Enhanced interrogation of DNA roadblocks and RNA polymerases
Pulling information fkom the individual to understand the masses

Berghuis, Bojk

DOI
10.4233/uuid:40d6eb23-6d6a-40a5-b814-4fd92ea8e1d9

Publication date
2016

Document Version
Final published version

Citation (APA)
Berghuis, B. (2016). *Enhanced interrogation of DNA roadblocks and RNA polymerases: Pulling information fkom the individual to understand the masses*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:40d6eb23-6d6a-40a5-b814-4fd92ea8e1d9

Enhanced interrogation of DNA roadblocks and RNA polymerases

Pulling information from the individual to understand the masses

Bojk A.
Berghuis

Propositions

accompanying the dissertation

ENHANCED INTERROGATION OF DNA ROADBLOCKS AND RNA POLYMERASES

PULLING INFORMATION FROM THE INDIVIDUAL TO UNDERSTAND THE MASSES

by

BOJAN ARTHUR BERGHUIS

1. A single protein can specifically block DNA replication in one direction, but not the other because the timescales at which small molecular rearrangements take place are vastly shorter than the timescales required to replicate a DNA base pair. (Chapters 8 and 9 of this thesis)
2. The dwell time distribution is a way to fingerprint different classes of polymerases. (Chapter 6 of this thesis)
3. The field of single-molecule biophysics is extremely fascinating and has proven its utility, but has not yet lead to fundamental shifts in everyday thinking as discoveries in other fields have done.
4. Roughly two character traits allow one to become a leading figure in science: relentless ambition and the possession of an unstoppable curiosity. Reality teaches us that having a bit of both traits is necessary for reaching the top of today's scientific food chain. The balance between the two traits determines whether someone is viewed as a shark, or as a source of inspiration.
5. Statistics should be at the basis of the teaching curriculum in mathematics, starting at primary school, together with the basics of programming.
6. The Belgian educational system is orders of magnitude better than the Dutch in preparation for a university study because of the rigorous focus on factual and theoretical knowledge. The Dutch university system however, is a much better preparation for life thereafter as a result of the focus on practical skills.
7. An experimental scientist needs mostly two things: common sense and pragmatism. Common sense is however – contrary to what the term suggests – not all that common.
8. Accuracy is often confused with precision: being very precise about every single step in an experiment creates a false sense of accuracy concerning the experimental outcome.
9. Life's path is stochastic, and success strongly correlates with how aware you are of this as well as your ability to assess your options.
10. The selfie stick is the embodiment of (everything that is wrong with) modern society.

These propositions are regarded as opposable and defendable, and have been approved as such by the promotor prof. dr. N.H. Dekker.

Stellingen

behorende bij het proefschrift

ENHANCED INTERROGATION OF DNA ROADBLOCKS AND RNA POLYMERASES

PULLING INFORMATION FROM THE INDIVIDUAL TO UNDERSTAND THE MASSES

door

BOJAN ARTHUR BERGHUIS

1. Een enkel DNA-gebonden eiwit kan DNA replicatie in één specifieke richting blokkeren doordat de tijdschaal waarop kleinschalige moleculaire herschikkingen plaatsvinden vele malen korter is dan de tijd benodigd voor het repliceren van een enkel DNA basepaar. (Hoofdstuk 8 en 9 van dit proefschrift)
2. Het polymerase waarschijnlijkheidsdistributiespectrum kan gezien worden als de polymerase-vingerafdruk, daarmee een middel om verschillende polymerases te classificeren. (Hoofdstuk 6 van dit proefschrift)
3. Enkel-molecuulbiofysica is een zeer fascinerend vakgebied dat vele vernieuwende inzichten heeft voortgebracht. Het vakgebied heeft echter nog niet onze kijk op de wereld op een fundamentele manier veranderd, zoals ontdekkingen in andere vakgebieden dat wel hebben gedaan.
4. Grofweg twee dominante karaktertrekken maken dat iemand de top van de wetenschappelijke voedselketen bereikt: een medogenloze ambitie en het hebben van een niet-aflatende nieuwsgierigheid. De verhouding tussen de twee bepaalt of een leidend figuur als een haai dan wel als bron van inspiratie gezien wordt.
5. Statistiek zou tot de basis van het wiskunde-onderwijsprogramma moeten behoren en zou, net als informatica, al aan de basisschool onderwezen moeten worden.
6. Het Belgische onderwijssysteem is ordes van grootte beter dan het Nederlandse als voorbereiding voor een universitaire opleiding door de rigoureuze nadruk die er op theoretische kennis gelegd wordt. Het Nederlands wetenschappelijk onderwijs is daarentegen veel beter dan het Belgische als voorbereiding op het postuniversitaire leven, door de nadruk op praktische vaardigheden.
7. Een experimenteel wetenschapper heeft met name een goede dosis gezond verstand en pragmatisme nodig. Het gezonde verstand is ook onder verstandige mensen soms ver te zoeken.
8. Precisie wordt vaak onbewust en onterecht verward met accuratesse. Het zeer precies uitvoeren van experimentele procedures wekt in dat geval de schijn dat de experimentele uitkomst accuraat zal zijn.
9. Elk levenspad is stochastisch, succesvol zijn is nauw verbonden met het besef hiervan alsmede de mate waarin iemand zijn opties kan inschatten.
10. De 'selfiestick' is de belichaming van (wat er mis is met) de moderne maatschappij.

Deze stellingen worden opponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotor prof. dr. N.H. Dekker.

ENHANCED INTERROGATION OF DNA ROADBLOCKS AND RNA POLYMERASES

PULLING INFORMATION FROM THE INDIVIDUAL TO UNDERSTAND THE MASSES

ENHANCED INTERROGATION OF DNA ROADBLOCKS AND RNA POLYMERASES

PULLING INFORMATION FROM THE INDIVIDUAL TO UNDERSTAND THE MASSES

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 5 april 2016 om 10:00 uur

door

BOJAN ARTHUR BERGHUIS

Ingenieur in de Moleculaire Wetenschappen
geboren te Breda, Nederland.

Dit proefschrift is goedgekeurd door de

promotor: Prof. dr. N.H. Dekker

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Onafhankelijke leden:	
Prof. dr. C. Dekker	Technische Universiteit Delft
Prof. dr. V. Croquette	Ecole Normale Supérieure, Paris, France
Prof. dr. G.J.L. Wuite	Vrije Universiteit Amsterdam
Prof. dr. P. McGlynn	University of York, United Kingdom
Prof. dr. M. Dogterom	Technische Universiteit Delft, reservelid
Overige leden:	
Prof. dr. N.E. Dixon	University of Wollongong, Australia
Dr. S.M. Depken	Technische Universiteit Delft

Keywords: magnetic tweezers, single-molecule, multiplexing, Tus-*Ter*, *E.coli* DNA replication, viral transcription, P2 RNA-dependent RNA polymerase

Printed by: Gildeprint

Front & Back: "Dwell times" by Jacob Kerssemakers

Casimir PhD Series, Delft-Leiden 2016-08

ISBN 978-90-8593-252-9

An electronic version of this dissertation is available at
`http://repository.tudelft.nl/`.

Opgedragen aan ir. Lukas Willem Lievense
de Opa die mij geleerd heeft
dat je een hand nodig hebt
om een vuist te kunnen maken.

CONTENTS

PREFACE

WHAT is life? At the very least it is a concept that we humans find surprisingly difficult to define. Though generally wet and dynamic[1], arguments about what defines life inevitably involve terms like 'reproduction', 'growth', and 'adaptation' – matters very common to cells and viruses alike, yet whether the latter belong to the category of living things is a matter of ongoing debate. For the purpose of this thesis (as well as for my own understanding), I adopt a less materialistic and more conceptual definition of life:

Life is a process brought forward by the self-organization of molecules, a process that seemingly violates the second law of thermodynamics[2] as it increasingly acquires and maintains information over timescales that vastly exceed the lifetime of the molecules holding this information.

In this manner the distinction between a virus and a cell becomes rather meaningless: viruses are just as much part of the process that we know as 'life' as a *homo sapiens* like you or me is.[3] In addition, it allows me to conveniently classify my thesis work as "an effort to gain better understanding of the molecular processes and building blocks of life". Though this classification is rather broad – as myriads of doctoral theses written over the past century or so belong to this category – my thesis belongs to a relatively small and novel subcategory of the 'gaining insight into the building blocks of life' class by making use of two concepts: *single-molecule* and *bottom-up approach.*

Gaining direct, real-time access to the behavior of single molecules was a great leap forward a couple of decades ago. Not only did direct access to the reaction kinetics of individual molecules prove to provide crucial insights into the stochastic behavior and transient dynamics unavoidably obscured in the ensemble-averaged outcomes of classical bulk experiments, the idea of watching an individual, nanometer-sized DNA polymerase performing life's work was simply fascinating. What made this possible to a large extent, was stray physicists wandering into the field of biology with in mind the idealized system of a particle in a box. Credit to the idea of starting from the bottom upwards with an as simple as possible system – a protein in a box – to study and quantify the fundamentals of life goes to the experimental physicists pioneering this field.

---

[1] To paraphrase Howard C. Berg in 'Random walks in biology'.

[2] The second law of thermodynamics states that, to quote Max Planck: "Every process occurring in nature proceeds in the sense in which the sum of the entropies of all bodies taking part in the process is increased." In other words: the amount of randomness or disorder must always increase.

[3] This definition might invoke cognitive dissonances for future generations that might co-exist with artificial intelligence.

Anno 2016, watching a single molecule in action is still fascinating, yet no longer a truly remarkable achievement. The ability to observe and manipulate single proteins, multi-enzyme complexes, or cells has become commonplace, and even accessible by light microscopy[4] through applying some clever tricks. But what do we learn from observing all these single events? Have single-molecule techniques become invaluable to the field of biology, fundamentally changing the way we think about life, like e.g. the discovery of the DNA double helix structure by Watson, Crick, and Franklin has done? Or is the field doomed to have an impact of lesser greatness: sexy, yet convicted to an existence in the shadows of the true fundamental insights that shaped the way we think about life – insights gained by the more traditional fields of biology, physics and chemistry?

Personally, I would say it is still too early to tell, but that the techniques developed in the field of single-molecule biophysics have proven useful in innumerable ways. Single-molecule experiments have the potential (and are well on their way) to become a tool used not only by scientists with a background in applied physics or engineering, but also by biologists – i.e. the scientists who generally have a better idea of which biology-related questions to ask. While knowing the exact behavior of single molecules is generally not pivotal to understanding entropic (i.e. non-living) systems, this does not hold true for the entropy-defying building blocks of life. However, as a result of the very different schools of thought that underlie the disciplines of biophysics and biology, both communities tend to communicate at their own specific wavelength – maybe keeping single-molecule biophysics more at the periphery of biology than it should be. Though there is definitely a trend towards increased interdisciplinarity, there is still sometimes a reciprocal lack of appreciation as a result.

With this in mind, I like to take the liberty of placing my thesis – also the result of the collective effort of the many great scientists I had the pleasure to learn from and work with over the past years – in a sub-subcategory by employing and developing an additional concept: *multiplexed force spectroscopy*. As the ensuing chapters will elaborate on, there is a huge gap between the behavior of a single molecule and the behavior of an ensemble of billions of molecules. Observing single events of protein binding or enzymatic complex activity makes poor statistics, especially if it is the rare, out of the ordinary behavior that is of interest. Repetition is key: crucial to fully appreciate the richness of detail that single-molecule observations add to the known ensemble-averaged behavior. Large scale repetition in parallel – multiplexing – makes it possible to gain these insights on such timescales that they can actually contribute to a successful career in science as well.

