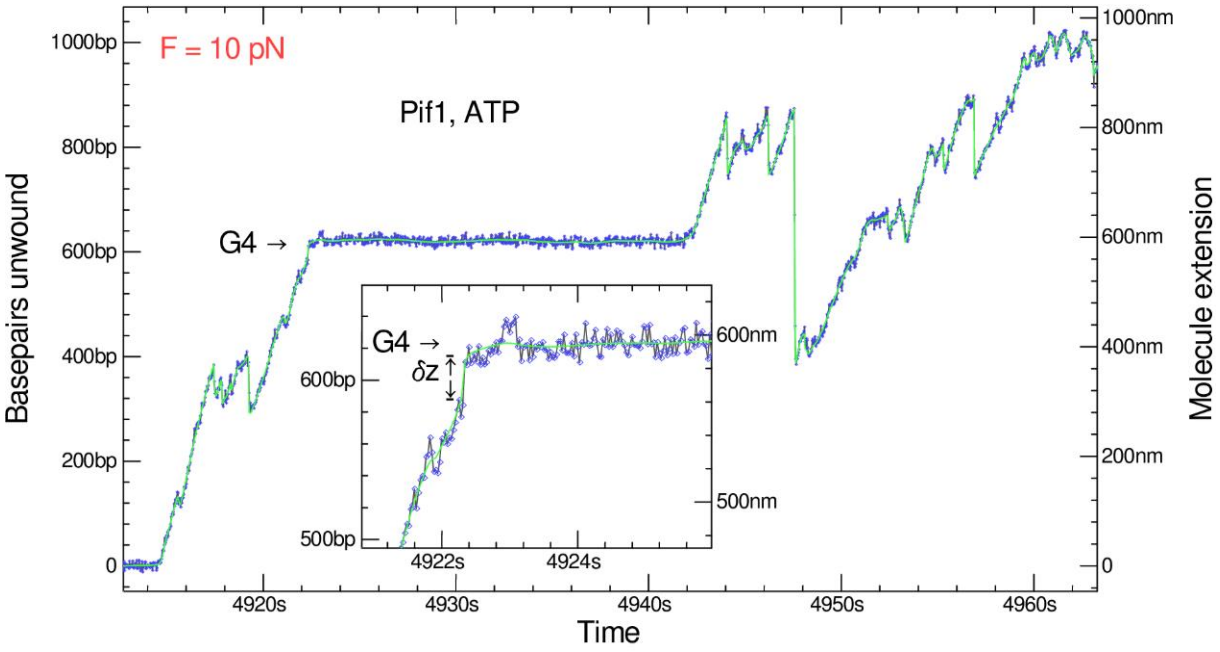


Fig. V.38: Zoom on the Pif1 pause on the G4. *The G4 structure is folded at the lagging strand. The Pif1 helicase loads on the single-stranded flap and translocates from 5' to 3'. It unwinds the base pairs before the G4, and when it bumps in the G4 structure, the c-rich sequence is released in one stroke, and the helicase enters a pause state. After this pause, the G4 structure is unfolded and Pif1 continues unwinding the rest of the duplex. This last segment alone is not enough to confirm that the Pif1 has unfolded the G4, since the further unwinding could be the result of another Pif1 helicase entering in the fork and unwinding it while the G4 structure could remain folded (as seen with Gp41). In order to ensure that the G4 has indeed unfolded, one should wait until the helicase passes the loop and detaches or translocates on the other strand, in order to see if the hairpin closes and block at the G4 level (G4 still folded) or closes completely (G4 is unfolded).*

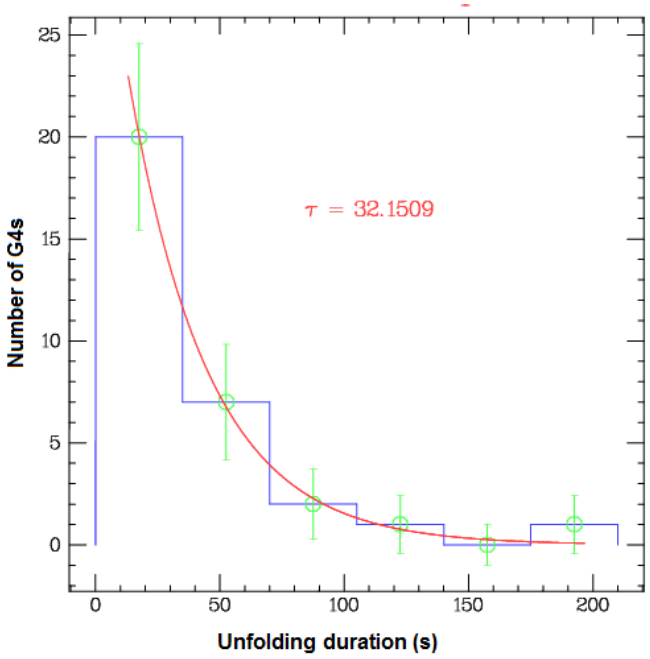


Fig. V.39: Pif1 pause duration on the c-MYC (14,23) G4. *G4 unfolding duration by Pif1 helicase: The G4 unfolding duration was calculated from the beginning of the pause at the G4 structure until the moment when the G4 unfolding is confirmed. This duration is obtained by the characteristic time of the exponential fit of the histogram.*

As we have seen, the Pif1 helicase opens dsDNA with a fast rate but when the helicase encounters the G4, it pauses before actually resolving the G4 (fig.V.39). The G4 is clearly a roadblock for the helicase that is able to remove it after some struggle. This unfolding time displays a Poisson distribution as seen in fig. V.40, this means that some G4 are unfolding very fast but some are removed after a long time. The mean characteristic time of 32 s is long compared with the cycle time required to unwind one base (~1/120 s) suggesting that in its struggle Pif1 has burned several ATPs before succeeding the G4 unfolding. Pif1 translocation rate was shown to be the same of its unwinding rate. Besides, Pif1 was shown to hydrolyse an ATP while translocating by a nucleotide [142]. Therefore, we can estimate that in order to unfold the c-MYC G4, Pif1 hydrolyses around 3800 molecules of ATP.

V.5.2.3 G4 on the leading strand

We also tested the Pif1 helicase with the hairpins having a G4 on the leading strand. This test was done with a hairpin blocked by the G4 structure at a force of 8 pN. Two cases can occur (fig.V.41):

a- The Pif1 opens the duplex, reach the apex, switch to the other strand, and when it arrives to the G4, it makes a pause and then unfold the G4.

b- The Pif1 opens the duplex (few base pairs of more), releasing the complementary bases on the other strand, making a single stranded region accessible before the G4. Another helicase can load on this region and unfold the G4. This case occurs in all the tested molecules.

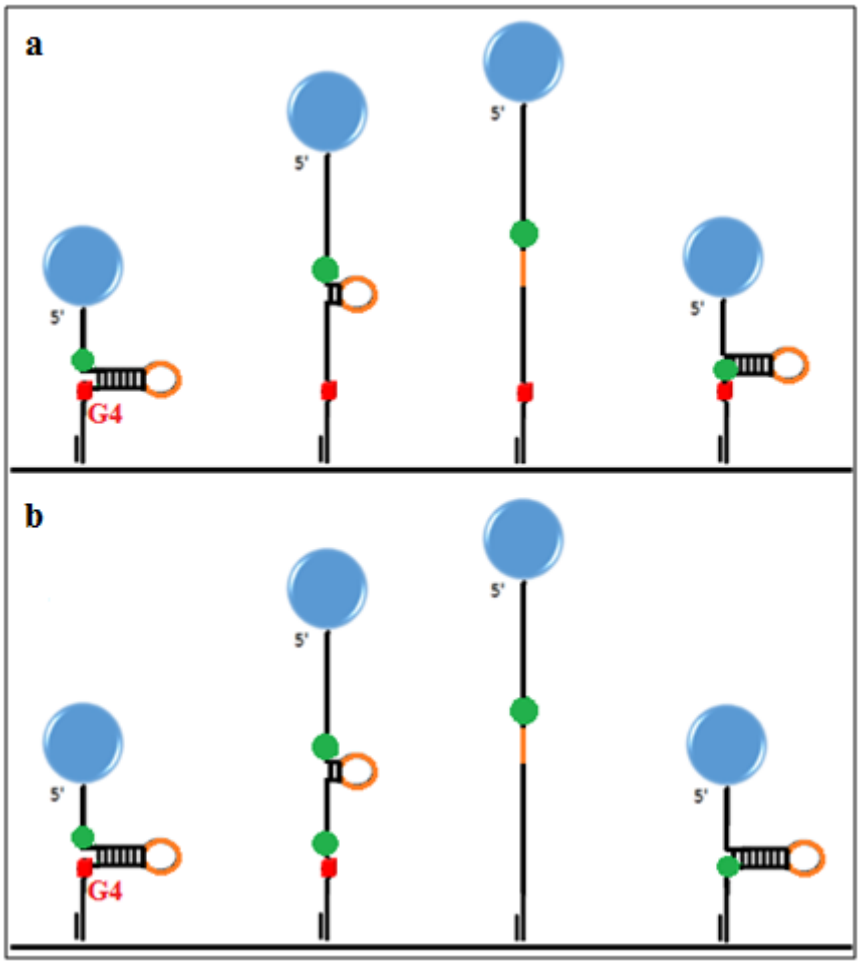


Fig. V.40 : Pif1 added to hairpins having a G4 on their leading strand. a) *The Pif1 opens the duplex, reach the apex, switch to the other strand, and when it arrives to the G4, it makes a pause and then unfolds the G4.* ***b)*** *The Pif1 opens the duplex (few base pairs of more), releasing the complementary bases on the other strand, making a single stranded region accessible before the G4. Another helicase can load on this region and unfold the G4. This case occurs in all the tested molecules.*

V.5.2.4 Does the Pif1 helicase have an affinity for the G4 structure?

The third test consists to check if Pif1 has an affinity for the G4 structure, and if it melts the G4 that is embedded in the double-stranded region by directly loading on it. So we fold the G4s, encircle them in the hairpin by lowering the force to ~ 4 pN. Then we add 10 nM Pif1 and 1 mM ATP in 60 mM k+. We leave Pif1 in the chamber for some time keeping the force at 4 pN at which Pif1 is basically inactive and we did not observe any hairpin opening. In order to check if the G4 is still folded or not, we need to open the hairpin and close it to check for the G4 blockage. But if Pif1 remains in the chamber, the helicase can unfold the G4 when the hairpin is open. To avoid these false positive results, we rinse Pif1 from the chamber

at 4pN with extensive buffer solution. We check then, using opening-closing cycles, if the G4 has been unfolded by Pif1 or not. Actually, 97% of the G4s were not unfolded. This indicates that Pif1 does not target the G4 structure directly but requires a single stranded flap to load and then to translocate towards the G4 and resolve it.

V.5.2.5 Does Pif1 resolve the c-MYC (14, 23) at a high concentration of k+?

At high potassium concentration (150 mM), the G4 stability against unwinding by Pif1 is strongly enhanced. Only 10% of the G4s were resolved by Pif1 (table 4). The high concentration of potassium was shown in previous studies to stabilize S-MYC and telomeric G4s against unwinding by Pif1 [100].

V.5.2.6 Does Pif1 resolve the c-MYC (14, 23) in the presence of Phen-DC3 ligand?

Phen-DC3 ligand enhances the G4 stability against unwinding by Pif1. Pif1 is stopped by a G4 structure stabilized by Phen-DC3 ligand. Only 2.5% of the G4s were resolved by Pif1 (table 4). PhenDC3 ligand has been previously reported to inhibit CEB1 unfolding by Pif1[86].

V.5.2.7 Does Sub1 protein enhance the G4 stability against unwinding by Pif1?

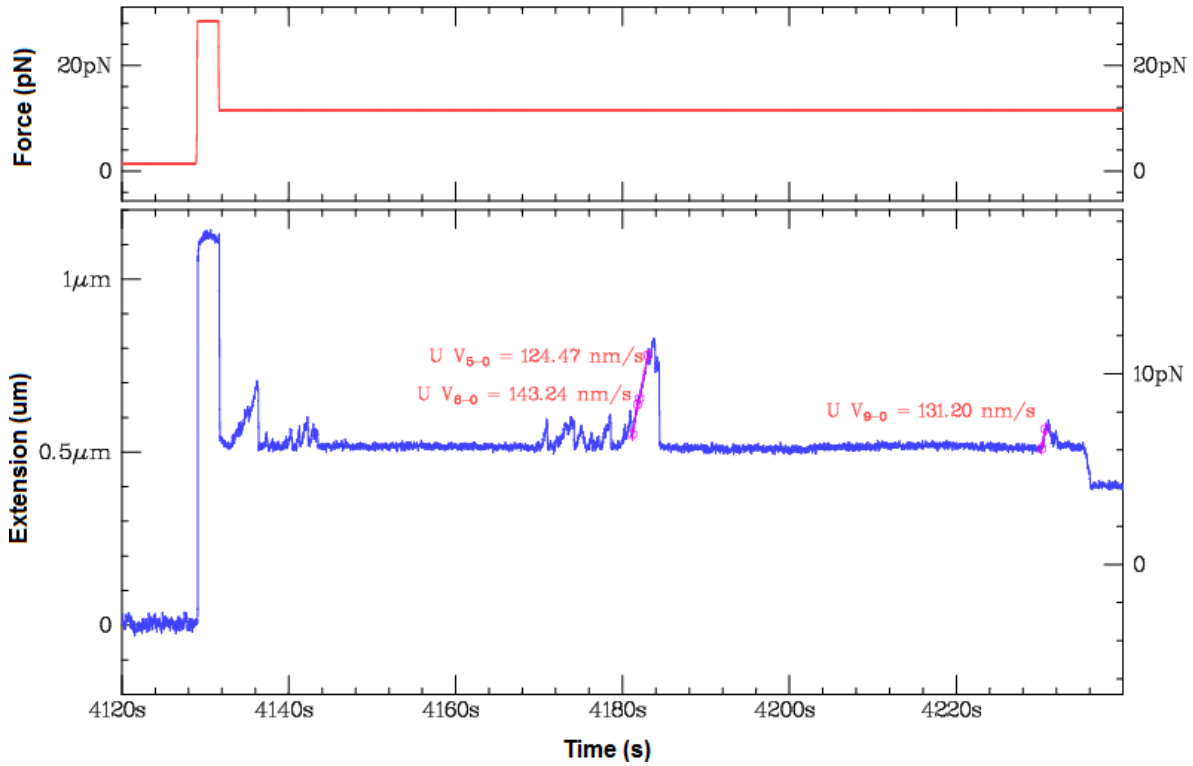


Fig. V.41: G4 unwinding by Pif1 in the presence of Sub1. *In the presence of Sub1, Pif1 is unwinding the G4 slower than in the absence of Sub1. The G4 was removed only on 30% of the molecules.*

We have reported that Sub1 stabilizes the G4 in a sodium buffer. We want to check if Sub1 can stabilize the G4 against Pif1 unwinding. We found that in the presence of Sub1, Pif1 was unable to resolve the G4 structure on 70% of the molecules. However on the other molecules, untangling the G4 was very slow with a mean duration of 700s (fig.V.42),. This shows that Sub1 stabilizes the G4 against Pif1 unwinding and confirms the result of biochemical study that has shown that Pif1-catalyzed melting of the MYC (14, 23) G4 occurred more slowly in the presence of Sub1 [144].

G4	K+ concentration	Stabilizing agent	Unwinding events by Pif1 (%)
Modified c-MYC	60 mM	-	90
Modified c-MYC	60 mM	Sub1 protein	30
Modified c-MYC	60 mM	Phen DC3 ligand	2.5
Modified c-MYC	150 mM	-	10

Table V.4 : The percentage of the G4 unwinding events by Pif1 in the absence and in the presence of a stabilizing agent as well as in a high potassium concentration.

V.5.3 RecQ helicase

RecQ Helicase is a 3'-5' helicase that has been said to unwind telomeric G4 in Na^+ but not in k+ [99]. RecQ has already been studied intensively in different laboratories and also in our laboratory. The wild type enzyme tested in the simple unzipping assay displays a complex behavior: it does unwind DNA but switches between two different rates in a random manner. One of these rates equals ~70 bp/s and the bursts of activity are typical of a helicase. The second rate is far slower and resembles a Brownian motion of the enzyme on its substrate. Biochemists have shown that one can prepare mutants of RecQ with contains a subset of the three domains of the wild type enzyme. RecQ-ΔC is a mutant that lacks the HRDC domain that interacts with dsDNA, while RecQ-ΔΔC just contains the helicase core domain. RecQ-ΔΔC has a very small processivity in our hand but RecQ-ΔC behaves has a perfect helicase unwinding dsDNA with a strong processivity and a homogeneous rate of 70 bp/s. Moreover RecQ-ΔC has been crystallized and its structure was solved by J. Keck [145].

We have add RecQ WT in 60 mM k+ to the hairpins having the G4 on the leading strand.

V.5.3.1 Does RecQ unfold the c-MYC(14,23) G4?

We block the Hairpins on the G4 structure and maintained the force at ~8-10 pN. Then, we add RecQ WT and leave it for some time in the chamber. After rinsing the helicase, 98% of the hairpins were still having the G4 structure. This shows that RecQ is unable to unwind

the c-MYC G4. The same result was also gotten with RecQ-ΔC, where the helicase jumps the G4 structure and unwinds the hairpin without resolving the G4 like what we have seen with gp41. In order to compare with Pif1, we repeat the same test after adding 10 nM Pif1 helicase, and we leave Pif1 in the chamber.After rinsing Pif1 all the G4s were unfolded.

V.5.3.2 Folding of G4s due to hairpin opening by helicases

In fact if most helicases do not unwind G4s, they actually help to fold G4. Starting from hairpins where no G4 was folded, we observe that on 22% of the molecules end up having a G4 folded (fig.V.43),. In fact, when the helicase opens the hairpin under a pulling force (of 8-10 pN), it provides the required conditions to form G4. This assay shows that RecQ does not untangle the c-MYC G4. In the presence of Pif1 helicase, some G4s folding due to the hairpins opening were seen, however, they were unfolded by Pif1 later.

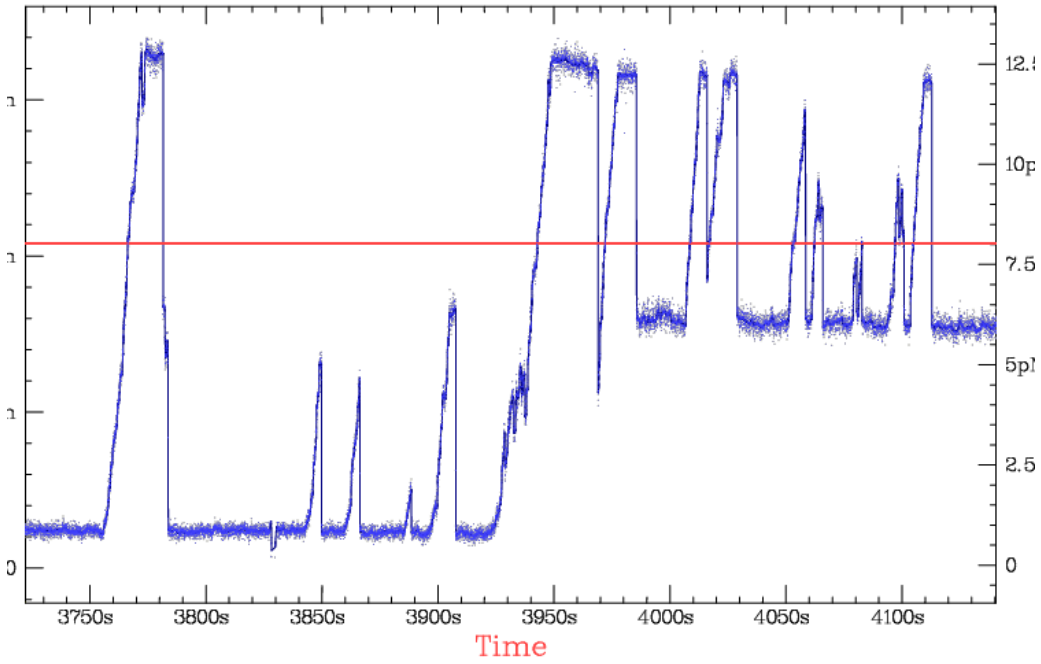


Fig. V.42: RecQ delta opening the hairpin and leading to the G4 folding.

helicase	Resolved G4s
Pif1	90%
RecQ	2%
BLM	-x resolved-
Rep	0%
Gp41	0%

Table V.5: percentage of G4 structures resolved by different helicases. *Owing to the Poisson distribution of the G4 lifetime, a few G4 may unfold spontaneously during an experiment without the specific action of the helicase.*

V.5.4 Does RPA or SSB unfold the G4?

Some studies have reported that RPA unfolds G4s such as telomeric G4 [73] and Gq23 [121] and that the unfolding was better more efficient in Na+ than in k+. We wanted to test the effect of the human RPA on the folding and the stability of c-MYC G-quadruplexes. And the same test was done using the *Escherichia Coli* single stranded binding protein SSB.

V.5.4.1 G4 folding in presence of single stranded proteins

We tried to fold the G4 in presence of the prokaryote single stranded protein SSB and the loop oligo in 60 mM k+. We could not fold any G4. We get the same result in the presence of the eukaryote replication protein RPA. These proteins inhibit the G4 folding.

V.5.4.2 Unfolding the G4 by the single-stranded proteins

a) RPA unwinds DNA duplex

When assisted with a force, RPA succeds to unwind DNA duplexes without ATP with an unwinding rate of 10 Bp/s (fig.V.44). This activity has been reported in previous studies.

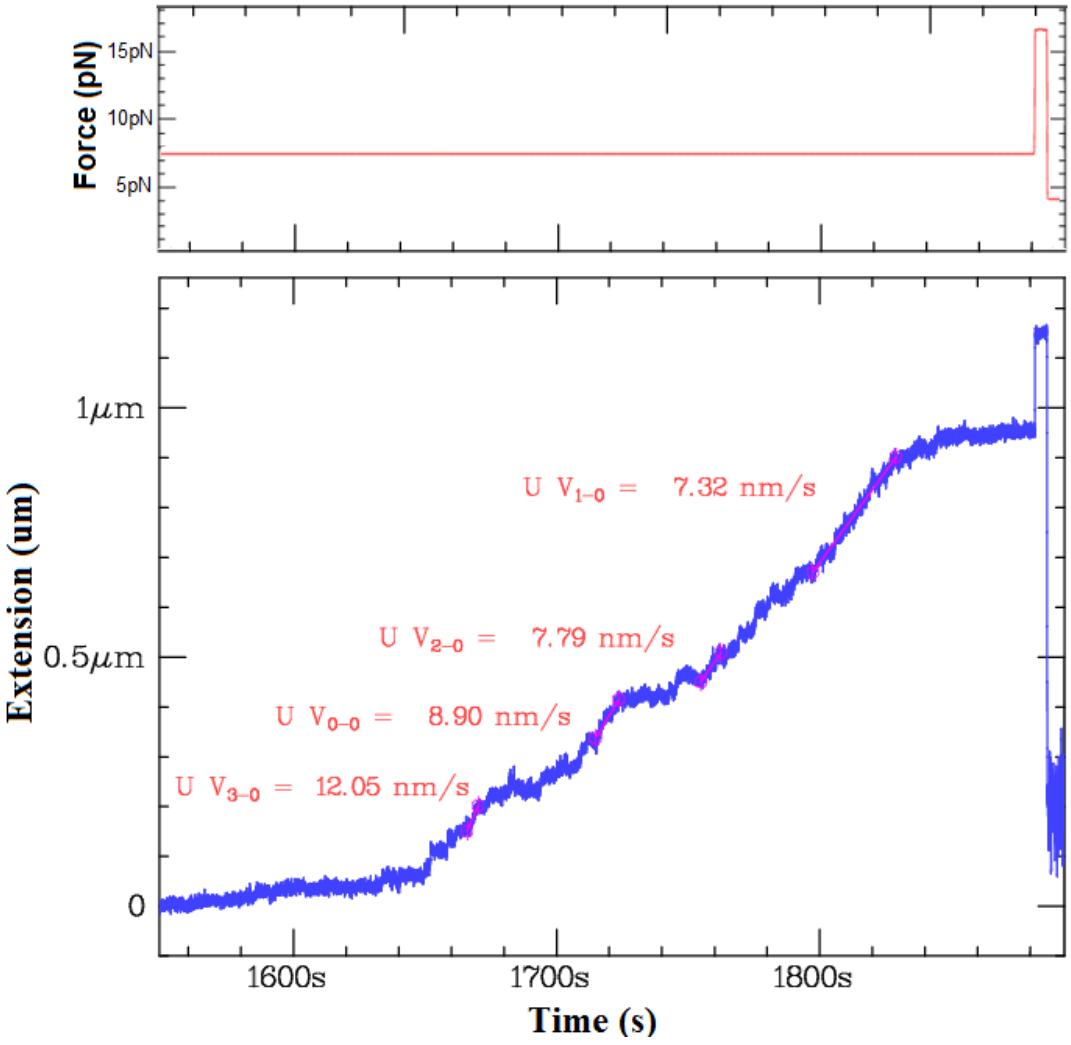


Fig. V.43: RPA unzipping DNA lex in the absence of ATP.

b) Does RPA unfold the c-MYC (14, 23) G4s?

The duplex unfolding hides any eventual G4 unfolding event by RPA. In fact in order to check if the G4 has been unfolded by RPA, one should decrease the force lower than 3 pN. However, at this force the hairpin will encircle the G4 and hinders the verification of the presence or absence of G4. Therefore the effect of RPA cannot be seen in real-time but we can check after rinsing the RPA from the chamber if it has or not unfolded the G4. So first, after folding the G4 on the leading (or lagging) strand of the hairpins, we block the hairpin closing by the G4 structure and we keep the force at ~8 pN thus leaving a 3' (or 5') single-stranded flap accessible to the RPA loading. Then we add RPA and wait for some time. After rinsing the RPA protein, all the G4s were still folded. The same result was obtained with the SSB protein.

This clearly demonstrates that RPA or SSB do not unfold the c-MYC (14, 23) G4 that is more stable of the telomeric G4 where these proteins have been tested before. As we have previously shown, the lifetime of telomeric G4 is in the range of 22 s, and thus the telomeric G4 fluctuates between open and close over the experimental time. Therefore, we think that in the previous study assuming the unfolding of the G4 by RPA [115], the protein does not actively unfold the G4, but stabilizes the unfolded ssDNA once the G4 has pop open.

c) RPA and Pif1

In order to check if RPA can help Pif1 to unwind an encircled G4, we have first done the negative test by adding only RPA to the encircled G4s on the leading strands keeping the force at ~4pN. Only 2% of the G4s were unfolded, this result may be easily explained by the Poisson distribution of the G4 lifetime. Now when we add Pif1 and RPA leaving the hairpin in the same configuration, 94% of the G4s were unfolded. The verification of the presence or absence of the G4 is always done by the same method: We open the hairpin and then lower the force to check if there is a blockage on the G4 level. The fact that Pif1 with RPA could unwind G4 can be explained by the binding of RPA to c-rich sequence complementary to the G4 sequence that is not paired when the G4 is folded (fig.V.45). This binding increases the single-stranded region and facilitates Pif1 loading on both strands. Thus when Pif1 is loaded on the ssDNA that is on the 5' side of the G4, it can unzip it.

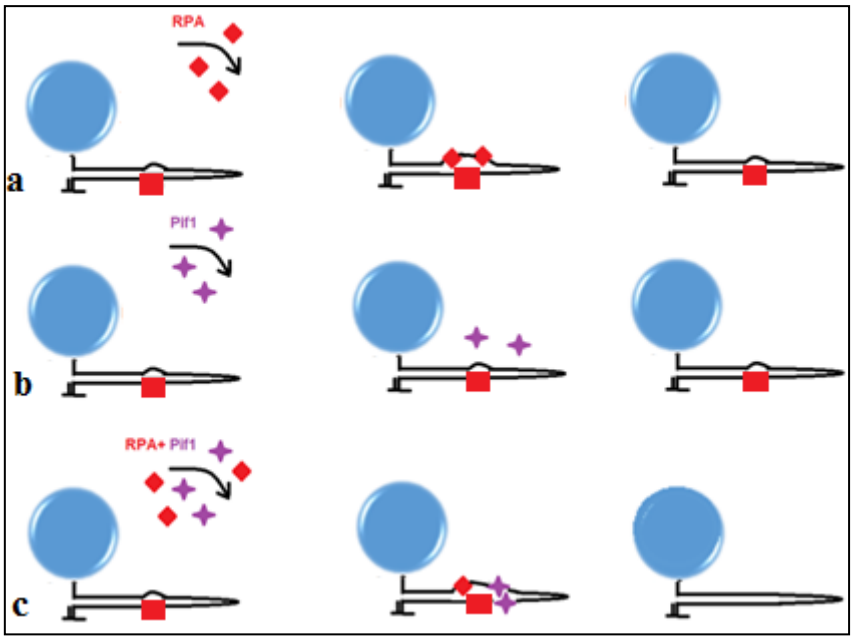


Fig. V.44 : Pif1 resolving a G4 embedded in a hairpin in the presence of RPA. *a) and b) Neither RPA nor Pif1 can resolve a G4 that is embedded in a double stranded DNA region. c) In the presence of RPA, Pif1 succeed to resolve the G4 structure that is encircled in a double stranded DNA.*

Protein	Unfolded G4s (%)
Pif1	3
RPA	2
Pif1 + RPA	94

Table V.6: percentage of unfolded G4s while the hairpin had been kept close.

V.6 Unfolding the G4 structure by polymerases

V.6.1 T7 Polymerase

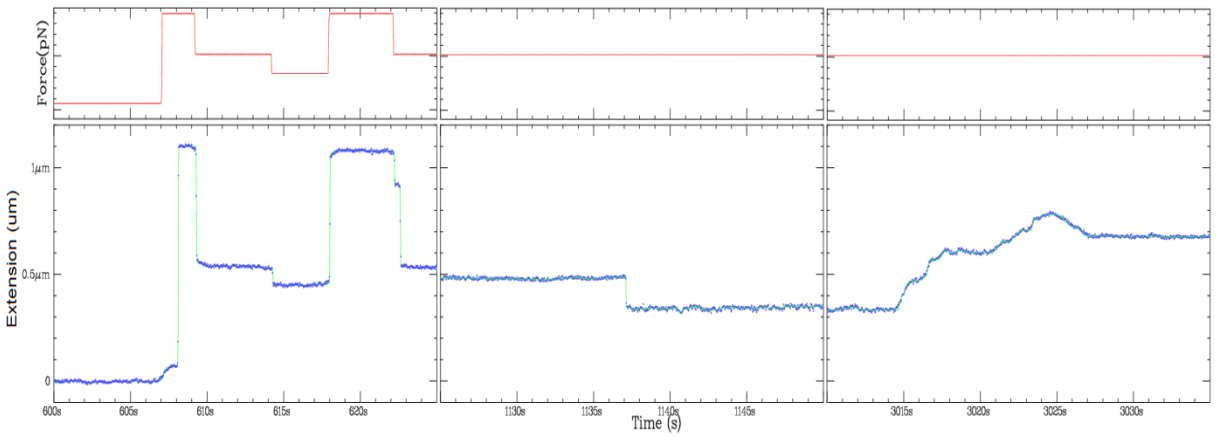


Fig. V.45: T7 polymerase replicating c-MYC (14, 23) G4. *In the first panel, we confirm the presence of the folded G4. In the second panel, the hairpin is first blocked on the G4, then the G4 is encircled and the hairpin is now blocked on the 50-mer primer. In the third panel, we see the T7 polymerase replicating the first strand of the hairpin (the ascendant curve) and arriving to the apex (the peak) and then replicating the other strand where the extension decreases while transforming the single-stranded DNA to a double-stranded DNA.*

The polymerase of bacteriophage T7 belongs to Pol I polymerases family. We have added the T7 DNA polymerase (at 20 nM) to the hairpins having a G4 on their leading strand in a k+buffer. 34% of the G4s were unfolded by the polymerase. An opening-closing cycle is done at the beginning of the experiment (fig V.46) to confirm the presence of the G4 structure. The hairpin remains blocked on the G4 before encircling the G4 and blocking the hairpin on the primer that was hybridized to it. Then when the T7 polymerase is added with dNTPs, a small decrease in length is seen, and that corresponds to the replication of the 200 bases separating the primer from the G4. Then an increase in the extension is seen in the third panel that corresponds to the opening of the fork and the replication of the bases at the same time. Once the polymerase arrives at the apex, the molecule is totally open and the replication of the other strand results thus in a decrease of the extension.

V.6.2 Manta polymerase

The manta polymerase that is reputed to have a strong strand displacement was also tested on the hairpins having the G4s on the leading strand. The manta polymerase was found to replicate the DNA until bumping in the G4 structure and does not progress further. This means that it cannot unfold the G4 structure. We did not get any G4 unfolding event with this polymerase.

V.6.3 Pol ε polymerase

The Pol ε, like gp43, belongs to the B family of polymerases. This polymerase is the leading strand polymerase of the yeast replisome. We thought that it could be interesting to check if Pol ε, like gp43, is capable of unfolding the G4 and replicate its G rich sequence. First we have tested Pol Epsilon exo$^-$ on hairpins that do not have a G4 structure, and calculated its replicating rate that was of 10Bp/s, and its processivity that was of 150 Bases (fig.V.47).

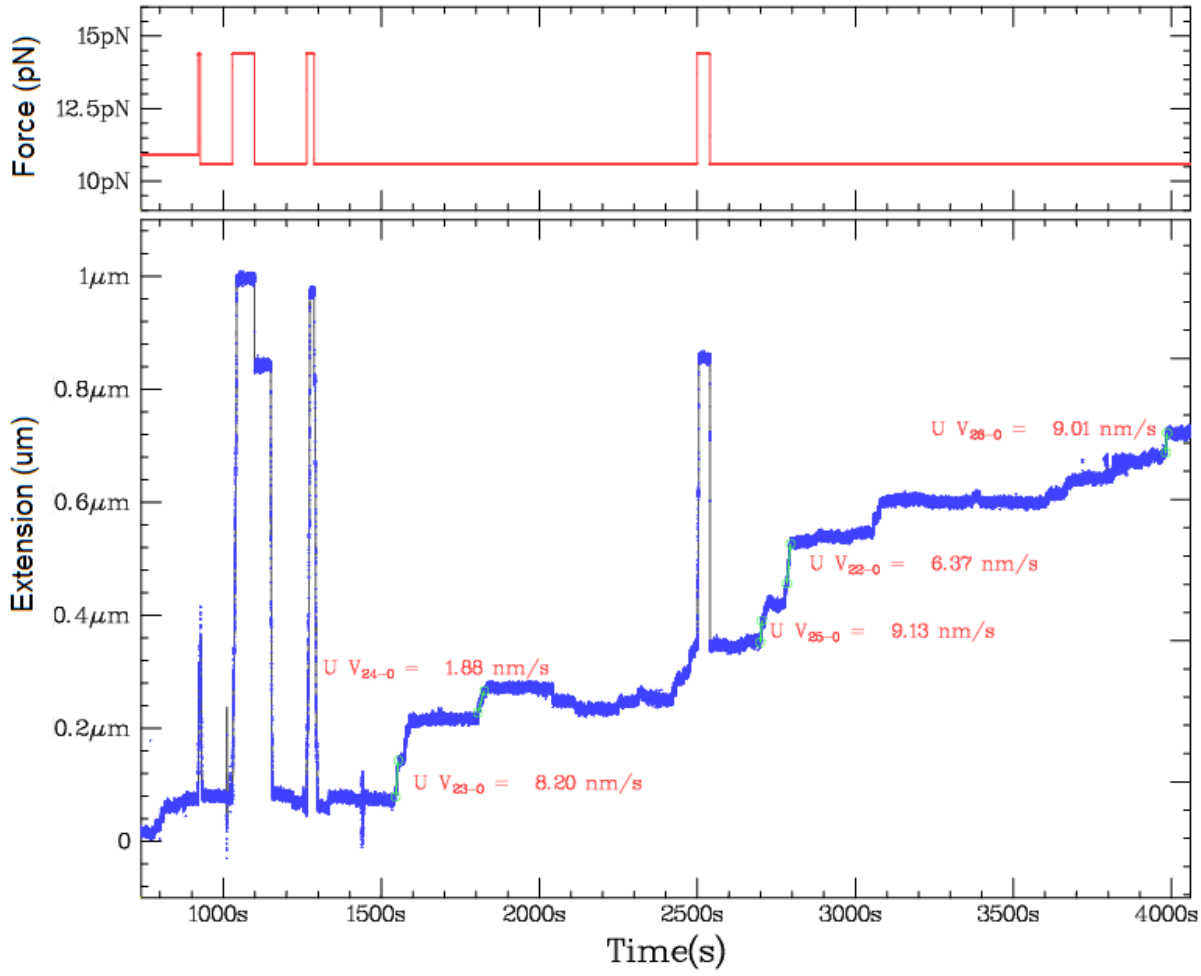


Fig. V.46 : The Pol ε exo- replicating a molecule that does not have a G4 structure. *The replication rate measured in nm/s can be converted to Bp/s using the conversion coefficient at 11 pN. In fact, when the polymerase replicates a base is the duplex region, it transforms the base to a base pair while releasing the complementary base on the other strand.*

Next, we tested the effect of the Pol ε WT on the G4 using two methods, and we got 67% of G4 unwinding events.

a) The hairpin is blocked on the G4 structure: Pol ε acting in strand displacement

20 nM of Pol ε WT is added to the hairpins that were blocked by the G4 structure along with 200 µm of dNTPs in 60 mM KAc (instead of KCl because of the inactivation of Pol ε in the presence of chloride). The polymerase is replicating the bases in a strand displacement manner, where it open the duplex to replicate the bases on its path. The Pol ε succeeds to replicate the G4 on 67% of the hairpins (fig.V.48).