---

[4] though super-resolution microscopy of living cells is typically a top-down approach

Gaining quantitative insights from single or rare events, while being able to convincingly place these in the bigger picture of their biological framework, is what multiplexed single-molecule force spectroscopy in essence aims to achieve. Multiplexed single-molecule force spectroscopy, in my opinion, can be regarded as a tool to bridge the gap between the single and the collective behavior of life, a tool to further enhance the contribution of the field of biophysics to understanding the processes of life. I hope this is what the careful reader will be able to appreciate throughout this thesis.

Bojan Arthur Berghuis
Delft, April 2016

1

GENERAL INTRODUCTION

Let me lay my cards on the table. If I were to give an award for
the single best idea anyone has ever had, I'd give it to Darwin,
ahead of Newton and Einstein and everyone else.
In a single stroke, the idea of evolution by natural selection
unifies the realm of life, meaning, and purpose
with the realm of space and time, cause and effect,
mechanism and physical law.

Daniel Dennett

Besides being an introduction to the ensuing chapters, this chapter is my version of an introduction to the field and concepts of single-molecule biophysics research - with a particular focus on the aspects of using magnetic tweezers as a technique of course. The topics covered here are by no means exhaustive; the ideas and concepts introduced are a selection of topics that proved crucial to my own understanding of the field of single-molecule biophysics.

1.1. THE PHYSICAL CORE OF LIFE

1.1.1. DEOXYRIBONUCLEIC ACID (DNA)

DEOXYRIBONUCLEIC acid or DNA is the molecule - or polymer rather - that is the central carrier of life's information. A pivotal moment in biological research is marked by Watson and Crick's proposal of DNA's molecular structure using the x-ray fiber diffraction images of Rosalind Franklin in 1953.[1] This provided direct insight into how a polymer composed of four bases (adenine, guanine, cytosine and thymine or A, G, C, and T, respectively), a pentose sugar ribose, and a phosphate backbone could contain and preserve bits of genetic, hereditary information.

Watson and Crick discovered the DNA double helical form so familiar to us now. The double helix is formed by two intertwining strands of DNA polymer with the bases directed inwards towards the other strand (**Fig. 1.1a**). Each base 'couples to' or 'pairs with' a specific other base through hydrogen bonding interactions: A pairs with T through two hydrogen bonds, G pairs with C through three hydrogen bonds. These base pairs (bp) are the bits of life - life's own 0's and 1's - and it is the redundancy of DNA as a result of base pairing that is pivotal. If one DNA strand breaks or loses a base, it can be easily repaired by making use of the complementary strand, essentially making the longer-term storage of information possible. The helix formed by DNA contains a major and a minor groove, the base pairs are stacked in an orderly fashion spaced ~0.34 nanometers (nm) apart. The diameter of the double helix is 2 nm and the length can vary from only a couple of base pairs to 4.6 million base pairs (Mbp) for the genome of *Escherichia coli*, to hundreds

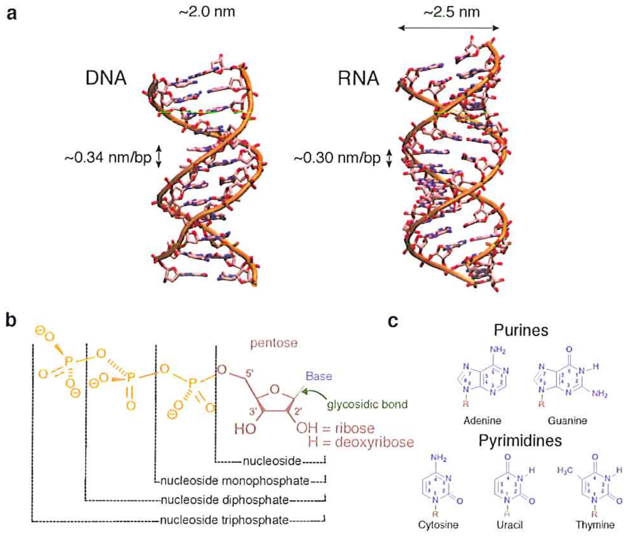


Figure 1.1: **The structure and composition of DNA and RNA. a)** The double helix structure as discovered by Watson and Crick. **b)** The composition of nucleotides. **c)** The different bases.

of millions of base pairs, like is the case for our own chromosomes. DNA strands have a specific directionality or polarity as a result of the molecular structure. The phosphate 'backbone' of a DNA strand covalently links two adjacent bases of the same strand via the sugar pentose ring on the third and fifth carbon atom. This creates an asymmetry where if one end of a DNA polymer has a vacancy at the third position on the ribose sugar ring - the 3-prime or 3' end - the other end will have a vacancy on the fifth position of the sugar ring, the 5' end.[1] This polarity is at the very essence of DNA, as all proteins and enzymes interacting with DNA for one reason or another can be characterized by the polarity of the strand with which they interact, as many chapters in this thesis will show.

1.1.2. RIBONUCLEIC ACID (RNA)

Ribonucleic acid or RNA is a biopolymer that is structurally very similar to DNA, yet its role is much more versatile than just being an information carrier. Structurally, the main difference between RNA and DNA is an additional hydroxy group on the second carbon atom of the ribose sugar ring (**Fig. 1.1b**). In addition, RNA uses uracil (U) as a base instead of DNA's thymine, which implies RNA base pairing involves A-U interactions. RNA is mostly known for its role as an intermediate single-stranded (ss) carrier of DNA information to protein function: in this role RNA is a mere messenger, also known as mRNA. Yet the role of RNA is much more versatile. In its ss form, it plays a role in gene expression as short interfering (si)RNA or micro (mi)RNA. In a double-stranded (ds) helical form RNA fulfills a role similar to that of genomic DNA, though this is mainly a feature of dsRNA viruses (see **Chapters 6 and 7**). In addition, RNA – much more than DNA – is a flexible polymer that can form elaborate secondary structures, such as hairpins, pan-handles or pseudo-knots. These secondary RNA structures are well known to be of such intricate complexity that they become metal ion bound catalytically active enzymes or ribozymes, facilitating an array of chemical reactions such as RNA-based RNA cleavage[2], peptide bond formation[3] and self-replication[4].

1.1.3. PROTEINS AND ENZYMES

If DNA is the hard disk of life, and RNA is what enables communication between the drive and main frame, then proteins are the main frame, the wiring, and most other components that contribute to physical structure. Enzymes are the processing units, the myriads of checkpoints, flips, switches, and apps of life. In other words, proteins are the building blocks of life, and enzymes are its workhorses. Both proteins and enzymes are chains of amino acids (peptides) that fold into well-defined structures based on the polypeptide sequence. As there are 21 different amino acids with a wide variety of properties – from bulky and hydrophobic to small and hydrophilic (and vice versa); positively, neutral or negatively charged; from acidic to basic – the number of possible structures with unique properties that can be conceived is seemingly endless.

Enzymes are a special class of proteins that contain a catalytic center which accelerates or catalyzes a specific chemical reaction. Virtually all metabolic processes of life require the activity of enzymes to ensure that specific chemical reactions happen at rates

---

[1] 3'-ACCG-5' does not contain the same information as 5'-ACCG-3'

fast enough to sustain life. Like non-biological catalysts, enzymes are not consumed in chemical reactions, they merely increase the probability of occurrence of a specific chemical reaction by lowering the the reaction activation energy. There are known examples of enzymes that enhance a chemical reaction to such an extent that a process that would on average happen only once every millions of years, happens on the timescale of milliseconds. Enzymes collectively create order out of randomness, thereby seemingly[2] violating the second law of thermodynamics. Enzymes are very specific catalysts, examples in this thesis are the RNA-dependent RNA polymerase of bacteriophage phi6 studied in chapter 6 and 7. As many proteins and molecules can enhance or inhibit enzyme activity (i.e. co-factors and inhibitors respectively), many (anti-viral) drugs specifically inhibit the activity of a certain enzyme.

1.1.4. THE CENTRAL DOGMA

Upon the discovery of DNA, Francis Crick proposed 'the central dogma of molecular biology'.[5] This states that information is transferred from DNA through RNA to proteins, and never in the reverse fashion (**Fig. 1.2**). Information encoded in the DNA is transcribed into RNA by a (DNA-dependent) RNA polymerase enzyme, the mRNA is then translated into a polypeptide sequence by the vastly complex protein-RNA complex called the ribosome. From one generation to a next, DNA is replicated by replicative DNA polymerases, enzymes that do not operate in an isolated fashion, but are a part of the multi-enzyme complex called the replisome. Since Crick's somewhat inappropriate use of the word 'dogma' and its introduction to the realm of molecular biology, the number of roles the central actors play in this process has been expanded, yet the basic concept of information flow from DNA to RNA to proteins still holds true.

1.1.5. THE CELL

Cells are the entities that allow for the processes of life to occur, they are the walled fortresses containing and protecting everything involved in the central dogma - and much more. A main reason why cells must exist for life to gain the complex and versatile form we are familiar with today, is that a small, closed cellular volume vastly in-

---

[2] Seemingly indeed, because the second law is never actually violated. To quote Richard Dawkins in his magnificent book 'The greatest show on earth': "The Second Law of Thermodynamics states that, although energy can be neither created nor destroyed, it can - must, in a closed system - become more impotent to do useful work. [...] Almost all the energy in the universe is steadily being degraded from forms that are capable of doing work into forms that are incapable of doing work. There is a levelling off, a mixing up, until eventually the entire universe will settle into a uniform, (literally) uneventful 'heat death'. But while the universe as a whole is hurtling downhill towards its inevitable heat death, there is scope for small quantities of energy to drive little local systems in the opposite direction. [...] While never actually disobeying the laws of physics and chemistry - and certainly never disobeying the Second Law - energy from the sun powers life, to coax and stretch the laws of physics and chemistry to evolve prodigious feats of complexity, diversity, beauty, and an uncanny illusion of statistical improbability and deliberate design. So compelling is that illusion that it fooled our greatest minds for centuries, until Charles Darwin burst onto the scene. Natural selection is an improbability pump: a process that generates the statistically improbable. It systematically seizes the minority of random changes that have what it takes to survive, and accumulates them, step by tiny step over unimaginable timescales, until evolution eventually climbs mountains of improbability and diversity, peaks whose height and range seem to know no limit [...]. Life evolves greater complexity only because natural selection drives it locally away from the statistically probable towards the improbable. And this is possible only because of the ceaseless supply of energy from the sun."

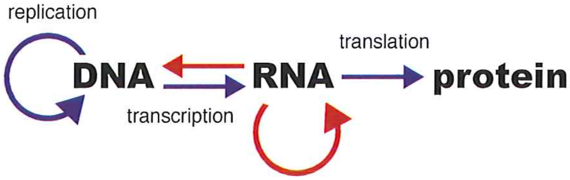


Figure 1.2: **The central dogma of molecular biology.** Information encoded in DNA is transcribed into RNA, which is translated into protein (blue arrows). From one generation to a next, DNA is replicated (blue circular arrow). Expansion of the central dogma (in red) followed the discovery of reverse transcription (from RNA to DNA) and RNA-dependent RNA replication mechanisms.

creases the likelihood of two molecules interacting or reacting with each other. In other words, cells concentrate (bio)molecules[3] to such an extent that reactions can occur at rates fast enough to sustain life.[4] Cells come in many shapes and sizes, and are classified by their property of lacking or having a compartmentalized genome-containing nucleus (or kernel, from the Greek *karyon*): prokaryotic or eukaryotic respectively. Multicellular organisms such as ourselves are eukaryotic, while unicellular organisms belonging to the bacterial and archeal kingdoms are prokaryotic. Typical dimensions of eukaryotic cells are 10 × 10 × 10 microns (10^{-6} m or a 100 femtoliter (1×10^{-13} liter) volume), a bacterium such as *Escherichia coli* (*E. coli*) typically has dimensions of 1 × 1 × 2 microns (2 femtoliter (2×10^{-15} liter) volume).

1.2. DNA REPLICATION AND TRANSCRIPTION

1.2.1. DNA REPLICATION

The survival of every organism inextricably depends on the extent to which it faithfully duplicates its genetic information. Genomic replication is generally catalyzed by DNA polymerases[5], aided by the assembly of various enzymes, proteins and co-factors that together form the replisome.[6, 7] The replisomal organization is conserved throughout all life forms, from viruses to higher order eukaryotes. The antiparallel nature of the double-stranded (ds) DNA helix necessitates simultaneous coordinated synthesis of two DNA strands with opposite polarity. This led the replisome to evolve to perform DNA synthesis of the two daughter strands in an asymmetric fashion: a continuous leading strand, versus a discontinuous lagging strand DNA polymerization activity. While the efficiency of DNA replication depends heavily on the coordinated action of all replisome components, the fidelity of DNA replication is a matter left mainly to the DNA polymerases. Replicative DNA polymerases are DNA-dependent DNA polymerases that extend a primer in the 5' to 3' direction, and typically have a 3' to 5' exonuclease or proofreading activity. Proofreading, the capability to transfer a faulty primer strand from the polymerization site to the polymerase exonuclease (exo) site, increases the fidelity of DNA replication as the erroneous nucleotide is excised at the exo site.

---

[3] Additional concentration enhancing effects of protein-and-DNA-filled closed volumes also take place: a phenomenon called molecular crowding.

[4] Reactions largely aided and controlled by enzymes, of course.

[5] With the exception of RNA viruses, which have RNA as a genome, thus requiring RNA polymerases.

1.2.2. E. COLI AND E. COLI DNA REPLICATION

E. coli is the workhorse of molecular biology, and generally a model system for prokaryotic organisms. As the lineage of prokaryotic organisms is evolutionarily much more ancient than that of eukaryotes, the enzyme complexes carrying out processes such as DNA replication are of a lesser complexity in prokaryotes than in eukaryotes. Much of what we know about DNA replication has been inferred from experiments performed with *E. coli*, as it is the most wide studied prokaryotic organism. *E. coli* is an important species in the fields of biotechnology and microbiology, as it can be easily engineered to produce vast quantities of a specific molecule or protein. In addition, *E. coli* populates our intestines in vast numbers and forms an essential part of the delicate biotope of our intestinal tract.

E. coli has a single, circular chromosome of approximately 4.6 million base pairs (Mbp) in length. A typical generation time of this bacterium is 40 minutes, implying that 4.6 Mbp need to be copied without error on average[6] within this time frame, an astonishing rate of ~ 1900 bp per second. *E. coli* achieves this through initiating DNA replication in a bidirectional fashion from a single origin of replication (*oriC*). This implies two replisomes assemble at the origin and move in opposite directions at average rates of 0.8-1 kilo bp (kbp) per second. The 'footprint' or space a replisome occupies on a DNA template is in the order of 30 bp. At ~900 bp/s the replisome thus moves 30 times its own length per second. If we would scale this up to the macroscopic world of cars, the velocities and especially the accuracy of the replication process becomes astonishing.[7]: a car-sized replisome – with a length of 5 m – would travel at 5 × 30 = 150 m/s or 540 km/h on average[8], while having to place a nucleotide tile with the correct color (having four colors to choose from) every 15 cm head-to-tail – making an error only roughly once every 15 km.[9] *Morituri te salutant.*

1.2.3. VIRUSES AND VIRAL TRANSCRIPTION

A virus is an agent, or a package of genetic information, that requires a living host cell to replicate and pass on this information. Viruses are as inextricably connected to the process of life as cells are[10], and they are known to infect and take hostage cells of all life forms: from bacteria to plant and animal cells **Fig. 1.3**). The first conception of the term 'virus' as a *contegium vivum fluidum* - i.e. an infectious agent that did not behave like any known bacterium - was upon discovery of the tobacco mosaic virus of the tobacco plant by Martinus Beijerinck in 1898[11].[10] Since then, virus research has mainly focussed on the control and prevention of viral infection, with notable successes through

---

[6] On average as a species that is, as cells with lethal mutations will die.

[7] I imply the scaling of spatial dimensions only, while leaving time as it is.

[8] With instantaneous stops or jumps to speeds twice as fast in this inertia-free world (see section 1.3)

[9] Based on the notion that the error rate of *E. coli* DNA pol III is roughly 10^{-5}, not including proofreading.[8]

[10] For instance, in the human genome, the amount of DNA from viral origin is roughly equal to the entire (protein) coding region [9]

[11] Interestingly, this Wageningen-based scientist describes how he made his discovery in a Delft-based laboratory as he describes in his 1898 paper: "Then, in 1897, the resources of the newly erected bacteriological laboratory of the Polytechnical Institute at Delft were put at my disposal. This included a greenhouse with heating facilities, which I started using that same year for further experiments on the spot disease. I was, therefore, able to carry out a series of incontestable infection experiments the results of which I shall now briefly describe."

vaccination strategies and anti-viral drugs.[11, 12] Nevertheless, many challenges remain, and the central role of viruses in the process of life can hardly be overestimated: the genomic diversity of viruses as well as their astronomical population size (estimated to be over 10^{31} viral particles[12], with the vast majority being bacterial viruses or bacteriophages) create a massive selective pressure on all cellular organisms alike.[13–15]

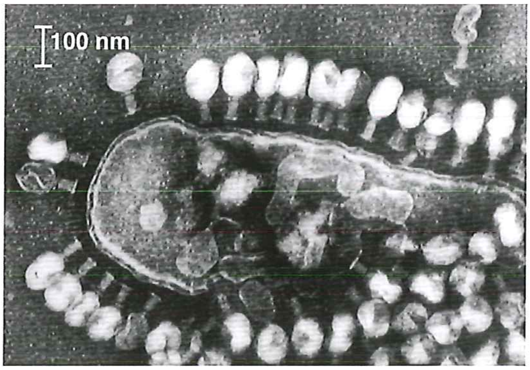


Figure 1.3: **Electron micrograph of T4 bacteriophage virions infecting an *E. coli* bacterium.** Adapted from Simon and Anderson [16]

The virus particle or virion is the central basic unit of every virus population. A virion consists of two or three elements: 1) The viral genome, which can be single-stranded or double-stranded DNA or RNA, 2) a protein coat or capsid, which encases and protects the viral genome, and in some cases 3) an envelope consisting of a lipid outer layer that surrounds the capsid. The complexity of virions ranges from the simple, two-gene system (one gene to replicate the genome, the other to build the capsid [17]) of porcine circoviruses (capsid diameter of 17 nm), to the extremely complex (> 900 genes) and large (0.65 micron diameter) virion of mimiviruses.[18] Viruses are generally classified through the nature of the genome[13], but classification based on the structure of the viral capsid has also been proposed.[19]

Chapters 6 and 7 focus on elucidating the mechanism of the RNA dependent RNA polymerase (RdRP) P2 of bacteriophage $\phi 6$, a virus known to infect a pathogenic plant bacterium *Pseudomonas syringae*. $\phi 6$ belongs to the family of *Cystoviridae*, a class of bacteriophages with a segmented, 3-piece dsRNA genome totaling ~14 kb in length and a lipid envelope surrounding the (65 nm diameter) capsid. Cystoviridae share common features with eukaryotic dsRNA viruses. The viral genome is both transcribed and replicated by P2 RdRP.[20] This polymerase is a compact, spherical molecule with an internally located active site, which is accessible from the outside by two positively charged tunnels - one large enough to accommodate diffusion of negatively charged NTPs, the

---

[12] To get at least some sense of this astronomical number: the average number of cells in a human body $\approx 10^{13}$. With $\sim 10^{10}$ humans on this planet, that makes ~ 100 million virions for every cell of every human alive.

[13] Also known as the Baltimore scheme.

other large enough to accommodate negatively charged ssRNA, but not dsRNA.[21] Purified P2 readily replicates ssRNA templates *in vitro* at a rate of ~120 nucleotides (nt) per second. P2 is less efficient at initiating transcription *in vitro*, yet once initiated the polymerase is highly processive[14] *in vitro*, producing full length transcripts at ~30 nt/s. The lower transcription velocity is a result of the additional work required by P2 to unpair the RNA base pairs ahead of the polymerase as it moves along a dsRNA template.

1.2.4. RDRP FIDELITY AND QUASISPECIES THEORY

A distinct feature of viral polymerases is the lower accuracy or fidelity than their non-viral counterparts. Besides being inherently more error prone than other polymerases, viral polymerases lack the proofreading mechanism of prokaryotic and eukaryotic replicative polymerases. Typical error rates of the *E. coli* DNA replication process for instance is a phenomenal 1 error in every billion to 100 billion ($10^{-9} - 10^{-11}$) base pairs.[15][8] Viral RdRPs are vastly more error prone, with error rates of typically in the order of 10^{-4}: one error per 10 kbp.[22] With typical genome sizes in the order of 10 kbp, this implies one error per replication cycle on average. Naively this still might seem relatively error free, but given sheer size of viral populations as well as their fast replication rates, this implies a huge genetic diversity in every virus population, placing viral populations on the brink of error catastrophe (**Fig. 1.4**).[23]

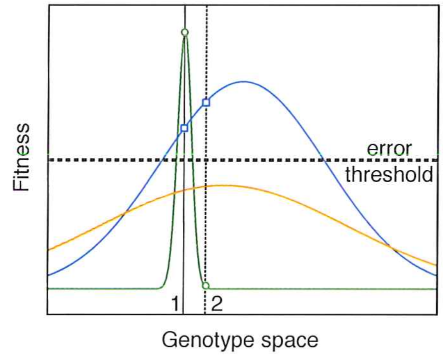


Figure 1.4: **The fitness of a (viral) species as a function of the population genotype.** A viral species with a low mutation rate will have a genetically homogenous population – thus a narrow spread (green) – with the higher fitness in environment 1 (solid vertical line) than a population with a high mutation rate (i.e. a quasispecies, blue). A shift in environmental conditions that would allow populations of genotype 2 to thrive (dashed vertical line) will render the fitness of the species below the error threshold, while the quasispecies fitness actually improves. A viral population with a mutation rate that is too high will succumb to error catastrophy (orange).

The large genetic diversity of viruses is what makes a viral population a quasispecies, as opposed to a genetically homogenous species like (multi-)cellular organisms are. The

---

[14] Processivity is a (hand waving) measure for the number of nucleotides an enzyme transcribes, replicates, unwinds or simply travels (translocates) along a nucleic acid template. Low processivity is typically a number in the hundreds of bp, high processivity can be anything from a couple of kb to hundreds of kb.