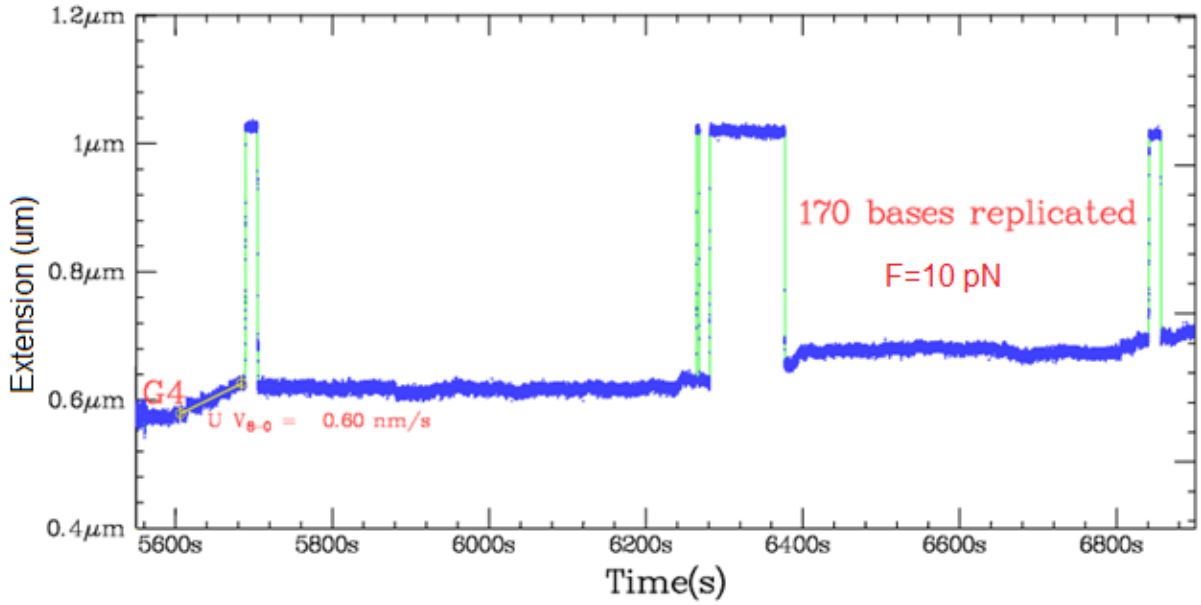


Fig. V.47: Polymerase Epsilon WT replicating a G4. *In this graph one can see the polymerase replicating the G4 and 170 bases downstream of it. One can also notice that the extension of the opened hairpin is reduced due to the repliaction of the bases to base pairs that have a smaller extension.*

b) The polymerase is injected when the hairpin is completely open: The Pol ε is acting in primer extension

The replication is seen by a decrease in the opened DNA molecule length. In this case the replication is known by primer extension where the polymerase is extending a primer and replicating the bases by transforming them into base pairs. The number of replicated bases is thus calculated by converting the extension decrease (nm) to Bp (Annexe 5). The polymerase succeeds to replicate the G4 on many molecules (fig.V.49).

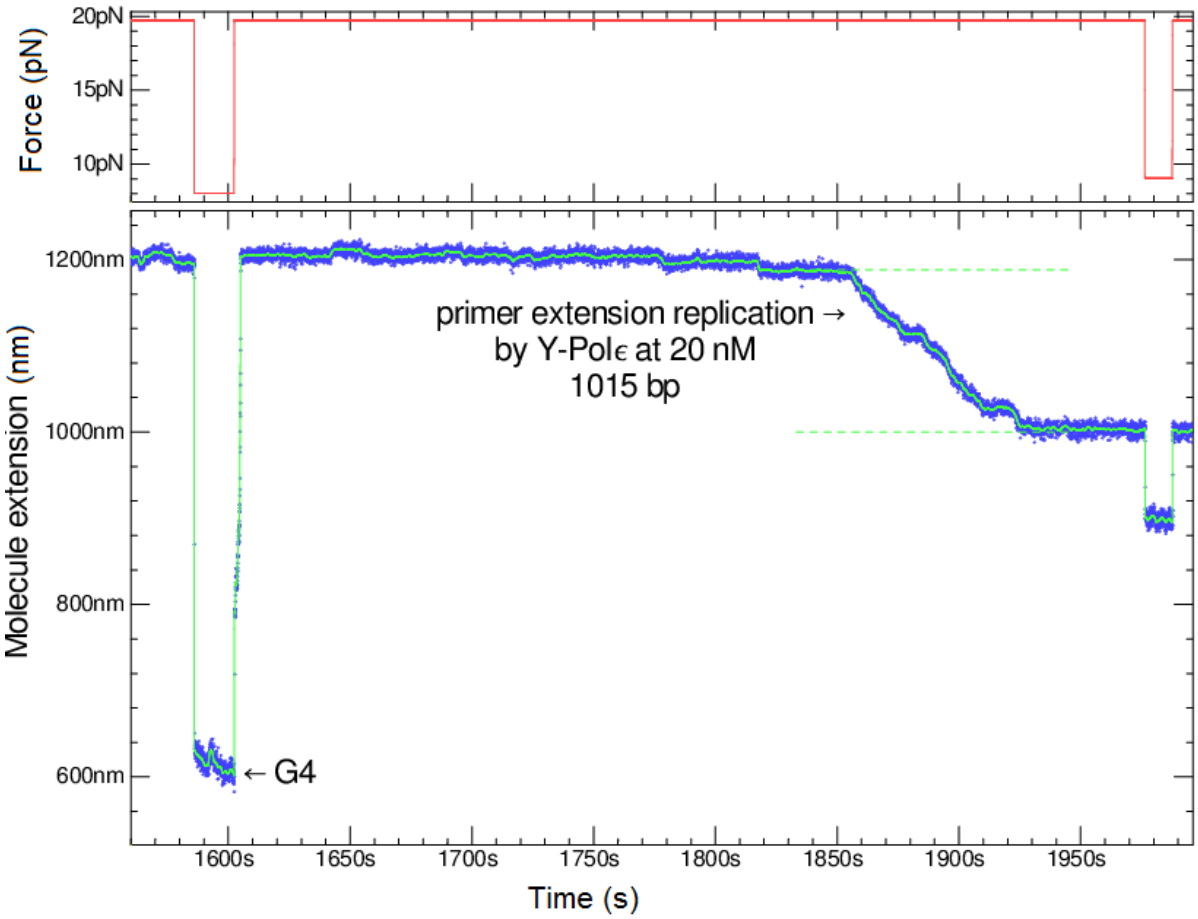


Fig. V.48: The Polymerase Epsilon replicating a G4 in a primer extension mode.

Polymerase	Resolved G4s (%)
Manta	0
T4 bacteriophage Polymerase	88%
T7 bacteriophage Polymerase	34%
Pol ymerase ε (Yeast)	67%

Table V.7 : Table: percentage of resolved G4s by different polymerases.

Conclusion

G-quadruplexes structures form the major impediments for replisome progression, and thus are considered as a treat for the genome stability. Bioinformatic studies reported the presence of more than 700 000 putative G4 forming sequences in the human genome. These studies have taken into account the G4s of low stability such as those formed by non – continuous guanine tracts and those having loops of more than 7 nucleotides. Because the G4s are mostly located in biologically relevant regions of the DNA, it is believed that they are implicated in gene expression regulation, telomeres capping and replication initiation. However, a previous study has reevaluated the G4 forming sequences and has reported that only the thermally stable G4s having short loops, such as the c-MYC G4, can lead to genomic instability [146]. This finding let suggest that the G4s having low stability do not form roadblocks on the enzymes path and do not arrest replication.

Nowadays, several helicases have been shown to resolve G4s, such as Pif1, FANCJ, WRN and BLM. Furthermore, Rev1 polymerase has been reported to replicate the G4, and PrimPol polymerase to prime after the G4 in order to restart the replication. But, *in vivo*, even in the absence of these factors, the cell succeeds to resolve a fraction of the G4s.

G4s DNA roadblocks have been extensively studied by bulk and single molecule assays. In this work, using single molecule magnetic tweezers technique, we have reported a new method to study a G4 structure embedded in a double-stranded DNA region, a situation that mimics *in vivo* G4 in a gene promoter. The presented method offers the advantages of measuring the G4 kinetics and stability in various buffer conditions, and of visualizing in real time the collision between a replicative molecular motor, such as a helicase or a polymerase, and the G4 structure.

Contrarily to biochemical studies, our method allows the direct measurement of the G4 folding and unfolding duration and therefore the kinetics and stability. However, unlike previous single molecule studies, our method gives a G4 folding signal that is stable and well distinguishable and thus it does not present false positive measurement.

Using this method, we have demonstrated that the G4 can be encircled in a double-stranded region for thousands of seconds without being unfolded by the C-rich complementary strand. This finding confirms the possible persistence of a G4 structure *in vivo* for multiple cell divisions if it has not been untangled by G4 resolvases. This is in agreement with a previous biochemical study that has reported that an unresolved G4 survives through multiple divisions without changing conformation [147].

The kinetic study of the c-MYC (14, 23) has shown a finite-time folding duration of 20 s in potassium buffers, that is in the same range as the results previously reported in a magnetic tweezers single molecule study [148]. The c-MYC unfolding in potassium buffer is much slower due to the high stability of this structure in potassium, where the G4 remains folded in average two hours.

The folding of the same G4 was not possible in sodium buffers. Furthermore, when the sodium buffer is added to a potassium-coordinated G4 it destabilizes it in approximately 50 s.

The destabilization of the c-MYC (14, 23) G4 in sodium was used in our study to investigate the G4 stabilizing effect of Sub1 and PC4 proteins as well as Phen-DC3 G4 ligand. In presence of these co-factors, the G4 lifetime in sodium has been increased to thousands of seconds. We suggest that the binding of these agents to the G4 might sequester the potassium cations in the vicinity of the structure and thus delay its unfolding. It is also possible that the binding of these proteins by their positively charged domains help reduce the electrostatic repulsion between the guanines of a sodium-coordinated G4.

Moreover, we have found that Sub1, PC4 and Phen-DC3 act as chaperones and allow the G4 folding at high force without the use of a loop oligonucleotide. Without these agents and in the absence of the loop oligonucleotide we have not seen any G4 folding at high force. Therefore the use of Phen-DC3 ligand can, besides stabilizing the G4s, promote their folding.

The telomeric G4 was found to be much less stable than the c-MYC G4. Its stability has been increased using Phen-DC3 ligand. However it is still below 100 s.

The effect of RPA and SSB proteins on G4 has been also assessed in this work. We have found that these proteins do not allow the G4 folding. But contrarily to what has been previously reported about RPA protein on telomeric G4s, we have not seen any unwinding event of the c-MYC G4 by RPA, neither when the G4 had a 3’ single-stranded flap nor when it had a 5’ single-stranded flap. Therefore, we suggest that in bulk assays, once the G4 is unfolded, the RPA protein keeps this G-rich sequence in a single-stranded conformation and shifts the G4 folding-unfolding equilibrium and therefore we propose that the G4 unwinding by RPA is not an active mechanism. The presence of these proteins *in vivo* is of high interest in order to reduce the occurrence of G4 folding and thus to avoid genomic instabilities.

We have tested the collision of different helicases with the c-MYC (14, 23) G4. Among the tested helicases, the Pif1 helicase was the only one having a G4 resolvase activity, where it unwinds around 90% of the G4s in a meantime of 30 s. In the presence of Sub1, the G4

unwinding by Pif1 was slowed and the pause duration of the G4 structure was of 700 s in average, but the Pif1 succeeds in displacing the G4 binding protein Sub1 after a struggle. However, the efficiency of G4 unfolding by Pif1 is much more reduced in a highly concentrated potassium buffer where only 10% of the G4s were unfolded. Pif1 was also tested on G4s embedded in double-stranded DNA, maintained at low force, so that Pif1 cannot unzipp the DNA. In this situation, Pif1 alone cannot resolve the G4s, whereas in the presence of RPA that increase the proportion of single stranded DNA around the G4, Pif1 can load and resolve the majority of G4s. This is in agreement with a biochemical study where RPA was found to stimulate Pif1-mediated G4 resolution [144].

RecQ helicase has been reported to unwind telomeric G4 in sodium buffer but not in potassium buffer. The RecQ assays that we have done on the c-MYC (14, 23) G4 have confirmed the disability of RecQ to process a very stable G4.

The T4 bacteriophage replicative helicase gp41 also does not resolve the G4. But like RecQ, gp41 is able to jump the G4 structure and access the duplex that is located downstream of it. This result is important to the design of DNA molecules that are tested with the helicases-mediated G4 processing. In fact, this finding suggests that the unwinding of a DNA duplex located downstream of the G4, in bulk or in other single-stranded methods, cannot confirm the unwinding of this structure.

Using the hairpin method, we have seen that a helicase can offer suitable conditions for the G4 folding while unwinding the DNA duplex. Even in the presence of Pif1 helicase the G4 was found to fold when the Pif1 unzipps the duplex and releases the G4 sequence from the DNA helix. However, in a previous study, Pif1 was reported to unfold the G4 structure and to keep it unfolded by patrolling on the G-rich sequence. Upon testing Pif1 we have found that Pif1 unfolds the G4 and pursue the unwinding of the rest of the hairpin. And therefore we suggest that it does not patrol on this sequence.

We have also tested the effect of some polymerases on the G4s. While we were expecting the arrest of polymerases by the G4, we found that some of them, even if they do not have a strand displacement activity, are not very disturbed by these structures. The T4 replicative polymerase gp43, the T7 replicative polymerase, and the Yeast replicative polymerase Pol ε were able to replicate the G4 and pursue replicating the duplex downstream of it. In a previous study however, the Pol ε was found to be stopped by the G4s.

It is important to point out that the data that we obtained with the polymerases assays on G4 are a little more complicated than those obtained with the helicases assays when the polymerase is stopped near to the G4. In fact while studying helicases, the extension of the molecule revealing the position of the G4 structure before and after the helicase activity does not change, however, when a polymerase replicates the DNA substrate, the extension of the molecule corresponding to the position of the G4 is changed because of the elasticity difference between single and double stranded DNA. Therefore we have only reported the results of the molecules where the polymerase has pursued the replication far after the G4 in order to ensure the G4 unfolding. On some molecules the polymerases have replicated some bases and stopped far from the G4. For the rest of molecules, the polymerase is stopped near

the G4 but it is hard to confirm if it is arrested by the G4 or stopped some bases before it. Therefore we do not reject the fact that Pol ε or other tested polymerases can be arrested by the G4 structures on some molecules but we confirm that they can replicate this structure on many molecules.

The perspectives of this work will be to extend the study to other putative single G4- and consecutive G4s forming sequences such as consecutive telomeric G4s. It will be of great interest to study the interaction between enzymes, such as the RNA polymerase, and G4 structures. It will be also interesting to test in bulk, and thus in the absence of force, whether the polymerases that we have reported to replicate the G4 do also replicate it in bulk.

CHAPTER VI: T4 replisome study using DNA rolling circle

VI.1 Rolling circle construction

We have designed a rolling circle DNA molecule in order to study the replication by the T4 replisome, and more specifically the lagging-strand replication.

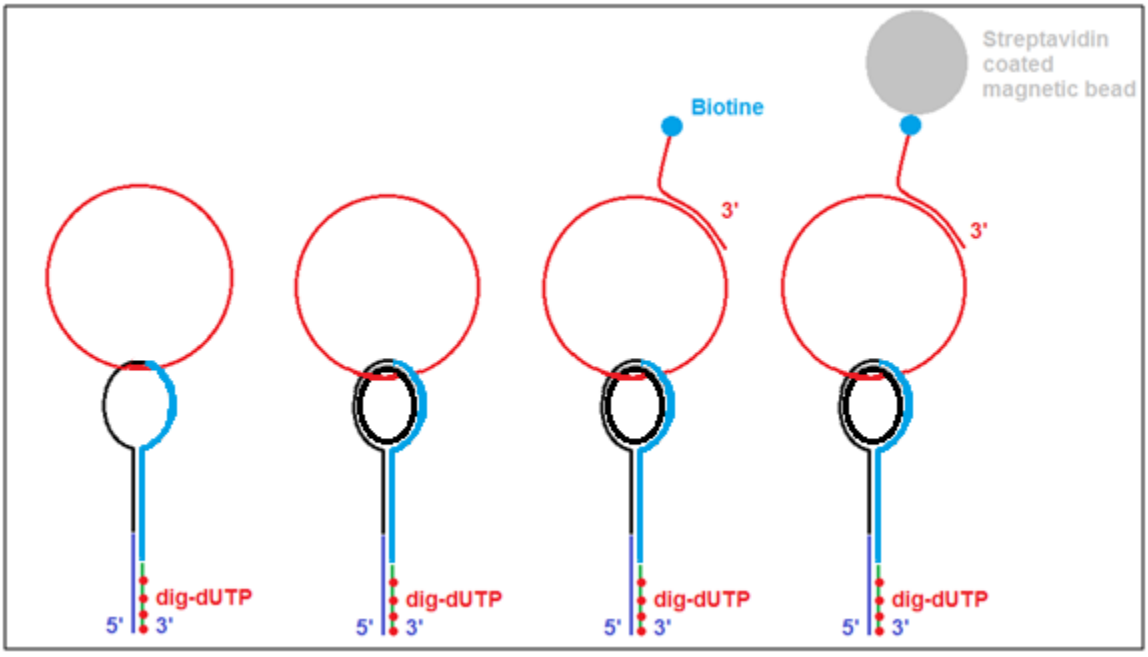


Fig. VI.1: Construction of a rolling circleDNA substrate. *A circular ssM13 DNA is hooked by the double-stranded loop of a DNA hairpin tether. A DNA primer with a 5' biotin is hydridized to the circular DNA. A streptavidin coated magnetic bead is attached to the biotin on the DNA substrate.*

To construct a rolling circle DNA substrate, we first annealed a circular ssM13 DNA and two oligonucleotides that are partially complementary to each others, and at the same time each one of the oligonucleotides has 8 bases complementary to the circular DNA. When the oligonucleotides are partially annealed and hybridized to the circular DNA and then ligated, they will form a hairpin having 16 loop-bases hybridized to the circular DNA. Thus the loop of the hairpin is making one and half turn around the circular DNA. Then in order to attach the DNA hairpin to an anti-digoxygenin coated surface, we ligate to its 5' end an oligonucleotide, and then we make a fill-in reaction using the Klenow 3'-5' exonuclease that uses the latter oligonucleotide as a matrix to add dig-dUTP to the 3' end of the hairpin. Now, in order to hook the circular DNA to the loop of the hairpin, we add 3 oligonucleotides that are complementary to the hairpin loop. Two mismatches of two thymine bases are added to have a small stretch of ssDNA in order to increase the flexibility of the attachement loop. The result is a circular DNA hooked by the double stranded loop of the DNA hairpin tether. Finally, we add a 5' biotin-modified oligonucleotide having 16 bases complementary to the circular DNA. The biotine will be used later to bind a streptavidin coated magnetic bead. And the oligonucleotide will be used later as a primer for the circular DNA replication.

VI.2 Rolling circle replication

We inject the rolling circle molecules into the chamber, and when they are tethered we apply a pulling force. We inject the polymerase to replicate the circular DNA. The polymerase extends the primer and when it makes the round of the circular DNA, the polymerase will displace the strand where the bead is attached and begin a new round of replication. This time, because the 5' end is pulled to the top by the force, the circular DNA will rotate and the polymerase will replicate it while staying at the junction. The circular DNA may be replicated many times depending on the processivity of the polymerase.

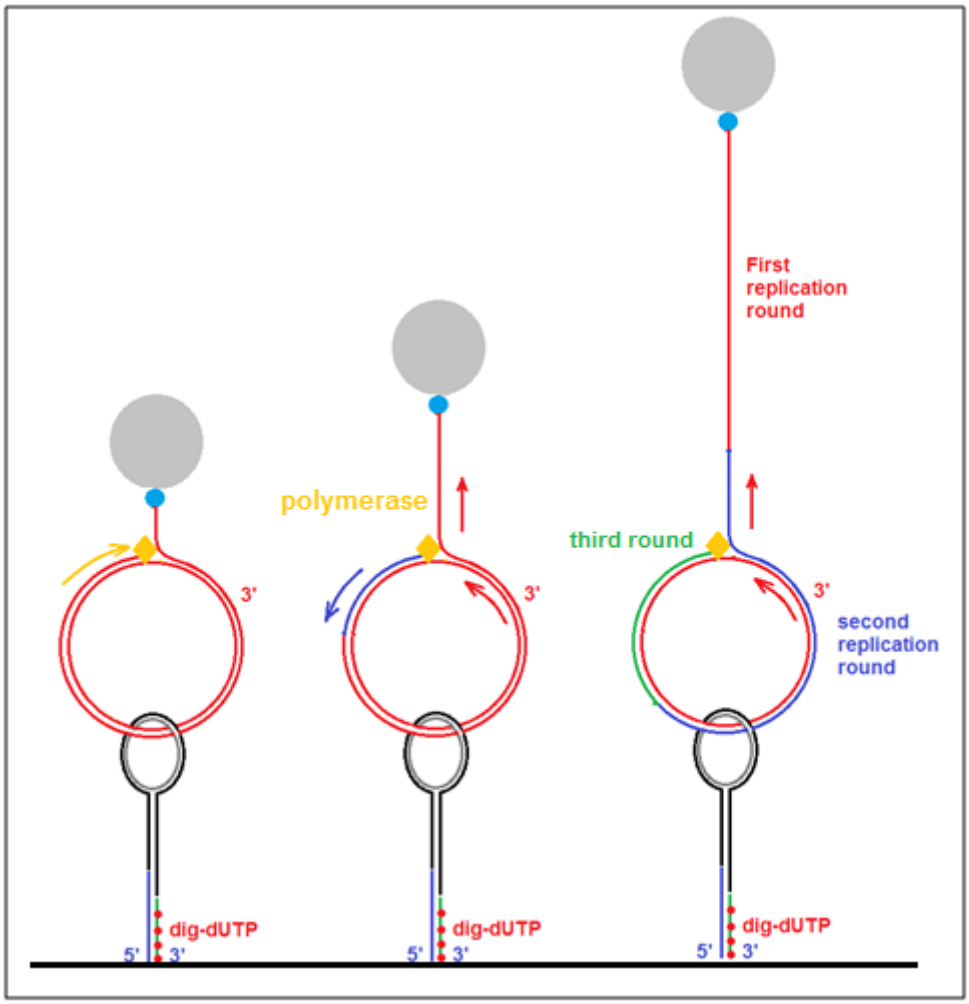


Fig. VI.2: Replication of a rolling circle DNA molecule by a polymerase. *The polymerase replicates the circular DNA, and when it arrives to the junction with the 5' single-stranded overhang, the polymerase displaces the DNA strand and stays at the junction and replicates the circular DNA while the latter rotates.*

VI.2.1 gp43 exo- polymerase

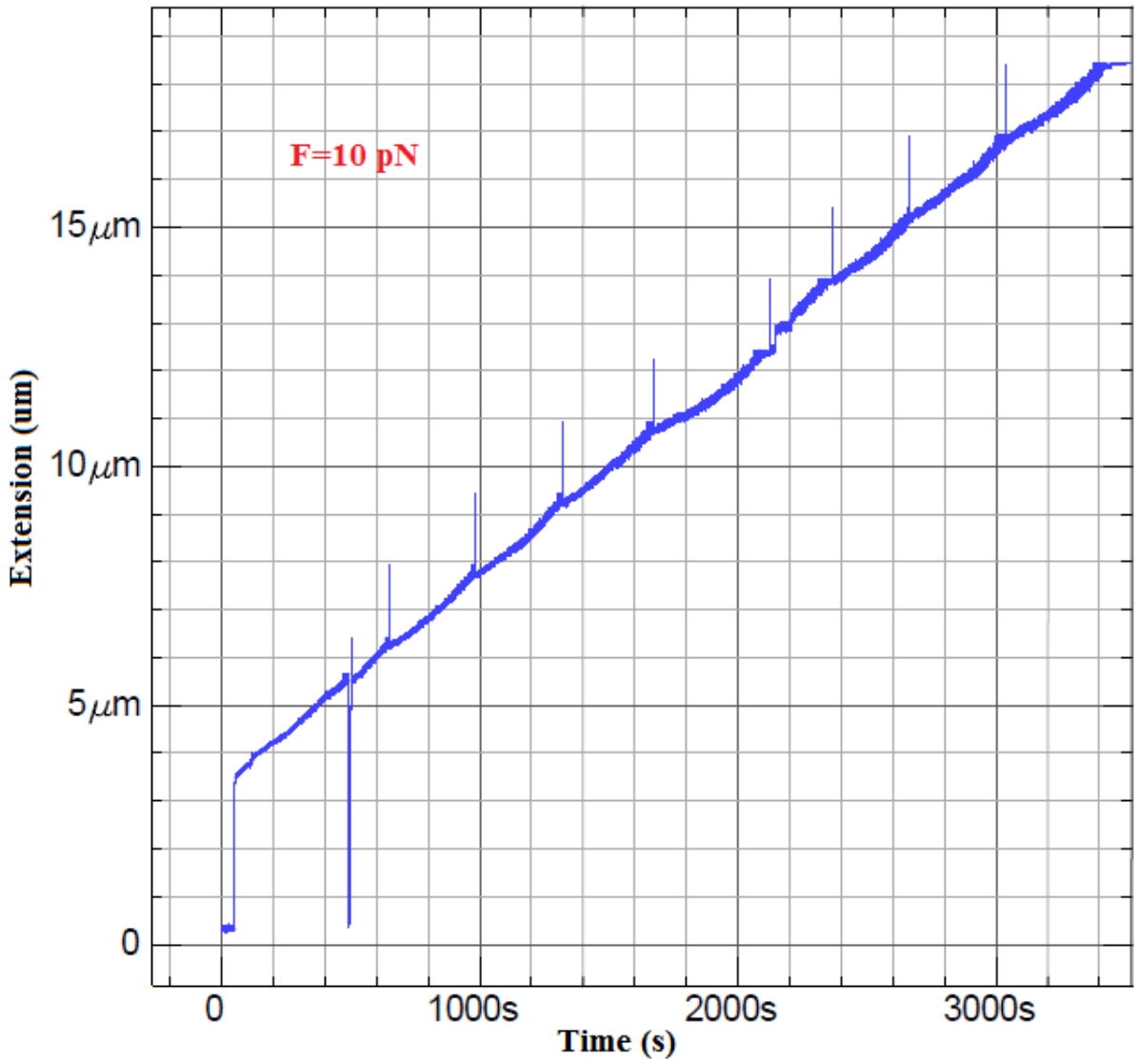


Fig. VI.3: gp43 exo- polymerase replicating the rolling circle DNA molecule. *It generates a single stranded DNA of 14 µm. The spikes are due to focus adjustments because the molecule length exceeds the calibration image extent.*

We add GP43 exo- to the rolling circle molecules made with the M13 circular DNA. As expected, some molecules have been replicated several times and the M13 has turned several times for an hour sliding in the loop without serious problems. The replication rate was found to be of 5 nm/s which corresponds to 20 bp/s.

VI.2.2 Manta polymerase

The same test was done with the manta polymerase that resulted in nearly continuous replication and synthesis of a traceable single – stranded DNA tail of 50 µm. In fact we

cannot follow an extension that is superior to this value, which is equal to the chamber double - sided tape, because at this level, the bead touches the ceiling of the chamber (the mylar sheet) and its position along the axis is constant. Even if the polymerase is still replicating the circular DNA, the generated single – stranded DNA is going to accumulate in the chamber without being visualized.

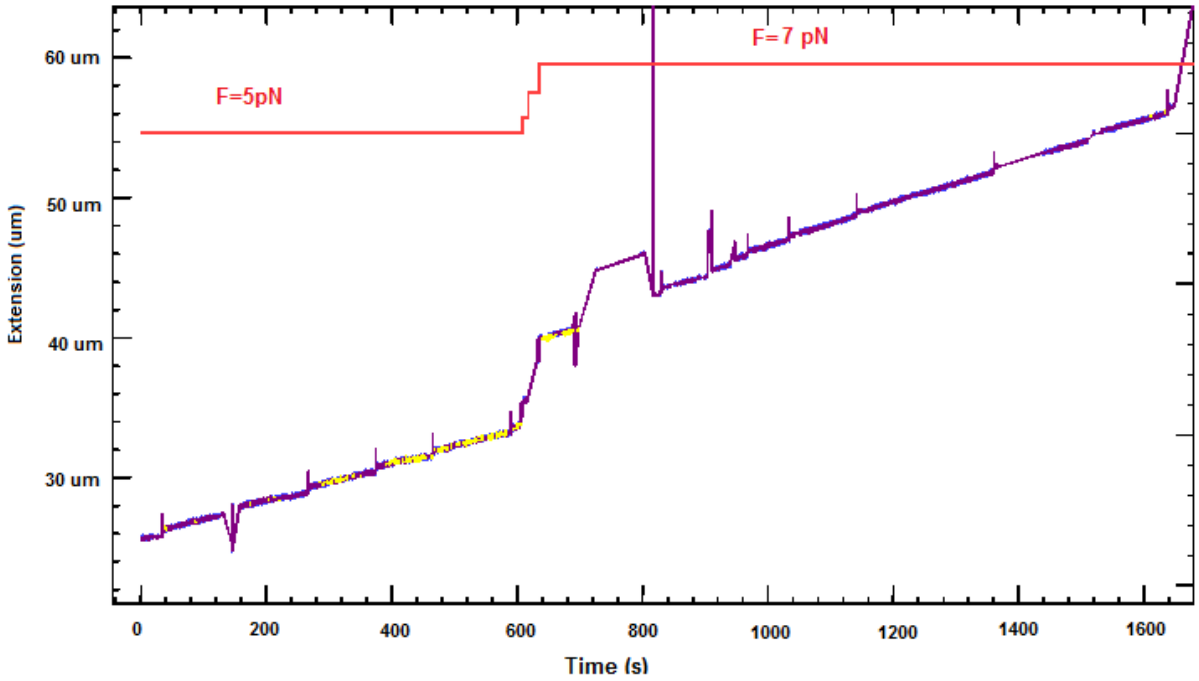


Fig. VI.4: Manta polymerase replicating the rolling circle DNA molecule. *It generates a single stranded DNA of 50 µm.*

VI.3 T4 coupled leading strand and lagging strand polymerases

During the replication, the T4 replisome is known to act by coupling the polymerases of the leading and the lagging strands. However, the leading-strand polymerase is moving in the same direction as the helicase, while the lagging-strand polymerase is moving in the opposite direction. Therefore, the lagging-strand polymerase forms a loop out of the single-stranded lagging strand, so that both polymerases can replicate the DNA while moving in the same direction. When the lagging strand polymerase encounters the previously synthetized Okazaki fragment it dissociates and initiates a new Okazaki ffragment from another primer.

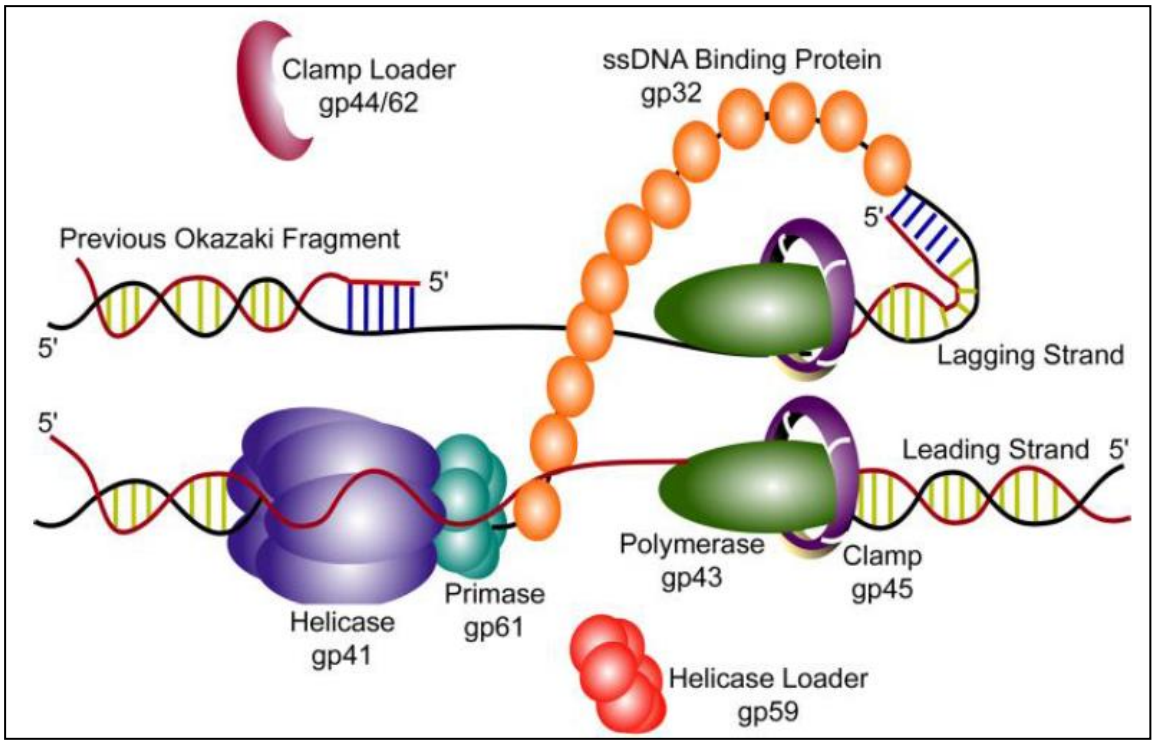


Fig. VI.5 : A model of the T4 bacteriophage DNA replisome*[149]. Replication of T4 genomic DNA is accomplished by a replication complex composed of eight proteins. The helicase (gp41) and primase (gp61) interact to form the primosome with the assistance of the helicase loader (gp59). The primosome complex encircles the lagging strand DNA, unwinding duplex DNA while synthesizing RNA primers for use by the lagging strand polymerase (gp43). DNA synthesis on both strands is catalyzed by a holoenzyme complex consisting of a polymerase (gp43) and a trimeric processivity clamp (gp45). The clamp is loaded onto the DNA by the clamp loader complex (gp44/62). The leading and lagging strand holoenzymes interact to form a dimer. Single-stranded DNA formed by the helicase is coated with single-stranded DNA-binding protein (gp32).*

VI.3.1 GP43 polymerase

We wanted to study the coupling between the leading and lagging strands polymerases in a replisome on rolling circle DNA substrate. The drawing below explains what we expect to happen on this type of substrate while maintaining a force to simulate the helicase activity.

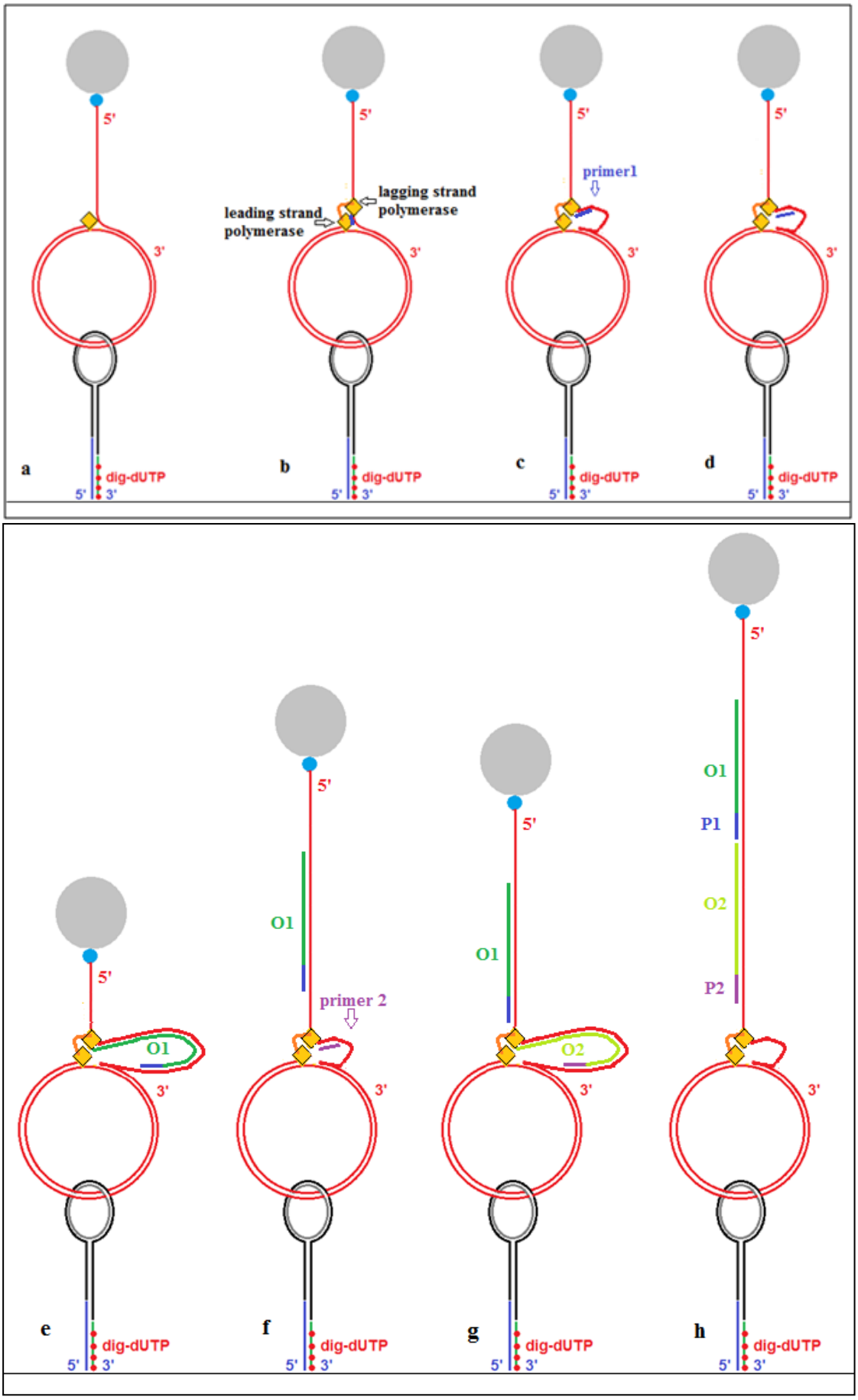

a
b
c
d
e
f
g
h
5'
3'
dig-dUTP
leading strand polymerase
lagging strand polymerase
primer1
primer 2
O1
O2
P1
P2

Fig. VI.6 : Rolling circle replication with the coupled lagging and leading strands polymerases. *a) The leading polymerase achieves the first replication round. b) The lagging polylerase loads on the lagging strand and intercats with the leading strand polymerase. c) The leading strand replicates the circular DNA that rotates but the lagging strand polymerase forms a loop with the resulting single stranded DNA. d) and e) The lagging strand polymerase synthetizes an Okazaki fragment which causes the extension of the bead to decrease as the single-stranded DNA is converted to double-stranded DNA. f), g) and h) a second primer is added and the process restarts. We can imagine the polymerase acting on the circular DNA as the needle of the sewing machine acting on textile.*

VI.3.2 Minimal replisome (gp41 helicase and gp43 polymerase)

We have added only the minimal replisome consisting of the helicase gp41 and the polymerase gp43. Below, we show some of the traces we have observed.