[15] The *E. coli* replicative polymerase has an error rate of 10^{-5}, proofreading adds another 10^{-2}, mismatch repair another 10^{-3}, leading to an overall error rate of $\sim 10^{-10}$.

high viral mutation rates would be catastrophic to any normal species and would lead to rapid extinction. Viruses, however, are thought to thrive exactly because of their high mutation rates combined with their extremely large population numbers. The existence of the viral population as a quasispecies makes it possible for the population as a whole to explore large regions of 'sequence space', increasing the probability that certain mutations will perform better than their direct ancestors[16] in an ever-changing fitness landscape (**Fig. 1.4**). In other words: a high mutation rate makes viruses extremely adaptable, resilient and able to overcome anti-viral defense mechanisms. It is thought that viruses in this way tune their error rates such that a population is diverse enough to adapt, yet homogeneous enough to succumb to error catastrophe. This is called quasispecies theory.[24]

1.3. ASPECTS OF LIFE AT THE MOLECULAR SCALE

Somewhere in going from the macroscopic scale of our daily lives to the nanoscopic scale of atoms, there exists a regime where gravity and inertial forces are no longer dominant forces, yet viscous forces, electrostatic forces and Brownian motion become increasingly important. This is the counterintuitive regime at which all molecules, cells, and particles used or investigated in the experimental assays of this thesis reside. Flows in micro-chambers are laminar; beads, cells and proteins are in continuous movement as a result of Brownian motion; chemical reactions, protein binding and enzyme polymerization events are not deterministic, but stochastic in nature - and therefore need to be described in terms of rates and probabilities. While a full description of these phenomena is beyond the scope of this introduction, some aspects do deserve a short mention here.

1.3.1. LIFE AT LOW REYNOLDS NUMBERS

The Reynolds number (Re) represents the ratio between inertial forces and viscous drag forces on a particle. While there is more that sets the microscopic to nanoscopic world apart from our macroscopic world, the difference in Re is a very characteristic one. We are very familiar with the moment of inertia of objects surrounding us: a ball set in motion by the kick of a football player tends to stay in motion for quite a while[17] before drag or frictional forces eventually bring it to a halt. How very different matters become at the microscopic scale of single cells: viscous forces dominate in this regime, an object's moment of inertia has become negligible at these length scales. For a body with length-scale l moving at a velocity v through a medium, Re is given by:

$$\mathrm{Re} = \frac{F_{\mathrm{inertial}}}{F_{\mathrm{viscous}}} = \frac{\rho v l}{\eta} \tag{1.1}$$

where ρ is the medium density, η the viscosity. From this it immediately becomes clear that Re for a macroscopic body in water ($v \cdot l \sim 1\ \mathrm{m}^2/\mathrm{s}$, $\mathrm{Re} \sim 10^6$) will vastly differ from Re

---

[16] At the expense of most other virions that will not - in this sense viruses truly adopt an 'all for one and one for all' survival strategy.

[17] Placing your head in the ball's way moments after kick off will make one experience the ball's moment of inertia first hand (though I generally would not recommend doing this).

of a micrometer-sized bacterium moving at a couple of micrometers a second in water ($v \cdot l \sim 10^{-12}$, $\mathrm{Re} \sim 10^{-4}$).[18]

1.3.2. BROWNIAN MOTION AND DIFFUSION

Brownian motion is the random and continuous movement of molecules or small particles due to temperature-driven collisions with other particles. Diffusion is the random migration of particles that occurs as a result of this Brownian motion. Directed diffusion occurs when a temperature or concentration gradient steers the average movement of particles from high to low temperature regions or from regions of high to low molecule concentrations.[19] The equipartition theorem states that a particle at an absolute temperature T has – on average – a kinetic energy associated with each motional degree of freedom of $\frac{1}{2}k_\mathrm{B}T$, with k_B being Boltzmann's constant.[20] Einstein showed that this holds true for all particles under influence of Brownian motion [25]: from the micron-sized magnetic beads in magnetic tweezers assays to the movement of individual NTP molecules in a cell. As the average kinetic energy of a particle with mass m and velocity v along one axis (x) is $\langle mv_x^2/2\rangle$, we can use the equipartition theorem to calculate the root mean square velocity of this particle along this axis:

$$\frac{1}{2}m\langle v_x^2\rangle = \frac{1}{2}k_B T \tag{1.2}$$

$$\Longleftrightarrow \langle v_x^2\rangle^{1/2} = \left(\frac{k_B T}{m}\right)^{1/2} \tag{1.3}$$

As the particle is in continuous collision with other particles, movement in one direction will be as likely as movement in the other such the average displacement over time along x will remain zero. The root mean square displacement however, will increase with time such that the probability of finding the particle along x is normally distributed with a width that increases (and a height that decreases) with the square root of time. The measure that characterizes this movement over time t is the diffusion coefficient $D = \langle v_x^2\rangle/(2t)$, in units of cm^2/s. A typical value for D of proteins in water at room temperature is 10^{-7} cm^2/s. That the displacement of a particle is proportional to the square root of time implies that diffusion is a very 'localized' phenomenon (**Fig. 1.5**).

1.3.3. REACTION KINETICS

Ever since publication of the studies of J.D. van der Waals [26], it became clear that a vast array of properties of matter (e.g. boiling points of liquids, viscosity, compressibility of gasses, etc.) can be described in terms of the forces that molecules exert on each other. These forces explain the chemical reactivity and affinity of molecules and proteins. Svante Arrhenius in turn proposed that the rate constant k of a chemical reaction (e.g. a substrate forming a product, [S] → [P]) depends on the temperature T and an energy of activation E_a of the participating molecules scaled by a pre-exponential factor A

---

[18] For the macroscopic body in this example to experience the same viscous environment as the swimming bacterium, the medium would have to be a million times more viscous than water(!)

[19] The higher number of molecular collisions among molecules with a higher temperature or in more concentrated regions is what drives the average motion to lower temperature of concentration regions.

[20] $k_\mathrm{B}T$ is a unit of energy expressed in Joules (J). At room temperature $k_\mathrm{B}T \approx 4.11 \times 10^{-21}$ J. The energy to break a DNA base pair ranges from $\sim 1-4k_\mathrm{B}T$.

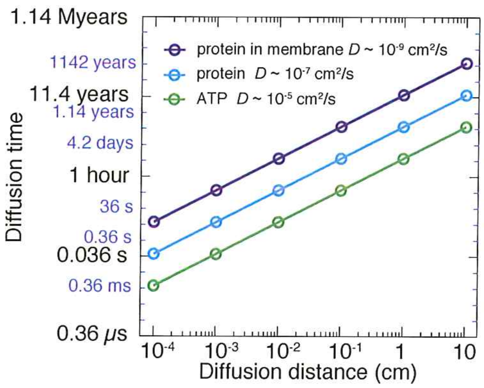


Figure 1.5: **Diffusion distance versus the time it takes to diffuse this distance.** For every increase in the order of magnitude of Δx, the time it takes any diffusive particle to cover such a distance increases with 2 orders of magnitude. This is why e.g. a small molecule such as ATP takes on average 0.5 milliseconds (ms) to diffuse through the average bacterium (10^{-4} cm), 50 ms to cross the average human cell (10^{-3} cm), yet 58 days to diffuse through a cup of coffee (~10 cm).

[27]:

$$k = Ae^{-E_a/k_B T} \tag{1.4}$$

thereby implying that only chemical species with a higher than population average activation energy can take part in chemical reactions, such that the energy required to physically transform a product into a substrate can be overcome. The pre-exponential factor A is not a well defined measure here, but Kramers' theory of reaction kinetics expands on this and arrives at an exact expression for the rate constant based (amongst other things) on the force exerted on the reacting molecules by the environment.[28] The advent of single-molecule manipulation techniques made it possible to measure these intermolecular forces directly, and to steer the equilibrium of a (bio)chemical reaction in a specific direction using externally applied force.

1.4. ASPECTS OF DATA ACQUISITION

1.4.1. SINGLE-MOLECULE MEASUREMENTS

Our current ability to track and manipulate single molecules in real time allows us to gain great additional insight into the behavior and kinetics of biological processes. Bulk biochemical assays typically observe the ensemble-averaged behavior of millions of identical copies of the molecules in question, while single-molecule experiments provide insight into the behavior of the single molecules that constitute the classical ensemble average. Though the principle of ergodicity holds for the biological systems studied in this thesis – i.e. the behavior of many molecules at a single moment in time is equivalent to the behavior of a single molecule over a long stretch of time – the latter allows us to observe the inherent stochastic nature of biochemical reactions in real time. A disadvantage is that a single observation lacks the statistical proof to infer all possible behaviors of the molecule – or the behavior of the population of molecules as a whole. This is why the efforts in this thesis are directed towards detecting as many single molecule events simultaneously as possible in a single experimental assay, a concept known as multiplexing. The technical efforts of this thesis are directed towards bridging the gap between single-molecule events and ensemble averaged behavior.

1.4.2. VISUALIZING THE INVISIBLE: SINGLE-MOLECULE MAGNETIC TWEEZERS

Magnetic tweezers (MT) is a technique that makes use of an externally applied magnetic field to manipulate the movement of and forces acting on magnetic particles in a fluid chamber. The magnetic particles are typically micron-sized beads that are made as homogeneously as possible to ensure homogeneity of the force acting on each single bead. The beads are tethered to the surface of the fluid chamber by means of a long polymer molecule: a DNA polymer in the vast majority of MT assays. The beads are visualized through a microscope objective, and the resulting image is recorded with a camera (**Fig. 1.6**). As the individual DNA molecules are too small to be visualized through conventional light microscopy[29], the properties of the individual [21] molecules are inferred through the movement of the beads: continuously in motion due to Brownian motion of the sur- rounding water molecules, but limited in their movement as a result of the anchoring via the molecule of interest. An increased upward pulling force induced by lowering a pair of magnets towards the fluid chamber stretches the tethers and limits bead movement even further. It is this force-extension relationship that is molecule-specific and can be used to e.g. infer the rate at which a helicase unwinds a dsDNA helix.[30, 31]

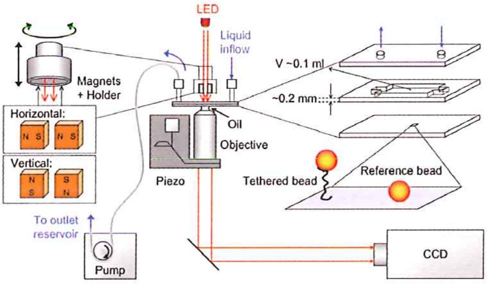


Figure 1.6: **The magnetic tweezers setup.** Adapted from [32].

1.4.3. FROM BEAD MOVEMENT TO APPLIED FORCE

The force exerted on DNA molecules tethering the beads to the surface depends on the magnitude of the magnetic field $\vec{\mathbf{B}}$ and the magnetic moment of the beads $\vec{\mathbf{m}}$ (**Eqn. 1.5**). $\vec{\mathbf{B}}$ is determined by the size of the pair of magnets above the fluid chamber, their orientation, the distance between the magnets (gap size), and the distance between the pair of magnets and the beads in the flow cell. The bead magnetic moment scales with the

---

[21] And for light microscopy invisible molecules

bead size. As is described in more detail in most chapters of this thesis, the magnet size, orientation and gap size, as well as bead size, have been standardized.

$$\vec{\mathbf{F}} = \frac{1}{2}\nabla\vec{\mathbf{m}}\cdot\vec{\mathbf{B}} \tag{1.5}$$

The bead and the DNA tether experience a constant force because the length scale of motion of the bead is small (micrometers) compared to the length scale of changes in the magnetic field $\vec{\mathbf{B}}$ (millimeters). The force exerted on the bead is calculated from the extent of bead x,y plane fluctuations – for forces lower than ~1 pN and a tether of known length – as the bead movement is a function of the upward pulling force. For higher forces, the force on the bead is calculated in Fourier space – i.e. from the frequency spectrum of the bead movement – as this allows for better correction of drift and finite integration times of the camera that set a limit to the range of measurable frequencies.[32] At any given time, there are two forces acting on the bead: the applied magnetic force (applied force, F) and a restoring force exerted by the DNA on the bead. The total potential energy of the system then equals

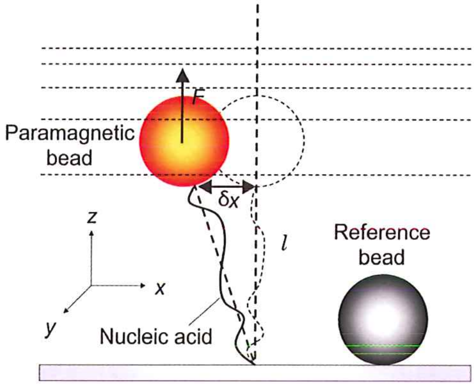


Figure 1.7: **The displacement of a tethered bead in a MT setup.** Adapted from [32].

$$E_p = E_{\text{mag}} + E_{\text{DNA}} = A(l) - F \tag{1.6}$$

where $A(l)$ is the mechanical energy stored in the tether, which is a function of the tether length l. E_{mag} can be taken to be F, as we argued that the force was constant over the length scales of changes in bead height. When this system is in equilibrium, all partial derivatives of the potential energy should equal to zero, which allows us to express the potential gradient in the z direction as:

$$\frac{\partial E_p}{\partial z} = \frac{\partial A(l)}{\partial z} - F = \frac{\partial A(l)}{\partial l} - F = 0 \tag{1.7}$$

In other words, we find $F = \partial A(l)/\partial l$. As Brownian motion continuously forces the tethered system out of equilibrium, the potential energy for small deviations from the equilibrium can be expressed as the Taylor expansion of E_p to leading order[22], yielding $\partial^2 E_p/\partial x^2 =$

---

[22] The term that contributes most significantly.

F/l for displacements in the direction of the magnetic field (i.e. the x direction) ∂x. By setting the potential energy to zero for the equilibrium position of the tether, we can conveniently express the potential energy in the direction of the magnetic field lines as the average potential energy

$$\langle E_x \rangle = \frac{1}{2}\frac{F}{l}\langle \partial x^2 \rangle \tag{1.8}$$

The equipartition theorem tells us that the energy for this degree of freedom is $E_x = \frac{1}{2}k_B T$, which yields a simple expression for the force acting on the bead

$$F = \frac{k_B T l}{\langle \partial x^2 \rangle} \tag{1.9}$$

Thus by knowing the length l of a tether and the variance of the displacement in the x direction, the force can be obtained for (long [23]) tethers in the low-force regime. For forces above 1 pN, forces most commonly applied in the work presented over the chapters of this thesis, the variance of the displacement in the x direction can be obtained by integrating the power spectrum $P(\omega)$ of the bead motion over all frequencies:

$$\langle \partial x^2 \rangle = \frac{1}{2\pi}\int P(\omega)d\omega = \int_0^\infty \frac{6k_B T\eta R}{k_x^2}\frac{1}{1+(\frac{\omega}{\omega_c})^2}d\omega \tag{1.10}$$

where $k_x = F/l$, which is a measure for the trap stiffness, η is the viscosity of the buffer, R is the bead radius, ω is the frequency and ω_c is the cutoff or corner frequency.[24] Thus eqn. 1.10 is an expression of the frequency spectrum as a function of the corner frequency. Experimentally, this integral can be obtained at an acquisition frequency (f_{ac}) that is naturally bounded by the inverse of the acquisition time t_{ac}^{-1} at the low frequency end of the spectrum, yet *must* have an upper boundary of twice the corner frequency (i.e. $t_{\text{ac}}^{-1} \lll 2 \cdot \omega_c < f_{\text{ac}}$), otherwise the recorded spectrum of frequencies becomes too narrow to accurately determine the force.

1.4.4. FROM FORCE TO DNA EXTENSION: THE WORM-LIKE CHAIN MODEL

MT setups make use of advanced bead-tracking algorithms to accurately determine the spatial coordinates of tethered beads in real-time [33]. From this we infer the extension z of a tethered bead, which corresponds to the length of the nucleic acid molecule tethering the bead to the surface. But since life at the scale of DNA polymers is dominated by electrostatic and Van der Waals forces, it may hardly come as a surprise that there is no simple relationship between the applied force and the end-to-end extension of a DNA molecule.

The behavior of DNA under an applied force can be divided into two regimes. In a regime of low force ($F << Bk_B T$) the molecule can be treated as inextensible, thus not stretchable beyond its contour length. The number of possible conformations of the molecule –

---

[23] The longer the molecule, the larger the displacement ∂x. Typically, 21 kb dsDNA tethers are used in our force calibrations.

[24] The corner frequency marks the freqency beyond which higher frequencies start decaying – and thus are no longer 'accessible' by the system.

thus the molecule entropy – becomes smaller as the force-induced extension z becomes larger, this regime is referred to as the *entropic* regime. At higher forces ($F \sim Bk_BT$, or $F >\sim 5$ pN) DNA is effectively stretched like a spring, requiring the need for a stretching parameter in any given model. This regime is called the *enthalpic* regime, bounded at high-forces by the overstretching transition, i.e. the force induced melting of the double helix to ssDNA, which occurs at ~65 pN at room temperature.

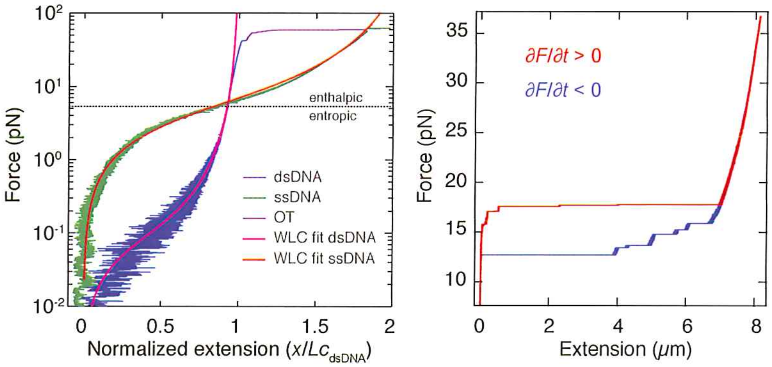


Figure 1.8: **Force extension characteristics of dsDNA, ssDNA (left), and hairpin DNA (right).** Left: Increasing the applied force on a linear dsDNA tether (blue, WLC fit in magenta) reduces the number of possible conformations (entropic stretching) until the DNA reaches its contour length (Lc). Above a force of ~5 pN the dsDNA stretches further, until at ~65pN the force disrupts the Watson-Crick basepairs, leading to an abrupt increase in length. This is the overstretching transition (OT, purple). In subsequent lowering of the applied force, the DNA can, transition back to its ds form with or without hysteresis, or remain single-stranded in elevated temperature conditions (34 °C shown here). The extension per base (pair) for ssDNA (green, WLC fit in orange) is longer than dsDNA in the enthalpic regime, and shorter in the entropic regime, as it is more flexible. Right: The force extension plot of a 7 kb DNA hairpin. Starting from low (< 10 pN) forces the hairpin remains closed until ~16-17 pN, when the force is high enough to disrupt base pairing (red), leading to a sudden jump in the extension. Once opened, the 14000 ssDNA bases are stretched further. In lowering the force the hairpin remains opened at forces below the opening force (as a result of thermal fluctuations that keep the system out of equilibrium (blue)) until it closes again at ~12-13 pN.

The worm-like chain (WLC) model, as devised by Bouchiat *et al.* [34], gives an approximation of the expected average extension $\langle z \rangle$ at a given force F in terms of the contour length l_c and the persistence length L_p by approximating any polymer as an continuously flexible [25] isotropic rod [26]. The former is the length (contour) of the molecule when it is fully extended [27], the latter is a measure of the molecule rigidity. In the entropic regime, the force versus extension is usually expressed as the difference between the exact numerical solution and the residuals of an earlier interpolation formula [35] in

---

[25] As opposed to the segmented flexibility of the freely-jointed chain model.

[26] Which makes this model an approximation, of course.

[27] And thus, in the case of DNA, a function of the inter-base distances.

a 7th order[28] polynomial:

$$F(z)=\frac{k_B T}{L_p}\Big[\frac{1}{4(1-z/l_c)^2}-\frac{1}{4}+\frac{z}{l_c}+\sum_{i=2}^{7}\alpha_i\big(\frac{z}{l_c}\big)^i\Big] \tag{1.11}$$

which is known as the *inextensible worm-like chain*[29]. In the enthalpic regime a stretching parameter K_0 needs to be introduced, as was first conceived by Odijk [36], and inspired by Wang [37]. In practice this implies replacing all z/l_c terms in eqn. 1.11 by $z/l_c - F/K_0$, leading to the expression

$$F(z)=\frac{k_B T}{L_p}\Big[\frac{1}{4(1-l)^2}-\frac{1}{4}+l+\sum_{i=2}^{7}\alpha_i l^i\Big] \tag{1.12}$$

with $l = z/l_c - F/K_0$, which is also known as the *extensible* WLC model. An explicit approximation for the extension under applied load also exists for the high-force regime ($10 < F < 60$ pN):

$$\frac{z}{l_c}=1-\left[\frac{l_c F}{k_B T}\right]^{-\frac{1}{2}}+\frac{F}{K_0} \tag{1.13}$$

1.4.5. MULTIPLEXING: CAN WE BRIDGE THE GAP BETWEEN SINGLE EVENTS AND ENSEMBLE AVERAGES?

As stated above, measuring the behavior of a single-molecule comes with the limitation that observation of a single event provides poor ground to infer all possible behaviors or the behavior of a population (ensemble) in a statistically solid way. Or, stated differently: the disadvantage of measuring on single molecules is, well, the fact that you are only measuring single molecules... To gain statistically solid data, the solution is either to measure a single molecule for extended periods of time, or measure many single molecules at a time. Gaining statistics in the former manner is time-consuming at the very least, yet often plain impossible to combine with a successful academic career.

Measuring multiple molecules simultaneously however – i.e. multiplexing – is an approach MT is very well equipped to do. The recorded image of a field of view size at a given magnification has increased over the past years as camera technologies improved (Fig. 1.9). Though now typically spanning stretches of flow cell surface in the order of 300 × 400 μm, the applied force on the beads over the field of view area (i.e. the x,y plane) is still considered to be homogeneous, as the curvature in the direction of the magnetic field[30] – a direction parallel to the x,y plane – spans millimeters. The homogeneity of the magnetic force on the beads across a typical MT field of view implies that - bead tracking algorithms permitting - the number of identical single-molecule events that can be observed simultaneously is far greater than one.

---

[28] The order of the polynomial is arbitrary, but chosen as this leads to an error smaller than 0.01% over the measured extension range.

[29] Note that the extension is expressed as a fraction of the contour length.