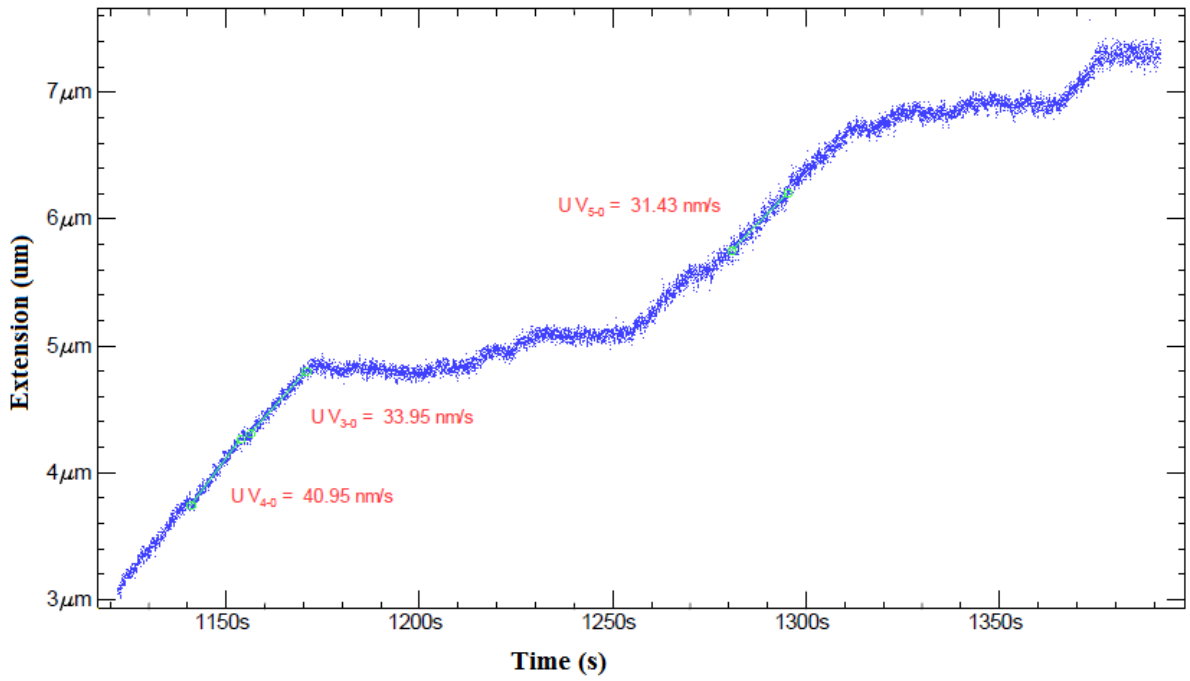


Fig. VI.7 : The rolling circle is replicated without coupling between the lagging and leading strands polymerases.

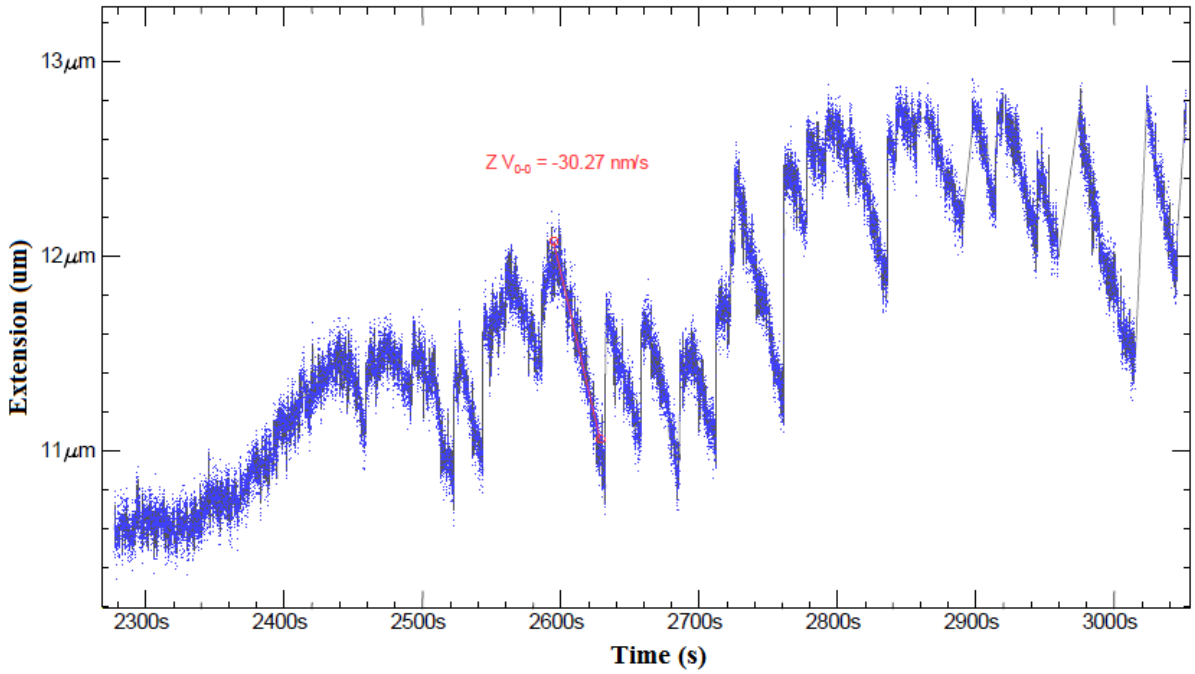


Fig. VI.8 : Rolling circle replication showing a coupling between the lagging and leading strands polymerases. *The abrupt extension jumps are due to the release of the lagging strand trapped by the lagging strand polymerase.*

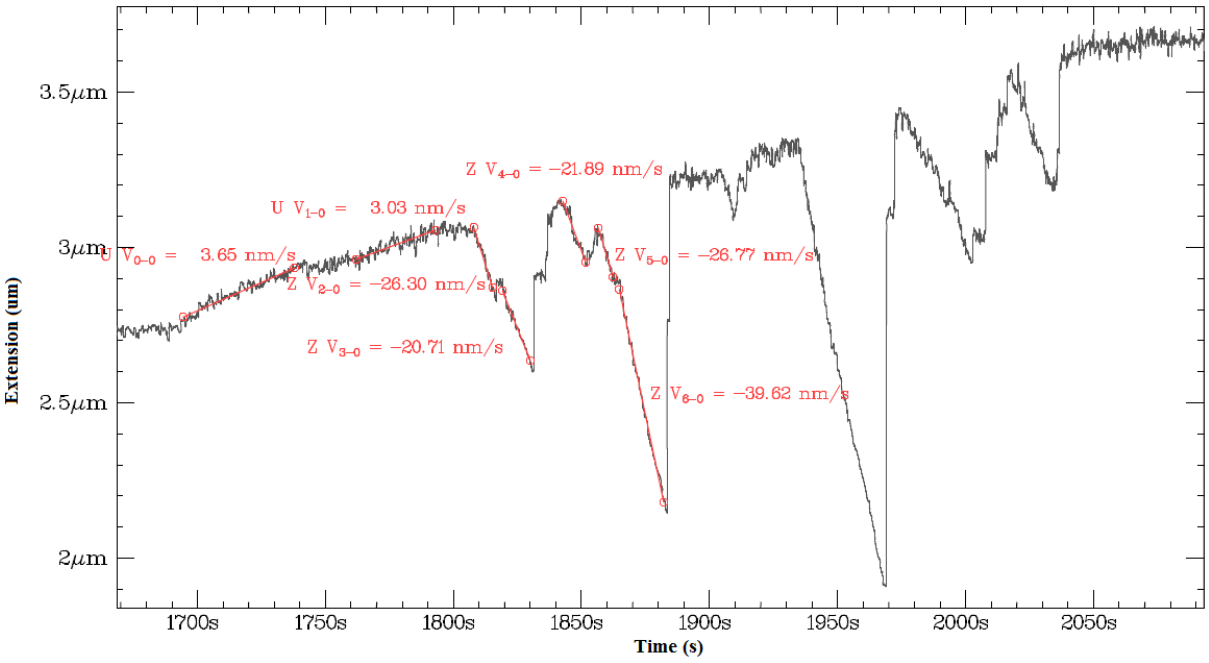


Fig. VI.9: Rolling circle replication with the coupled polymerases.

Resumé du travail de thèse en français

Résumé du travail de thèse en français

I- Introduction

La forme de l'ADN la plus présente *in vivo* est une double hélice formée par deux brins complémentaires où une adénine se lie à une thymine et une guanine se lie à une cytosine par des liaisons Watson-Crick. Cependant, les séquences d'ADN riches en guanines (G) peuvent former des structures secondaires d'ADN, intramoléculaires ou intermoléculaires, appelés G-quadruplexes (G4s) en présence de cations monovalents ou divalents comme le k+ et le Na+. Au moins quatre répétitions d'au moins deux guanines sont requises pour la formation de G4s intramoléculaires. Un G-quadruplex est une structure d'ADN non-canonique possédant quatre brins formée par l'empilement de quartets de guanines coordonnés par des cations, où un quartet est composé de quatre guanines liés d'une façon cyclique par des liaisons Hoogsteen. Les G4s ont une grande variabilité topologique qui dépend du nombre de brins qui les constituent et de leurs orientations ainsi que de la forme des boucles qui sont formées par les bases qui séparent les séries de guanines. Ces structures ont été découvertes *in vitro* et puis retrouvées in vivo.

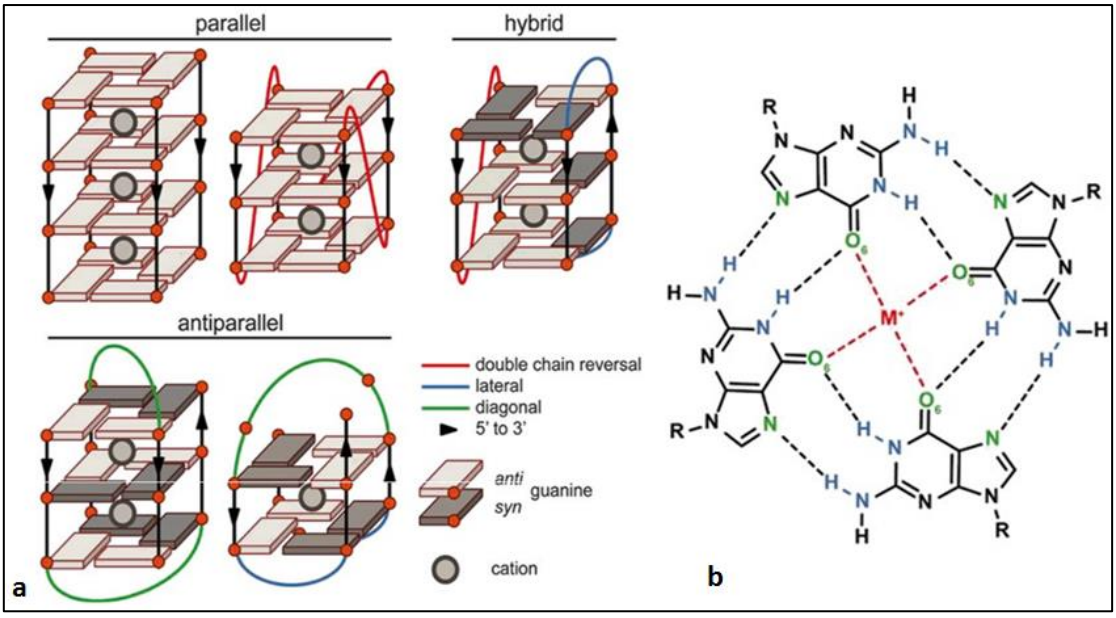


Fig. 1: Les G-quadruplexes. a) *Les structures G4s sont formés par un empilement de quartets de guanines. Les différentes topologies des G4s dépendent du nombre de brins qui les constituent, de leur orientation et de la forme des boucles.* ***b)*** *Les quartets de guanines comprennent huit liaisons Hydrogène et sont coordonnés par un cation.*

Les études bioinformatiques rapportent la présence de plus de 700 000 séquences riches en G qui sont susceptibles de former des G-quadruplexes dans le génome humain. Cependant, le nombre de séquences qui forment un G4 in vivo est encore inconnu. Dans le génome humain, les séquences susceptibles de former des G4 ne sont pas éparpillées mais localisées dans des régions qui ont une signification biologique comme les extrémités simple brin des télomères et les promoteurs des oncogènes qui sont des régions d'ADN double brin. Ces séquences sont aussi présentes dans l'ADN des mitochondries et dans l'ARN. Les séquences de G4s ont été aussi retrouvées dans les virus et les bactéries.

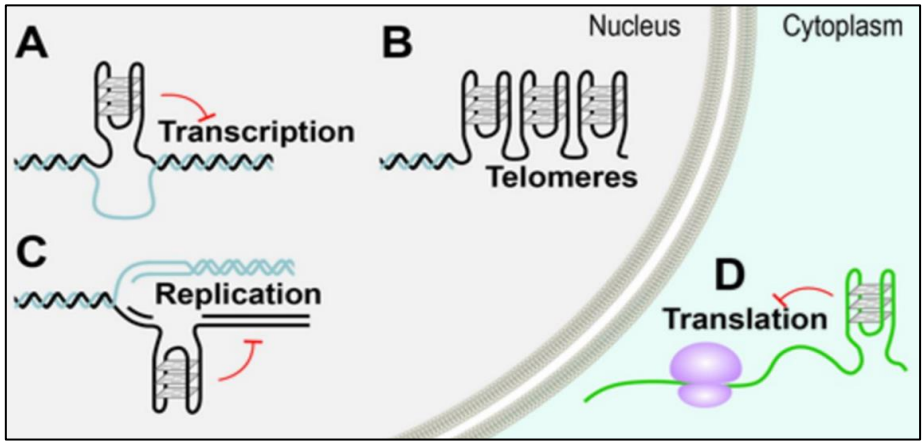


Fig.2: La distribution des séquences susceptibles de former les G quadruplexes in vivo. ***(A)*** *G4 dans un promoteur de gène,* ***(B)*** *G4 dans la région simple brin des télomères* ***(C)*** *G4 dans un origine de réplication,* ***(D)*** *G4 dans un ARN.*

Pour qu'une structure G4 enfouie dans une double hélice d'ADN puisse se replier, l'ADN doit être en simple brin. Ceci peut donc arriver lorsque la double hélice d'ADN est transitoirement ouverte au cours de la réplication et la transcription, à ce niveau il y a un équilibre dynamique entre l'ADN simple brin, l'ADN double brin et la structure G4. Lorsqu'une structure G4 se replie sur un des deux brins de la fourche de réplication, elle peut constituer un obstacle qui perturbe la progression des enzymes de la réplication comme l'hélicase et la polymérase. Ces structures peuvent donc arrêter la réplication et conduire à des délétions génétiques. De la même façon, au cours de la transcription, la formation d'une

structure G4 dans le promoteur d'un gène peut supprimer son expression. La stabilisation des G4s dans les proto-oncogènes est donc visée dans les traitements anti-cancer. La formation d'un G4 sur l'extrémité des télomères peut inhiber l'activité de la télomérase qui est surexprimée dans 80% des cancers. Ceci peut donc empêcher l'extension des télomères. La stabilisation des G4 télomériques constitue une nouvelle approche pour limiter la croissance de cellules tumorales. Pour cette raison, depuis quelques décennies, il y a eu un grand intérêt dans l'étude des G4s comme des cibles pharmaceutiques dans les traitements du cancer. Les G4s ont été donc largement étudiés par les différentes méthodes biochimiques et biophysiques, en solution ainsi qu'en molécule unique. Plusieurs papiers rapportent une étude thermodynamique et cinétique de la formation et la déformation des G4s. De plus, plusieurs études ont rapportés la stabilisation des G4s par des protéines ou des ligands ainsi que la résolution de ces structures par certaines hélicases comme RHAU, WRN, BLM et Pif1.

La surexpression du gène c-MYC est associée à plusieurs cancers comme les leucémies myéloïdes, le cancer du sein, les ostéosarcomes… Dans la région hypersensible à la nucléase de ce gène qui est impliquée dans 80-95% de la transcription du c-MYC, il existe une séquence susceptible de former un G-quadruplex (**TGGGGAGGGTGGGGAGGGTGGGGAAGG**). Cette séquence peut former plusieurs isomères de structure G-quadruplexes selon les guanines impliquées dans les quartets. Nous avons voulu étudier l'isomère majeur de ce G4 qui sera appelé par la suite G4(14,23) où les G14 et G23 font partie des boucles et non pas des quartets. Il a été rapporté que ce G4 est très stable et a une température de fusion de 75,5°C et 3 boucles en forme de N.

Dans cette étude, nous avons mis en place une nouvelle méthode pour étudier les structures G4 et leur interaction avec les enzymes par le dispositif des pinces magnétiques. La nouveauté de cette méthode c'est qu'on va pouvoir visualiser en direct la formation et la déformation d'une structure G4, ce qui va permettre d'étudier sa cinétique et sa stabilité ainsi que sa résolution par les enzymes.

II- Méthodes

II.1 Construction de hairpins d'ADN contenant une séquence G4

Nous avons réalisé une série d'expériences en molécule unique pour étudier les G4s insérés dans une région d'ADN double brin notamment un hairpin d'ADN où la séquence complémentaire riche en cytosine entre en compétition avec la formation de la structure du G4. Cette situation imite un G4 dans un promoteur ou aussi un G4 dans la région double brin des télomères situés en amont de la région simple brin des télomères. Un hairpin d'ADN simule une fourche de réplication et permet d'étudier l'effet des enzymes. Nous avons construit deux types de hairpins, le premier contient une séquence G4 du côté de l'extrémité 5' pour simuler un G4 sur le brin tardif, et l'autre hairpin contient la même séquence G4 sur le brin complémentaire, du côté de l'extrémité 3' pour simuler un G4 sur le brin précoce. Les

hairpins ont une biotine sur l'extrémité 5' afin d'accrocher une bille paramagnétique couverte de streptavidine et ont plusieurs molécules digoxygénines sur leur extrémité 3' afin de les fixer sur une surface couverte d'anti-digoxygénines (fig.3).

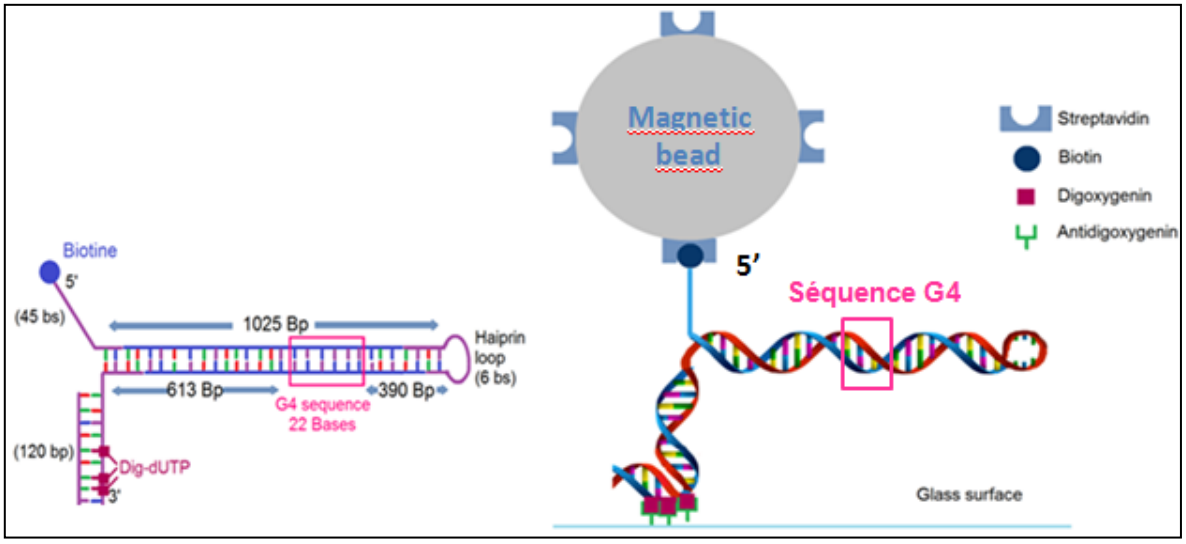


Fig.3: Le hairpin d'ADN contenant la séquence G4. *Cet hairpin est construit en assemblant plusieurs oligonucléotides : Le duplex d'ADN contenant la séquence G4 sur un brin, la boucle du hairpin, l'oligo biotine, et le duplex contenant les digoxygénines. Les hairpins sont ensuite accrochés à une bille couverte de streptavidine et puis fixés dans la cellule fluidique (fig.4).*

II.2 le dispositif de pinces magnétiques

Sur le dispositif de pinces magnétiques, nous fixons une cellule fluidique formée par : une lamelle de microscope couverte d'anti-digoxygénines et un papier milar, les deux sont joints par un scotch double-face (fig.4.A). Un canal rectangulaire découpé dans le scotch délimite la chambre fluidique où l'échantillon est injecté à l'aide d'un système de pousse seringue. L'échantillon est éclairé par dessus par une LED rouge, et observé par-dessous par un objectif de microscope 40x. L'image est ensuite transférée à une caméra CCD reliée à l'ordinateur. Des centaines de billes appartenant au même champ de vue sont observées et suivies simultanément. L'image d'une bille est un ensemble d'anneaux concentriques qui proviennent de l'interférence entre la lumière directe de la LED et la lumière diffractée par les billes. Les pinces magnétiques sont formées par deux aimants antiparallèles qui, une fois approchés de l'échantillon, exercent une force verticale qui va tirer les billes vers le haut. Cette force va permettre de manipuler les billes magnétiques et donc les hairpins d'ADN (fig.4.B,C). Une image de calibration est enregistrée pour chaque bille au début de l'expérience. Ceci est fait en déplaçant l'objectif verticalement afin de changer la distance entre l'objectif et les billes ce qui change le diamètre des anneaux de diffraction. Cette relation entre la distance objectif-bille et le diamètre des anneaux est utilisée au cours du suivi pour déduire, de la forme des anneaux, la position des billes et donc l'extension de la molécule à une résolution de quelques nanomètres.

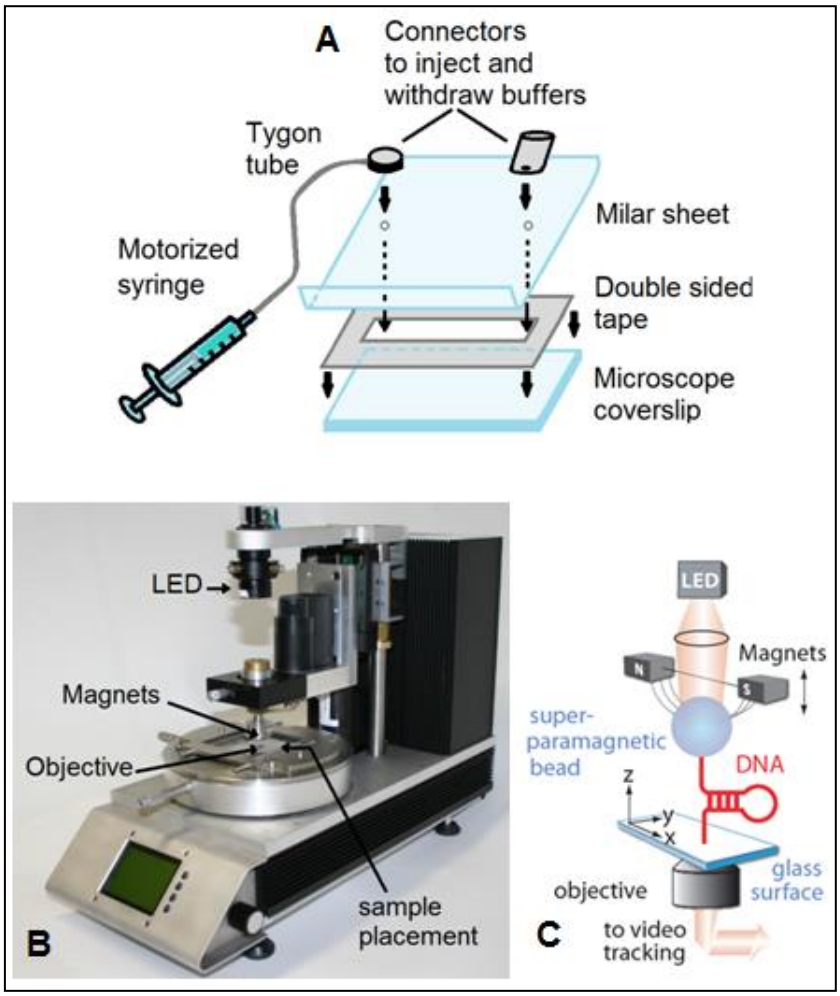


Fig.4: Le dispositif des pinces magnétiques. *A) La cellule fluidique formée en assemblant une lamelle de microscope et un papier milar. B) Le dispositif des pinces magnétiques. C) Le hairpin d'ADN est tout d'abord accroché à une bille magnétique et puis fixé sur la lamelle de microscope. L'échantillon est éclairé du dessus par la LED et observé du dessous par un objectif de microscope qui envoie l'image à une camera CCD. Les pinces magnétiques appliquent une force verticale vers le haut sur la bille et par suite sur la molécule d'ADN.*

La force nécessaire pour ouvrir un hairpin d'ADN est de 15 pN. Pour refermer le hairpin, la force doit être baissée à 10 pN, pour reformer le centre de nucléation du hairpin qui conduit à la fermeture complète de ce dernier (fig.5).

La grandeur mesurable dans notre dispositif est la position de la bille et donc l'extension de la molécule d'ADN. Afin de calculer le nombre de paires de bases ouverts (Bp), on utilise les courbes de conversion nm/Bp (fig.6).

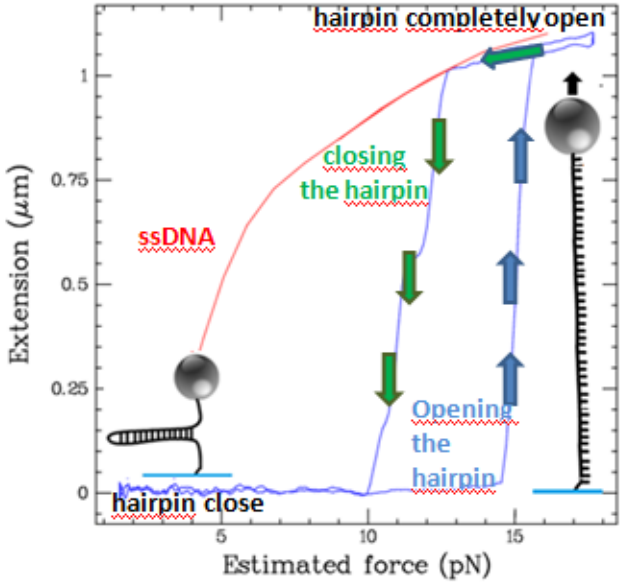


Fig.5: Courbes d'extension de l'ADN en fonction de la force. *L'extension d'un ADN simple brin ayant 2000 bases est représenté par la courbe rouge. L'ouverture d'un hairpin d'ADN de 1000 paires a lieu à 15 pN. La fermeture du même hairpin a lieu vers 10 pN.*

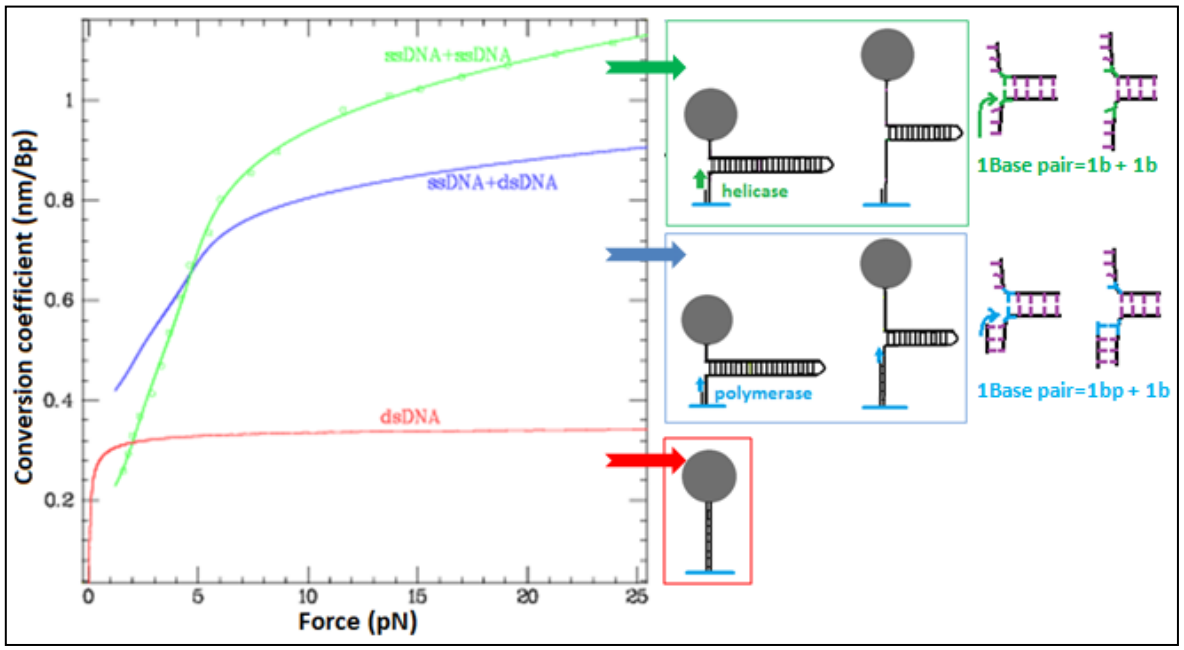


Fig. 6: Courbes de conversion nm-paire de base. *Courbe verte: Variation de l'extension d'une paire de base d'ADN ouverte par une hélicase suivant la force. L'hélicase génère deux bases simple brin. Courbe bleue: Variation de l'extension d'une paire de base d'ADN ouverte par une polymérase suivant la force. La polymérase réplique une base et libère la base sur*

l'autre brin. Courbe rouge: Variation de l'extension d'une paire de base suivant la force appliquée.

II.3 Blocage de la fermeture du hairpin d'ADN par un oligonucléotide

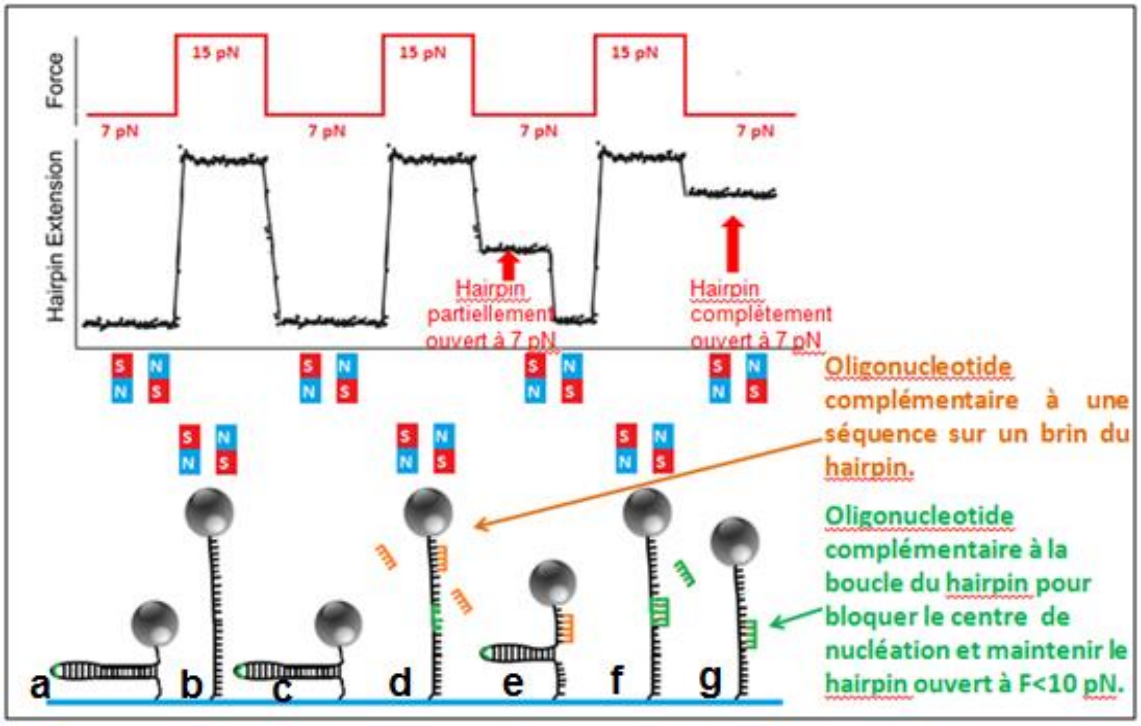


Fig. 7: Ouverture et fermeture d'un hairpin en présence d'un oligonucléotide complémentaire à une séquence du hairpin. *Les cycles de force sont représentés en haut de l'image. En bas de l'image sont représentés les hairpins correspondants en absence et en présence d'un oligonucléotide ainsi que les courbes d'extension* ***(a)*** *à 7 pN, le hairpin est complètement fermé,* ***(b)*** ***à*** *15 pN le hairpin est ouvert,* ***(c)*** *le fait de baisser la force à 7 pN entraîne la fermeture complète du hairpin* ***(d)*** *En ajoutant un oligonucléotide complémentaire à un segment du duplex d'ADN, cet oligonucléotide s'hybride quand le hairpin est ouvert à 15 pN,* ***(e)*** *Quand la force est baissé à 7 pN, le hairpin se ferme partiellement et reste bloqué par l'oligo* ***(f)*** *Quand un oligonucléotide complémentaire à la boucle du hairpin est ajouté, ce dernier va s'hybrider à la boucle qui est le centre de nucléation du hairpin et* ***(g)*** *empêche la refermeture du hairpin quand la force est baissée à 7 pN, en gardant le hairpin complètement ouvert.*

Si on ajoute un oligonucléotide complémentaire à une séquence de la région double brin du hairpin, cet oligonucléotide va s'hybrider à sa séquence complémentaire lorsque le hairpin est ouvert. Cet oligonucléotide va bloquer la fermeture complète du hairpin pendant une certaine durée qui dépend de la longueur de cet oligonucléotide, sa séquence, et la force appliquée. Lorsqu'on diminue la force, le hairpin va se refermer partiellement jusqu'à bloquer

sur l'oligonucléotide. Cependant, si l'oligonucléotide est complémentaire à la boucle du hairpin, lorsqu'il s'hybride pendant l'ouverture du hairpin, il va bloquer la fermeture un hairpin. Dans ce cas, lorsqu'on diminue la force, le hairpin va rester complètement ouvert mais son extension va diminuer à cause de la diminution de la force et donc la variation de l'extension des bases (fig.7).

III- Résultats

III.1 Repliement d'une structure G4 dans un hairpin

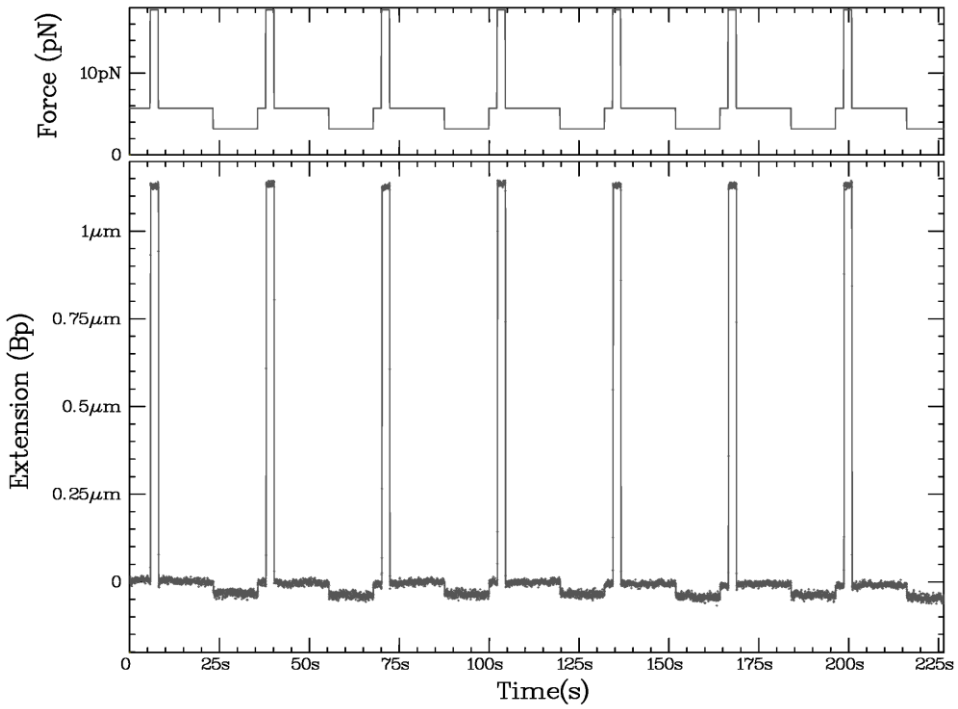


Fig.8: Une série de cycles d'ouverture - fermeture du hairpin ayant une séquence de G4 sur un brin. *En haut de l'image sont représentés les cycles de force faits en 60 mM qui consistent à une alternance de haute force pour ouvrir le hairpin et une basse force pour le refermer. En bas de l'image est représentée l'extension du hairpin. Ce dernier s'ouvre à haute force puis se referme complètement en baissant la force à moins que 10 pN. Clairement, la structure de G4 ne s'est pas formée.*

Pour que la séquence de G4 qui est présente sur l'un des deux brins du hairpin se replie en une structure G4, le hairpin doit être ouvert. Une fois repliée, cette structure va entraîner le blocage de la fermeture du hairpin, similairement à l'hybridation d'un

oligonucléotide sur un brin. Ceci devrait se voir dans la courbe d'extension de la molécule d'ADN. Nous avons donc essayé de former cette structure en faisant des cycles de forces pour ouvrir et fermer le hairpin en présence de 60 mM k+. Cependant aucun blocage n'a été observé et donc aucune structure G4 n'a été formée. Nous avons donc suggéré que lorsque le hairpin est complètement ouvert, la force appliquée (15 pN) est très élevée et empêche la formation du G4. Nous avons donc besoin de maintenir le hairpin complètement ouvert à une force plus basse (7-8 pN) sauf que si on diminue la force à moins que 10 pN le hairpin va se refermer et le G4 ne peut plus se former. Nous avons donc utilisé un oligonucléotide de 7 bases complémentaire à la boucle du hairpin, qui une fois hybridé pendant l'ouverture du hairpin, va permettre de baisser la force à 7 pN en maintenant le hairpin complètement ouvert. Effectivement cette méthode, en présence de 60 mM k+, nous a permis de former le G4 sur la majorité des molécules (fig.9-10).