[30] a curvature as a result of finite magnet pair size

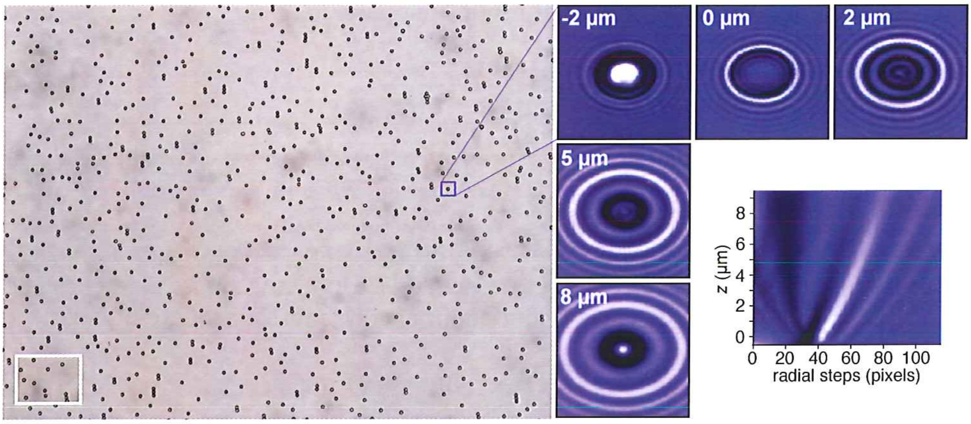


Figure 1.9: **The field of view during a multiplexed MT experiment.** Left:The size of a current field of view (2014), compared to a field of view at the same magnification in 2009 (inset). Right: Beads are tracked in x, y, and z by recording the diffraction pattern of the bead in the x, y plane in a region of interest centered around a bead, and comparing that to a pre-made z-stack of the change in diffraction pattern over well-defined changes in the height of the objective focal plane (i.e. a look-up table).

In this light, it is hard to overestimate how the efforts of Jelmer Cnossen *et al.* have contributed to the ability of gaining the statistically solid datasets presented throughout most of this thesis [33]. In short, the CUDA-based algorithms of Cnossen *et al.* have made it possible to move on from tracking a few beads in real-time (online) or tens of beads offline, to tracking hundreds of beads simultaneously online, at frequencies in the order of ~ 100 Hz. It is these efforts that pave the way towards experiments that simultaneously capture single-molecule events as well as ensemble-averaged population behavior.

1.5. QUANTITATIVE APPROACHES IN BIOPHYSICS

A central, unifying notion to the field of (single-molecule) biophysics[31] is the aim of describing biological phenomena in terms of fundamental physical parameters, and the use of these parameters to quantify the probability[32] of observed phenomena. Though most quantitative methods used throughout this thesis require no further introduction than what is stated in the respective chapter, there are a few guiding and recurring principles/concepts of particular interest that deserve a short, additional introduction.

1.5.1. BOLTZMANN DISTRIBUTIONS

A Boltzmann distribution is the probability distribution over the states of an equilibrated system. Whether it is the probability of finding a reversibly folding DNA hairpin under load in the closed or open state in Chapter 5, or the probability of finding Tus-*Ter* in the fully locked state in Chapters 8 and 9, all state i probabilities p_i can be expressed as a

---

[31] or rather perhaps the physical sciences in general

[32] as stochastic behavior

function of the state free energy E_i and absolute temperature T as

$$p_i = \frac{e^{-E_i/k_B T}}{\sum_{i=1}^{M} e^{-E_i/k_B T}} \tag{1.14}$$

with total number of possible states M. The negative term in the exponent implies that a system with the lowest free energy will always have a highest probability of being occupied. Which state actually has the lowest energy depends not only on entities as the temperature and pressure, but also on the mechanical force acting on the system, as the field of single-molecule biophysics has amply shown.[38] The indivitual discrete states can be seen as corresponding to minima in a continuous energy landscape (**Fig. 1.10a,b**).

1.5.2. THE ENERGY LANDSCAPE AND THE EFFECT OF FORCE

A reaction free-energy landscape is the portrayal of a (bio)chemical reaction along a specific reaction coordinate – a physical length scale – in terms of the continuous free-energy energy associated with any position along this coordinate. Free-energy minima correspond to discrete states, and the intermediary maxima correspond to the (unstable) transition states[33] (**Fig. 1.11b**). As stated earlier, all (bio)chemical reactions can be described by the magnitude and the direction of the forces that molecules exert on each other (see section 1.3.3). This implies that force can be a parameter that can be used to determine the fate of a chemical reaction, just like e.g. temperature can.[34]

Application of force along this reaction coordinate 'tilts' the energy landscape by adding the work term $F\Delta x$ to the free-energy (**Fig. 1.10b**). The application of a force thereby changes the relative free-energy of the states and the transition state, thus affecting the populations (i.e. the state probabilities) of the former, as well as the rates between the states.[38] The probability of a state is given by the Boltzmann weights, which in the case of a two-state system becomes:

$$p_2 = \frac{e^{-E_2/k_B T}}{e^{-E_1/k_B T} + e^{-E_2/k_B T}} = \frac{1}{1 + e^{(E_2 - E_1)/k_B T}} \tag{1.15}$$

where $E_2 - E_1 = \Delta G = F\Delta x$. Expressed as a function of the equilibrium force F_{eq} the probability of the high-energy state then becomes $p_2 = 1/(1 - \exp[(F_{eq} - F)\Delta x/k_B T])$ (**Fig. 1.10c**).

1.5.3. DISSECTING AND QUANTIFYING SINGLE POLYMERASE ACTIVITY

Despite what we can infer from reaction equilibria, biological processes are generally not in equilibrium. The movement of a polymerase along a nucleic acid template is an example of such a non-equilibrium system. Polymerase activity – i.e. the movement along a nucleic acid template while lacing nucleotides to an ever-growing complementary strand – is stochastic in nature, yet can be tracked using single-molecule techniques

---

[33] An 'uneasy' conformation often denoted with the double dagger (‡) symbol. A macroscopic example of a 2-state reaction with a high-energy, unstable intermediate state would be getting a stiff ruler bent one way (state 1) – by 'squeezing' the two ends between your hands – to bend the other way (state 2) without changing the distance between your hands.

[34] For the effect of temperature on the free-energy landscape of a two-state system, see Chapter 5.

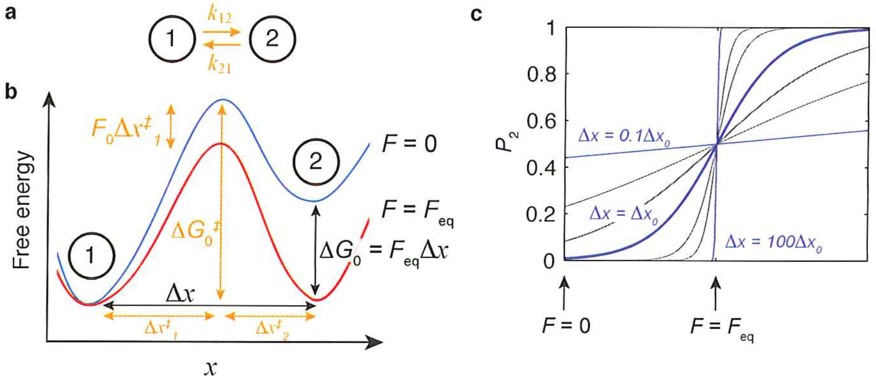


Figure 1.10: **Using mechanical force as a reaction coordinate in (bio)chemical reactions.** **a)** The equilibrium of a (first-order) chemical reaction between molecular species 1 and 2 will be determined by the rates of formation of product 1 (k_{12}) and 2 (k_{12}). **b)** This can be visualized as the energy of the species as it physically transforms from one conformation to another, inevitably through the formation of an unfavorable transition state. If one species has a higher energy (species 2, blue), application of force can compensate for this energy difference, thereby 'tilting' the energy landscape to e.g. such an extent that both states have equal energies (red). **c)** The ratio of species 1 and 2, as given by the Boltzmann probability distribution, is in other words, influenced by mechanical force, yet the extent to which force has effect heavily depends on the distance Δx between the two states (In the limit of $\Delta x \rightarrow 0$ the states become equally probable, as there is no longer a physical distance: species 1= species 2. In the limit of $\Delta x \rightarrow \infty$, a gradual and continuous (measurable) transition becomes very improbable: For $\Delta x < F_{eq}$, $P_2 = 0$ and $P_2 = 1$ otherwise. As we can observe in our experiments with long DNA hairpins (e.g. Chapters 4 and 8). In single-molecule force spectroscopy experiments, the energy landscape parameters in black can be obtained with time-averaged data, parameters in orange require time series analysis.

to yield discrete trajectories. The polymerase trajectory along the template will be a product of a number of states the polymerase can find itself in. The Gillespie algorithm is known to give exact statistics[35], thus can be used to model polymerase behavior.[39] As stochastic polymerase trajectories must be a result of a specific number of polymerase states, a probability distribution of dwell times[36] arising from the time a polymerase spent dwelling at each discrete position along the template will be composed of a specific number of exponential distributions.[37] In other words, as long as we have discrete single-molecule trajectories that yield a distribution of dwell times, we can find stochastic equations – i.e. kinetic models – that 'fit' the polymerase movement along a nucleic acid template. Choosing which model is the correct representation of polymerase behavior however, will inevitably depend on evidence other than dwell time distributions alone, as chapter 6 will show.

Chapters 6 and 7 quantify the movement of the viral RdRP P2 of bacteriophage $\phi 6$ by

---

[35] By randomly drawing a number of discrete rates from a same number of exponential distributions.

[36] This time will be long when the polymerase is paused (i.e. not moving along the template) and relatively short for normal active polymerization (i.e. stepping from one bp to the next).

[37] As noted before, the exponential distribution is the only distribution that is a result of a process occurring at a constant, single rate

fitting a kinetic model that has a discrete number of (pause) states to a probability distribution of dwell times originating from a collection of single-molecule trajectories (i.e. traces).[38] Transitions between the states are characterized by rates which, given a model, can be obtained by fitting the model to the dwell time distribution (DTD). Knowing the rates, one can then calculate the probability of each state.[40] A discrete stochastic simulation of a single (simplified) polymerase along a nucleic acid template – using the Gillespie algorithm and the most simple kinetic polymerase stepping model devisable – can visualize the basic features of the DTDs encountered in this thesis.

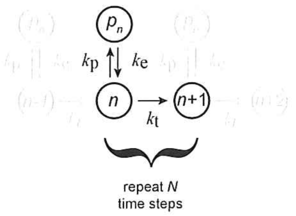


Figure 1.11: A kinetic model of a polymerase progressing along a nucleic acid template from basepair n to $n+1$ at a constant rate k_t, while there is an ever present chance of entering a pause state at a rate k_p, implying these two rates are in direct competition with each other. If paused, exiting the pause will occur at a rate k_e. Parallel processes (k_t, k_p) will give rise to a single exponential distribution of dwell times, serial processes will give rise to multiple exponentials. Note that single exponential distributions arising from discrete steps will only appear as such if each position along the template can be resolved exactly. Otherwise a process of multiple serial steps seen as one (e.g. the resolution is 10 bp) will yield a gamma (Γ) distribution with a width that corresponds to the number of underlying steps. We will further elaborate on this in chapter 2.

At each position, the polymerase must be in one of two states: an active polymerization state n, or a pause state p_n. When in the active state, the polymerase can make a transcription step (i.e. incorporating a nucleotide and moving along to the next base along the template) or switch to the pause state. When in the pause state it can only switch back to the active state. This implies 'choice' is imposed on the polymerase for each iteration in the active state only. The time between each transition is random and without memory, and drawn from exponential distributions: one distribution characterized by the relevant transition rates. This yields three rates for the polymerase stepper model: a transcription rate k_t, a rate of pause entry k_p, and a pause exit rate k_e. When in the active state, the probability of either transcribing (p_t) or pausing (p_p) is a result of a competition between the pause entry rate and the transcripiton rate, such that $p_t = k_t/(k_t + k_p)$ and $p_p = 1 - p_t = k_p/(k_t + k_p)$. This probability can be modeled by random selection between state probabilities p_t and p_p when in the active state, in the pause state it always goes back to the active state. Time intervals are drawn from exponential distributions with characteristic lifetimes $(k_t + k_p)^{-1}$, and k_e^{-1}, respectively.