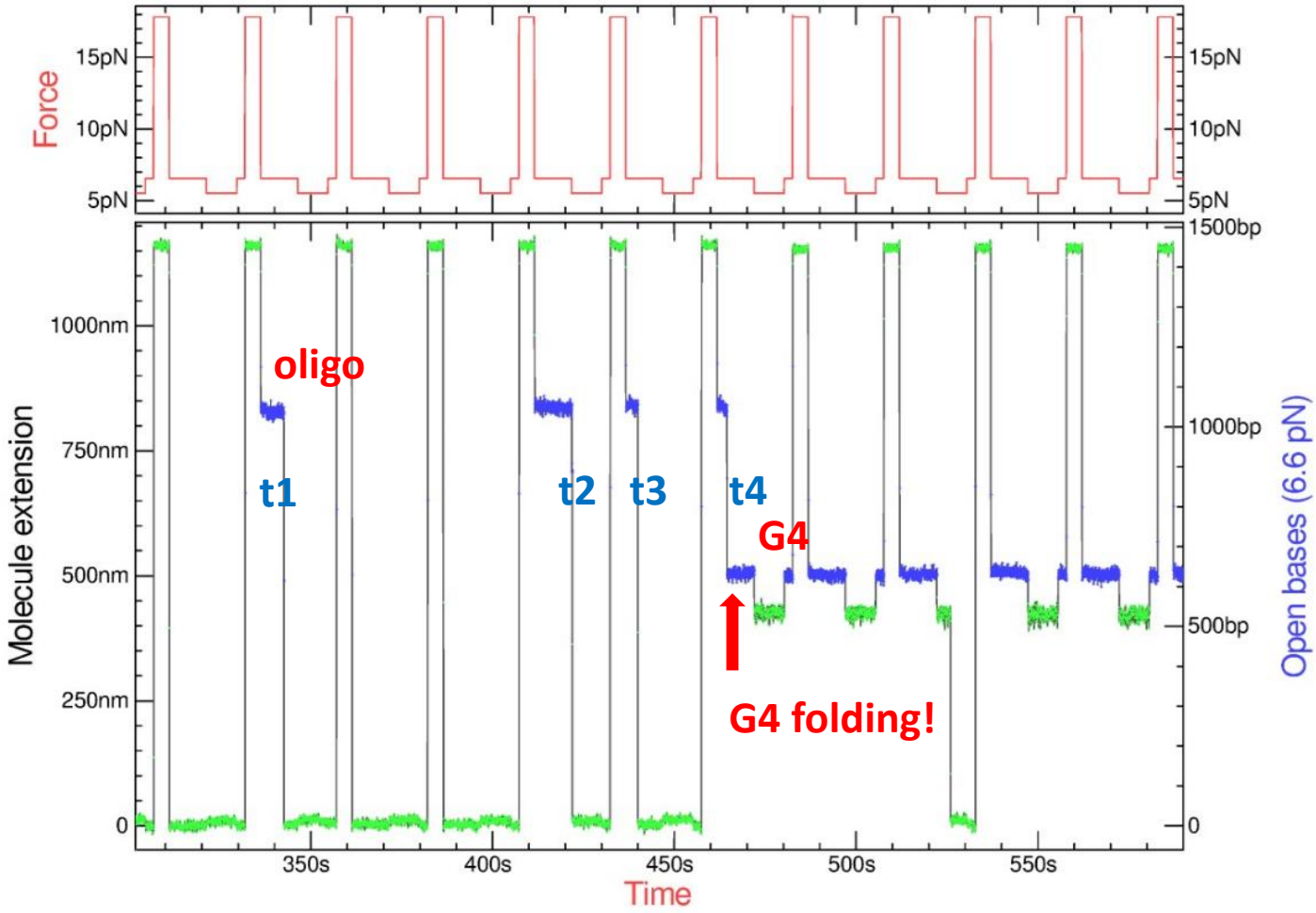


Fig. 9: Repliement du G4 dans un hairpin d'ADN en utilisant l'oligo complémentaire à la boucle. Ce graph monter les cycles de force et l'extension du harpin. *L'extension maximale représente le hairpin complètement ouvert, le blocage de fermeture du hairpin à 850 nm représente le blocage le l'oligo complémentaire à la boucle où le hairpin est maintenu complètement ouvert à la force de 6.6 pN. Les cycles qui ne conduisent pas à un second blocage ne conduisent donc pas à la formation de la structure G4. Une fois le G4*

formé un deuxième blocage est bloque la fermeture du hairpin lorsque l'oligo de la boucle se détache. Durant ce blocage, le hairpin n'est que partiellement fermé. Seuls les bases entre le G4 et la boucle sont fermées. Seules les durées du blocage de l'oligonucléotide complémentaire à la boucle sont comptés dans le calcul de la durée de formation du G4. Dans ce cas, t, la durée de formation du G4 avec t1+t2+t3<t< t1+t2+t3+t4.

Lorsque l'oligonucléotide s'hybride à la boucle, il peut ou pas conduire à la formation du G4. Si le G4 ne se forme pas pendant le temps d'hybridation de l'oligonucléotide, lorsque ce dernier se détache, le hairpin va se refermer complètement. Cependant lorsque le blocage de l'oligonucléotide conduit à la formation du G4 et puis se détache, le hairpin va seulement se refermer partiellement. Le fait de baisser la force encore plus, à 4 pN va refermer complètement le hairpin en encerclant le G4. Nous avons trouvé que ce G4 peut être encerclé par le hairpin sans être résolu.

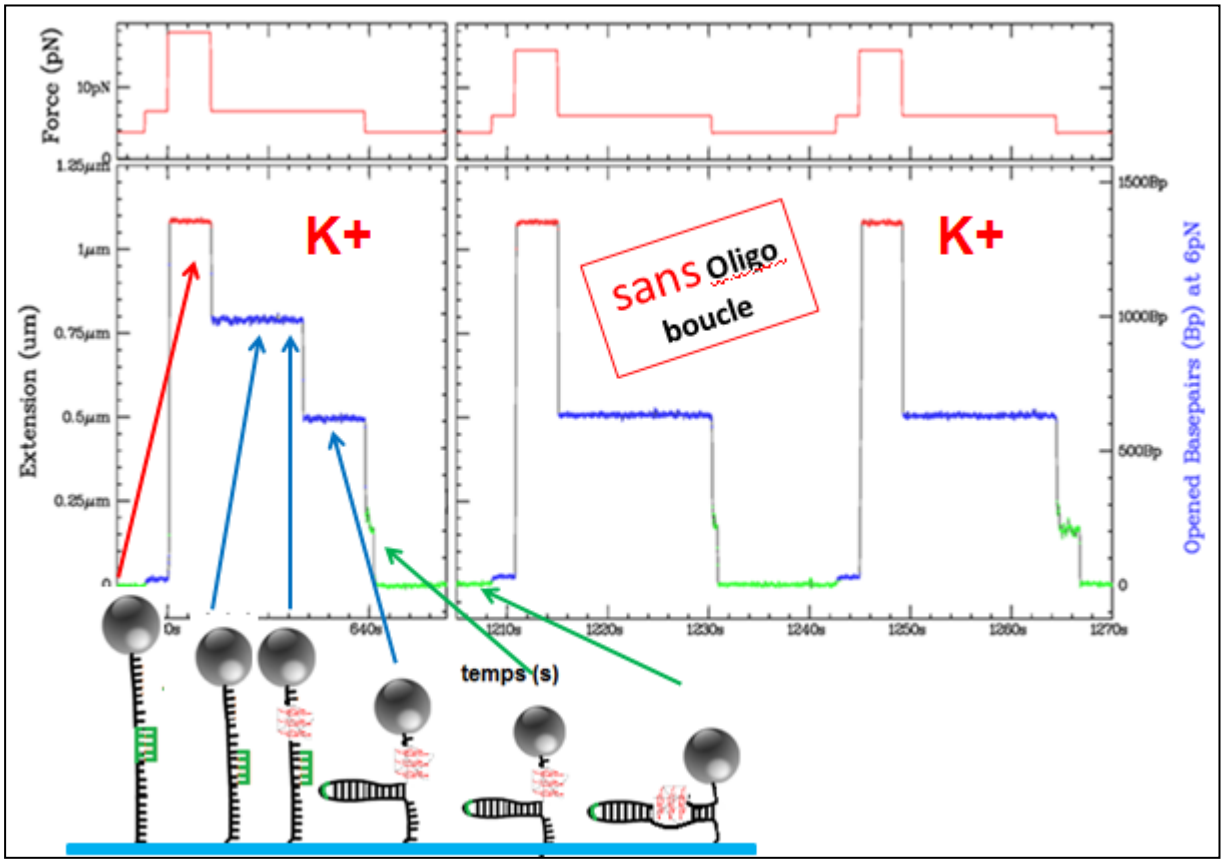


Fig. 10: Repliement du G4 dans le hairpin. *Dans le premier quadrant le plateau rouge représente l'extension du hairpin complètement ouvert à une force supérieure à 15 pN. Le plateau bleu suivant représente l'extension de la molécule maintenue complètement ouverte*

par l'oligonucléotide hybridé sur la boucle à 6.6 pN. Puis lorsque ce dernier se détache, le hairpin se referme partiellement en se bloquant sur le G4 et son extension diminue. En baissant la force encore plus à 4 pN, le hairpin reste dans la même configuration mais son extension diminue à cause de la variation de l'extension de la base d'ADN avec la force. Enfin le hairpin se referme complètement en encerclant le G4 sans le résoudre. Dans le deuxième quadrant, après le rinçage de l'oligo, le hairpin se bloque toujours sur le G4 au cours de la fermeture du hairpin.

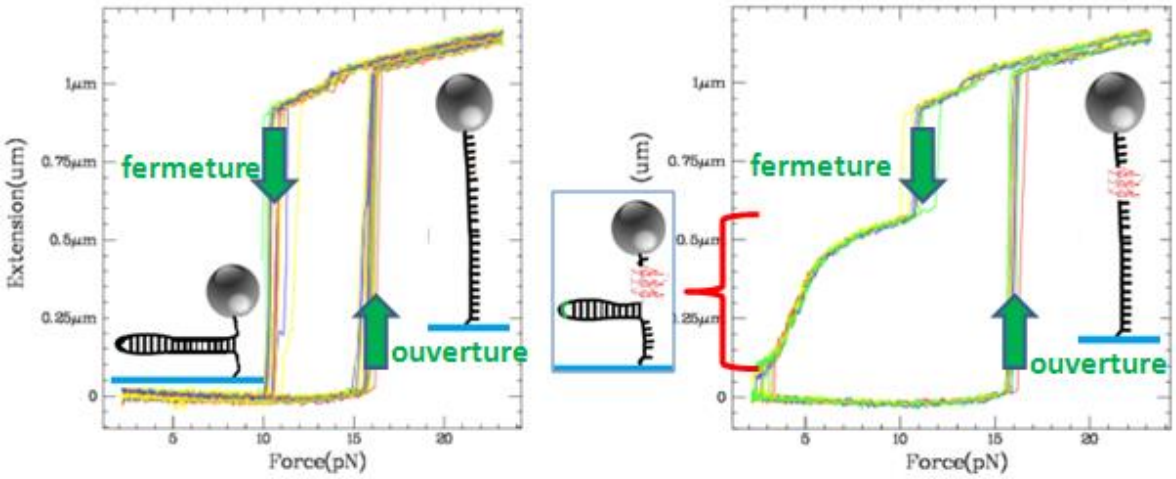


Fig. 11: Cycles d'ouverture et fermeture d'un hairpin d'ADN contenant une séquence de G4 C-MYC (14,23) A) avant B) et après la formation de la structure G4 dans 60 mM de k+.

III.2 Cinétique et stabilité des structures G4s

A partir des cycles de formation et de résolution des structures G4s, nous pouvons calculer le temps de formation des G4s et aussi leur stabilité (le temps de résolution des G4s). A l'aide de cette méthode nous avons mesuré les propriétés cinétiques de différentes structures de G4. Nous avons observés que le T_{off} varie fortement avec les différentes séquences de G4. Alors que le G4 du c-MYC est resté replié pendant des heures, les G4s télomériques sont stables pour quelques dizaines de secondes seulement (fig.12-13).

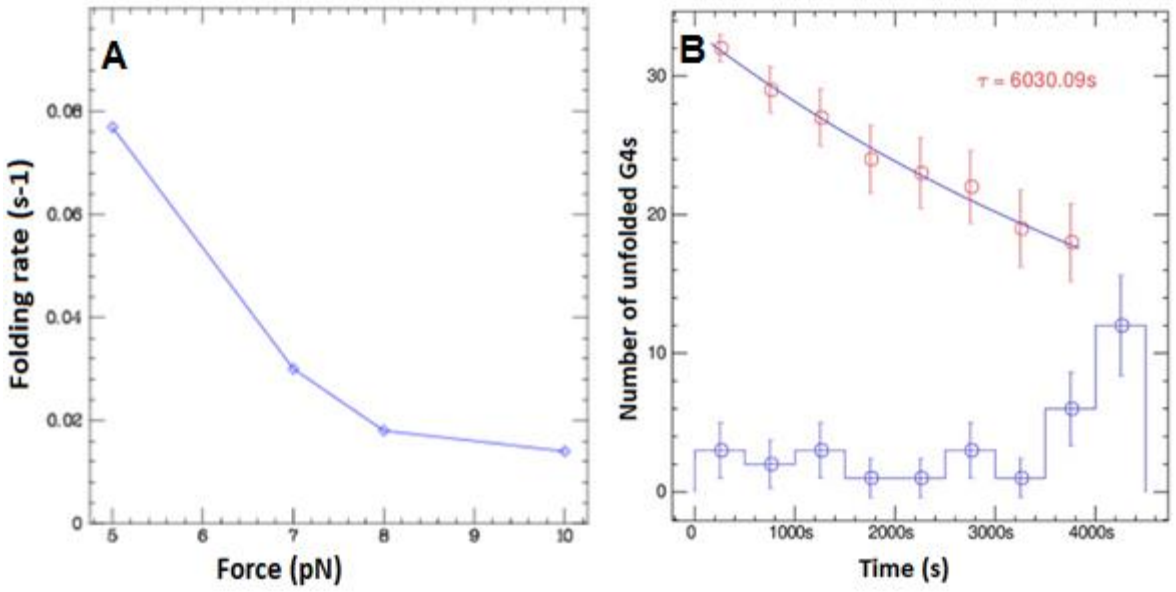


Fig. 12: A) Vitesse de repliement du G4 C-MYC (14, 23) : *La vitesse du repliement est représentée en fonction de la force à laquelle l'oligo est hybridé à la boucle.* **B) Stabilité du G4 C-MYC (14, 23) :** *Les points rouges représentent les G4 qui sont restés repliés au cours du temps.*

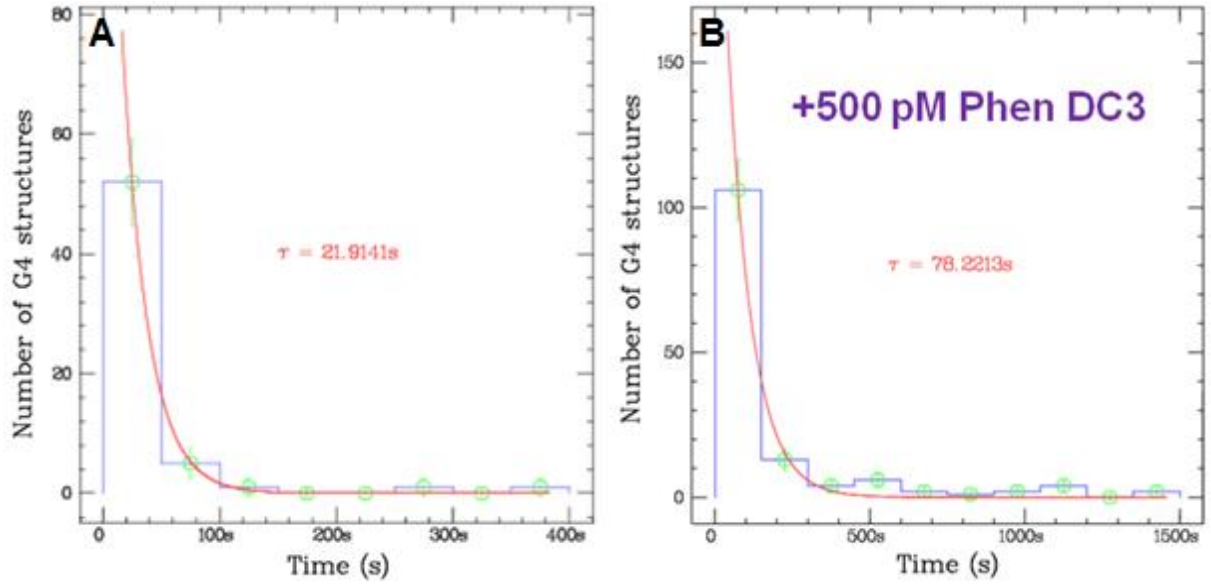


Fig. 13 : Stabilité d'une structure G4 du télomère humain A) sans B) et avec le ligand G4 Phen-DC3.

III.3 Repliement et stabilité de la structure G4 c-MYC(14,23) en présence des ions Na+

Nous avons essayé de former le G4 c-MYC(14,23) avec la même méthode mais en présence de 60 mM de cations Na+. Dans ces conditions aucun G4 n'a été replié.

Nous avons aussi voulu tester l'effet des ions Na+ sur les G4s qui ont été déjà repliés en k+. Pour se faire, nous avons repliés les G4s en k+ puis nous avons rincé la cellule fluidique avec une solution contenant 60 mM Na+. 50s après, nous avons observé la résolution de la plupart des G4s (fig.14).

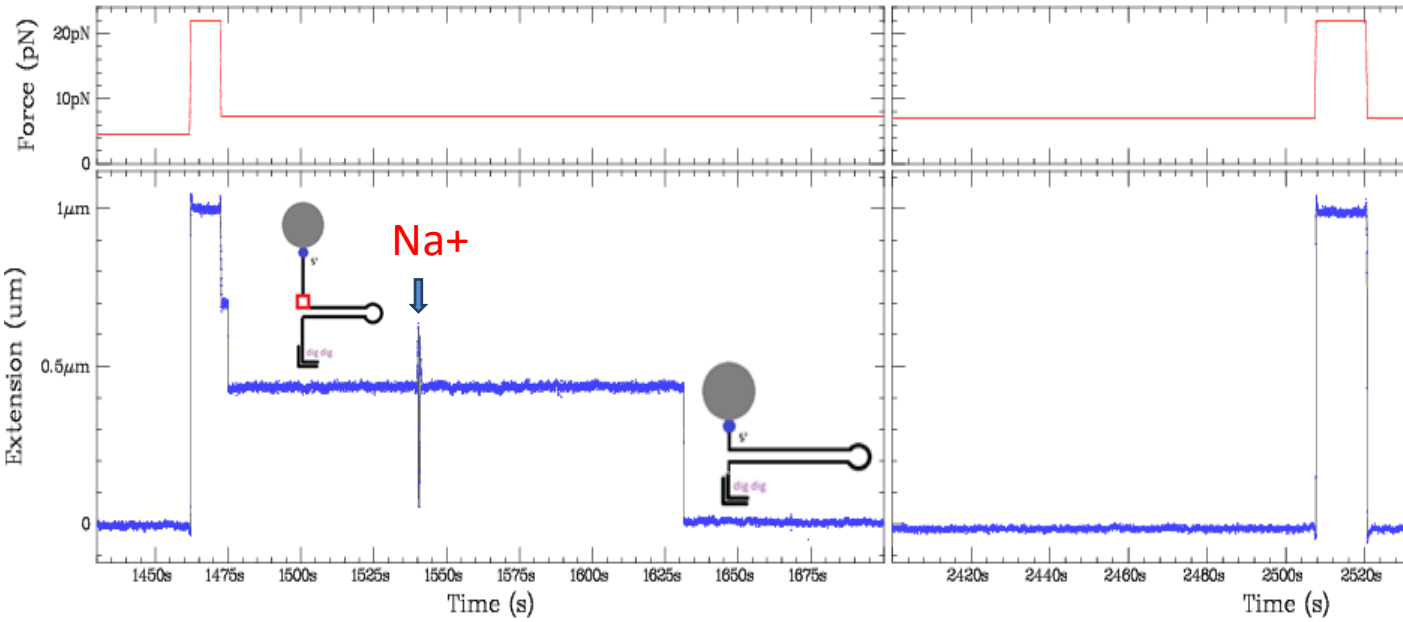


Fig. 14 : Le rinçage des G4s formés en k+ par une solution contenant 60 mM Na+ résout les G4s en une durée moyenne de 50s.

III.4 Agents stabilisant les structures G4s

Afin de tester si la protéine Sub1 et son homologue humain PC4, ainsi que le ligand Phen-DC3 peuvent stabiliser le repliement du G4, nous avons choisi de travailler en Na^+ puisque dans ces conditions salines le G4 c-MYC (14,23) n'est pas très stable. Donc nous avons réalisé 3 expériences différentes, où dans chacune on ajoute l'agent à tester dans une solution de Na+ aux G4s qui ont été formés dans une solution de k+. Nous avons trouvé que chacun de ces agents a un effet stabilisant sur la structure G4 (tableau 1).

Stabilizing agent	G4 unfolding duration in Na+ (s)	Ratio
-	50	
Sub1 protein	1000	20
PC4 protein	~2500	50
Phen DC3 G4 ligand	3000	60

Tableau 1 : Stabilité des structures G4s du c-MYC(14,23) dans une solution Na^+ en présence de Sub1, PC4 ou Phen-DC3.

III.5 Agents favorisant les structures G4s

Afin de tester si la protéine Sub1 et son homologue humain PC4, ainsi que le ligand Phen-DC3 favorisent le repliement du G4, nous avons ajouter ces molécules dans 3 tests différents et nous avons essayé de former le G4 en faisant des cycles de force sans ajouter l'oligo de la boucle. Effectivement, en présence de ces agents, le G4 a été formé (fig. 15). Nous pouvons donc considérer que ces molécules ont un rôle de chaperonnes dans le repliement du G4.

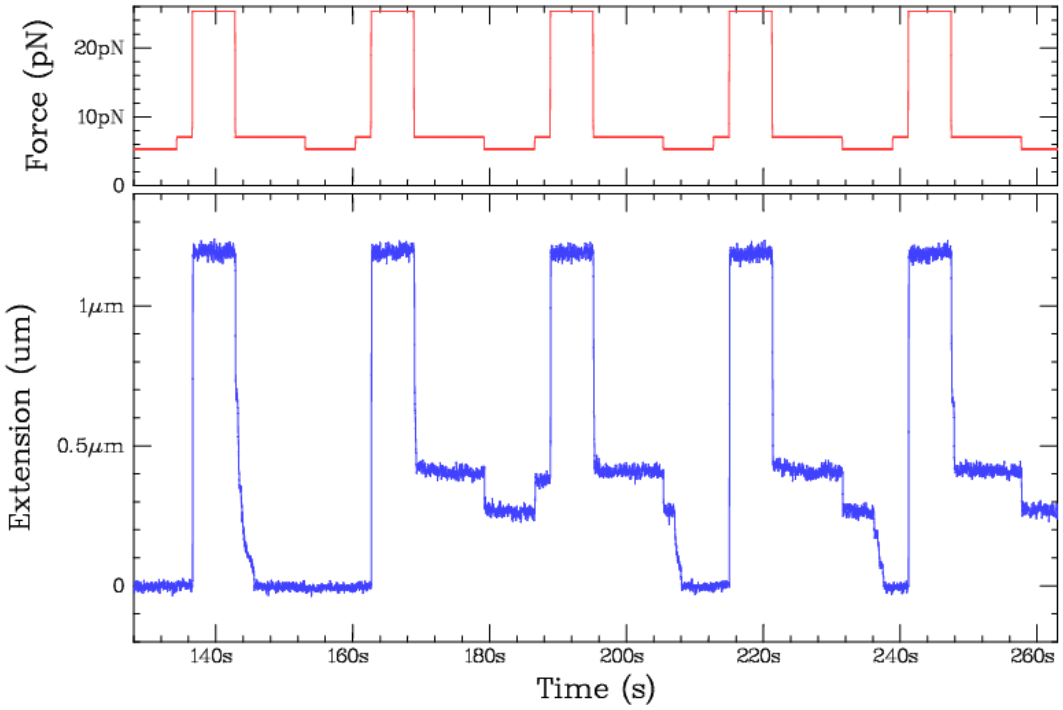


Fig. 15 : Formation de la structure G4 c-MYC(14,23) en présence de la protéine Sub1 sans l'utilisation de l'oligo complémentaire à la boucle du hairpin.

Stabilizing agent	folded G4s
-	0 %
Sub1 protein	30%
PC4 protein	12%
Phen DC3 G4 ligand	12%

Tableau 2 : Proportion des G4s repliés en présence de Sub1, PC4 ou Phen-DC3 sans l'utilisation de l'oligo complémentaire à la boucle du hairpin.

III.6 Interaction entre les hélicases et le G4 c-MYC(14,23)

La résolution d'un G4 par une hélicase est directement relié au Toff : Si le toff est inférieur au temps requis pour qu'une hélicase arrive au G4, l'hélicase n'est donc pas vraiment nécessaire pour résoudre cette structure. Nous avons donc testé l'effet de quelques hélicases sur le G4 du c-MYC(14,23) grâce à sa grande stabilité. Nous avons pu observer en temps réel comment les hélicases se comportent quand elles rencontrent un G4 sur leur chemin tout en ayant la signature sur le G4 replié. Nous avons trouvé que Pif1 résout les structures stables de G4 après avoir fait une pause de quelques dizaines de secondes (fig.16). Alors que RecQ, gp41 (fig.17) qui l'hélicase du bactériophage T4, et la protéine de réplication RPA ne résolvent pas le G4 du c-MYC mais elles peuvent le sauter et ouvrir l'ADN double brin situé en aval.

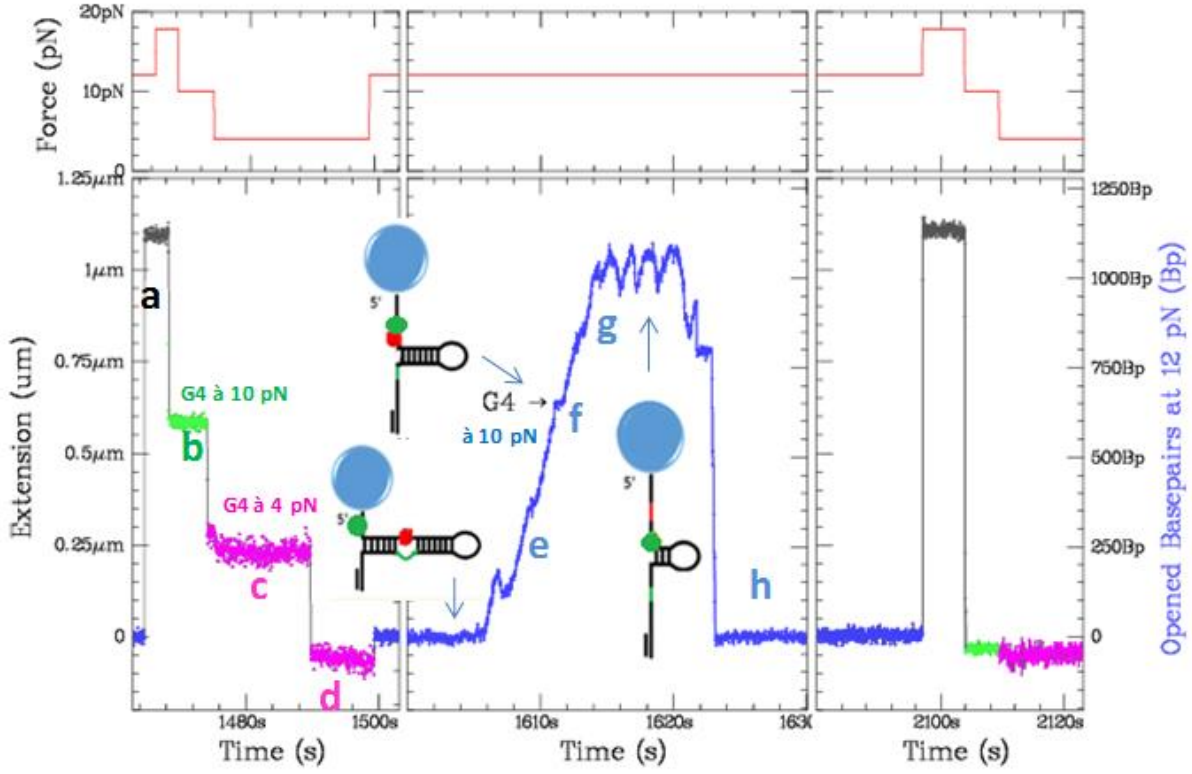


Fig. 16: l'hélicase Pif1 resolvant un G4 c-MYC(14,23) dans un hairpin d'ADN. ***(a,b,c)*** *Dans le premier quadrant nous vérifions que la structure G4 est bien présente* ***(d)*** *puis nous l'encerclant dans le hairpin* ***e)*** *Nous injectant l'hélicase Pif1 à 10 nM dans 60 mM de k+ et nous augmentons la force à 12 pN, la Pif1ouvre le hairpin jusqu'à* ***f)*** *atteindre le G4, la Pif1 fait une petite pause et* ***g)*** *puis continue à ouvrir le reste du hairpin, jusqu'à atteindre la boucle.h) Puis le hairpin se referme complètement sans se bloquer sur la structure G4. Dans le troisième quadrant nous faisons un cycle de force pour confirmer la résolution du G4.*

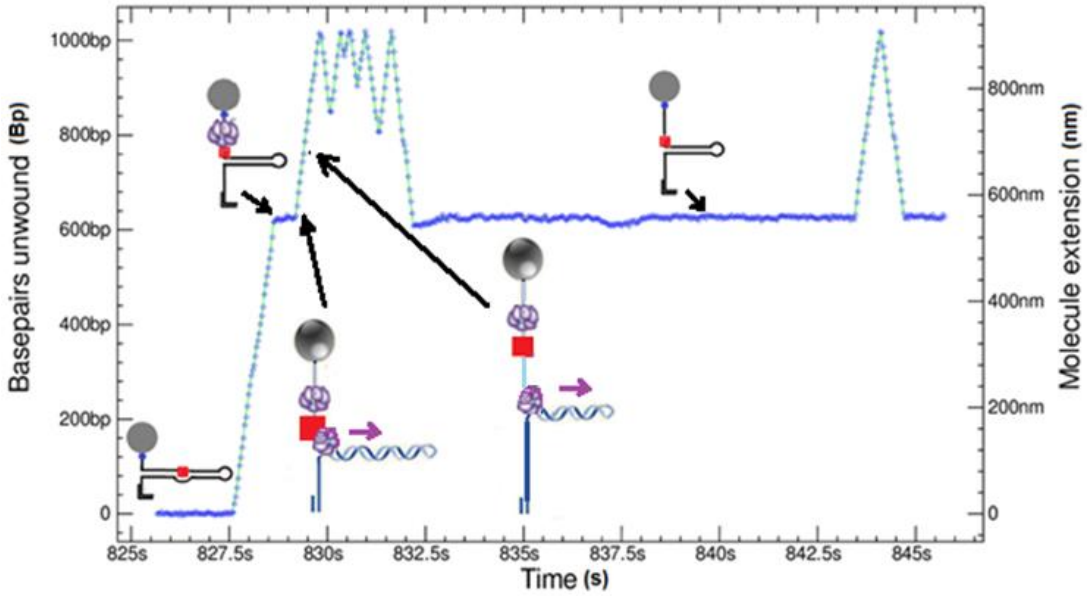


Fig. 17 : L'hélicase gp41 ouvrant un hairpin ayant une structure de G4 c-MYC(14,23). *L'hélicase gp41 ouvre le hairpin jusqu'à atteindre la structure G4 sur laquelle elle marque une pause. Ensuite l'hélicase ouvre les bases suivantes jusqu'à ce qu'elle atteint la boucle du hairpin. L'hélicase passe sur l'autre brin et le hairpin se referme derrière mais bloque de nouveau sur le G4 ce qui confirme que la gp41 n'avait pas résolu cette structure. Il est très probable que lorsque l'hélicase s'arrête sur le G4, une autre hélicase se charge après le G4 en profitant des fluctuations thermiques et continue à ouvrr le hairpin sans avoir résolu le G4.*

helicase	Resolved G4s
Pif1	90%
RecQ	2%
BLM	-x resolved-
Rep	0%
Gp41	0%

Tableau 3 : La proportion des structures G4s qui ont été résolues en présence des différentes hélicases.

III.7 Interaction entre les polymérases et le G4 c-MYC(14,23)

Nous avons testé l'effet des différentes polymérases sur le G4 c-MYC(14,23) en utilisant les hairpins ayant un G4 sur le brin précoce.

Polymerase	Resolved G4s
T4 bacteriophage Polymerase	88%
T7 bacteriophage Polymerase	34%
Polymerase ε (Yeast)	67%
Manta	0%

Tableau 4 : La proportion des structures G4s qui ont été résolues en présence des différentes polymérases.

III.8 Interaction entre le réplisome du bactériophage T4 et le G4 c-MYC(14,23)

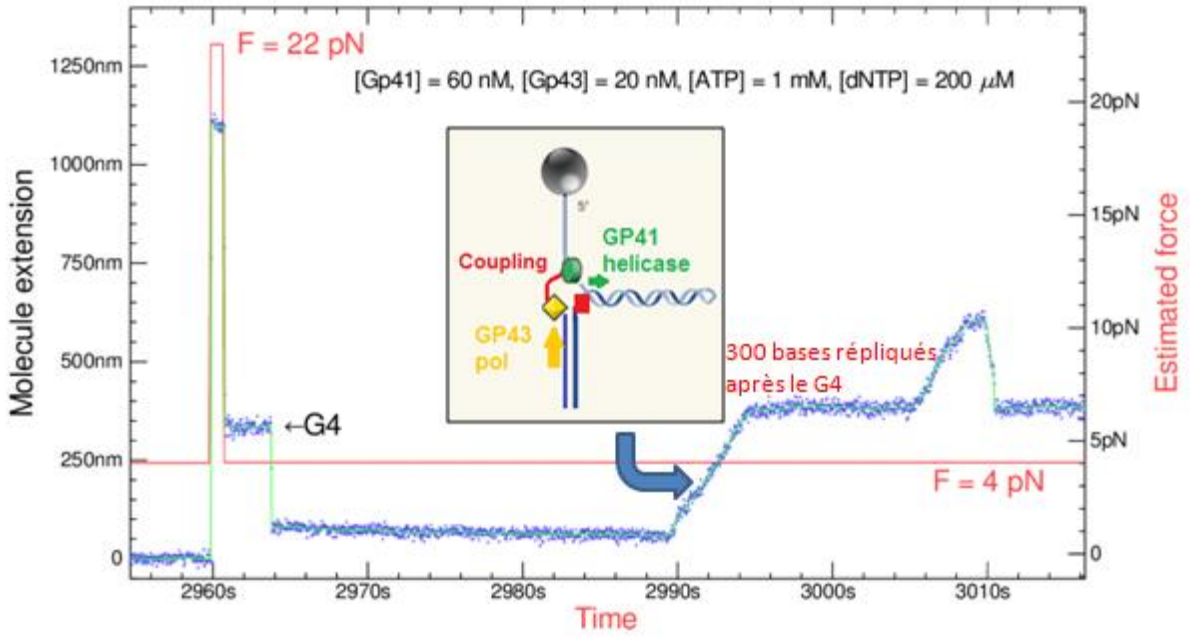


Fig. 18: Le réplisome T4 résolvant un G4. *Un oligo de 50 bases est ajouté au hairpin d'AND ayant ayant un G4 c-MYC(14,23) sur son brin précoce. Lorsque le hairpin est ouvert, l'oligo s'hybride en bas du hairpin. Lorsque la force est baissée à 4 pN, la refermeture du hairpin est arrêtée par la structure G4 pour un certain temps puis le G4 est encerclé et le hairpin bloque sur l'oligo hybridé. En ajoutant l'hélicase gp41 et la polymérase gp43 dans la cellule, les enzymes couplés ouvrent le hairpin en répliquant les bases sur le brin précoce en passant par le G4.*

En ajoutant le réplisome du T4 formé par l'hélicase gp41 et la polymérase gp43 aux hairpins ayant un G4 sur le brin précoce à une force de 4 pN. A cette force l'hélicase et la polymérase couplées ont pu résoudre le G4 sur quelques molécules. De plus, la polymérase

couplée à l'hélicase ouvre le hairpin d'une façon plus processive que lorsqu'elle est seule (fig.18).

III.9 RPA facilite l'ouverture d'une structure de G4 c-MYC(14,23) encerclé dans un hairpin

Nous avons tout d'abord testé si l'hélicase Pif1 peut résoudre un G4 encerclé dans un hairpin. Pour ceci nous avons injecté Pif1 en maintenant une très basse force. Ensuite nous avons rincé Pif1 et vérifié la présence des G4s. Ces derniers n'ont pas été résolus. Le même test a été fait en présence de la RPA. Et similairement, les G4s n'ont pas été résolus. En ajoutant la Pif1 et la RPA ensemble, la grande majorité des G4s ont été résolus. Donc apparemment, la RPA se charge sur la séquence riche en cytosine qui, lorsque le G4 est replié, reste en simple brin, et elle favorise le chargement de la Pif1 sur le simple brin..

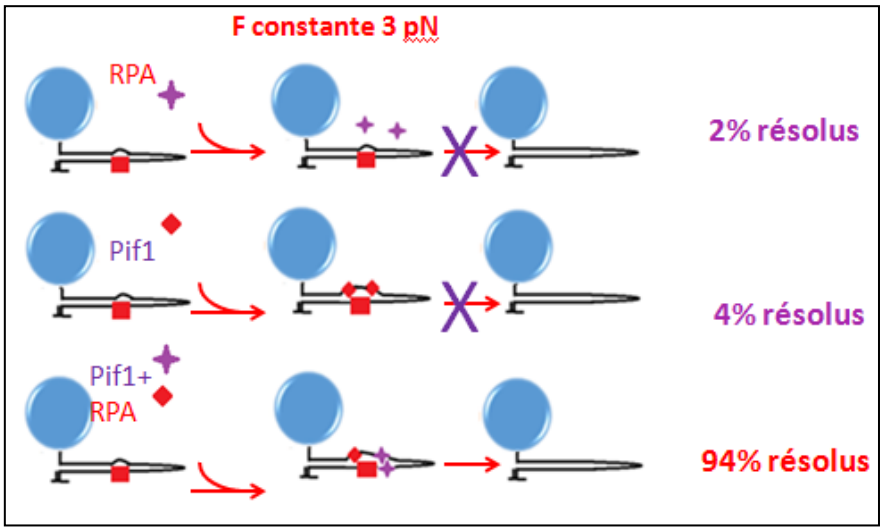


Fig. 19 : Pif1 résolvant un G4 dans un hairpin en présence de la RPA.

IV- Conclusion

Les structures G4s sont considérés comme des obstacles qui empêchent la progression du réplisome et entraine des instabilités génétiques. Dans ce travail de thèse, nous avons développé une nouvelle méthode ulilisant les pinces magnétiques pour étudier les G4s dans des régions d'ADN double brin ce qui imite la présence d'un G4 dans les promoteurs des gènes. En utilisant des hairpins d'ADN contenant une séquence de G4, nous avons pu étudier la cinétique de formation et de déformation des structures G4 en temps réel ainsi que leur stabilité. Nous avons trouvé

que le G4 c-MYC peut rester replié quand le hairpin l'encercle pour des heures. Nous avons aussi étudié l'interaction entre les enzymes et une structure G4 en temps réel. Nous avons trouvé que le G4 c-MYC ne constitue pas un vrai obstacle sur le chemin de certaines hélicases comme Pif1, mais il n'est pas résolu par d'autres hélicases comme la gp41. Cependant, plusieurs polymérases ont été trouvées efficaces à résoudre le G4, ce qui signifie que les G4s ne constituent pas toujours un « roadblock » pour les polymérases et le réplisome. Et en conséquence, seuls les G4s qui ne sont pas résolus par le réplisome vont provoquer des problèmes in vivo.

Les perspectives de ce travail seraient d'utiliser cette méthode pour étudier d'autres séquences de G4 et l'effet d'autres enzymes sur ces structures. Il sera intéressant d'étudier aussi des structures de G4s télomériques consécutives, puisque les premiers tests que nous avons faits montrent que la présence de deux structures G4s télomériques successives dans le hairpin empêchement l'ouverture du hairpin comme si elles forment un nœud impliquant les deux brins de ce dernier. Il sera aussi intéressant de tester s'il y a une hélicase spécifique pour chaque structure G4. L'étude de l'interaction entre l'ARN polymérase et les structures G4s s'avère aussi très intéressante pour comprendre mieux ce qui se passe au cours de la transcription de l'ADN.

Annexes

Annexe 1

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 1

THE PROTEIN SOCIETY

REVIEWS

A mechanistic study of helicases with magnetic traps

Samar Hodeib,[1,2] Saurabh Raj,[1,2] Maria Manosas,[3,4] Weiting Zhang,[1,2] Debjani Bagchi,[1,2] Bertrand Ducos,[1,2] Francesca Fiorini,[5] Joanne Kanaan,[2] Herve Le Hir,[2] Jean-François Allemand,[1,2] David Bensimon,[1,2,6] and Vincent Croquette[1,2]*

[1]Laboratoire de physique statistique, Departement de physique de l'ENS, Ecole normale superieure, PSL Research University, Universite Paris Diderot, Sorbonne Paris Cite, Sorbonne Universites, UPMC Univ. Paris 06, CNRS, 75005 Paris, France

[2]Institut de Biologie de l'Ecole Normale Superieure (IBENS), Departement de biologie, Ecole normale superieure, CNRS, INSERM, PSL Research University, 75005 Paris, France

[3]Departament de Fisica Fonamental, Facultat de Fisica, Universitat de Barcelona, Barcelona, 08028, Spain

[4]CIBER-BBN de Bioingenieria, Biomateriales y Nanomedicina, Instituto de Sanidad Carlos III, Madrid, Spain

[5]Univ Lyon, Molecular Microbiology and Structural Biochemistry, MMSB-IBCP UMR5086 CNRS/Lyon1, Lyon Cedex 7, 69367, France

[6]Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, California, 90095

Received 8 February 2017; Accepted 2 May 2017
DOI: 10.1002/pro.3187
Published online 00 Month 2017 proteinscience.org

Abstract: Helicases are a broad family of enzymes that separate nucleic acid double strand structures (DNA/DNA, DNA/RNA, or RNA/RNA) and thus are essential to DNA replication and the maintenance of nucleic acid integrity. We review the picture that has emerged from single molecule studies of the mechanisms of DNA and RNA helicases and their interactions with other proteins. Many features have been uncovered by these studies that were obscured by bulk studies, such as DNA strands switching, mechanical (rather than biochemical) coupling between helicases and polymerases, helicase-induced re-hybridization and stalled fork rescue.

Keywords: magnetic traps; helicases; polymerase; primase; fork regression; DNA unwinding; Holliday junction migration; ds-DNA fork; active unwinding; helicase/polymerase coupling; replisome; primosome; strand annealing

Debjani Bagchi current address is Physics Department, Faculty of Science, The M.S. University of Baroda, Vadodara, Gujarat 390002, India

Abbreviations: ATP, adenosine triphosphate; bp, base pair; DNA, deoxyribonucleic acid; dsDNA, double-stranded DNA; HJ, Holliday junction; k_B, Boltzmann constant; k_{cat}, turnover number under optimum conditions (saturated enzyme); k_M, Michaelis constant; LNA, locked nucleic acid; NA, nucleic acid; NTP, nucleoside triphosphate; mm, micrometer; nm, nanometer; nts, nucleotides; P_m, maximum processivity under saturated condition; pN, pico-Newtons; RNA, ribonucleic acid; SF, superfamily; SM, single molecule; ssDNA, single stranded DNA; ss-dsNA, nucleic acid forming a junction between single-stranded and double stranded molecule; SSB, single strand binding; v_m, maximum speed under saturated condition.

Grant sponsor: ERC grant Magreps 267 862 (to V.C.), by the Fondation Pierre-Gilles de Gennes, program FPGG032 (to V.C. and H.L.H.), by ANR-14-CE10-0014 CLEANNMD, by ANR HeliDEAD; Grant number: ANR-13-BSV8-0009-02; Grant sponsor: Centre National de la Recherche Scientifique.

*Correspondence to: Vincent Croquette, Laboratoire de physique statistique, Departement de physique de l'ENS, Ecole normale superieure, PSL Research University, Universite Paris Diderot, Sorbonne Paris Cite, Sorbonne Universites, UPMC Univ. Paris 06, CNRS, 75005 Paris, France. E-mail: Vincent.Croquette@lps.ens.fr

ID: jwweb3b2server Time: 06:17 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 2

Introduction

Helicases are enzymes whose role is to unwind the double helix of nucleic acid to provide other enzymes with an ssDNA or ssRNA substrate. DNA helicases are involved in DNA replication, recombination and repair,[1–5] whereas RNA helicases are involved in transcription termination, translation initiation and RNA splicing among other roles.[6] Since helicases translocate along their substrate using the energy of NTP hydrolysis, they are molecular motors: they use chemical energy to perform mechanical work. Their activity is so crucial in many key elements of living systems, that they are ubiquitous in the tree of life[3] and their genetic sequences are hypothesized to represent about 1% of each individual genome.[7]

Complete descriptions of DNA and RNA helicase families can be found respectively in the books of M. Spies[8] and E. Jankowsky[9] and the citations therein. Helicases can be classified according to diverse parameters. One classification is based on the polarity of translocation along the DNA strand, which has classically been deduced from the overhang required to load the enzyme at the ss-dsNA (nucleic acid) junction where unwinding is initiated. Helicases that translocate along NAs from the 3^0 to the 5^0 end are designated as 3^0-5^0 helicases and often need a 3^0-overhang to load onto the junction. Similarly, enzymes that translocate and unwind in the 5^0 to 3^0 direction are designated as 5^0-3^0 helicases and often require a 5^0-overhang loading site.

Several groups also suggested wider and more unifying classifications of DNA and RNA helicases encompassing unwinding polarity with sequence and structural features.[10–13] The most recent classification thus sorts helicases into superfamilies 1 to 6 (SF1 to 6), with toroidal ring forming enzymes sorted into SF3 to 6, while non-ring forming are found in SF1 and SF2.

SF1 and SF2 helicases share a conserved helicase core formed of two similar RecA-like domains resembling the fold of RecA recombination protein. These folds notably comprise at least 12 signature helicases motifs shared by both SF1 and SF2 but not present in each family and group. Such characteristic motifs are either expected or experimentally proven to share similar functions despite sequence variability among different helicase families. Highest conservation between SF1 and SF2 is observed in the residues coordinating binding and hydrolysis of triphosphates (motifs I, II and VI). The Q-motif, which coordinates the ATP adenine base is less conserved between both SFs, and not present in all families. Other motifs coordinating NTP and nucleic acid binding are highly conserved within each superfamily but not across both, indicating mechanistic variations in the communication between the two functional sites. Several motifs are also known to contact nucleic acids on the face opposite the ATP binding site, and are well conserved between both SFs.

SF1 enzymes, with members such as UvrD, Rep or Upf1, exhibit a large contact area with ssNA. For this reason, they are classified as ssNA translocases that melt downstream dsNA as they proceed along one of the strands, see below.

Conversely, SF2 enzymes, with members such as RecG, RecQ, NS3, or DEAD-box helicases, interact nonspecifically with the NA phosphodiester backbone and are assumed to be dsNA translocases, see below.

SF3 comprises helicases from small DNA and RNA viruses, while SF4 and SF5 contain 5^0-3^0 hexameric helicases related either to the E. coli replicative helicase DnaB (SF4) or the bacterial termination factor, Rho (SF5).

Beyond such classifications, a full understanding of the activity of these motors in a living cell requires an understanding of their assembly, structure and conformation(s) of their nucleic acid target, their sequence specificity if any, and their binding modality. Equally important is the mechanism used during their translocation, their directionality, the number of bases they can unwind before dissociating, their velocity, their step size and stoichiometry of NTP hydrolysis (how many base-pairs are unwound per NTP consumed) and, of course, their coupling with other proteins. Many of these elements are related to the mechanical aspects of their activity and thus single molecule micromanipulations are very useful tools to investigate these issues together with classical bulk assays.

From a mechanistic point of view the most important issue concerning the function of helicases is the coupling between ATP hydrolysis and DNA unwinding.

The simplest model assumes that the helicase is "passive": ATP hydrolysis is required only for translocation of the helicase which advances upon unbinding of the DNA junction [see Fig. 1(A)]. DNA fraying maybe thermally induced, or the junction may be destabilized by its interactions with the helicase or other enzymes (e.g., polymerases active with the helicase at the replication fork). F1

In other models the helicase is "active": it uses part of the energy from ATP hydrolysis to melt the dsDNA and translocate [see Fig. 1(B,C)]. One such model assumes that ATP hydrolysis switches the affinity of the binding sites of a dimeric enzyme from a single strand to double strand DNA. In this "rolling model" of DNA unwinding [see Fig. 1(B)], in the absence of ATP, both subunits have a high affinity for ssDNA. Binding of ATP to the subunit upstream of the junction causes it to detach from ssDNA and rebind to the adjacent (downstream) dsDNA, which melt apart when ATP is hydrolyzed.

ID: jwweb3b2server Time: 06:17 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 3

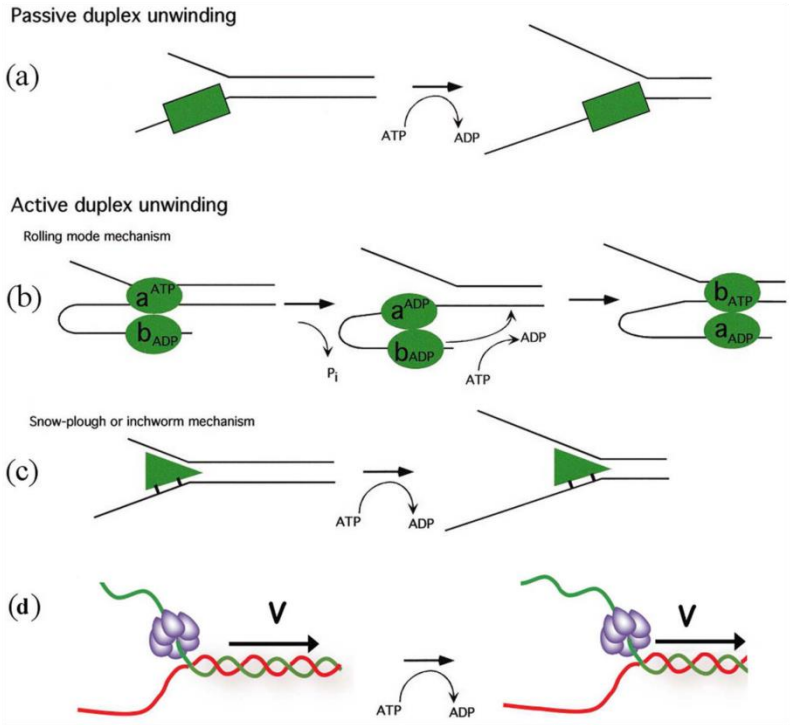


Figure 1. Possible translocation mechanisms of helicases. (A) The helicase can be passive, trapping fork fluctuations through an ATP-driven progression. Alternatively, helicases can actively unwind DNA (B) as they progress along by a "rolling model" where each monomer takes turn at unwinding the molecule or (C) by driving a "plow" through the DNA fork or (D) pulling the DNA through the helicase as in some hexameric helicases. So far all these mechanisms are found in helicase with the exception of the rolling model which had not been demonstrated experimentally for any helicase.

The release of products revert the system to its initial state with the subunits having switched order. In this model, the enzymatic step size is at least as large as the number of DNA bases contacted by each monomer. However it seems that this mechanism is not used by any of the helicases studied so far.[14–18]

A second active model, the "inchworm model," assumes that ATP hydrolysis is used to translocate the enzyme along single or double stranded DNA, with the opening of the junction a consequence of that translocation. For example, if the helicase translocates along ssDNA (SF1 enzymes), it will unwind the junction like a snowplow [see Fig. 1(B)]. If the enzyme translocates on dsDNA (SF2 enzymes), it may use a wedge domain positioned at the junction to open the DNA like a wire-stripper. This model is supported by observations on various helicases: RecBCD.[18–21] Another possibility to form an active helicase is to force a kink in the NA duplex that act as a wedge destabilizing locally increasing the unpeeling fluctuations.[22]

In spite of these precise descriptions of the enzymatic mechanism, whether an enzyme is active or passive, has rested on problematic estimations of the free energy of the interaction of the enzyme with the DNA fork. To cut this Gordian knot, we have defined a passive enzyme as one whose unwinding rate is very sensitive to the DNA sequence and increases significantly when a force is acting to unzip the two DNA strands (see below). By contrast the unwinding rate of an active enzyme is insensitive to the DNA sequence and unchanged (or even possibly decreased) when a force is acting to separate the two strands.

When unwound by the helicase the two single-stranded nucleic acids are in close proximity and may thus re-anneal in the wake of the enzyme. In vivo to prevent this re-annealing cofactors act on the ssDNA generated by the helicase activity:[1,2,23] proteins that bind ssDNA to stabilize it (e.g., Single Strand Binding (SSB) proteins) or enzymes that process the ssDNA (e.g., Single Strand Nucleases or the

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 4

cell replication machinery). In fact most helicases function poorly (inefficiently and with reduced processivity) when isolated from the macromolecular machinery and coupling factors with which they are intended to operate.[23] The inefficiency of isolated helicases is not entirely surprising, as unwinding dsDNA could be potentially dangerous for the cell if it was not tightly regulated. The cell further regulates the activity of its helicases in two ways: by using specific enzymes to load the helicase on its template (this is the case for replicative helicases for example) or by using a cofactor to switch the helicase between different modes. One of the challenges of studying helicases as molecular machines is thus to understand not only the mechanisms of these helicases but also their coupling with other proteins and their alteration in different in vitro and in vivo contexts.

The re-annealing of the unwound strands behind the helicase makes it challenging to understand their mechanism in vitro. Indeed bulk in vitro assays are limited by the fact that once dsDNA has been opened by the helicase under study the re-hybridization of the two strands proceeds spontaneously and "erases" the action of the helicase unless very specific actions are taken to prevent this. The re-hybridization problem is a major issue which has been solved only partially. A classical approach is to use Single Strand Binding (SSB) proteins that are supposed to stabilize ssDNA but as we shall see single molecule approaches show this strategy to be problematic: the characterization of a helicase depends on the experimental conditions.[23] Due to this re-hybridization issue, most bulk assays of helicases report a lower processivity than expected from their role in vivo. The only exceptions are helicases that work in conjunction with a polymerase that synthesizes a strand complementary to the unwound template[2] or destroy one of the strands via the action of a coupled endonuclease as in RecBCD[24] which we discuss first. In these two cases, the rehybridization issue has been abolished leading to strong processivity.

Mechanism of RecBCD

RecBCD consists of two helicases RecB and RecD. The RecB subunit is a helicase with 3^0-5^0 directionality. Its first two domains define it as a member of the SF1 family. Its third domain is a nuclease domain, connected to the remainder of the protein by a long linker region of about 70 amino acids.[15,25]

The RecD subunit is also a SF1 helicase but with a rather uncommon 5^0-3^0 unwinding activity. The RecC protein contacts both strands of DNA and splits them (like the shovel of a snowplow) before feeding the 3^0 strand to the RecB protein and the 5^0 tail to the RecD subunit. The two helicase motors apparently work independently of each other and because of their opposite polarity they both pull-in the anti-parallel DNA strands. A detailed picture of the mechanism of coupling between these units has emerged from single-molecule experiments and structural data.

Bulk measurements of the activity of RecBCD (estimated by the amount of hydrolyzed DNA) yield a very large unwinding rate[26] (v_m 930 bps/s), a huge processivity[27] (about P_m5 27000 6 3000 bps) and a rate of ATP hydrolysis[28] (with k_{cat} 740 $s^{2\,1}$) all characterized by the same Michaelis-Menten coefficient[26,28–30] (k_M 100 mM). These results imply that the unwinding rate and processivity are controlled by ATP hydrolysis. Assuming a tight coupling between ATP consumption and translocation, one deduces that about 1 ATP is required to unwind a single (v_m/k_{cat}) base-pair. Furthermore, single-turnover DNA unwinding experiments[31] yield an estimate of the helicase step size of about 3.96 1.3 bps. However in these bulk experiments an estimated 30% of the enzymatic population was inactive (either because of misfolding or for not being on their substrate). Therefore, it was not clear how the previously quoted numbers were affected by the presence of this population of inactive enzymes.

To address that issue Bianco et al.[32] have directly visualized the motion of RecBCD on single DNA molecules (see Fig. 2). In their experiment, a DNA molecule was tethered at one end to a small polystyrene bead held in an optical tweezers. The DNA molecule was stained with a fluorescent dye (YOYO1), stretched by the flow of the enzyme containing buffer and visualized on a sensitive camera. F2

As it is unwound and degraded by the RecB nuclease domain, its extension shortens (see Fig. 2). By monitoring the change in extension as a function of time, one can directly measure the rate and processivity of RecBCD. A surprising result of these single-molecule investigations was the extreme variability in the unwinding rates measured for various enzymes on the same substrate. Although the rate of DNA unwinding by any individual RecBCD was uniform (within experimental error) on any given DNA molecule, the rate for different enzyme molecules deviated by up to fivefold. Subtle changes in the conformation of the complex as it is loaded on DNA may result in this large difference in unwinding rates. In spite of these large enzyme to enzyme fluctuations, the average rate (v_m 970 bps/s at 378C), average processivity (P_m5 38,0006 5700 bps) and Michaelis-Menten coefficient (k_M 150 mM) were all compatible with bulk measurements. This is a comforting result since for this enzyme, no coupling factors are required to observe its unwinding activity, i.e., to prevent reannealing of the two DNA strands. Hence direct observations of DNA unwinding and degradation by a single enzyme and bulk measurements of DNA unwinding are expected to yield similar results (as we shall see that need not always be the case).

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 5

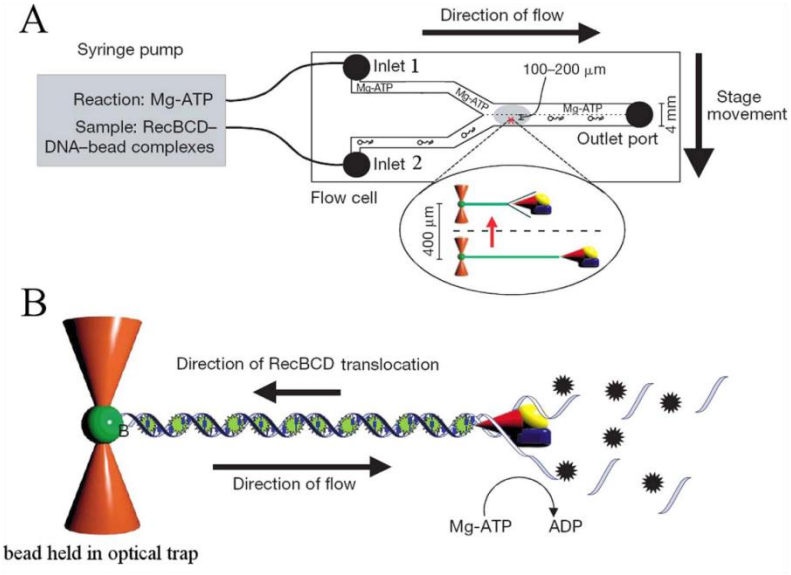


Figure 2. (A) The flow cell introduced by the Kowalczykowski group to study the interaction between DNA and various motors. It consists of a microfluidic cell with two (or more) inlets. Through one inlet a solution of ATP, dye, Mg^{11}, etc. can be introduced and through the other, DNA bound to transparent beads and possibly proteins (such as RecBCD). In this laminar flow regime, the two solutions do not mix. One bead is captured by optical tweezers. The cell is then translated and the trapped DNA molecule brought in contact with ATP (to launch the unwinding reaction) and/or a dye (to stain the molecule). (B) as the molecule is unwound by the progression of RecBCD, it is degraded by the RecB exonuclease activity. Its length can be deduced by the fluorescence of the stained and stretched remaining dsDNA. Hence the progression of the helicase complex on the molecule can be monitored in real time. (Figure taken from Figure 1 of [32] with permission)

Another result from these studies concerns the activity of RecBCD after it encounters a v sequence (5^0-GCTGGTGG-3^0). Recognition of v was known to reduce both nuclease activity and translocation speed of RecBCD and to activate RecA-loading. What single-molecule experiments have shown is that upon encountering a v sequence, the enzyme paused for a few seconds and then continued to translocate and unwind DNA as a full trimeric complex,[33] but on average at about half the previous rate (for some enzyme molecules the rate after v recognition could be one tenth of the prior rate). While before v, both single strands were degraded, after v the 3^0 strands is not degraded: its end is sequestered by the enzyme and as RecBCD translocates a ssDNA loop is extruded. In the presence of DNA specific dyes in solution this increasing loop is visible as an increasingly bright spot translocating along a fluorescently labeled dsDNA.[34]

A full mechanistic picture of the working of RecBCD emerges when these single-molecule results are combined with the crystallographic data of RecBCD.[15] The location of the nuclease active site (the third domain of RecB) allows processive hydrolysis of the 3^0 tail, as it emerges from the RecC subunit. However, proximity of the 5^0 tail would enable it to compete with the 3^0 tail for binding at the nuclease. Thus, before v, both strands can be degraded by the nuclease activity of RecB. Upon encountering v, the RecC subunit binds tightly to the 3^0 tail, preventing its further digestion. This induces a conformational change in RecD that slows it down. The 5^0 tail is now able to access the nuclease site more frequently and is degraded more fully. The enzyme continues to advance along the DNA extruding a loop out from the RecB subunit that can be loaded with RecA proteins.

Single Molecule Mechanical Assays of Helicases

The RecBCD study is a rather singular example as it is not plagued by the re-hybridization of the separated strands in the wake of the enzyme. As we shall see below, single molecule micro-manipulation approaches offer ways to study a variety of helicases, by avoiding re-hybridization using a stretching force that hinders the reannealing of the

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 6

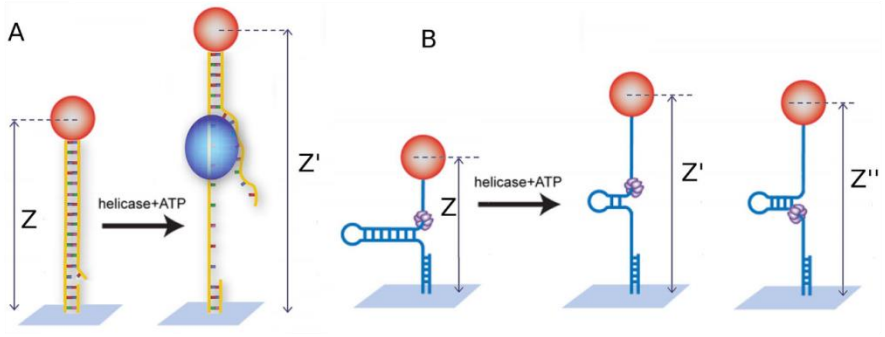


Figure 3. (A) Schematics of the unpeeling configuration used to study helicases: a helicase (blue blob) loads on a nick or gap in the dsDNA molecule under tension. Unwinding of the molecule results in an increase (Dz 5 $z^0 - z$) of the overall extension. The two ssDNA (one under tension and one free) are unable to match in the wake of the enzyme due to a mismatch in their extension. (B) Schematics of the unzipping assay: a helicase (the violet hexamer) loads at the fork of a DNA hairpin under tension. Unwinding of the hairpin results in an increase (Dz 5 $z^0 - z$) of the molecule's extension. The tension on the released ssDNA strands prevents their reannealing in the wake of the enzyme. As the helicase reaches the hairpin apex, its continuing translocation on a ssDNA template allows for reannealing of hairpin in its wake, monitored by the decreasing change in extension (Dz 5 $z^{00} - z$)

strands. This force, however, is not completely neutral and may influence the enzymatic behavior by, for example, favoring unpeeling of DNA. Extrapolation of the results to zero force is a way to address that issue.

The general approach for single molecule mechanical assays of helicases is to stretch a DNA molecule and observe the change in its extension resulting from the transformation of double-stranded DNA (dsDNA) into single-stranded DNA (ssDNA). This can be implemented mainly by two different configurations: the unpeeling configuration where a tension is applied between the two ends of the DNA molecule and the unzipping configuration where the force is applied between the two complementary strands at the same DNA extremity.

Force, the mechanical parameter introduced in these studies of helicases by micromanipulation techniques, has two roles:

1. It prevents re-hybridization of the strands in the wake of the helicase. By separating the two complementary strands, the force hinders their re-hybridization. As this effect increases with the force, the probability of re-hybridization is strongly dependent on the force, displaying an all or none behavior. Above a critical force F_r (25 pN for the unpeeling assay, 3 to 5 pN for the unzipping assay), re-hybridization is virtually impossible while it occurs readily below F_r (as reannealing in the wake of the enzyme becomes more frequent). This sets a lower limit to the force that can be used in these assays.
2. The force determines the molecule's extension and its fluctuations via the bead's Brownian motion, thus it sets the signal to noise ratio of the assay. The larger the difference between the ssDNA and dsDNA extensions the better is the signal (Fig. 3), in the same spirit reducing the bead size also reduces Brownian fluctuations. As the extension of both DNA forms is zero if no force is applied, the signal is null at zero force (and in the unpeeling situation at a force F 5 pN where the extensions of both are equal, see below). Considering the elastic response of ssDNA and dsDNA the sensitivity of the measurements increases with the force, so it is tempting to apply strong forces to improve the signal to noise ratio of the experiment but this comes at the cost of applying a mechanical tension that might be out of the physiological range. F3

Indeed a large applied tension may destabilize the dsDNA. The mechanical unfolding of a hairpin construct in the absence of a helicase can be characterized, as seen in Figure 4: above a critical force of F_u 5 156 1 pN the hairpin spontaneously unfolds, while it remains otherwise stably folded at forces $F_c < 126$ 1 pN. Thus at forces $F < F_c$, any unfolding observed in the presence of a helicase is a result of its activity. Indeed in its absence, the extension of the DNA molecule remains constant and equals to the folded hairpin. In the unpeeling configuration the destabilizing force is F_u 60 pN. However, in both configuration, there exists a substantial range of forces ($F_r < F < F_u$) where the opening of the F4

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 7

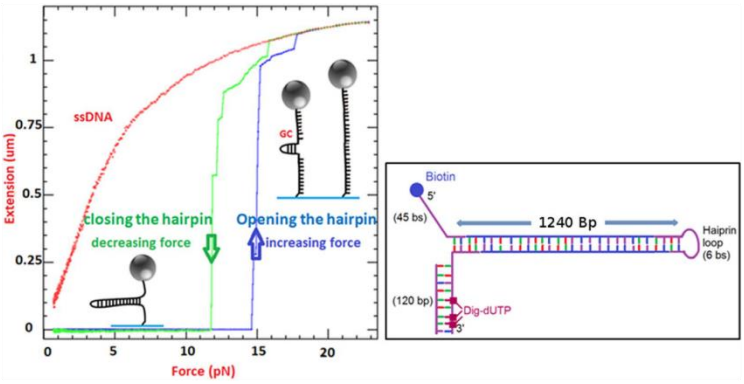


Figure 4. Extension of a DNA hairpin molecule versus the applied force in the unzipping configuration. The molecular design is given in the box on the right: a 1.2 kb hairpin is made of a dsDNA stem closed at one end by a loop and with a fork at the second end. The arm ending in 5^0 has a biotin while the other arm holds multiple digoxigenin. This molecule is attached to a 1-micron size magnetic bead coated with streptavidin and to its second extremity to the flow cell via a digoxigenin/anti-dig bound. At low force this molecule remains closed and its extension is null. As the magnets are brought closer to the bead, the force can exceed 15 pN leading to the molecule unfolding. The DNA sequence of this molecule presents a GC rich zone close to the apex: this region is more difficult to open and remains folded until the force reaches 17 pN. At that force the molecule is fully open and its extension is nearly 1.2 l m. Further increase of the force leads to a small stretching of the ssDNA molecule. Upon decreasing the force, the molecule refolds with a hysteresis of typically 3 pN. The refolding process nucleates at the molecule apex. If one introduces a 18 nts oligonucleotide that hybridizes to the apex, the refolding of the hairpin is hindered and one observes the red curve corresponding to ssDNA elasticity. When the force is decreased to very low values, the oligonucleotide can be expelled and the hairpin refolds

dsDNA can only be caused by the helicase and where the tension on the molecule prevents re-hybridization of the strands in the wake of the enzyme.

In the unpeeling configuration, the molecule is a stretched dsDNA [Fig. 3(A)]. A nick or a gap in one of the strands of the DNA serves as a loading site for the helicase which proceeds by unpeeling one strand from the other. To prevent the reannealing in the wake of the helicase a large enough tension on the molecule has to be applied ($F > 25$ pN 5 25 3 $10^{2\,12}$ N). In these conditions the mismatch in the inter-nucleotide distance in the peeled (free) strand and in the strand under tension is large enough: the product of the force (25 pN) by the distance (about 1 nm) between the just separated complementary strands blocked by the helicase leads to an estimate of the energy cost that thermal fluctuations should overcome to re-hybridize the strands in the wake of the helicase. At $F > 25$ pN this energy exceeds several times the typical scale of thermal energies, k_BT (4.1 3 $10^{2\,21}$ J, at room temperatures) resulting in a rare encircling of the helicase. As the extension of ssDNA is longer than that of dsDNA (for force > 5 pN) the result of helicase activity is an increase in the distance between the bead and the surface (the extension of the molecule) proportional to the amount of peeled ssDNA. Typically, the change of extension is a fraction of the extent of a base-pair (0.34 nm) depending on the force. Thus, in this configuration one has to apply quite a large force to get a good signal albeit with a poor sensitivity. The main advantage of this assay is its simplicity.

In the unzipping assay a tension is applied between the two ends of a hairpin structure. This configuration mimics a DNA replication fork structure [Fig. 3(B)]. The tension applied on the molecule acts on the two arms of the fork. The molecule's opening by the helicase increases the length of these arms while the tension prevents their re-hybridization in the wake of the enzyme. When the enzyme reaches the apex of the hairpin the molecule is completely unzipped. Since nothing prevents helicase translocation on the extended ssDNA strand, further translocation of the enzyme along that ssDNA allows the hairpin to refold in its wake, regenerating a fork that may push on the enzyme. In this configuration, the separation between the two arms of the fork is about twice the extension of ssDNA: 0.93 nm/bp at 10 pN. Consequently, the signal (i.e., the change in the bead/surface distance) is very sensitive to DNA unwinding. With a typical precision in the measurement of extensions of a few nanometers, this unzipping assay approaches single base resolution. In addition the critical force to prevent re-hybridization, F_r is low, typically 5 pN which

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 8

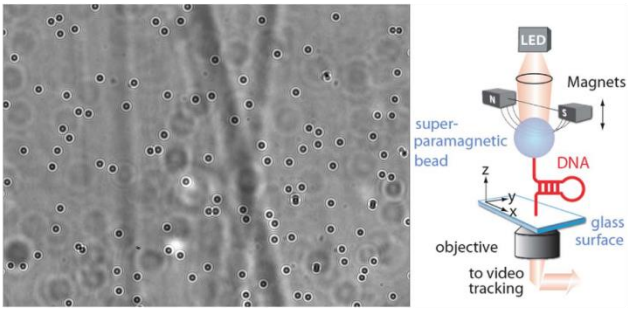


Figure 5. Right, schematic representation of a magnetic trap. Small super-paramagnetic beads are bound to the surface of a flow cell by one (or a few) DNA molecules. Magnets positioned above the sample exert a force on the beads and thus on the tethering molecules. DNA hairpins attached at their free ends by one strand to a magnetic bead and at the opposite one to a surface can be unzipped at high enough force (typically F > 15pN). About 50 beads are simultaneously tracked on an inverted microscope. A typical image is shown on the left. Analysis of the successive images of the beads on a camera allows deducing their 3D position and from it the distance of the bead to the surface (i.e., the extension of the tethering molecule) and the exerted force (from the bead's fluctuations, see text)

allows an investigation of these enzymes in a more reasonable range of forces.

The magnetic trap

To stretch the DNA molecule in the unpeeling or unzipping configurations, the magnetic trap is a very convenient method for high throughput single molecule assays.[35] Briefly, it consists of stretching a single DNA molecule bound at one end to a surface and at the other to a magnetic micro-bead (1 to 4.5 mm in diameter, a bigger size allowing to reach higher forces but with more noise) (see Fig. 5). F5

Small rare earth magnets, whose position is controlled, are used to pull on the micro-bead and thus stretch the molecule. This system allows applying and measuring forces ranging from a few femtoNewtons (f_N 5 $10^{2\ 15}$N) to more than 25 pN (25 pN for 1 mm beads, 200 pN for 2.8 mm beads and 1nN for 4.5 mm beads),[35] with a relative absolute accuracy of 10%. In contrast with other techniques, the force measurement is absolute and does not require a calibration of a force sensor (once the microscope magnification has been calibrated). It is based on the analysis of the Brownian fluctuations of a tethered bead, whose 3D position can be measured by an analysis of its image on a CCD camera at frequencies up to few hundreds Hertz and in some cases to kHz.[36–38] The center of the bead allows for the determination of its x, y-coordinates, while the size of its diffraction rings allows for a determination of its distance along the optical axis to the focal point, i.e., its z-coordinate. The DNA-bead system is completely equivalent to a damped pendulum of length l 5 <z> pulled by a magnetic force F (along the vertical axis). Its longitudinal fluctuations ($dz^2$5 < z^2> − < z >^{2}) and transverse fluctuations dx^2 are characterized by effective rigidities k_{jj}5 o_zF and $k_?$5 F/l. By the equipartition theorem, they satisfy:[39,40]

$$dz^2 5 k_B T = k_{jj} 5 k_B T = @F$$

$$dx^2 5 k_B T = k_? 5 k_B T \ l = F$$

Thus from the bead's Brownian fluctuations (dx^2; dy^2), one can deduce the force pulling on the molecule (the smaller the fluctuations, the larger F) and from dz^2 one deduces its first derivative o_zF. This measurement method is valid with magnetic (but not optical) traps since the variation of the trapping gradients occurs on a scale (few hundreds of mm, linked to the separation of the two pulling magnets) much larger than the scale on which the elasticity of the molecule changes (0.1 mm)). In other words, the stiffness of a magnetic trap is very small compared to F/l (the reverse is true for optical tweezers). A further advantage of the magnetic trap technique is that measurements of DNA at constant force are trivial (simply fix the position of the magnets).[41] In contrast, for optical tweezers to work at a constant force an appropriate feedback is required to ensure that the displacement of the sensor is kept constant. The accuracy of the absolute force measured by Brownian motion is typically 10%. This number depends on the accuracy of the measurement of dx^2: better accuracy is possible but at the expense of longer experiments (more data points). Conversely, for a fixed magnet position the force is constant and its magnitude can be adjusted with extreme accuracy: it is possible to increase the force F by 0.1%, though its precise value is more difficult to determine.