If we start off with three polymerases - A, B, and C - with identical stepping rates k_t of 150 bp/s, identical pause exit rates k_e of $0.1k_t$, yet different pause rates: $k_p(\mathrm{A}) = 0.2k_t$,

---

[38] For those unfamiliar with dwell time distributions, I refer to Appendix A of this thesis.

$k_p(\mathrm{B}) = 0.05k_t$, and $k_p(\mathrm{C}) = 0.02k_t$, we obtain three discrete traces with differing apparent linear stepping velocities over long trajectories (**Fig. 1.12a**). If we zoom in by decreasing the number of iterations of the algorithm[39], it becomes clear that there is no such thing as a linear stepping velocity (**Fig. 1.12b**), yet that there merely is an average rate $\langle k_{\mathrm{tot}} \rangle$ for single traces over extensive periods of time. Typical single-molecule traces cover lengths ranging from 0.1 to several kb (**Fig. 1.12c**). If the short experiment is repeated often enough, we obtain an ensemble-averaged stepping rate k_{tot} (**Fig. 1.12d**, $N = 5000$) equal to the average rate of a single trace.[40] This average rate is typically the only information accessible with classical bulk experiments, so how do we best extract the additional information that single-molecule traces have? How do we find the actual stepping rate $k_t = 150$ bp/s?

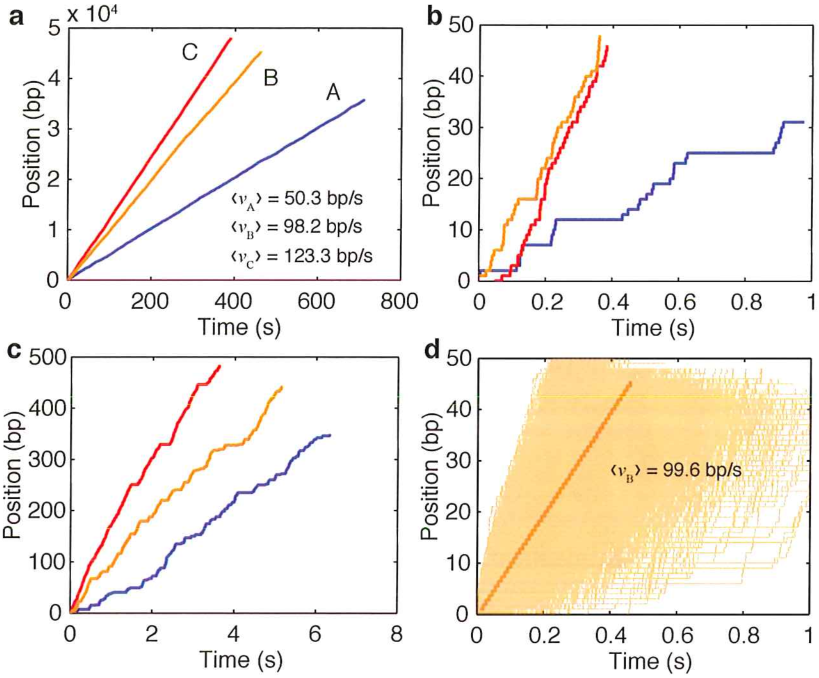


Figure 1.12: Discrete trajectory outcomes of the two-state polymerase model (Fig. 1.11) simulated using the Gillespie algorithm. (See text for explanation.)

Calculating the average velocity actually becomes complex surprisingly fast. Even for this simple polymerase model, the inverse of the average dwell time probability $P(\tau)$ has an exact solution that is too complex to serve as a clarifying example here. It becomes

---

[39] In single-molecule MT experiments, this would imply either a decreased DNA/RNA construct length, a decrease in polymerase processivity (which is a function of the competing rates of stepping and enzyme dissociation), or both.

[40] Ergodicity, visualized.

complicated because the rates k_t and k_p are in direct competition with each other such that it is in principle possible for the polymerase to, by chance, enter and exit the pause state forever without moving a single step forward. For a single-step model (with rate k_t) though, the probability of a dwell time τ at each step has the Arhennius-like form of:

$$P(\tau) = k_t \mathrm{e}^{-k_t \tau} \tag{1.16}$$

If we consider the limiting case of having a pause entry rate much smaller than the the stepping rate ($k_p \lll k_t \rightarrow$ either 0 or 1 pause at each step) and a pause exit rate much smaller than the stepping rate ($k_e \lll k_t \rightarrow$ if paused, $\tau_{\text{step}} \approx k_e^{-1}$), a simplified version of the dwell time probability $P(\tau)$ becomes:

$$\begin{aligned} P(\tau) &= p_{\text{step}} \cdot p(\tau_{\text{step}}) + p_{\text{pause}} \cdot p(\tau_{\text{pause}}) && (1.17) \\ &= \frac{k_t}{k_t + k_p} \cdot k_t \mathrm{e}^{-k_t \tau} + \frac{k_p}{k_t + k_p} \cdot k_e \mathrm{e}^{-k_e \tau} && (1.18) \end{aligned}$$

If the polymerase moves along without pausing, the dwell times will be short and will have a lifetime corresponding to the inverse of the overall stepping rate $\tau_{\text{step}} = (k_t + k_p)^{-1}$. If the polymerase was paused, the dwell time will be longer and have a lifetime corresponding to the (inverse of) the pause exit rate $\tau_{\text{pause}} = k_e^{-1}$. Given enough dwell times and a big enough difference between the rates of pausing and stepping[41], we can plot the distribution of dwell times as a probability density[42] for polymerases A, B, and C discussed above (**Fig. 1.13a**). Here we see a collection of exactly two exponential distributions[43] – i.e. two kinetic states – brought forward by our steppers! The exponential distribution that short dwell times has a lifetime corresponding to the stepping rate (τ_{step}), and the second distribution that covers longer dwell times has a characteristic lifetime τ_{pause}.[44] Qualitatively we can see[45] that it is the pause state probability that is influenced by changing the pause entry rate k_p (**Fig. 1.13a**) and that the pause length is equal to the pause exit rate k_e (**Fig. 1.13b**). A more complex pattern can also be observed, if we for instance decrease the stepping rate to similar orders of magnitude as the pause rates (**Fig. 1.13c**). Fitting the model to our DTD datasets[46] allows us to extract the rates from the DTDs. We can see that this works as the found rates are in good agreement with the initial rates the simulation was performed with (**Fig. 1.13d**).[47]

Reality is always more complex of course, as chapters 2, 6, 7, and – to a certain extent – chapter 8 will show, discuss, and elaborate on. First of all, in the simulation we know the number of states and rates beforehand. This does not hold true for real polymerases. In

---

[41] Which in reality holds true as pauses are infrequent as they are a result of something out of the ordinary happening, like the incorporation of an incorrect nucleotide (e.g. the polymerase error rate).

[42] This is normalizing to histogram bin height and width, see Appendix A.

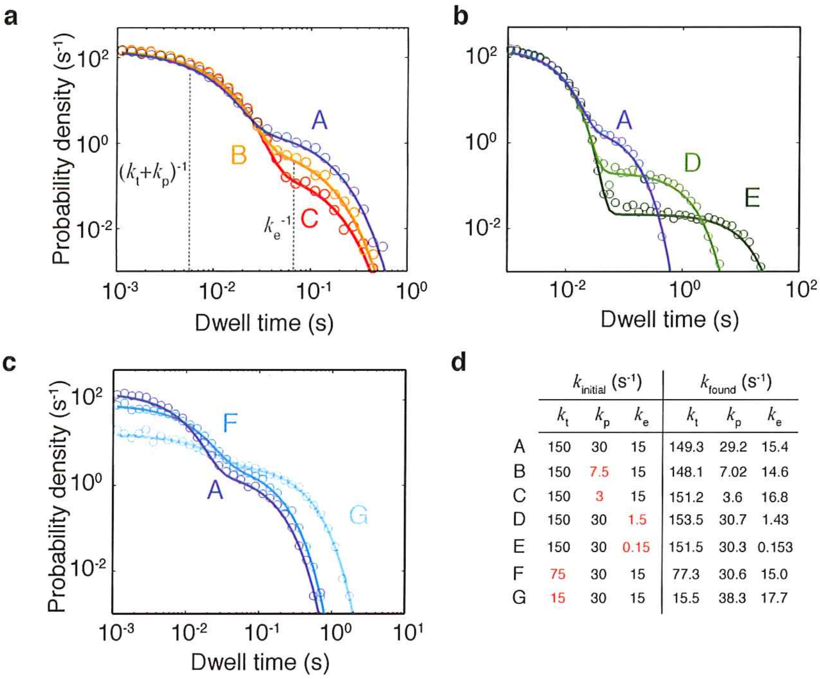
[43] On a log-log scale, an exponential distribution has the form of an arc, we plot it like this because of the orders of magnitude the data spans in (p, t) space

[44] This beats ensemble averages on so many levels.

[45] and intuitively understand

[46] Using maximum likelihood estimation, this is discussed in chapter 6.

[47] These found rates are in no way statistically solid! There will be variation in the fit: we can get confidence intervals on the fit parameters by repeating this experiment many times, or by bootstrapping the datasets.



	k_{initial} (s^{-1})			k_{found} (s^{-1})		
	k_t	k_p	k_e	k_t	k_p	k_e
A	150	30	15	149.3	29.2	15.4
B	150	7.5	15	148.1	7.02	14.6
C	150	3	15	151.2	3.6	16.8
D	150	30	1.5	153.5	30.7	1.43
E	150	30	0.15	151.5	30.3	0.153
F	75	30	15	77.3	30.6	15.0
G	15	30	15	15.5	38.3	17.7

Figure 1.13: **Dwell time distributions extracted from discrete trajectories (10^4 dwell times per distribution) of the kinetic polymerase model of figure 1.11. a-c)** A maximum likelihood fit (solid lines) of the model to the data (circles) shows that the found rates (k_{found}) are a good approximation of the exact rates (**d**, k_{initial}) used for the simulation.

addition, the number of states are meaningless if we are unable to clarify the physical origin of the underlying rate limiting process. For given a model with enough states, we can fit any distribution of dwell times.[48] Secondly, a simulated trajectory yields the *exact* position along the template for each moment in time, thus has unlimited resolution. The Brownian motion dominated world of single-molecule polymerase has of course a limited resolution (**Fig. 1.14**). This is the result of the extent to which spatiotemporal[49] camera specifications[50] and magnification strength allow for gathering enough information per unit time to average out non-specific errors introduced by Brownian motion (thermal noise).[51] In addition, averaging ideally only makes sense if it is done for the time the polymerase is on an unchanging position along the template. Instruments also suffer from signal drift ($1/f$ noise)[52] which, like polymerase movement, generally sets an

---

[48] Or, as John von Neumann allegedly put it: "With four parameters I can fit an elephant, with five I can make him wiggle his trunk."

[49] i.e. spatial and temporal resolution of the camera – given by the number of pixels (and size) and the maximum obtainable acquisition rate, respectively.