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 9

The force deduced from Brownian motion assumes that the measuring device (the camera) has a bandwidth larger than the characteristic frequency of the Brownian fluctuations which may not be the case when the force is large and/or the DNA tether is small. In such instances, special care must be taken in analyzing these fluctuations. Analyzing them in Fourier's space is usually the best strategy.[42–44]

A great advantage of magnetic traps is their intrinsic parallelism as the magnets apply a constant force on a large surface and as a consequence, on a large number of beads, i.e., DNA molecules. The increasing size and speed of the new CCD and CMOS cameras offer the possibility to visualize a large number of beads, while fast acquisition cards and parallel computing allow one to track their position (i.e., the molecules' extension) simultaneously.[45,46] It is thus possible to track several hundreds to thousands of beads with a few nanometers accuracy. The inherent parallelism of magnetic traps is extremely valuable when monitoring enzymatic activity. Helicases can be delicate molecular motors which do not work 100% of the time in these in vitro conditions. With hundreds of parallel assays (DNA bound beads), it is far easier to characterize them. Magnetic tweezers are now used by many groups who have extended and refine the technic, the reader may find complementary useful information in Refs. 38 and 47.

Although the magnetic tweezers offer many interesting features they also have limitations: some of them are related to single molecule micromanipulation in general, others are specific to magnetic tweezers. Devices achieving single molecule micromanipulation rely on the measure of the molecule extension, thus they require a finite force to lead to good measurements. This force applied to the NA molecule does not correspond to the in vivo situation and may also help (or hinder in rare case) the enzyme activity. Although the force constrain helps understanding molecular motors function, extrapolating results to the zero force regime is always a challenge. A second issue is the requirement for single molecule events: all the helicase activity bursts present in this article correspond to well identify events surrounded by signals at rest. That is events represent at most 10% of the signal. This condition implies that a single enzyme is working, this is fine but it constrains enzyme concentration to be low and again in some case not close to in vivo conditions. A third issue appears when investigating several enzymes collaborating in a biological process (see below), one would like to have a signal characterizing each enzyme but we only have a single signal: the molecule extension. This implies supplemental control experiments to analyze this signal and to find which enzyme this signal is related. An alternative to this issue is to couple micromanipulation and fluorescence. The beauty of the magnetic tweezers is to be a very simple force clamp system applied to many molecules in parallel, but in some occasions one may want to move things around at a specific position, then optical tweezers are better adapted. Finally, good signal resolution is often difficult to achieve: while you typically get a resolution of three bases in the molecule extension measurement from one video frame to the next, one would like to reach the single base resolution. This is not impossible to achieve[36,38,47] but this is not done with simple and easy to use systems, technical improvements might help in this direction.

DNA Unwinding in the Unpeeling Configuration

UvrD, a member of the SF1 superfamily, plays a crucial role in nucleotide excision repair and methyl directed mismatch repair.[3–5] It has been shown to initiate unwinding from a 3^0 end ssDNA tail, a gap or a nick and to translocate along ssDNA in a 3^0-5^0 direction,[48,49] a loading configuration that can be used in the unpeeling assay previously described. Most bulk assays of its activity require the presence of single-stranded stabilizing factors, proteins such as SSB or a high concentration of UvrD (stoichiometric ratio of enzyme/nucleotide[49]) which by binding to ssDNA in the wake of the enzyme prevent reannealing of the separated strands behind the advancing helicase complex. Clearly, in these conditions, it is difficult to assess the intrinsic unwinding rate, processivity or efficiency of an active enzymatic complex.

As already mentioned, in the unpeeling configuration [see Fig. 3(A)], at a high enough tension ($F > F_r$ 5 25 pN) the reannealing of the strands in the wake of the enzyme is greatly slowed down.[50] This allows one to study the activity of a single enzymatic complex at very low concentration (well below the enzyme's dissociation constant Kd 10 nM) and without processivity factors. As explained before [see Fig. 3(A,B)] and displayed in Figure 6, DNA unwinding (U) results in a continuous (ATP dependent) increase with time of the molecule's extension. The slope of this signal (the derivative of the extension with respect to time) yields the unwinding rate v_U, its spatial extent—the number of base-pairs unwound NU and its duration—the on-time s_U of the enzymatic complex. As the tension on the dsDNA is not strong enough to destabilize it, upon dissociation, the DNA fork serves as a nucleation point for the quick re-hybridization (H) of the two unwound strands, resulting in a rapid (ATP independent) shortening of the molecular extension. Bursts of increase in the molecule's extension (upon enzymatic Unwinding) followed upon enzyme dissociation by rapid shortening (re-Hybridization) are thus often observed. We call these activity bursts, UH bursts. Figure 6 displays such a burst observed in the unzipping configuration. The

F6

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 10

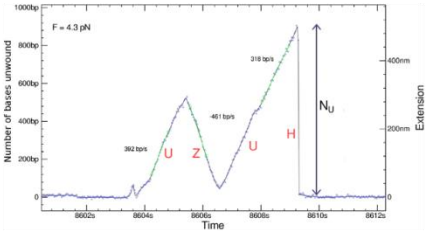


Figure 6. A typical signal observed with UvrD in the unzipping configuration: Unwinding (U) of the dsDNA results in a continuous increase of the extension by N_U bps during a time s_U from which the mean processivity $< N_U >$, mean rate of unwinding v_U 5 $<N_U/s_U>$ can be determined. Unbinding of the enzyme results in re-hybridization (H) of the two strands. Strand switching by UvrD results in an ATP dependent reZipping (Z) of the two strands in the wake of the helicase

large temporal separation between activity bursts (relative to their duration) makes it very unlikely that they are due to the simultaneous action of many enzymatic complexes.

These single-molecule experiments yield the probability distribution of unwinding rates, which is Gaussian with a mean v_U 248 bps/s and standard deviation r_U 100 bps/s at 258C. By varying the ATP concentration, the various parameters characterizing enzymatic activity (rate, processivity, on time, step-size, etc.) may vary. While the measurement of these parameters in single-molecule experiments can be time consuming, the parallelism of magnetic traps greatly facilitates that task.

For UvrD helicases the average rate follows a Michaelis-Menten (MM) kinetics:

$$<v_U> 5 v_u^{max} [ATP] = (K_M 1 [ATP])$$

With v_u^{max} 5 275 bp/s and K_M5 53 l M. The value of v_u^{max} is about three times larger than quenched flow estimates,[51] a significant but not unusual difference when comparing single-molecule and bulk assays. Considering the caveats accompanying bulk measurements (for example the proportion of inactive enzymes or reannealing of the ssDNAs in the back of the complex) and the fact that the SM data were obtained at a rather large stretching force (F 5 35pN) this discrepancy is reasonable. More to the point, the value of KM (which should be less sensitive to systematic experimental errors) is consistent with the value deduced from bulk ATPase assays.

UvrD Strands Switching and Complex Unwinding Dynamics

Surprisingly and unexpectedly, UvrD helicase exhibits another type of unwinding bursts. In many events, unwinding (U) is followed by a slow, ATP dependent re-zipping (Z) of the two separated strands (see Fig. 6). The re-zipping rate v_Z obeys the same Michaelis-Menten kinetics as the unwinding rate v_U, suggesting that these (UZ) events are due to the enzyme switching strands, moving 3^0-5^0 on the complementary ssDNA and limiting thereby the reannealing rate of the two separated strands in its wake. The similarity of the unwinding and re-zipping rates suggests that the rate of UvrD on DNA is only slightly affected by the enzyme having to open the strands or by the fork closing in its wake and points to an active mechanism of unwinding,[52] as for RecQ (see below).

The strand switching mechanism could not be deduced from bulk assays. It was first considered as a possible peculiarity of UvrD, but in fact many of the helicases studied so far in single molecule conditions display similar behavior, the exceptions being hexameric enzymes that encircle a single strand of DNA and thus cannot switch strands.[53] Single molecule analysis has also revealed that helicases may lose their grip on the ssDNA on which they move leading to a slippage or rapid backward motion.[54]

The strand switching behavior of helicases is compatible with an inchworm mechanism: the structure of the helicase involves two RecA domains articulated with a hinge domain. The enzyme binds to ssDNA in two regions whose distance varies upon ATP binding. The inchworm mechanism implies that the binding sites of the two helicase domains on ssDNA alternate. In the normal situation, the enzyme steps on the same strand. However, if one of the domains binds to the opposite strand, the enzyme will switch strands and invert its propagation direction. Though these events are rare on a per cycle basis, they become common when the number of enzymatic cycles grows (i.e., as the enzyme processivity increases). Occasionally, some complex types of events are observed, such as Unwinding, followed by slow re-Zipping(Z), then Unwinding and finally a rehybridization may also occur, see

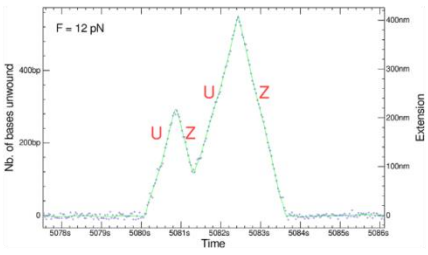


Figure 7. (A) More complex signals observed with UvrD resulting from a combination of the elementary processes described in the text: Unwinding (U), re-zipping (Z). Signal obtained in the unzipping configuration

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 11

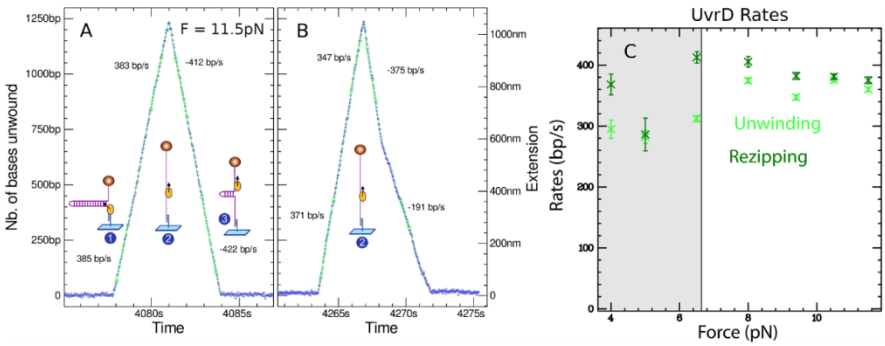


Figure 8. (A) Signal produced by a UvrD helicase unwinding a 1.2 kbps DNA hairpin at T 5 298C and F 5 11.5 pN. Before t 5 4078 s the hairpin is closed. Then (1) a UvrD helicase loads on the hairpin and starts unwinding it in a processive and fast rate (about 380 bp/s). The UvrD helicase reaches the hairpin apex (2) and pursues its translocation as it essentially travels on a ssDNA molecule. Once the helicase has passed the apex, the hairpin starts refolding until the fork bumps on the helicase (3). The extension in this phase reproduces the position of the helicase as it translocates along ssDNA at a rate of about 400 bp/s. Note that the motion is extremely regular (each point corresponds to a video frame acquired at 30 Hz). (B) A trace obtained in conditions similar to (A) but where the helicase changes speed as it re-zips the molecule. The events with lower speed are not very frequent but do occur repeatedly, suggesting that a helicase might display different rates perhaps due to a change in its conformation. (C) Unzipping and re-zipping rates show no significant dependence on the applied force (the gray area corresponds to forces where the statistics on traces is weaker)

F7 Figure 7. Single molecule approaches have thus revealed that helicases are molecular motors working in a very stochastic manner.

DNA Unwinding in the Unzipping Configuration

As said in paragraph 2, in the unzipping assay the helicase activity results in a larger signal than in the unpeeling assay which allows one to study helicases at much lower tensions on the molecule, in a force range that is likely more compatible with in vivo assays. Under these conditions DNA tethers also last much longer, which helps to collect more data.

UvrD has also been studied in the unzipping configuration as shown in Figures 6–8. The unwinding signal to noise is larger than in the unpeeling F8

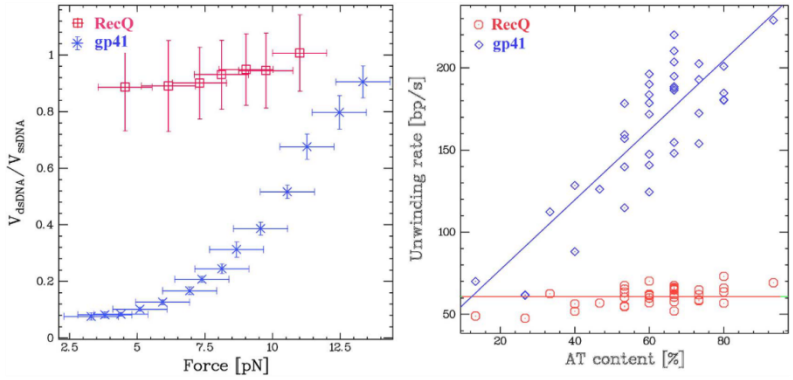


Figure 9. (A) Comparison of the rates of unwinding of RecQ and gp41 as a function of the tension on the hairpin molecule in the unzipping assay at saturating ATP concentrations. (B) Comparison of the rates of unwinding of RecQ and gp41 as a function of the AT content of the hairpin molecule (at F 5 9 pN). The independence of the rates of unwinding by RecQ on force and AT content are compatible with an active unwinding mechanism. On the other hand, the strong dependence of the rates of unwinding by gp41 on force and AT content are indicative of a passive unwinding mechanism, whereby the helicase progression is limited by the probability of spontaneous fork opening.[52]

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 12

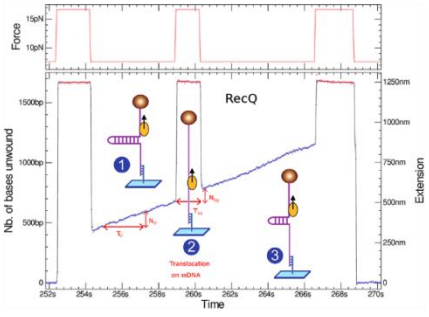


Figure 10. To investigate the translocation of a helicase on ssDNA the following assay has proved useful. Once the enzyme has started to unwind the hairpin fork, the force is increased such as to mechanically unzip the hairpin. The enzyme thus proceeds by translocating on ssDNA for a time T_{SS} determined by the time span over which the hairpin is maintained open. Upon reducing the tension of the DNA molecule, the hairpin rewinds until its fork encounters the helicase. At that point the increase in the DNA's extension, N bpS, allows one to deduce the rate of translocation of the enzyme on ssDNA: v_{SS}5 N bpS/T_{SS}

configuration since the sensitivity of the former is larger by a factor 2 to 5 as compared to the latter. In Figure 8, the trace in which the DNA extension increases as a function of time corresponds to hairpin unwinding by UvrD (see sketch 1 in Fig. 8), while the trace in which the DNA extension decreases as a function of time (see sketch 3 in Fig. 8) is due to the enzyme translocating on ssDNA after reaching the hairpin apex [see sketch 2 in Figs. 8 and 3(B)]. Note that in the extension decreasing trace the helicase blocks the fork that is reforming in its wake which could push the enzyme forward. The two traces thus correspond to different thermodynamic situations: in the increasing trace the helicase is working to unwind the dsDNA while in the decreasing trace the helicase delivers no work but is pushed in the back by the fork. For UvrD we observe that the rates in both situations are similar which means that the enzyme is not slowed down as it unwinds the molecule. Moreover, both rates are independent of the applied force which can help or hinder dsDNA opening. Such a behavior is the hallmark of an active helicase.[52]

These traces also illustrates the very processive behavior of UvrD. In this experiment the enzyme has traveled along 2.4 knts without falling off and with a remarkable regularity. The strand-switching ability of UvrD has also been observed in this configuration as the enzyme can switch directions before reaching the hairpin apex. Complex bursts were also observed. The unzipping assay with UvrD also demonstrates an interesting feature of this helicase: its ability to unbind biotin from streptavidin.[55] As the enzyme translocates past the hairpin apex, it can reach the end of the molecule where biotins present on the DNA are bound to streptavidin present on the magnetic bead. We observed that many beads that were anchored to a surface by a DNA hairpin would detach in the presence of UvrD. When tracking several tens of beads, a small subset remains bound and provides useful and reproducible unwinding signals (in these instances the biotin link is presumably inaccessible to UvrD).

The possibility of observing on the same trace helicase-mediated DNA unwinding in the rising part of the extension signal and helicase-translocation on ssDNA on the decreasing part opens interesting possibilities. The replicative helicase of the T4 replisome, gp41, behaves in a very different way[56] [Fig. 9(A)] upon unwinding and upon translocation on ssDNA. The rezipping (translocation) rate is constant and independent of the tension on the DNA molecule, while the unwinding rate increases exponentially with the force, equaling the ssDNA translocation rate at a force (F 14 pN) large enough to open the hairpin. This behavior is compatible with the behavior of a passive helicase which is not capable of melting the DNA fork but relies on spontaneous (thermal) fluctuations of the fork to proceed forward.[57] F9

The magnetic tweezer helicase assay has also been used to study RecQ, a 3^0-5^0 helicase belonging (as NS3, PriA and RecG) to the SF2 superfamily of helicases. Adding RecQ in a solution with hairpin molecules under moderate tension (F_r F $<F_c$), results in processive unwinding of the hairpin at a constant rate v_{RecQ} 80 bps/s (Fig. 9). As for UvrD, the unwinding rate of RecQ varies little with force [see Fig. 9(A)] or enzymatic concentration.[52] Conversely, and in contrast with UvrD, RecQ is never observed in a situation where the fork pushes on the back of the enzyme once it has passed the hairpin apex. As the enzyme passes the apex, the force applied by the closing fork induces the enzyme to switch strands reverting into an unwinding mode.[58]

The rate of DNA unwinding by helicases has also been studied as a function of the local AT (or GC) content of the hairpin.[59–62] In that case one measures the local rate of unwinding (averaged over many unwinding traces, such as the increasing trace in Fig. 8) as a function of the molecule's extension (i.e., the position of the enzyme along the molecule). One can use that assay to distinguish [see Fig. 8(B)], between an active helicase (such as RecQ) whose rate is insensitive to the AT-content and a passive helicase (such as gp41) whose rate increases with the local AT content of the molecule. Similar results have been obtained on the E. coli replicative helicase DnaB in the two unpeeling and unwinding configuration.[63,64] E1 helicase has also been characterized by single molecule FRET assay.[65]

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 13

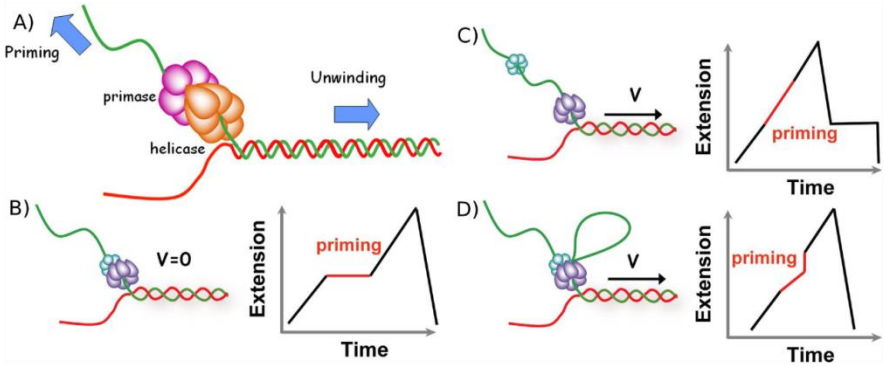


Figure 11. Models of primosome behavior during primer synthesis. (A) In the T4 virus, the helicase and the primase work as a complex encircling the lagging strand. However, their respective displacement directions are just opposite raising the question of how they collaborate. (B, C, and D) Schematic representation of three possible models for helicase and primase interaction during primer synthesis (left) and the real-time DNA extension traces expected for each model (right). (B) In the pausing model the helicase temporarily stop translocating during priming. (C) In the disassembly model the primase dissociates from the helicase to synthesize a primer while the helicase continues unwinding DNA. (D) In the DNA looping model the primosome remains intact and DNA unwound during priming forms a loop

Helicase Translocation on ssDNA

One can use the magnetic trap force modulations on a hairpin substrate to measure the rate of translocation of a helicase on an ssDNA substrate (without pushing by the closing fork as before). By mechanically and transiently increasing the force (to a value $F > Fu$), the hairpin can be completely opened during an unwinding event allowing the enzyme to translocate on ssDNA at mean rate v_{ss} (see Fig. 10). As the force is reduced (down to its initial value $F < Fc$), the hairpin rewinds. However it is blocked by the helicase and its length is shorter by an amount corresponding to the distance traveled by the enzyme. Notice that in the case of RecQ the translocation rate on ssDNA is the same as the dsDNA unwinding rate (Fig. 10), confirming the interpretation of this enzyme as an active helicase which can melt the DNA fork.[56,66,67]

F10

Coupling between Helicase and Primase in the Primosome

DNA replication is a fundamental process in the cell, the fact that the double helix implies two strands that are anti-parallel and that the polymerase enzyme can only copy a template along one direction leads to a complex process. In the replication fork that is opened by the replicative helicase, one strand (the leading strand) is in the correct orientation to be synthesized continuously, but the other (the lagging) strand is in the opposite orientation. To

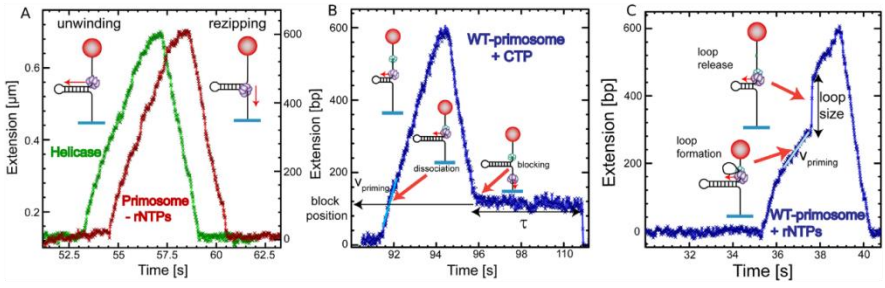


Figure 12. Primer synthesis by primosome depends on rNTP concentration. (A) Experimental traces corresponding to the gp41 helicase activity (green) and the wt primosome activity (red) in the absence of rNTPs. (B) Experimental trace displaying characteristics (unwinding velocity during priming, position and lifetime of the block) for the primosome disassembly model. DNA looping mechanism for primer synthesis. (C) Experimental trace displaying characteristics (unwinding velocity during priming and the loop size) for the DNA looping model.[68]

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 14

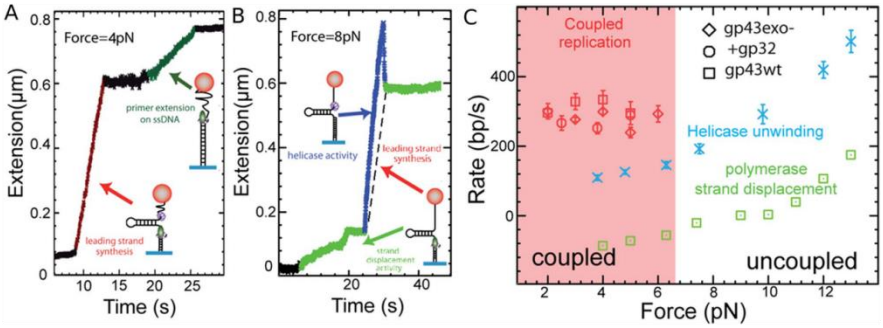


Figure 13. Single molecule studies of the coupling between the T4 helicase (gp41) and the T4 polymerase (gp43). (A) At low forces (F < 7pN), the polymerase synthesizes the complementary strand as the helicase unwinds the molecule, resulting in a fast increase (red) of the molecule extension. As the helicase and polymerase meet head-on at the hairpin apex, they stall (black). The helicase then falls off and the polymerase resumes its activity on the remaining ssDNA (the upper complementary strand of the hairpin). (B) At higher forces (F > 7pN), the helicase unwinds DNA (blue) independently of the proximity of the polymerase which synthesizes the complementary strand on the unwound ssDNA. As the helicase falls off (at t5 30 s), the hairpin re-hybridizes up to the point where the polymerase is (the dashed line represents the polymerase rate during the uncoupled helicase unwinding). The polymerase resumes polymerization in the very inefficient strand displacement mode (green). (C) Measured rates of unwinding and strand-synthesis by gp41 and gp43 alone or when they are coupled (with or without the SSB protein, gp32). Notice that at low forces (F < 5pN) and in the absence of helicase the polymerase switches into exonuclease mode (its rate is negative).[76]

replicate it the cell uses a discontinuous synthesis process whereby it is synthesized in small stretches known as Okazaki segments. These segments are synthetized by a polymerase working opposite to the direction of the replication fork. This situation cannot proceed for long and the lagging polymerase detaches after finishing one segment to start a next one. However, the polymerase is unable to initiate the replication of its template, it needs a primer that it elongates. This primer is synthesized by a special enzyme, a primase that in the T4 virus is part of a complex with the helicase (the primosome) that encircles the lagging strand. In the T7 virus, the primase is actually a subdomain of the helicase. Whenever the helicase opens a specific sequence, the primase detects this event and starts synthesizing an RNA primer complementary to the specific sequence. This primer will eventually be used by the polymerase to start an Okazaki segment. The primase is thus a type of RNA polymerase which surprisingly synthesizes its primer opposite to direction of the helicase motion. How these two opposite activities are coupled has long been an open question. Several models had been proposed (see Fig. 11) that suggested that either the activity of the primase was blocking the translocation of the helicase [pausing model, Fig. 11(B)], or that the two enzymes dissociated transiently [disassembly model, Fig. 11(C)] or that an ssDNA loop was extruded between the two enzymes during priming [looping model, Fig. 11(D)].

The unzipping assay has provided a means to address that issue. As we have seen, the T4 replicative helicase gp41 unwinds a DNA hairpin in a few seconds. Since the helicase encircles the DNA it displays a burst with a rising edge corresponding to the helicase unwinding the hairpin and a falling edge corresponding to the helicase translocating on the ssDNA while a DNA fork reforms in its wake. The primase is performing a task that is less easy to observe: if we supply NTPs, the primase will lay down a 5mer RNA oligonucleotide each time it encounters a DNA sequence starting by GTT or GCT. We have constructed two hairpins with a small number of priming sites located in only one of the strands.[68] If we add the primase to a buffer containing helicase and ATP only, the primase is unable to initiate a primer since that requires CTP (Cytidine triphosphate). In this situation, we observe bursts of helicase activity similar to those observed with helicase alone. This test shows that a primase does not alter the helicase motion when it is inactive.

If we add the NTPs and use a hairpin with priming sites in the lagging strand only that is with priming sites that are encountered by the primosome when opening the DNA, we observe two noticeable features in the activity bursts. As shown in Figure 12(B), the rising edge of the burst is similar to a helicase burst but on the falling edge we observe a blockage where the hairpin remains open at a given position for sometimes. This phenomenon is not seen if we use a hairpin having no priming sites on the lagging strand. The blockage positions occur randomly in time but at positions corresponding to the expected priming sites. The interpretation of this observation is as follows: as the helicase

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 15

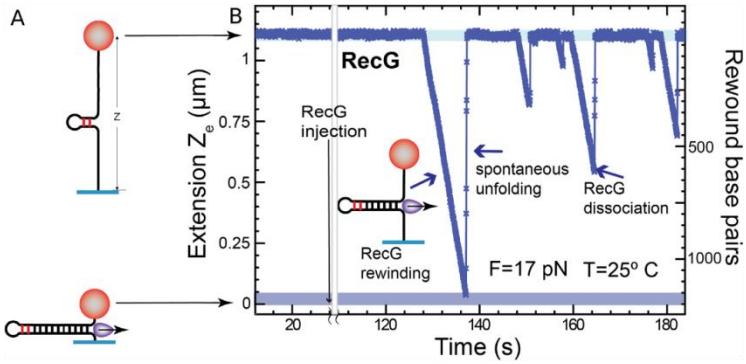


Figure 14. (A) The substrate for the single molecule study of the rewinding activity of a helicase is an opened hairpin with a small stem loop at its end (that serves as a loading site for the enzyme). (B) Upon enzymes rewinding the extension of the hairpin (maintained under a high tension) is observed to decrease. Sudden increases in extension are due to dissociation of the enzyme from its substrate and subsequent force-induced unfolding of the hairpin.[80]

opens the hairpin, whenever a priming site is detected, the primase synthesizes a primer that remains hybridized to its template and is blocking the rehybridization of the hairpin in the wake of the helicase (once it has passed the apex). The hybridized primer is in fact linked to a part of the primase that stabilizes it on its DNA template. This observation validates the disassembly model for the primosome whereby the primase dissociates from the helicase as it synthesizes its primer. This priming activity is clearly a stochastic process where the possible priming sites randomly trigger the priming activity: in many instances a priming site is not associated with a priming event. Thus, the blockage positions although very well localized appear in random order and the duration of the blockages follows a Poisson distribution with a mean duration of a few seconds.

We also observe a second type of events where the rising edge of the burst displays a special feature shown in Figure 14(C). Starting at a priming site, the measured increase in extension slowed down by a factor two, before presenting a jump in extension and recovering its original rate of lengthening. This event may be explained by the extrusion of a loop between the helicase and the primase. During this loop extrusion, the lengthening of the molecule is only due to the other opened strand, as the sequestered loop does not participate in the extension signal. As the primase dissociates from the helicase, the extension increases suddenly. This event is compatible with the looping model. The size of the loop is typically 200 or 300 nts.

This single molecule assay of the primosome has provided valuable information which nicely complement bulk assays, in particular we can definitely rule out the pausing model which was never seen. The stochastic nature of the primosome was not fully expected. While it seems reasonable that not all priming sites lead to a priming event, it is surprising that two priming processes occur in a random manner. We found that in our conditions, upon priming a primase will more often dissociate from the helicase that extrude a loop, but the ratio between the two behaviors may depend on replisome cofactors.

Coupling between Helicases and Polymerases during Replication

During the replication of a DNA molecule, the rate of DNA unwinding and DNA polymerization must be tightly coupled within the replication complex. The situation has been first investigated by De Lagoutte and Von Hippel, more recently the T4 replisome description is discussed in Refs. 69 and 70, while the T7 replisome has been also extensively studied.[71–74] In the case of the T4 replisome[1,2] using as a template, a dsDNA molecule ending with an open fork and a primer on the leading strand (mimicking the natural replication fork) they have shown the following surprising result:

The helicase gp41 alone is very inefficient in opening DNA, its rate is a few tens of bp/s and its processivity is reduced to 100 bases
The T4 polymerase alone is not active since it lacks strand-displacement activity
But the combination of the helicase and the polymerase (holoenzyme) together results in a rapid elongation of the primer at 400 bp/s over several kbs.

While the inactivity of the T4 holoenzyme alone is understood, the inefficiency of the helicase alone is more surprising: one would imagine that a replicative

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 16

helicase which eventually opens the entire genome, is a heavy-duty enzyme that does not require the collaboration of the polymerase to unwind DNA. An understanding of the coupling between helicase and polymerase was made possible using single molecule assays.

Single molecule manipulations, particularly in the unzipping configuration, offers a convenient and versatile assay to study that coupling. First, each component of the complex (e.g., helicase, polymerase, primase, etc.) can be studied in isolation from the others. As shown previously the replicative helicase from phage T4, gp41, is a passive helicase with a slow unwinding rate at low tensions, a rate that can be 10 times slower than its unhindered translocation rate on ssDNA of 600nt/sec [see Fig. 9(B)]. Thus, the single molecule approach confirmed the bulk results that the helicase alone is inefficient at low forces, but it also shows that the full speed of the helicase can be recovered provided it is assisted by a strong force.[75] Conversely, the T4 polymerase, gp43, cannot work in a strand displacement mode. However, studying the T4 polymerase in the unzipping configuration with a primer hybridized on one strand revealed a surprising result: If a pulling force $F > 10$ pN is applied, the polymerase can elongate the primer with a substantial rate of 250bp/s, which suggests that if the fork is slightly destabilized (by the tension) the polymerase can synthesize a new complementary strand by displacing the old one in front of it. At lower tension on the hairpin the polymerase stalls and can even go into exonuclease mode, removing the strand it has just synthesized
F13 [see Fig. 13(C)]. When the two enzymes (helicase and polymerase) are assembled on a DNA hairpin,[76] at low forces their joint action allows the helicase to unwind at a rate similar to the maximal rate of the gp43 polymerase [Fig. 13(A)]. At higher tensions, the two enzymes decouple and each proceeds at the rate measured in isolation [Fig. 13(B)].

A simple mechanical model accounts for these observations. It assumes that at low forces the polymerase destabilizes the fork enough to allow the passive gp41 helicase to proceed with DNA unwinding. Its rate is then determined by the maximal rate of polymerization (it cannot unwind faster as it will decouple from the polymerase and slow down). At higher tensions, the fork is sufficiently destabilized that the helicase does not need the polymerase to unwind the molecule. It unwinds DNA at a rate which is the one measured at that tension in the absence of polymerase [Fig. 13(B)]. In the wake of the helicase, on the generated ssDNA the polymerase synthesizes the complementary strand [see Fig. 13(B)]. As the helicase unbinds from its substrate, the hairpin reforms up to the polymerase whose action is slowed down or stalled by the fork ahead.

This purely mechanical coupling between helicase and polymerase, in contrast with a biochemical (for example allosteric) coupling, suggests that helicases and polymerases from different species could work together. This prediction has been validated by showing that the phage T4 helicase and the phage T7 polymerase could indeed complement each other to form an efficient replicative complex. These results suggest the presence in the T4 polymerase of a specific region that destabilizes the fork, increasing unpeeling fluctuations that speed up the gp41 helicase.

Finally, this mechanical coupling provides a rational for the difference existing between passive and active helicases: the proximity of the T4 polymerase effectively switches the gp41 passive helicase into an active mode, a coupling that prevents opening of the fork without DNA synthesis activity. Very similar coupling has been observed in the T7 replisome.[73,77,78]

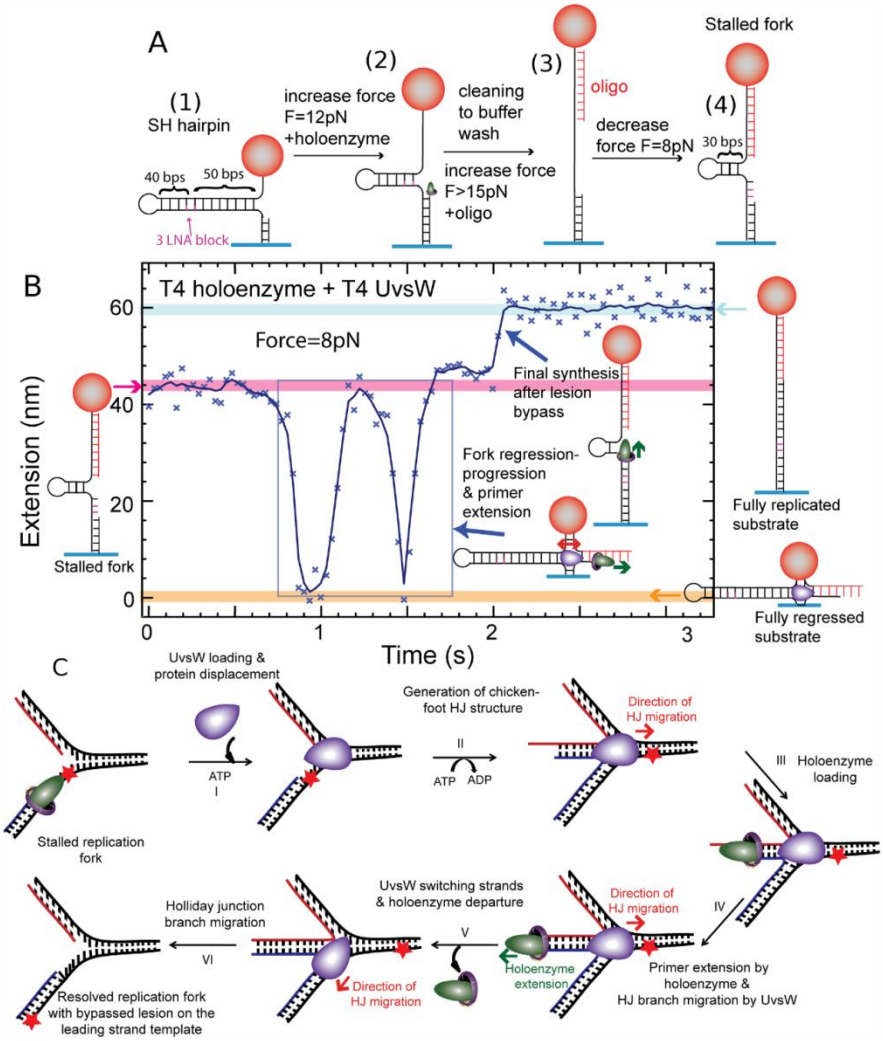
Helicase Rewinding of DNA Hairpin

Some helicases, such as RecG and UvsW, possess the surprising ability to rewind an unzipped DNA hairpin against a considerable force (up to about 30 pN).[79,80] This activity is required in the rescue of a stalled replication fork. If a DNA damage exists on the leading strand of a replication fork, it will stop the leading strand polymerase posing a serious threat to the cell. Even a phage such as T4 has a special helicase devoted to address that issue. If such damage blocks the leading polymerase, one way to bypass the problem is to replace the defective template by the copy of its complementary strand done by the lagging strand polymerase, this implies regressing the stalled fork and generating a so-called "chicken-foot" DNA structure. Regenerating the DNA fork allows the replication complex to bypass the damaged area and to proceed with replication. This fascinating mechanism led to a strong debate as it seemed that the coordination of the different events was so crucial that its occurrence would be highly improbable. Indeed, it was difficult to understand how an enzyme would reverse the replication fork when it is stalled, how the reverse process will happen once the elongation of the critical bases had occurred and how many different enzymes were required. Using a single molecule approach, we found that this complex process is in fact the result of the collaboration of just two enzymes engaged in a stochastic ballet. For the T4 virus, these enzymes are the UvsW helicase and the gp43 polymerase (in E. coli, RecG replaces UvsW)

To study these enzymes an unzipping assay was used. The DNA hairpin was designed to have a GC-rich segment close to its apex. The tension in the hairpin ($F > 14$ pN) was adjusted such as to partially open it (up to the more stable GC-rich segment close

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 17



COLOR IN ONLINE AND PRINT

Figure 15. Reconstruction of the template switching pathway. (A) Construction of a stalled fork substrate with a LNA block at the leading strand, mimicking a damage. (B) Experimental trace at F 5 8 pN displaying the extension z(t) with ATP, dNTPs, UvsW and T4 holoenzyme starting with the stalled fork and ending with the fully replicated substrate. (C) Schematic of the repair of a stalled replication fork by its regression.[79]

to the apex, see the rise from 15 to 17 pN in the hairpin opening trace of Fig. 4) leaving a small stem loop at its end that serves as a loading site for the helicase [Fig. 14(A)]. The rewinding (re-zipping) activity of the enzyme against the tension pulling on the DNA can be monitored via the decrease in extension of the molecule. Detachment of the enzyme from its substrate is observed as a rapid increase in extension resulting from the tension-induced opening of the hairpin [Fig. 14(B)].

Except for the inversion in the extension signal, the processivity and the rate of these enzymes can be deduced in a similar manner as for the unwinding activity of more common helicases.

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 18

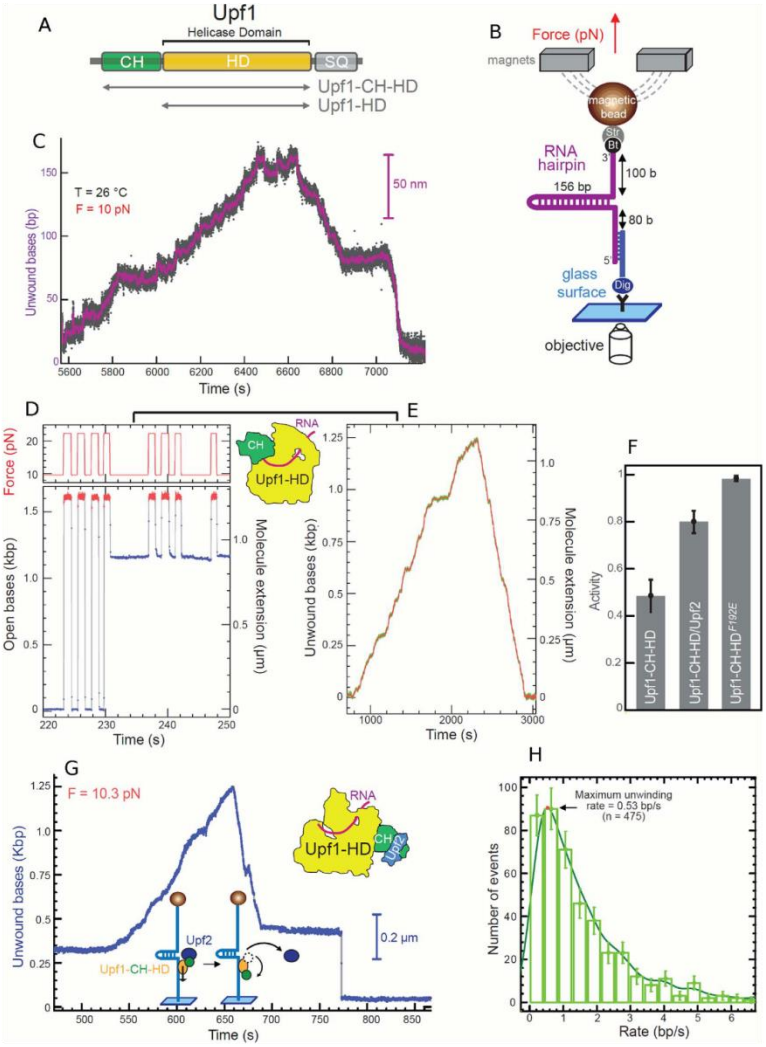


Figure 16. Upf1 helicase activity on RNA. (A) Organization of Upf1 and truncated versions (gray arrows) used in our study, Upf1-HD and Upf1-CH-HD. (B) Schematic representation of the RNA used for the MT set up. (C) Experimental trace showing the activity of Upf1-HD in saturating concentration of ATP. The number of unwound bases is deduced from the molecular extension Z(t) obtained at F 5 10pN. From 5600 s to 6440 s, the helicase unwound the 156bp RNA hairpin. From 6640 s to 7200 s, the RNA hairpin refolded, while Upf1-HD translocated on ssRNA reaching the 3′extremity. (D, E, and F) The binding of Upf2 to CH domain activates Upf1-CH-HD unwinding and translocation. (D, E) Experimental traces corresponding to the two types of enzymatic activity detected for Upf1 on a DNA hairpin. (D) Upf1 binds on ssDNA (starting at 231 s), blocking the re-zipping of the hairpin. (E) The enzyme is active. (F) Histogram of relative activity of Upf1-CH-HD, Upf1-CH-HD/Upf2 complex and Upf1F192E mutant. (G) Trace of human Upf1-CH-HD/Upf2 complex unwinding steadily and completely the DNA molecule, passing the apex, pursuing its translocation and refolding the DNA hairpin. At t 5 675 s, the complex makes two strand switching events before finally stopping its activity leaving the hairpin blocked (at t 5 687 s). This trace is atypical since the Upf1-CH-HD/Upf2 complex translocates for some time before stalling; for most bursts, the complex stops very quickly after passing the apex. (H) Distribution of instantaneous unwinding rate of the Upf1-CH-HD/Upf2 complex.[87]

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 19

Converting a Three-Way Junction to a Holliday Junction

Helicases not only translocate on ssDNA but are also involved in the conversion of a three-way junction into a "chicken-foot" Holliday junction (HJ) and in the migration of this junction. This process is a key aspect of the fork regression error correction mechanism. Using a variant of the unzipping configuration it has been possible to study this process in real time.

The preparation of a DNA template mimicking a stalled replication is sketched in Figure 15(A). First a DNA hairpin with a built-in primer (1) is used to anchor a bead to a surface. The DNA hairpin contains three consecutive Locked Nucleic Acids (LNA) bases a distance 40 bps from its apex. Adding the T4 polymerase and increasing the force (2) allows for primer extension until the polymerase bumps into the LNA bases that stop the elongation. An oligonucleotide with a sequence complementary to the lagging strand (3) is then injected. As the tension on the molecule is reduced the hairpin refolds partially mimicking a stalled replication fork (4) with the LNA playing the role of the DNA damage.

To test the efficiency of a helicase in the rescue of this stalled replication fork by converting this three-way junction into a Holliday junction, the UvsW helicase, the gp43 polymerase and ATP were flowed in. As in the previous experiment, UvsW starts refolding the hairpin. But here the two arms are in fact dsDNA and the helicase first converts the three-way junction into a Holiday junction that it displaces. This is evidenced by the rapid decrease of the molecule's extension as the bead approaches the glass surface. During this process, the strand which elongation was blocked by the presence of improper LNA bases, is re-hybridized to a strand that has no damage. This offers the possibility for the elongation to restart by polymerase elongation of a proper template. UvsW displays an interesting behavior: it reverts the direction of the Holiday junction migration in a random manner. Thus, the bead goes up again after its first downwards motion. On Figure 15(B), we observe two consecutive motions up and down illustrating this behavior. There is no special coordination between UvsW and the T4 polymerase, so that the polymerase will not load automatically as the fork is reverted but it can do so. In Figure 15(B), on the second fork reversal, the T4 polymerase was able to load and elongate its template so that the lesion was actually passed. As UvsW switches back the chicken foot to a three-way junction, the fork replication can restart.

This error correction mechanism is a perfect illustration of a stochastic collaboration between two enzymes: by moving the Holiday junction forward and backward, UvsW provides an opportunity for the gp43 polymerase to elongate the strand that was blocked. When the proper complementary strand has been synthesized the fork is ready to reload the replication helicase gp41 and to re-initiate replication on a proper template. We observed that the gp41 helicase has priority over UvsW so that in normal fork progression, UvsW does not revert the fork. When a damage blocks the replication fork, the uncoupling between the polymerase and the replicative helicase leads to the unbinding of the helicase, which allows for the action of UvsW.[79,81]

Helicase Unwinding of RNA/DNA Duplexes

The world of RNA helicases is more diverse than the DNA helicases homologs. The RNA helicase activities target specific RNA structures which are not simply double-stranded RNA, a structure which is rather uncommon in the cell. The RNA helicase substrates might be specific secondary structures (local hairpins or bulges) but also ribonucleoprotein complexes. In eukaryotes, RNA helicases are implicated in every molecular process involving RNAs and notably transcription and all the post-transcriptional events including pre-mRNA processing, mRNA transport, translation and degradation.[82,83] Some RNA helicases are really surprising as they may have sometimes extremely low processivities,[84] for instance, DEAD box helicases are involved in disrupting just a few nucleotides and do not translocate on their template while others simply serve as RNA clamp.[85]

We can also test RNA helicases with our magnetic tweezers using the same protocol as that for DNA helicases. However, preparing an RNA template either in the peeling or unzipping configuration is more difficult than the equivalent DNA template. The panel of enzymes handling RNA in the laboratory is not as rich as for DNA, no restriction enzymes are available, ligases are less efficient, and incorporating modified bases is not obvious, though one can find interesting recipes.[86] RNA also has the ability to form more complex secondary structures than those expected from Watson-Crick pairings, leading to a difficult molecular design. Finally, RNAs are prone to degradation by very ubiquitous ribonucleases. Preparing a complete RNAse-free single molecule flow cell is a challenge. Consequently, the lifetime of an RNA construct in a magnetic tweezers set-up is usually substantially shorter than a DNA one, at least in our hands. A useful alternative when the helicase works on a RNA/DNA hybrid is to use a DNA template molecule as a holding device and to hybridize RNA single stranded molecules to the DNA.

We have recently studied the human Upf1 RNA helicase[87] that is essential to survey the integrity of eukaryotic mRNAs by the process of nonsense-mediated mRNA decay (NMD).[88] This helicase has

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 20

some interesting features: it translocates slowly but F16 presents a surprisingly high processivity (Fig. 16). It works both on RNA and DNA templates without significant differences. Upf1 translocates at a few bases per second with an irregular rate. The Upf1 enzyme is regulated by several factors including the internal CH domain [Fig. 16(A)] which inhibits the helicase activity when not in the correct configuration and the NMD cofactor Upf2 which is needed for helicase activity. The existence of these two mechanisms is presumably responsible for the irregular activity of Upf1. The single molecule approach allowed us to characterize the regulatory effect of Upf2. First, in the absence of the regulatory CH domain, Upf1-HD binds to its template very firmly and keeps on moving for very long distance: it is able to completely unwind a 156 bp RNA hairpin [Fig. 16(B)] as well as a 1.2kbps DNA hairpin. When we tested the enzyme with its regulatory CH domain, Upf1-CH-HD was seen mostly binding ssDNA and blocking the DNA fork (a small fraction of the enzymes shown some helicase activity presumably because the CH domain may have different configurations and thus did not block all the enzyme). Interestingly, adding Upf2 clearly changed the pattern as most complexes became active with a normal helicase behavior. The activity is compatible with Upf2 flipping the CH domain in the active configuration. This was confirmed by using the Upf1-CH-HD-F192E mutant where the CH domain is flipped in the active configuration permanently by the mutation.[89] The ability of testing helicase co-factors is fundamental to better characterize RNA helicases which often act as part of protein complexes in which RNA helicases are regulated by one or several cofactors. Using our single molecule assay, we estimated that the Upf1 processivity is extremely large reaching 16 kbps. This very high processivity suggests that Upf1 might remodel numerous protein-RNA interactions to trigger the degradation of aberrant mRNAs by NMD.[87]

Conclusions and Perspectives

Single molecule manipulations are now used either to characterize helicase stepping[90] or the role of their subdomains and their link[91] or to understand the coupling between enzymes required in many biological situations. Here we have shown examples where two enzymes interact: this is the case of a primase and a helicase, a helicase and a polymerase in leading strand synthesis and in the fork regression mechanism. The obvious advantage of these single molecule studies is the ability to follow in real time the different interactions, revealing their underlying dynamics. More complex systems have also been investigated which pave the way for further deeper studies. The Van Oijen group has studied the entire replisome of T7[92] and E. coli. These experiments involve many enzymatic partners and the probability of their successful assembly is low. Hence these authors have used extensive parallelism (tracking several 10,000 beads) to observe a small set of clear events where all the partners in the complex were active. This has allowed them to study the collaborative behavior of the different elements of the replisome.

In the same spirit, the collision of the replication fork with the Tus-Ter replication termination has been investigated using single molecule micromanipulation.[93] DNA repair mechanism is also a subject where micromanipulation is bringing valuable information such as the way used by polymerase to overcome DNA damage[94] or how stalled RNA-pol are restarted.[95] The observation of helicase activity on a specific DNA quadruplex[96] opens new avenues of investigation.[97] However, as soon as several enzymes are involved in a biological process, the probability of assembly of a functional complex in single molecule assays decreases dramatically. The success of this approach thus relies on the development of high throughput single-molecule manipulation assays[98] which the inherent parallelism of magnetic traps is ideally poised to provide.

Acknowledgments

We thank N. Desprat, J. Ouellet, M.M. Spiering, S.J. Benkovic and C. Andre, for useful comments on the manuscript.

References

1. Delagoutte E, von Hippel PH (2002) Helicase mechanisms and the coupling of helicases within macromolecular machines Part 1: Structures and properties of isolated helicases. Quart Rev Biophys 35:431–478.
2. Delagoutte E, von Hippel PH (2003) Helicase mechanisms and the coupling of helicases within macromolecular machines - Part II: Integration of helicases into cellular processes. Quart Rev Biophys 36:1–69.
3. Lohman TM, Bjornson KP (1996) Mechanisms of helicase-catalyzed unwinding. Annu Rev Biochem 65: 169–214.
4. West SC (1996) DNA helicases: New breeds of translocating motors and molecular pumps. Cell 86:177–180.
5. Soultanas P, Wigley DB (2001) Unwinding the 'Gordian knot' of helicase action. Trends Biochem Sci 26:47–54.
6. Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R (2004) Molecular biology of the gene. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
7. Patel SS, Picha KM (2000) Structure and function of hexameric helicases. Ann Rev Biochem 69:651.
8. DNA Helicases and DNA Motor Proteins, Maria Spies, Volume 973 of the series Advances in Experimental Medicine and Biology; Springer (2012)
9. Jankowsky E (2012) Methods in enzymology. Preface. Methods Enzymol 511:xix–xxx. doi:10.1016/B978-0-12-396546-2.00029-2.
10. Gorbalenya AE, Koonin EV (1993) Helicase amino acid comparisons and structure function relationship. Curr Opin Struct Biol 3:419–429.

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 21

11. Singleton MR, Dillingham MS, Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem 76:23–50.
12. Fairman-Williams ME, Guenther U-P, Jankowsky E (2010) SF1 and SF2 helicases: family matters. Curr Opin Struct Biol 20:313–324.
13. Benkovic SJ, Raney KD (2011) Mechanisms: molecular machines. Curr Opin Chem Biol 15:577–579.
14. Velankar SS, Soultanas P, Dillingham MS, Subramanya HS, Wigley DB (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell 97:75–84.
15. Wigley DB (2007) RecBCD: the supercar of DNA repair. Cell 131:651–653.
16. Nanduri B, Byrd AK, Eoff RL, Tackett AJ, Raney KD (2002) Pre-steady-state DNA unwinding by bacteriophage T4 Dda helicase reveals a monomeric molecular motor. Proc Natl Acad Sci USA 99:14722–14727.
17. Raney KD, Byrd AK, Aarattuthodiyil S (2013) Structure and mechanisms of SF1 DNA helicases. Adv Exp Med Biol 767:17–46.
18. Lee JY, Yang W (2006) UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. Cell 127:1349–1360.
19. Dillingham MS, Kowalczykowski SC (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiol Mol Biol Rev 72:642–671.
20. Yang W (2010) Lessons learned from UvrD helicase: mechanism for directional movement. Annu Rev Biophys 39:367–385.
21. Jia H, Korolev S, Niedziela-Majka A, Maluf NK, Gauss GH, Myong S, Ha T, Waksman G, Lohman TM (2011) Rotations of the 2B sub-domain of E. coli UvrD helicase/translocase coupled to nucleotide and DNA binding. J Mol Biol 411:633–648.
22. Manthei KA, Hill MC, Burke JE, Butcher SE, Keck JL (2015) Structural mechanisms of DNA bnding and unwinding in bacterial RecQ helicases. Proc Natl Acad Sci USA 114:4292–4297.
23. von Hippel PH, Delagoutte E (2001) A general model for nucleic acid helicases and their coupling" within macromolecular machines. Cell 104:177–190.
24. Liu B, Baskin RJ, Kowalczykowski SC (2013) DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate. Nature 500:482.
25. Wu CG, Bradford C, Lohman TM (2010) Escherichia coli RecBC helicase has two translocase activities controlled by a single ATPase motor. Nat Struct Mol Biol 17:1210–1217.
26. Roman LJ, Kowalczykowski SC (1989) Characterization of the helicase activity of the E. coli RecBCD enzyme using a novel helicase assay. Biochemistry 28: 2863–2873.
27. Roman LJ, Eggleston AK, Kowalczykowski SC (1992) Processivity of the DNA helicase activity of Escherichia-coli RecBCD enzymze. J Biol Chem 267: 4207–4214.
28. Roman LJ, Kowalczykowski SC (1989) Characterization of the ATPase activity of the E. coli RecBCD enzyme: relationship of ATP hydrolysis to the unwinding of duplex DNA. Biochemistry 28:2873–2881.
29. Lucius AL, Vindigni A, Gregorian R, Ali JA, Taylor AF, Smith GR, Lohman TM (2002) DNA unwinding stepsize of E. coli RecBCD helicase determined from single turnover chemical quenched-flow kinetic studies. J Mol Biol 324:409–428.
30. Lucius AL, Jason Wong C, Lohman TM (2004) Fluorescence stopped-flow studies of single turnover kinetics of E. coli RecBCD helicase-catalyzed DNA unwinding. J Mol Biol 339:731–750.
31. Lucius AL, Vindigni A, Gregorian R, Ali JA, Taylor AF, Smith GR, Lohman TM (1989) DNA unwinding stepsize of E. coli RecBCD helicase determined from single turnover chemical quenched-flow kinetic studies. Biochemistry 28:2863–2873.
32. Bianco PR, Brewer LR, Corzett M, Balhorn R, Yeh Y, Kowalczykowski SC, Baskin RJ (2001) Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. Nature 409:374–378.
33. Handa N, Bianco PR, Baskin RJ, Kowalczykowski SC (2005) Direct visualization of RecBCD movement reveals cotranslocation of the RecD motor after Chi recognition. Mol Cell 17:771–750.
34. Spies M, Bianco PR, Dillingham MS, Handa N, Baskin RJ, Kowalczykowski SC (2003) A molecular throttle: The recombination hotspot chi controls DNA translocation by the RecBCD helicase. Cell 114:647–654.
35. Strick T, Allemand JF, Bensimon D, Bensimon A, Croquette V (1996) The elasticity of a single supercoiled DNA molecule. Science 271:835–1837.
36. Lansdorp BM, Tabrizi SJ, Dittmore A, Saleh OA (2013) A high-speed magnetic tweezer beyond 10,000 frames per second. Rev Sci Instrum 84:044301.
37. Lee KS, Balci H, Jia H, Lohman TM, Ha T (2013) Direct imaging of single UvrD helicase dynamics on long single-stranded DNA. Nat Commun 4:1878.
38. Dulin D, Cui T, Cnossen J, Docter MW, Lipfert J, Dekker NH (2015) High spatiotemporal resolution magnetic tweezers for single-molecule force spectroscopy: Calibration and applications to DNA dynamics. Biophys J 109:2113–2125.
39. Einstein A (1956) Investigation of the Brownian theory of movement. Mineola, New York: Dover Publication.
40. Reif F (1965) Fundamentals of statistical and thermal physics. New York: McGraw-Hill.
41. Lionnet T, Allemand JF, Revyakin A, Strick TR, Saleh OA, Bensimon D, Croquette V, Single-molecule studies using magnetic traps. In: Ha T, Selvin PR, Eds. (2008) Single-molecule techniques: A laboratory manual. Cold Spring Harbor, New York: CSHL Press.
42. Allemand JF (1997) Micromanipulation de molecules d'ADN isolees, Thèse Universite Pierre et Marie Curie.
43. Velthuis AJ, Kerssemakers JW, Lipfert J, Dekker NH (2010) Quantitative guidelines for force calibration through spectral analysis of magnetic tweezers data. Biophys J 99:1292–1302.
44. Lansdorp BM, Saleh OA (2014) Power spectrum and Allan variance methods for calibrating single-molecule video-tracking instruments. Rev Sci Instrum 83: 025115.
45. De Vlaminck W, Henighan T, van Loenhout MTJ, Pfeiffer I, Huijts J, Kerssemakers JWJ, Katan AJ, van Langen-Suurling A, van der Drift E, Wyman C, Dekker C (2011) Highly parallel magnetic tweezers by targeted DNA tethering. Nano Lett 11:5489–5493.
46. Plenat T, Tardin C, Rousseau P, Salome L (2012) High-throughput single-molecule analysis of DNA–protein interactions by tethered particle motion. Nucleic Acids Res 40:389.
47. Kemmerich FE, Kasaciunaite K, Seidel R (2016) Modular magnetic tweezers for single-molecule characterizations of helicases. Methods 108:4–13.
48. Matson SW (1986) Escherichia coli helicase ii (uvrd gene product) translocates unidirectionnaly in a 3' to 5' direction. J Biol Chem 261:10169–10175.
49. Runyon GT, Lohman TM (1989) Escherichia coli helicase ii (UvrD) protein can completely unwind fully

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 22

duplex linear and nicked circular DNA. J Biol Chem 264:17502–17512.
50. Dessinges M-N, Lionnet T, Xi X, Bensimon D, Croquette V (2004) Single molecule assay reveals strand switching and enhanced processivity of UvrD. Proc Natl Acad SCI USA 101:6439–6444.
51. Ali JA, Lohman TM (1997) Kinetic measurement of the step size of DNA unwinding by Escherichia coli UvrD helicase. Science 275:377–380.
52. Manosas M, Xi X, Bensimon D, Croquette V (2010) Active and passive mechanisms of helicases. Nucleic Acids Res 38:5518–5526.
53. Xie P (2016) Dynamics of monomeric and hexameric helicases. Biophys Chem 211:49–58.
54. Sun B, Johnson DS, Patel G, Smith BY, Pandey M, Patel SS, Wang MD (2011) ATP-induced helicase slippage reveals highly coordinated subunits. Nature 478: 133.
55. Morris PD, Byrd AK, Tackett AJ, Cameron CE, Tanega P, Ott R, Fanning E, Raney KD (2002) Hepatitis C virus NS3 and simian virus 40 T antigen helicases displace streptavidin from 5′-biotinylated oligonucleotides but not from 3′-biotinylated oligonucleotides: evidence for directional bias in translocation on single-stranded DNA. Biochemistry 7:2372–2378.
56. Lionnet T, Spiering MM, Benkovic SC, Bensimon D, Croquette V (2007) Real-time observation of t4 gp41 helicase reveals a passive unwinding mechanism. Proc Natl Acad Sci USA 104:19790–19795.
57. Betterton MD, Jülicher F (2005) Opening of nucleic-acid double strands by helicases: active versus passive opening. Phys Rev E Stat Nonlin Soft Matter Phys 71: 011904. Erratum in: Phys Rev E Stat Nonlin Soft Matter Phys 72:029906.
58. Klaue D, Kobbe D, Kemmerich F, Kozikowska A, Puchta H, Seidel R (2013) Fork sensing and strand switching control antagonistic activities of RecQ helicases. Nat Commun 4:2024.
59. Qi Z, Pugh RA, Spies M, Chemla YR (2013) Sequence-dependent base pair stepping dynamics in XPD helicase unwinding. eLife 2:e00334.
60. Cheng W, Dumont S, Tinoco I, Bustamante C (2007) NS3 helicase actively separates RNA strands and senses sequence barriers ahead of the opening fork. Proc Natl Acad Sci USA 104:13954–13959.
61. Syed S, Pandey M, Patel SS, Ha T (2014) Single-molecule fluorescence reveals the unwinding stepping mechanism of replicative helicase. Cell Rep 6:1037–1045.
62. Byrd AK, Matlock DL, Bagchi B, Aarattuthodiy S, Harrison D, Croquette V, Raney KD (2012) Dda helicase tightly couples translocation on single-stranded DNA to unwinding of duplex DNA: Dda is an optimally active helices. J Mol Biol 420:141–154.
63. Ribeck N, Kaplan DL, Bruck I, Saleh OA (2010) DnaB helicase activity is modulated by DNA geometry and force. Biophys J 99:2170–2179.
64. Ribeck N, Saleh OA (2013) DNA unwinding by ring-shaped T4 helicase gp41 is hindered by tension on the occluded strand. PLoS One 8:e79237.
65. Lee S-J, Syedd S, Enemark EJ, Schuckc S, Stenlundc A, Ha T, Joshua-Tor L (2014) Dynamic look at DNA unwinding by a replicative helicase. Proc Natl Acad Sci USA 111:E827–E835.
66. Johnson DS, Bai L, Smith BY, Patel SS, Wang MD (2007) Single molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. Cell 129: 1299–1309.
67. Ha T, Kozlov AG, Lohman TM (2012) Single-molecule views of protein movement on single-stranded DNA. Annu Rev Biophys 41:295–319.
68. Manosas M, Spiering MM, Zhuang Z, Benkovic SJ, Croquette V (2009) Coupling DNA unwinding activity with primer synthesis in the bacteriophage T4 primosome. Nat Chem Biol 5:904–912.
69. Chen D, Yue H, Spiering MM, Benkovic SJ (2013) Insights into okazaki fragment synthesis by the T4 replisome the fate of lagging-strand holoenzyme components and their influence on Okazaki fragment size. J Biol Chem 288:20807–20816.
70. Noble E, Spiering MM, Benkovic SJ (2015) Coordinated DNA replication by the bacteriophage T4 peplisome. Viruses 7:3186–3200.
71. Patel SS, Pandey M, Nandakumar D (2011) Dynamic coupling between the motors of DNA replication: hexameric helicase, DNA polymerase, and primase. Curr Opin Chem Biol 15:595–605.
72. Lee S-J, Richardson CC (2011) Choreography of bacteriophage T7 DNA replication. Curr Opin Chem Biol 15: 580–586.
73. Pandey M, Patel SS (2014) Helicase and polymerase move together close to the fork junction and copy DNA in one-nucleotide steps. Cell Rep 6:1129–1138.
74. Geertsema HJ, van Oijen AM (2013) A single-molecule view of DNA replication: The dynamic nature of multi-protein complexes revealed. Curr Opin Struct Biol 23: 788–793.
75. Manosas M, Spiering MM, Ding F, Bensimon D, Allemand JF, Benkovic SJ, Croquette V (2012) Mechanism of strand displacement synthesis by DNA replicative polymerases. Nucleic Acids Res 40:6174–6186.
76. Manosas M, Spiering MM, Ding F, Croquette V, Benkovic SJ (2012) Collaborative coupling between polymerase and helicase for leading-strand synthesis. Nucleic Acids Res 40:6187.
77. Nandakumar D, Pandey M, Patel SS (2015) Cooperative base pair melting by helicase and polymerase positioned one nucleotide from each other. eLife 4:e06562.
78. Nandakumar D, Patel SS (2016) Methods to study the coupling between replicative helicase and leading strand DNA polymerase at the replication fork. Methods 108:65–78.
79. Manosas M, Perumal SK, Croquette V, Benkovic SJ (2012) Direct observation of stalled fork restart via fork regression in the T4 replication system. Science 338:1217–1220.
80. Manosas M, Perumal SK, Bianco PR, Ritort F, Benkovic SJ, Croquette V (2014) RecG and UvWcatalyse robust DNA rewinding critical for stalled DNA replication fork rescue. Nat Commun 4:2368.
81. Betous R, Couch FB, Mason AC, Eichman BF, Manosas M, Cortez D (2013) Substrate-selective repair and restart of replication forks by DNA translocases. Cell Rep 3:1958–1969.
82. Jankowsky E (2011) RNA helicases at work: binding and rearranging. Trends Biochem Sci 36:19–29.
83. Bourgeois CF, Mortreux F, Auboeuf D (2016) The multiple functions of RNA helicases as drivers and regulators of gene expression. Nat Rev Mol Cell Biol 17:426–438.
84. Garcıa-Garcıa C, Frieda KL, Feoktistova K, Fraser CS, Block SM (2015) Factor-dependent processivity in human eIF4A DEAD-box helicase. Science 348:1486.
85. Pyle AM (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. Annu Rev Biophys 37:317–336.

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

J_ID: PRO Customer A_ID: PRO3187 Cadmus Art: PRO3187 Ed. Ref. No.: PRO-17-0048.R1 Date: 8-June-17 Stage: Page: 23

86. Vilfan ID, Kamping W, van den Hout M, Candelli A, Hage S, Dekker NH (2007) An RNA toolbox for single-molecule force spectroscopy studies. Nucleic Acids Res 35:6625–6639.
87. Fiorini F, Bagchi D, Le Hir H, Croquette V (2015) Human Upf1 is a highly processive RNA helicase and translocase with RNP remodeling activities. Nat Commun 6:7581.
88. Hug N, Longman D, Caceres JF (2016) Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic Acids Res 44:1483–1495.
89. Chakrabarti S, Jayachandran U, Bonneau F, Fiorini F, Basquin C, Domcke S, Le Hir H, Conti E (2011) Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. Mol Cell 41:693–703.
90. Myong S, Ha T (2010) Stepwise translocation of nucleic acid motors. Curr Opin Struct Biol 20:121–127.
91. Arslan S, Khafizov R, Thomas CD, Chemla YR, Ha T (2015) Engineering of a superhelicase through conformational control. Science 348:344.
92. Duderstadt KE, Geertsema HJ, Stratmann SA, Punter CM, Kulczyk AW, Richardson CC, van Oijen AM (2016) Simultaneous real-time imaging of leading and lagging strand synthesis reveals the coordination dynamics of single replisomes. Mol Cell 64:1035–1047.
93. Elshenawy MM, Jergic S, Xu ZQ, Sobhy MA, Takahashi M, Oakley AJ, Dixon NE, Hamdan SM (2015) Replisome speed determines the efficiency of the Tus-Ter replication termination barrier. Nature 525:394.
94. Sun B, Pandey M, Inman JT, Yang Y, Kashlev M, Patel SS, Wang MD (2015) T7 replisome directly overcomes DNA damage. Nat Commun 6:10260.
95. Fan J, Leroux-Coyau M, Savery NJ, Strick TR (2016) Reconstruction of bacterial transcription-coupled repair at single-molecule resolution. Nature 536:234–237.
96. Paeschke K, Bochman ML, Garcia PD, Cejka P, Friedman KL, Kowalczykowski SC, Zakian VA (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. Nature 497:718–762.
97. Mendoza O, Bourdoncle A, Boule JB, Brosh RM, Jr, Mergny JL (2016) G-quadruplexes and helicases. Nucleic Acids Res 44:1989–2006.
98. Cnossen JP, Dulin D, Dekker NH (2014) An optimized software framework for real-time, high-throughput tracking of spherical beads. Rev Sci Instrum 85: 103712.

ID: jwweb3b2server Time: 06:18 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/PRO#/Vol00000/170076/Comp/APPFile/JW-PRO#170076

Annexe 2

Hairpin construction

A plasmid of 5KBp is linearized using restriction enzymes (BamHI-HF and EcoRI-HF from NEB). The sample is kept at 55°C for an hour. The used restriction enzymes leads to a double stranded DNA of 4 KBp having two 3' overhangs of four bases each. After purifying the DNA fragments using 1% agarose gel electrophoresis, we ligate to the resulting DNA molecule a double-stranded DNA of 30 bp containing the sequence of the G4 and having two 5' overhangs complementary to the mentioned 3' overhangs in order to seal the plasmid. The ligation reaction containing T4 DNA ligase (from NEB) and ATP is kept at room temperature for an hour. It is important to point out that in order to anneal the complementary oligonucleotides of 30 bases before ligating them to the plasmid, we first heat the sample for 2 minutes at 95°C then leave it to cool at room temperature. Then, in order to amplify the plasmid, we perform a transformation in DH5α competent cells (From Invitrogen). After cells culture, we extract the plasmids and we send a sample to sequencing (GATC) in order to ensure having the right G4 sequence. The received sequence is compared to the right sequence using serial cloner software. The plasmids are then cut in two positions using restriction enzymes (BsmBI from NEB) in order to have a double-stranded DNA of 1KBp. After DNA purification using 1% agarose gel electrophoresis, we ligate to the 1 KBp DNA fragment:

1- The oligonucleotide containing the loop (L). This oligonucleotide folds in a stem-loop structure with a 5' single-stranded overhang of four bases complementary to the 3' overhang of the 1 KBp double-stranded fragment. The loop contains 6 bases.
2- The stem of the hairpin that consists of two oligonucleotides that are partially annealed together leaving two single stranded flaps (a 5' overhang (O1) and a 3' overhang (O2)) on the same side and a 5' flap of four bases on the other side in order to be ligated to the 1KBp fragment. The O1 has a biotin on its 5' end that will be used to bind streptavidine coated magnetic beads. And the O2 is partially annealed to another oligonucleotide (T) that will be used as a template by the 3'-5' Klenow polymerase exo- (from NEB) in the fill-in reaction in order to add digoxygenin-dUTP (from Jena Bioscience). The fill-in reaction is done at 37 °C during 30 minutes.

In order to bind the hairpins to the magnetic beads (from Invitrogen), first the beads are washed in the called binding buffer (2x) (annexe 4). The beads are purified by magnetic separation. DNA hairpins are then added to the beads in the binding buffer and kept at room temperature. Twenty minutes later, the beads are washed with Passivation buffer (annexe 4) and suspended in 15 uL of the same buffer.

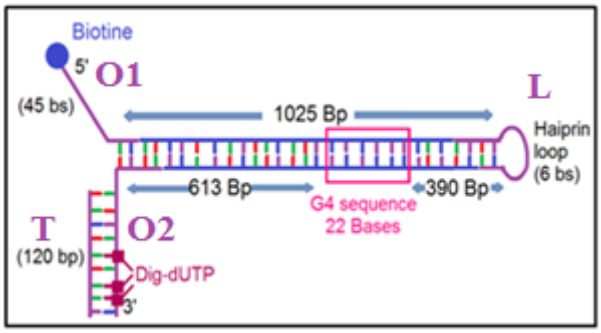


Fig 1: DNA hairpin

Annexe 3

Fluid chamber construction

First, a microscope slide is dipped in a 4M NaOH buffer during one minute. The slide is then rinsed by H_2O and heated until water evaporation. We apply a thin film of sigmacote on one side of the slide in order to have a hydrophobic surface. This treatment is necessary to bind the antibodies to the glass surface. Once the sigmacote becomes dry, we heat the slide again for 30 minutes and we rinse it with water. The cell is then assembled by sandwiching a double-sided tape having an empty channel between the hydrophobic side of the slide and the mylar sheet. The input and output connectors are then sticked to the top of the mylar sheet, after adjusting the connectors holes on the holes made in the mylar located on the entry and the exit of the double-sided tape channnel. The anti-digoxygenin (from Jackson Immuno Research) is then injected into the chamber. One hour later the passivation buffer (Annexe 4) is injected into the chamber and left inside for some hours or overnight. The passivation buffer contains 1mg/mL of BSA protein (Bovine Serum Albumin) that covers the holes between the antibodies and covers the Mylar sheet in order to avoid non-specific binding event during the experimental assays.

Annexe 4

Buffers composition

A) Binding buffer (2x) composition

10 mM Tris-HCl (pH 7.5)

1 mM EDTA

2 M NaCl

B) Passivation buffer

PBS

1 mM EDTA

1 mg/mL BSA

0.1% (w/v) pluronic F127

10 mM azide

C) Reaction buffer

20 mM Tris-HCl (pH 7.5)

3 mM $MgCl_2$

60mM KCl

1 mg/mL BSA

1 mM DTT

1 mM ATP (only for helicases assays)

200 um dNTP (only for polymerases assays)

Annexe 5

Nanometer to Base pair conversion

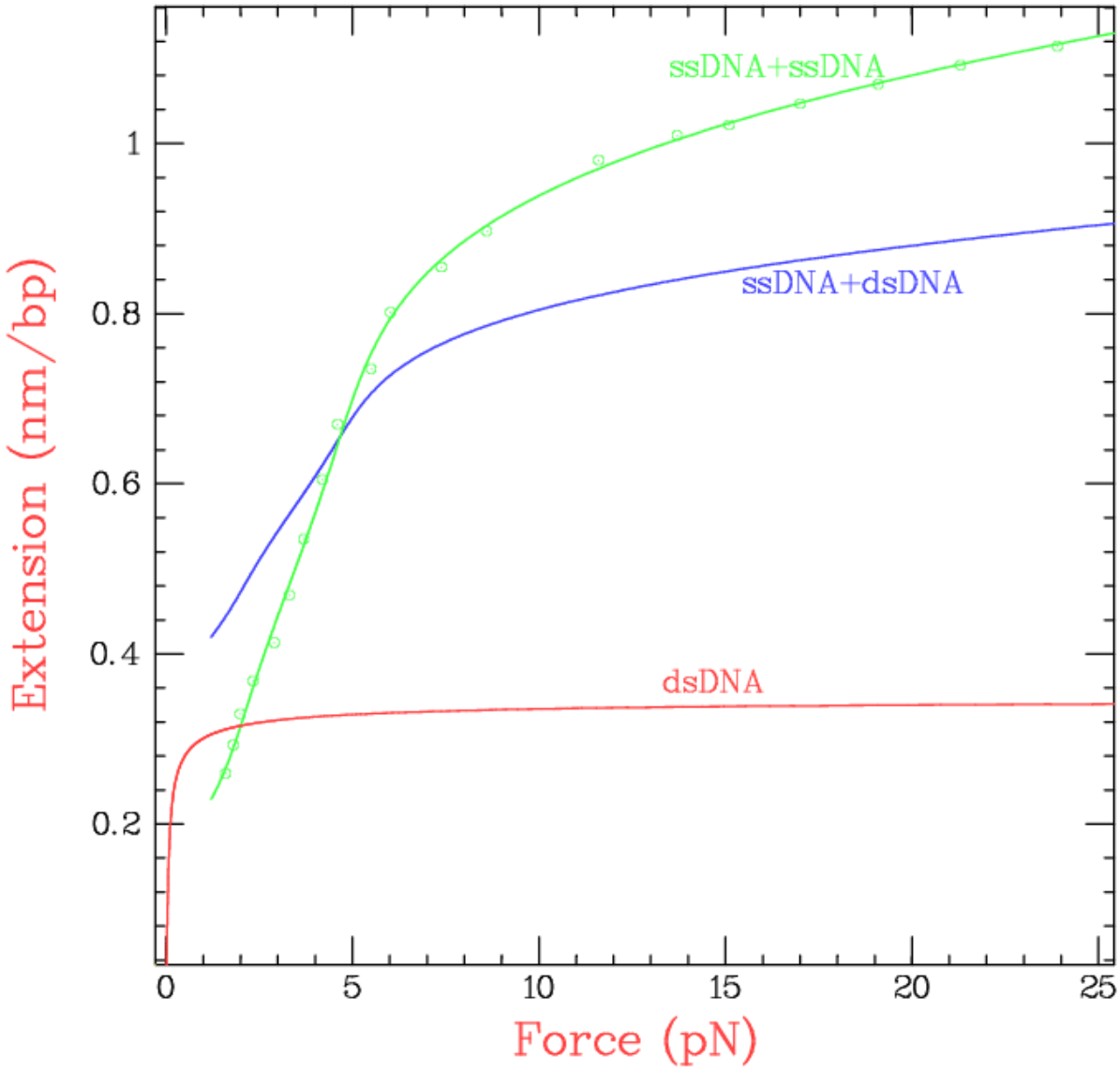


Fig. 1: Nanometer to base pair conversion. *Green curve: Extension of a DNA base pair opened by a helicase that generates two single-stranded bases. Blue curve: Extension of*

a base replicated by a polymerase and its complementary base (on the other strand) released. This curve is used when a polymerase is acting in strand displacement mode. Red curve: It is used when a polymerase is acting in a primer extension mode. A base is replicated and gives a double-stranded base.

References

[1] P. A. Levene, “The Structure of Yeast Nucleic Acid Iv. Ammonia Hydrolysis,” *J. Biol. Chem.*, vol. 40, no. 2, pp. 415–424, Dec. 1919.

[2] O. T. Avery, C. M. MacLeod, and M. McCarty, “Studies on the chemical nature of the substance inducing transformation of pneumococcal types,” *J Exp Med*, vol. 79, no. 2, pp. 137–158, Feb. 1944.

[3] E. Chargaff, “Structure and function of nucleic acids as cell constituents,” *Fed. Proc.*, vol. 10, no. 3, pp. 654–659, Sep. 1951.

[4] E. Chargaff and E. Vischer, “The composition of the desoxypentose nucleic acids of thymus and spleen,” *J. Biol. Chem.*, vol. 177, no. 1, pp. 405–416, Jan. 1949.

[5] L. Pauling and R. B. Corey, “A Proposed Structure For The Nucleic Acids,” *Proc Natl Acad Sci U S A*, vol. 39, no. 2, pp. 84–97, Feb. 1953.

[6] R. E. Franklin and R. G. Gosling, “Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate,” *Nature*, vol. 172, no. 4369, pp. 156–157, Jul. 1953.

[7] J. D. Watson and F. H. C. Crick, “Genetical Implications of the Structure of Deoxyribonucleic Acid,” *Nature*, vol. 171, no. 4361, pp. 964–967, May 1953.

[8] “A structure for deoxyribose nucleic acid.,” *Nature*, vol. 3, pp. 737–738, Apr. 1953.

[9] Bang Ivar, “Untersuchungen über die Guanylsäure. Biochem,” *Biochemische Zeitschrift*, vol. 26, pp. 293–231, 1910.

[10] M. Gellert, M. N. Lipsett, and D. R. Davies, “HELIX FORMATION BY GUANYLIC ACID,” *Proc Natl Acad Sci U S A*, vol. 48, no. 12, pp. 2013–2018, Dec. 1962.

[11] J. R. Williamson, M. K. Raghuraman, and T. R. Cech, “Monovalent cation-induced structure of telomeric DNA: the G-quartet model,” *Cell*, vol. 59, no. 5, pp. 871–880, Dec. 1989.

[12] J. L. Huppert, “Structure, location and interactions of G-quadruplexes,” *FEBS Journal*, vol. 277, no. 17, pp. 3452–3458, Sep. 2010.

[13] E. Largy, J.-L. Mergny, and V. Gabelica, “Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability,” *Met Ions Life Sci*, vol. 16, pp. 203–258, 2016.

[14] W. H. Taylor and P. J. Hagerman, “Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure. II. NaCl-dependence of DNA flexibility and helical repeat,” *J. Mol. Biol.*, vol. 212, no. 2, pp. 363–376, Mar. 1990.

[15] W. Palm and T. de Lange, “How shelterin protects mammalian telomeres,” *Annu. Rev. Genet.*, vol. 42, pp. 301–334, 2008.

[16] H. J. Lipps and D. Rhodes, “G-quadruplex structures: in vivo evidence and function,” *Trends Cell Biol.*, vol. 19, no. 8, pp. 414–422, Aug. 2009.

[17] C. Bouchiat, M. D. Wang, J.-F. Allemand, T. Strick, S. M. Block, and V. Croquette, “Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements,” *Biophysical Journal*, vol. 76, no. 1, pp. 409–413, Jan. 1999.

[18] M. Rief, H. Clausen-Schaumann, and H. E. Gaub, “Sequence-dependent mechanics of single DNA molecules,” *Nat Struct Mol Biol*, vol. 6, no. 4, pp. 346–349, Apr. 1999.

[19] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, “Mechanical separation of the complementary strands of DNA,” *PNAS*, vol. 94, no. 22, pp. 11935–11940, Oct. 1997.

[20] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, “Molecular Stick-Slip Motion Revealed by Opening DNA with Piconewton Forces,” *Phys. Rev. Lett.*, vol. 79, no. 22, pp. 4489–4492, Dec. 1997.

[21] D. D. Leipe, Y. I. Wolf, E. V. Koonin, and L. Aravind, “Classification and evolution of P-loop GTPases and related ATPases,” *J. Mol. Biol.*, vol. 317, no. 1, pp. 41–72, Mar. 2002.

[22] V. Anantharaman, E. V. Koonin, and L. Aravind, “Comparative genomics and evolution of proteins involved in RNA metabolism,” *Nucleic Acids Res*, vol. 30, no. 7, pp. 1427–1464, Apr. 2002.

[23] A. M. Pyle, “Translocation and unwinding mechanisms of RNA and DNA helicases,” *Annu Rev Biophys*, vol. 37, pp. 317–336, 2008.

[24] E. Jankowsky and M. E. Fairman, “RNA helicases--one fold for many functions,” *Curr. Opin. Struct. Biol.*, vol. 17, no. 3, pp. 316–324, Jun. 2007.

[25] O. Cordin, J. Banroques, N. K. Tanner, and P. Linder, “The DEAD-box protein family of RNA helicases,” *Gene*, vol. 367, pp. 17–37, Feb. 2006.

[26] G. Kadaré and A. L. Haenni, “Virus-encoded RNA helicases.,” *J Virol*, vol. 71, no. 4, pp. 2583–2590, Apr. 1997.

[27] P. umate, N. Tuteja, and R. Tuteja, “Genome-wide comprehensive analysis of human helicases,” *Commun Integr Biol*, vol. 4, no. 1, pp. 118–137, Jan. 2011.

[28] M. Abdelhaleem, “Do human RNA helicases have a role in cancer?,” *Biochim. Biophys. Acta*, vol. 1704, no. 1, pp. 37–46, Jul. 2004.

[29] K. Hanada and I. D. Hickson, “Molecular genetics of RecQ helicase disorders,” *Cell. Mol. Life Sci.*, vol. 64, no. 17, pp. 2306–2322, Sep. 2007.

[30] M. R. Singleton, M. S. Dillingham, and D. B. Wigley, “Structure and mechanism of helicases and nucleic acid translocases,” *Annu. Rev. Biochem.*, vol. 76, pp. 23–50, 2007.

[31] A. E. Gorbalenya and E. V. Koonin, “Helicases: amino acid sequence comparisons and structure-function relationships,” *Current Opinion in Structural Biology*, vol. 3, no. 3, pp. 419–429, Jun. 1993.

[32] A. E. Gorbalenya, E. V. Koonin, A. P. Donchenko, and V. M. Blinov, “A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination,” *FEBS Letters*, vol. 235, no. 1, pp. 16–24, Aug. 1988.

[33] A. M. Olovnikov, “[Principle of marginotomy in template synthesis of polynucleotides],” *Dokl. Akad. Nauk SSSR*, vol. 201, no. 6, pp. 1496–1499, 1971.

[34] R. C. Allsopp *et al.*, “Telomere length predicts replicative capacity of human fibroblasts,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 89, no. 21, pp. 10114–10118, Nov. 1992.

[35] C. I. Nugent and V. Lundblad, “The telomerase reverse transcriptase: components and regulation,” *Genes Dev.*, vol. 12, no. 8, pp. 1073–1085, Apr. 1998.

[36] A. M. Olovnikov, “A theory of marginotomy,” *Journal of Theoretical Biology*, vol. 41, no. 1, pp. 181–190, Sep. 1973.

[37] J. L. Osterhage and K. L. Friedman, “Chromosome End Maintenance by Telomerase,” *J Biol Chem*, vol. 284, no. 24, pp. 16061–16065, Jun. 2009.

[38] S. B. Smith, Y. Cui, and C. Bustamante, “Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules,” *Science*, vol. 271, no. 5250, pp. 795–799, Feb. 1996.

[39] C. Gosse and V. Croquette, “Magnetic tweezers: micromanipulation and force measurement at the molecular level.,” *Biophys J*, vol. 82, no. 6, pp. 3314–3329, Jun. 2002.
[40] S. B. Smith, L. Finzi, and C. Bustamante, “Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads,” *Science*, vol. 258, no. 5085, pp. 1122–1126, Nov. 1992.
[41] T. R. Strick, J. F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, “The elasticity of a single supercoiled DNA molecule,” *Science*, vol. 271, no. 5257, pp. 1835–1837, Mar. 1996.
[42] S. Hodeib *et al.*, “Single molecule studies of helicases with magnetic tweezers,” *Methods*, vol. 105, pp. 3–15, Aug. 2016.
[43] N. Beaume *et al.*, “Genome-wide study predicts promoter-G4 DNA motifs regulate selective functions in bacteria: radioresistance of D. radiodurans involves G4 DNA-mediated regulation,” *Nucleic Acids Res.*, vol. 41, no. 1, pp. 76–89, Jan. 2013.
[44] W. I. Sundquist and S. Heaphy, “Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA.,” *Proc Natl Acad Sci U S A*, vol. 90, no. 8, pp. 3393–3397, Apr. 1993.
[45] J. Norseen, F. B. Johnson, and P. M. Lieberman, “Role for G-Quadruplex RNA Binding by Epstein-Barr Virus Nuclear Antigen 1 in DNA Replication and Metaphase Chromosome Attachment,” *J Virol*, vol. 83, no. 20, pp. 10336–10346, Oct. 2009.
[46] B. Tuesuwan *et al.*, “Simian Virus 40 Large T-Antigen G-Quadruplex DNA Helicase Inhibition by G-Quadruplex DNA-Interactive Agents,” *Biochemistry*, vol. 47, no. 7, pp. 1896–1909, Feb. 2008.
[47] H. J. Lipps, W. Gruissem, and D. M. Prescott, “Higher order DNA structure in macronuclear chromatin of the hypotrichous ciliate Oxytricha nova.,” *Proc Natl Acad Sci U S A*, vol. 79, no. 8, pp. 2495–2499, Apr. 1982.
[48] W. I. Sundquist and A. Klug, “Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops,” *Nature*, vol. 342, no. 6251, pp. 825–829, Dec. 1989.
[49] C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H. J. Lipps, and A. Plückthun, “In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, no. 15, pp. 8572–8577, Jul. 2001.
[50] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, and L. H. Hurley, “Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription,” *PNAS*, vol. 99, no. 18, pp. 11593–11598, Sep. 2002.
[51] V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, and S. Balasubramanian, “High-throughput sequencing of DNA G-quadruplex structures in the human genome,” *Nat. Biotechnol.*, vol. 33, no. 8, pp. 877–881, Aug. 2015.
[52] J. L. Huppert and S. Balasubramanian, “Prevalence of quadruplexes in the human genome,” *Nucleic Acids Res.*, vol. 33, no. 9, pp. 2908–2916, 2005.
[53] A. K. Todd, M. Johnston, and S. Neidle, “Highly prevalent putative quadruplex sequence motifs in human DNA,” *Nucleic Acids Res.*, vol. 33, no. 9, pp. 2901–2907, 2005.
[54] J. L. Huppert and S. Balasubramanian, “G-quadruplexes in promoters throughout the human genome,” *Nucleic Acids Res.*, vol. 35, no. 2, pp. 406–413, 2007.
[55] A. Bugaut and S. Balasubramanian, “5’-UTR RNA G-quadruplexes: translation regulation and targeting,” *Nucleic Acids Res*, vol. 40, no. 11, pp. 4727–4741, Jun. 2012.
[56] Y. Xu, Y. Suzuki, K. Ito, and M. Komiyama, “Telomeric repeat-containing RNA structure in living cells,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 33, pp. 14579–14584, Aug. 2010.
[57] K. J. Woodford, R. M. Howell, and K. Usdin, “A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes.,” *J. Biol. Chem.*, vol. 269, no. 43, pp. 27029–27035, Oct. 1994.
[58] E. Kruisselbrink, V. Guryev, K. Brouwer, D. B. Pontier, E. Cuppen, and M. Tijsterman, “Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans,” *Curr. Biol.*, vol. 18, no. 12, pp. 900–905, Jun. 2008.

[59] B. Lemmens, R. van Schendel, and M. Tijsterman, “Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers,” *Nature Communications*, vol. 6, p. 8909, Nov. 2015.
[60] C. E. Nesbit, J. M. Tersak, and E. V. Prochownik, “MYC oncogenes and human neoplastic disease,” *Oncogene*, vol. 18, no. 19, pp. 3004–3016, May 1999.
[61] S. J. Berberich and E. H. Postel, “PuF/NM23-H2/NDPK-B transactivates a human c-myc promoter-CAT gene via a functional nuclease hypersensitive element,” *Oncogene*, vol. 10, no. 12, pp. 2343–2347, Jun. 1995.
[62] T. Simonsson, P. Pecinka, and M. Kubista, “DNA tetraplex formation in the control region of c-myc.,” *Nucleic Acids Res*, vol. 26, no. 5, pp. 1167–1172, Mar. 1998.
[63] C. L. Grand *et al.*, “The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo,” *Mol. Cancer Ther.*, vol. 1, no. 8, pp. 565–573, Jun. 2002.
[64] J. Seenisamy *et al.*, “The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4,” *J. Am. Chem. Soc.*, vol. 126, no. 28, pp. 8702–8709, Jul. 2004.
[65] A. Ambrus, D. Chen, J. Dai, R. A. Jones, and D. Yang, “Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human c-MYC Promoter. Implications for G-Quadruplex Stabilization,” *Biochemistry*, vol. 44, no. 6, pp. 2048–2058, Feb. 2005.
[66] C. Granotier *et al.*, “Preferential binding of a G-quadruplex ligand to human chromosome ends,” *Nucleic Acids Res*, vol. 33, no. 13, pp. 4182–4190, 2005.
[67] Y. Wang and D. J. Patel, “Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex,” *Structure*, vol. 1, no. 4, pp. 263–282, Dec. 1993.
[68] G. N. Parkinson, M. P. H. Lee, and S. Neidle, “Crystal structure of parallel quadruplexes from human telomeric DNA,” *Nature*, vol. 417, no. 6891, pp. 876–880, Jun. 2002.
[69] J. Dai, M. Carver, C. Punchihewa, R. A. Jones, and D. Yang, “Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence,” *Nucleic Acids Res.*, vol. 35, no. 15, pp. 4927–4940, 2007.
[70] A. T. Phan, V. Kuryavyi, K. N. Luu, and D. J. Patel, “Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K+ solution†,” *Nucleic Acids Res*, vol. 35, no. 19, pp. 6517–6525, Oct. 2007.
[71] A. Bugaut and P. Alberti, “Understanding the stability of DNA G-quadruplex units in long human telomeric strands,” *Biochimie*, vol. 113, pp. 125–133, Jun. 2015.
[72] A. J. Zaug, E. R. Podell, and T. R. Cech, “Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 31, pp. 10864–10869, Aug. 2005.
[73] T. R. Salas *et al.*, “Human replication protein A unfolds telomeric G-quadruplexes,” *Nucleic Acids Res*, vol. 34, no. 17, pp. 4857–4865, Oct. 2006.
[74] A. T. Phan, Y. S. Modi, and D. J. Patel, “Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter,” *J. Am. Chem. Soc.*, vol. 126, no. 28, pp. 8710–8716, Jul. 2004.
[75] J. Dai, D. Chen, R. A. Jones, L. H. Hurley, and D. Yang, “NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region,” *Nucleic Acids Res.*, vol. 34, no. 18, pp. 5133–5144, 2006.
[76] F. Hamon *et al.*, “An Acyclic Oligoheteroaryle That Discriminates Strongly between Diverse G-Quadruplex Topologies,” *Angew. Chem. Int. Ed.*, vol. 50, no. 37, pp. 8745–8749, Sep. 2011.
[77] D. Sun *et al.*, “Inhibition of human telomerase by a G-quadruplex-interactive compound,” *J. Med. Chem.*, vol. 40, no. 14, pp. 2113–2116, Jul. 1997.
[78] J. L. Li *et al.*, “Inhibition of the Bloom’s and Werner’s syndrome helicases by G-quadruplex interacting ligands,” *Biochemistry*, vol. 40, no. 50, pp. 15194–15202, Dec. 2001.

[79] R. Halder, J.-F. Riou, M.-P. Teulade-Fichou, T. Frickey, and J. S. Hartig, “Bisquinolinium compounds induce quadruplex-specific transcriptome changes in HeLa S3 cell lines,” *BMC Res Notes*, vol. 5, p. 138, Mar. 2012.

[80] D. Koirala *et al.*, “A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands,” *Nat Chem*, vol. 3, no. 10, pp. 782–787, Aug. 2011.

[81] P. Wang, L. Ren, H. He, F. Liang, X. Zhou, and Z. Tan, “A phenol quaternary ammonium porphyrin as a potent telomerase inhibitor by selective interaction with quadruplex DNA,” *Chembiochem*, vol. 7, no. 8, pp. 1155–1159, Aug. 2006.

[82] G. W. Collie and G. N. Parkinson, “The application of DNA and RNA G-quadruplexes to therapeutic medicines,” *Chem Soc Rev*, vol. 40, no. 12, pp. 5867–5892, Dec. 2011.

[83] W. C. Drosopoulos, S. T. Kosiyatrakul, and C. L. Schildkraut, “BLM helicase facilitates telomere replication during leading strand synthesis of telomeres,” *J. Cell Biol.*, vol. 210, no. 2, pp. 191–208, Jul. 2015.

[84] A. De Cian, E. DeLemos, J.-L. Mergny, M.-P. Teulade-Fichou, and D. Monchaud, “Highly Efficient G-Quadruplex Recognition by Bisquinolinium Compounds,” *J. Am. Chem. Soc.*, vol. 129, no. 7, pp. 1856–1857, Feb. 2007.

[85] D. Monchaud *et al.*, “Ligands playing musical chairs with G-quadruplex DNA: a rapid and simple displacement assay for identifying selective G-quadruplex binders,” *Biochimie*, vol. 90, no. 8, pp. 1207–1223, Aug. 2008.

[86] A. Piazza *et al.*, “Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae,” *Nucleic Acids Res.*, vol. 38, no. 13, pp. 4337–4348, Jul. 2010.

[87] W. J. Chung, B. Heddi, F. Hamon, M.-P. Teulade-Fichou, and A. T. Phan, “Solution Structure of a G-quadruplex Bound to the Bisquinolinium Compound Phen-DC3,” *Angew. Chem.*, vol. 126, no. 4, pp. 1017–1020, Jan. 2014.

[88] M. Bončina *et al.*, “Thermodynamic fingerprints of ligand binding to human telomeric G-quadruplexes,” *Nucleic Acids Res*, vol. 43, no. 21, pp. 10376–10386, Dec. 2015.

[89] M. C. Chen, P. Murat, K. Abecassis, A. R. Ferré-D’Amaré, and S. Balasubramanian, “Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase,” *Nucleic Acids Res.*, vol. 43, no. 4, pp. 2223–2231, Feb. 2015.

[90] E. Largy, J.-L. Mergny, and V. Gabelica, “Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability,” *Met Ions Life Sci*, vol. 16, pp. 203–258, 2016.

[91] Y. Zhao, Z. Kan, Z. Zeng, Y. Hao, H. Chen, and Z. Tan, “Determining the folding and unfolding rate constants of nucleic acids by biosensor. Application to telomere G-quadruplex,” *J. Am. Chem. Soc.*, vol. 126, no. 41, pp. 13255–13264, Oct. 2004.

[92] H. Qin, J. Ren, J. Wang, N. W. Luedtke, and E. Wang, “G-quadruplex-modulated fluorescence detection of potassium in the presence of a 3500-fold excess of sodium ions,” *Anal. Chem.*, vol. 82, no. 19, pp. 8356–8360, Oct. 2010.

[93] C. Saintomé, S. Amrane, J.-L. Mergny, and P. Alberti, “The exception that confirms the rule: a higher-order telomeric G-quadruplex structure more stable in sodium than in potassium,” *Nucleic Acids Res.*, vol. 44, no. 6, pp. 2926–2935, Apr. 2016.

[94] Y.-Y. Yan *et al.*, “Selective recognition of oncogene promoter G-quadruplexes by Mg2+,” *Biochem. Biophys. Res. Commun.*, vol. 402, no. 4, pp. 614–618, Nov. 2010.

[95] H. You, J. Wu, F. Shao, and J. Yan, “Stability and kinetics of c-MYC promoter G-quadruplexes studied by single-molecule manipulation,” *J. Am. Chem. Soc.*, vol. 137, no. 7, pp. 2424–2427, Feb. 2015.

[96] H. You *et al.*, “Dynamics and stability of polymorphic human telomeric G-quadruplex under tension,” *Nucleic Acids Res.*, vol. 42, no. 13, pp. 8789–8795, Jul. 2014.

[97] C. Ribeyre *et al.*, “The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo,” *PLoS Genet.*, vol. 5, no. 5, p. e1000475, May 2009.

[98] H. Sun, J. K. Karow, I. D. Hickson, and N. Maizels, “The Bloom’s syndrome helicase unwinds G4 DNA,” *J. Biol. Chem.*, vol. 273, no. 42, pp. 27587–27592, Oct. 1998.

[99] J. B. Budhathoki, S. Ray, V. Urban, P. Janscak, J. G. Yodh, and H. Balci, “RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP,” *Nucleic Acids Res.*, vol. 42, no. 18, pp. 11528–11545, Oct. 2014.

[100] O. Mendoza, N. M. Gueddouda, J.-B. Boulé, A. Bourdoncle, and J.-L. Mergny, “A fluorescence-based helicase assay: application to the screening of G-quadruplex ligands,” *Nucleic Acids Res.*, vol. 43, no. 11, p. e71, Jun. 2015.

[101] S. D. Creacy, E. D. Routh, F. Iwamoto, Y. Nagamine, S. A. Akman, and J. P. Vaughn, “G4 Resolvase 1 Binds Both DNA and RNA Tetramolecular Quadruplex with High Affinity and Is the Major Source of Tetramolecular Quadruplex G4-DNA and G4-RNA Resolving Activity in HeLa Cell Lysates,” *J Biol Chem*, vol. 283, no. 50, pp. 34626–34634, Dec. 2008.

[102] J.-B. Boulé and V. A. Zakian, “The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates,” *Nucleic Acids Res.*, vol. 35, no. 17, pp. 5809–5818, 2007.

[103] K. Paeschke *et al.*, “Pif1 family helicases suppress genome instability at G-quadruplex motifs,” *Nature*, vol. 497, no. 7450, pp. 458–462, May 2013.

[104] R. Zhou, J. Zhang, M. L. Bochman, V. A. Zakian, and T. Ha, “Periodic DNA patrolling underlies diverse functions of Pif1 on R-loops and G-rich DNA,” *eLife*, vol. 3, p. e02190, Apr. 2014.

[105] X.-L. Duan *et al.*, “G-Quadruplexes Significantly Stimulate Pif1 Helicase-catalyzed Duplex DNA Unwinding,” *J. Biol. Chem.*, p. jbc.M114.628008, Jan. 2015.

[106] M. D. Huber, M. L. Duquette, J. C. Shiels, and N. Maizels, “A Conserved G4 DNA Binding Domain in RecQ Family Helicases,” *Journal of Molecular Biology*, vol. 358, no. 4, pp. 1071–1080, May 2006.

[107] L. Crabbe, R. E. Verdun, C. I. Haggblom, and J. Karlseder, “Defective telomere lagging strand synthesis in cells lacking WRN helicase activity,” *Science*, vol. 306, no. 5703, pp. 1951–1953, Dec. 2004.

[108] S. Chatterjee *et al.*, “Mechanistic insight into the interaction of BLM helicase with intra-strand G-quadruplex structures,” *Nat Commun*, vol. 5, p. 5556, Nov. 2014.

[109] L. S. Kaguni and D. A. Clayton, “Template-directed pausing in in vitro DNA synthesis by DNA polymerase a from Drosophila melanogaster embryos,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 79, no. 4, pp. 983–987, Feb. 1982.

[110] M. N. Weitzmann, K. J. Woodford, and K. Usdin, “The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation,” *J. Biol. Chem.*, vol. 271, no. 34, pp. 20958–20964, Aug. 1996.

[111] A. S. Kamath-Loeb, L. A. Loeb, E. Johansson, P. M. Burgers, and M. Fry, “Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)n trinucleotide repeat sequence,” *J. Biol. Chem.*, vol. 276, no. 19, pp. 16439–16446, May 2001.

[112] D. Schiavone *et al.*, “PrimPol Is Required for Replicative Tolerance of G Quadruplexes in Vertebrate Cells,” *Mol. Cell*, vol. 61, no. 1, pp. 161–169, Jan. 2016.

[113] S. Prakash, R. E. Johnson, and L. Prakash, “Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function,” *Annu. Rev. Biochem.*, vol. 74, pp. 317–353, 2005.

[114] D. N. Edwards, A. Machwe, Z. Wang, and D. K. Orren, “Intramolecular Telomeric G-Quadruplexes Dramatically Inhibit DNA Synthesis by Replicative and Translesion Polymerases, Revealing their Potential to Lead to Genetic Change,” *PLoS One*, vol. 9, no. 1, Jan. 2014.

[115] L. Safa *et al.*, “5’ to 3’ unfolding directionality of DNA secondary structures by replication protein A: G-quadruplexes and duplexes.,” *J. Biol. Chem.*, p. jbc.M115.709667, Jul. 2016.

[116] G. G. Oakley and S. M. Patrick, “Replication protein A: directing traffic at the intersection of replication and repair,” *Front Biosci (Landmark Ed)*, vol. 15, pp. 883–900, Jun. 2010.

[117] N. Arnoult, C. Saintome, I. Ourliac-Garnier, J.-F. Riou, and A. Londoño-Vallejo, “Human POT1 is required for efficient telomere C-rich strand replication in the absence of WRN,” *Genes Dev*, vol. 23, no. 24, pp. 2915–2924, Dec. 2009.

[118] S. Liu *et al.*, “Distinct roles for DNA-PK, ATM and ATR in RPA phosphorylation and checkpoint activation in response to replication stress,” *Nucleic Acids Res.*, vol. 40, no. 21, pp. 10780–10794, Nov. 2012.
[119] Y. Kobayashi *et al.*, “Expression of mutant RPA in human cancer cells causes telomere shortening,” *Biosci. Biotechnol. Biochem.*, vol. 74, no. 2, pp. 382–385, 2010.
[120] S. Cohen, E. Jacob, and H. Manor, “Effects of single-stranded DNA binding proteins on primer extension by telomerase,” *Biochim. Biophys. Acta*, vol. 1679, no. 2, pp. 129–140, Aug. 2004.
[121] A. Prakash, F. Kieken, L. A. Marky, and G. E. O. Borgstahl, “Stabilization of a G-Quadruplex from Unfolding by Replication Protein A Using Potassium and the Porphyrin TMPyP4,” *J Nucleic Acids*, vol. 2011, p. 529828, 2011.
[122] M. H. Qureshi, S. Ray, A. L. Sewell, S. Basu, and H. Balci, “Replication protein A unfolds G-quadruplex structures with varying degrees of efficiency,” *J Phys Chem B*, vol. 116, no. 19, pp. 5588–5594, May 2012.
[123] S. Ray, M. H. Qureshi, D. W. Malcolm, J. B. Budhathoki, U. Celik, and H. Balci, “RPA-mediated unfolding of systematically varying G-quadruplex structures,” *Biophys. J.*, vol. 104, no. 10, pp. 2235–2245, May 2013.
[124] J. Gao *et al.*, “Yeast transcription co-activator Sub1 and its human homolog PC4 preferentially bind to G-quadruplex DNA,” *Chem. Commun. (Camb.)*, vol. 51, no. 33, pp. 7242–7244, Apr. 2015.
[125] N. V. Hud, F. W. Smith, F. A. Anet, and J. Feigon, “The selectivity for K+ versus Na+ in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by 1H NMR,” *Biochemistry*, vol. 35, no. 48, pp. 15383–15390, Dec. 1996.
[126] B. G. Kim, H. M. Evans, D. N. Dubins, and T. V. Chalikian, “Effects of Salt on the Stability of a G-Quadruplex from the Human c-MYC Promoter,” *Biochemistry*, vol. 54, no. 22, pp. 3420–3430, Jun. 2015.
[127] A. De Rache and J.-L. Mergny, “Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay,” *Biochimie*, vol. 115, pp. 194–202, Aug. 2015.
[128] E. Hatzakis, K. Okamoto, and D. Yang, “Thermodynamic stability and folding kinetics of the major G-quadruplex and its loop isomers formed in the nuclease hypersensitive element in the human c-Myc promoter: effect of loops and flanking segments on the stability of parallel-stranded intramolecular G-quadruplexes,” *Biochemistry*, vol. 49, no. 43, pp. 9152–9160, Nov. 2010.
[129] C. M. Olsen, W. H. Gmeiner, and L. A. Marky, “Unfolding of G-quadruplexes: energetic, and ion and water contributions of G-quartet stacking,” *J Phys Chem B*, vol. 110, no. 13, pp. 6962–6969, Apr. 2006.
[130] P. S. Shirude, L. Ying, and S. Balasubramanian, “Single molecule conformational analysis of the biologically relevant DNA G-quadruplex in the promoter of the proto-oncogene c-MYC,” *Chem Commun (Camb)*, no. 17, pp. 2007–2009, May 2008.
[131] A. K. Byrd and K. D. Raney, “A parallel quadruplex DNA is bound tightly but unfolded slowly by pif1 helicase,” *J. Biol. Chem.*, vol. 290, no. 10, pp. 6482–6494, Mar. 2015.
[132] S. Eddy, A. Ketkar, M. K. Zafar, L. Maddukuri, J.-Y. Choi, and R. L. Eoff, “Human Rev1 polymerase disrupts G-quadruplex DNA,” *Nucleic Acids Res.*, vol. 42, no. 5, pp. 3272–3285, Mar. 2014.
[133] J. Brčić and J. Plavec, “Solution structure of a DNA quadruplex containing ALS and FTD related GGGGCC repeat stabilized by 8-bromodeoxyguanosine substitution,” *Nucleic Acids Res.*, vol. 43, no. 17, pp. 8590–8600, Sep. 2015.
[134] N. Hayashi and S. Murakami, “STM1, a gene which encodes a guanine quadruplex binding protein, interacts with CDC13 in Saccharomyces cerevisiae,” *Mol. Genet. Genomics*, vol. 267, no. 6, pp. 806–813, Aug. 2002.
[135] E. Tosoni *et al.*, “Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription,” *Nucleic Acids Res*, vol. 43, no. 18, pp. 8884–8897, Oct. 2015.

[136] H. Han, C. L. Cliff, and L. H. Hurley, “Accelerated Assembly of G-Quadruplex Structures by a Small Molecule,” *Biochemistry*, vol. 38, no. 22, pp. 6981–6986, Jun. 1999.
[137] A. De Cian and J.-L. Mergny, “Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation,” *Nucleic Acids Res.*, vol. 35, no. 8, pp. 2483–2493, 2007.
[138] R. Giraldo and D. Rhodes, “The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA.,” *EMBO J*, vol. 13, no. 10, pp. 2411–2420, May 1994.
[139] M. M. Elshenawy *et al.*, “Replisome speed determines the efficiency of the Tus-Ter replication termination barrier,” *Nature*, vol. 525, no. 7569, pp. 394–398, Sep. 2015.
[140] M. K. Gupta *et al.*, “Protein-DNA complexes are the primary sources of replication fork pausing in Escherichia coli,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, no. 18, pp. 7252–7257, Apr. 2013.
[141] S. K. Perumal, K. D. Raney, and S. J. Benkovic, “Analysis of the DNA translocation and unwinding activities of T4 phage helicases,” *Methods*, vol. 51, no. 3, pp. 277–288, Jul. 2010.
[142] R. Ramanagoudr-Bhojappa *et al.*, “Yeast Pif1 helicase exhibits a one-base-pair stepping mechanism for unwinding duplex DNA,” *J. Biol. Chem.*, vol. 288, no. 22, pp. 16185–16195, May 2013.
[143] J.-H. Li *et al.*, “Pif1 is a force-regulated helicase,” *Nucleic Acids Res.*, vol. 44, no. 9, pp. 4330–4339, May 2016.
[144] Shubeena Chib, “Investigation into the biochemical mechanism of nucleic acid unwinding and protein displacement by Saccharomyces cerevisiae helicase Pif1,” University of Arkansas for Medical Sciences, USA, 2016.
[145] D. A. Bernstein, M. C. Zittel, and J. L. Keck, “High-resolution structure of the E.coli RecQ helicase catalytic core,” *EMBO J.*, vol. 22, no. 19, pp. 4910–4921, Oct. 2003.
[146] A. Piazza *et al.*, “Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites,” *EMBO J.*, vol. 34, no. 12, pp. 1718–1734, Jun. 2015.
[147] B. Lemmens, R. van Schendel, and M. Tijsterman, “Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers,” *Nature Communications*, vol. 6, p. 8909, Nov. 2015.
[148] H. You, J. Wu, F. Shao, and J. Yan, “Stability and kinetics of c-MYC promoter G-quadruplexes studied by single-molecule manipulation,” *J. Am. Chem. Soc.*, vol. 137, no. 7, pp. 2424–2427, Feb. 2015.
[149] E. Noble, M. M. Spiering, and S. J. Benkovic, “Coordinated DNA Replication by the Bacteriophage T4 Replisome,” *Viruses*, vol. 7, no. 6, pp. 3186–3200, Jun. 2015.

Résumé

Les structures G-quadruplexes (G4) sont considérées comme des obstacles qui s'opposent à la progression du réplisome. Les séquences capables de former des G4 dans le génome humain se trouvent dans les régions d'ADN double brin au niveau des oncogènes et des proto-oncongènes et sur l'extrémité simple brin des télomères. La plupart des études biochimiques et biophysiques ont caractérisé les propriétés thermodynamiques des G4 en utilisant par exemple la température de fusion Tm pour déduire la thermodynamique de la formation/résolution du G4. Cependant, les expériences en solution donnent seulement une information indirecte concernant la dynamique du G4. Dans ce travail de thèse en molécule unique utilisant la technique des pinces magnétiques, nous avons pu caractériser la cinétique de la formation et résolution des G4s ainsi que la stabilité d'une structure G4 insérée dans une région d'ADN double brin : Une situation qui ressemble aux G4 dans les promoteurs de gènes, où la séquence complémentaire est en compétition avec la formation de la structure de G4. Nous avons trouvé que le G4 télomérique a une très courte durée de vie (~20 s) et donc ce G4 se résout sans qu'une hélicase soit nécessaire. Au contraire, ce n'est pas le cas pour le G4 du c-MYC qui est très stable (~2h). Nous avons observé en temps réel la collision entre les hélicases et les polymérases et le G4 du c-MYC. Nous avons trouvé que l'hélicase Pif1 ouvre l'ADN puis résout le G4 après avoir effectué une pause et reprend l'ouverture de l'ADN, alors que l'hélicase RecQ et l'hélicase réplicative du bactériophage T4 ne peuvent pas le résoudre, mais peuvent le sauter. Nous avons aussi trouvé que la RPA ne peut pas résoudre le G4 du c-MYC. D'autre part, nous avons observé que la polyémrase du virus T4, la gp43, ainsi que la polymérase de T7, et la polymérase ε de la levure peuvent répliquer le G4 du c-MYC qui de façon étonnante ne constitue pas une barrière infranchissable.

Mots clés

G-quadruplexes, hélicase, polymérase, pinces magnétiques, ADN, réplication

Abstract

G-quadruplex (G4) structures are considered as the major impediments for the replisome progression. The putative G4 forming sequences in the human genome are mostly located in the double-stranded DNA regions of oncogenes and proto-oncogenes and on the single-stranded overhangs of telomeres. Most of the biochemical and biophysical studies have characterized the G4 thermodynamics properties using melting temperature Tm as a proxy to infer thermodynamics of G4 folding/unfolding energetic. However, these thermodynamics properties give only indirect information about G4 dynamics. In this work, using single molecule magnetic tweezers technique, we first characterize the kinetics of folding and unfolding and thus the stability of a single G4 inserted in a dsDNA: a situation that mimics the G4s in promoters, where the complementary sequence competes with the G-rich structure. We find that the lifetime of telomeric G4 is short (~20 s) and thus that this G4 unfolds without the need of a helicase. This is not the case for the very stable c-MYC G4 (~2 hr). We observe in real time how helicases or polymerases behave as they collide with the c-MYC G4 on their track. We find that the Pif1 helicase unwinds dsDNA, resolves this G4 after pausing and resume unwinding, while RecQ helicase and the bacteriophage T4 replicative helicase do not resolve the G4 but may jump it. We also find that RPA does not unfold the c-MYC G4. Besides, we find that T4 bacteriophage gp43 polymerase, T7 polymerase and Yeast Pol ε can replicate the G4 which surprisingly does not appear as a major roadblock for them.

Keywords

G-quadruplexes, helicase, polymerase, magnetic tweezers, DNA, replication