[50] and, related to that, the strength of the light source and light-sensitivity of the camera

[51] Given that the frequency spectrum of the beads has a high enough bandwidth such that averaging at high acquisition rates actually helps to reduce the noise.

[52] Which sometimes resembles polymerase activity to a remarkable extent...

upper limit to the timespan that averaging improves the resolution[53].

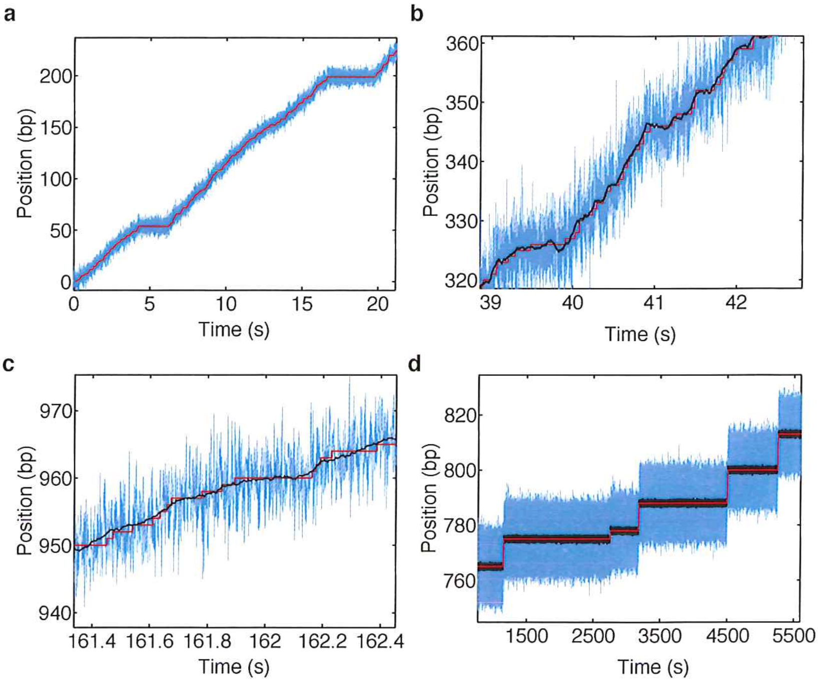


Figure 1.14: **Polymerase trajectories in a (drift-free) Brownian motion dominated world. a)** Addition of noise to a discrete simulated trajectory (red) yields a polymerase trajectory very similar to the traces obtained in an MT experiment (blue, the noise is Gaussian with a standard deviation of 4 bases, the sampling frequency simulated to be 1 kHz). **b,c)** So the noise obscures the data to such an extent that the exact position (red) is obscured. Filtering the data improves the resolution (black, 100 ms moving average – or 10 Hz low pass filter), though the exact position will not be known with 100% certainty. **d)** Finding the exact position with almost 100% certainty becomes possible for either very long dwells (e.g. slow stepping or long pausing (shown here)), or an acquisition rate many orders of magnitude faster than the stepping rate along the template. Note that in MT experiments, the bead movement is skewed (i.e. the noise not entirely Gaussian) as tethering sets a more stringent boundary to the upward directed motion of the bead than the upward pulling force sets to the downward bead motion.

Thirdly, even with unlimited resolution it will be impossible to state with absolute certainty that a certain event is the result of stochastic process a or b if the underlying rates of a and b occur on similar timescales (as e.g. starts occurring in **Fig. 1.13c**). In other words, given a DTD, we can get the most likely model parameters[54]. Whether the model is correct should depend on other arguements than the 'goodness of fit' and final conclusions will always be limited to a statement on the likelihood of the model being correct. This because the states in the model are the product of underlying distributions that can

---

[53] A way to quantify the effect of averaging thermal and $1/f$ noise is the Allan variance, discussed in Chapter 3.

[54] This is where the very useful Bayesian approach to statistics comes in.

(and will) overlap. There is a clear parallel with the macroscopic world here: life is a continuous grey-scale, not a discrete black-or-white phenomenon.

1.6. THIS THESIS

The first part of this thesis **Methods for multiplexed magnetic tweezers** (**Chapters 2–5**) is what can be seen as a collective effort towards improving both the quality and quantity of single-molecule information extracted from DNA-protein or DNA-enzyme interactions per magnetic tweezers experiment. The ensuing chapters (**Chapters 6–9**) are the fruit of these efforts, as we study and discuss the polymerase activity of bacteriophage ϕ6 RdRP P2 in **part II, RNA polymerases** as well as the sequence of events that lead to blocking of strand separation by a Tus protein at a specific *Ter* DNA site in **part III, DNA roadblocks**. Both systems would not have revealed their respective intricate mechanisms without the ongoing developments in MT multiplexing capacity.

In **Chapter 2** we summarize and discuss the developments in high-throughput measurements that lead to the successful completion of the P2 and Tus studies in chapters 6–8. The chapter introduces the concepts of dwell time distributions and subsequent interpretation more in the specific case of real life single-molecule experiments than is done in this chapter.

Chapter 3 describes the efforts of improving the tether strength and lifetime under an applied force by covalently anchoring DNA constructs to the magnetic beads and flow cell surface while keeping the non-specific adhesion of beads to the surface to a minimum, in an ever continuing effort to minimize tether loss.

Chapter 4 basically presents the myriads of refinements in experimental protocols developed over the course of the P2 and Tus-*Ter* projects. Here we present the experimental procedures in a detailed, step-by-step fashion, starting at the intricacies of DNA construct design and construction, and on through experimental preparation, with a focus on how to maximize the throughput of MT at high-force (> 20 pN) conditions.

In **Chapter 5** we demonstrate of a home-built MT temperature control system by measuring the temperature-dependence of the energetics of reversible unfolding of a short DNA hairpin. We also present the calibration and accuracy of the temperature control system in detail.

As touched upon several times throughout this chapter, **Chapter 6** elaborates on the intricacies of P2 RdRP transcription activity by creating dwell time distributions from extensive datasets comprised of of single-molecule polymerase trajectories along a dsRNA template. We propose a kinetic model for the bacteriophage polymerase based on the extensive amount of structural information available as well as trends in extracted rates and state probabilities as we perform a series of force and concentration sweeps. We also investigate the effect of the nucleotide analogs and an antiviral drug on P2 activity.

In **Chapter 7** we present a more obscure feature of P2 activity: backtracking. Because of the nature of the RNA tether design, we are able to gain some insight into the specifics of this phenomenon.

Chapter 8 encloses what has been the main effort of my graduate studies. A hypothesis dating from 2006 suggested that Tus-*Ter*, a protein-DNA complex found to play a role in the termination of DNA replication in *E. coli*, blocks replisome progression by forming a tight interaction between a *Ter* DNA base and specific amino acid residues of Tus that is induced by strand separation invoked by the approaching replication machinery.[41] As this hypothesis was surrounded by controversy (others favored a replisome-Tus interaction mechanism) and MT is a technique well suited to invoke the force-induced strand separation of DNA hairpins that mimicks replisome progression, we set out to test the strand-separation hypothesis. Soon enough we discovered that the Tus-*Ter* lock forms readily in hairpin unzipping experiments, such that we set out to characterize Tus-*Ter* lock formation. This chapter presents a detailed analysis of the Tus-*Ter* lock by measuring the lifetime of the lock under an applied constant force for a range of experimental conditions, Tus mutants, and *Ter* variants. We validate the 2006 hypothesis, thereby providing a clear-cut answer to an existing controversy. However, new controversies emerge as two other single-molecule studies shed light on different aspects of the Tus-*Ter* system.[42, 43]

This is why I set out to re-evaluate the status of the Tus-*Ter* in light of the findings presented by the three single-molecule studies in **Chapter 9**. The three studies turn out to be complementary in many aspects, allowing us to gain additional insights by combining the information available. I touch upon the remaining outstanding questions surrounding the Tus-*Ter* system and its role in *E. coli* as well as the technical challenges lying ahead.

Though having worked with single-molecule magnetic tweezers for four consecutive years still makes me a rookie in many ways, I will use the concluding **Chapter 10** to express my view on the current position of this technique in the field of single-molecule biophysics as well as in the field of biology in general – and speculate about its position in the future.

REFERENCES

[1] J. D. Watson, F. H. Crick, *et al.*, *Molecular structure of nucleic acids,* Nature **171**, 737 (1953).

[2] W. G. Scott, J. T. Finch, and A. Klug, *The crystal structure of an aii-rnahammerhead ribozyme: a proposed mechanism for rna catalytic cleavage,* Cell **81**, 991 (1995).

[3] T. A. Steitz, *A structural understanding of the dynamic ribosome machine,* Nature Reviews Molecular Cell Biology **9**, 242 (2008).

[4] T. A. Lincoln and G. F. Joyce, *Self-sustained replication of an rna enzyme,* Science **323**, 1229 (2009).

[5] F. Crick *et al.*, *Central dogma of molecular biology,* Nature **227**, 561 (1970).

[6] S. J. Benkovic, A. M. Valentine, and F. Salinas, *Replisome-mediated dna replication,* Annual review of biochemistry **70**, 181 (2001).

[7] S. M. Hamdan and C. C. Richardson, *Motors, switches, and contacts in the replisome,* Annual review of biochemistry **78**, 205 (2009).

[8] I. J. Fijalkowska, R. M. Schaaper, and P. Jonczyk, *Dna replication fidelity in escherichia coli: a multi-dna polymerase affair,* FEMS microbiology reviews **36**, 1105 (2012).

[9] C. Feschotte, *Virology: Bornavirus enters the genome,* Nature **463**, 39 (2010).

[10] M. Beijerinck, *Concerning a contagium viwm fluidum as cause of the spot disease of tobacco leaves,* (1898).

[11] R. Rappuoli, H. I. Miller, and S. Falkow, *The intangible value of vaccination,* Science **297**, 937 (2002).

[12] M. E. Halloran, I. M. Longini, A. Nizam, and Y. Yang, *Containing bioterrorist smallpox,* Science **298**, 1428 (2002).

[13] Ø. Bergh, K. Y. BØrsheim, G. Bratbak, and M. Heldal, *High abundance of viruses found in aquatic environments,* Nature **340**, 467 (1989).

[14] S. Srinivasiah, J. Bhavsar, K. Thapar, M. Liles, T. Schoenfeld, and K. E. Wommack, *Phages across the biosphere: contrasts of viruses in soil and aquatic environments,* Research in microbiology **159**, 349 (2008).

[15] C. A. Suttle, *Marine viruses—major players in the global ecosystem,* Nature Reviews Microbiology **5**, 801 (2007).

[16] L. D. Simon and T. F. Anderson, *The infection of escherichia coli by t2 and t4 bacteriophages as seen in the electron microscope i. attachment and penetration,* Virology **32**, 279 (1967).

[17] G. M. Allan and J. A. Ellis, *Porcine circoviruses: a review,* Journal of Veterinary Diagnostic Investigation **12**, 3 (2000).

[18] J.-M. Claverie, H. Ogata, S. Audic, C. Abergel, K. Suhre, and P.-E. Fournier, *Mimivirus and the emerging concept of "giant" virus,* Virus research **117**, 133 (2006).

[19] N. G. Abrescia, D. H. Bamford, J. M. Grimes, and D. I. Stuart, *Structure unifies the viral universe,* Annual review of biochemistry **81**, 795 (2012).

[20] E. V. Makeyev and D. H. Bamford, *The polymerase subunit of a dsrna virus plays a central role in the regulation of viral rna metabolism,* The EMBO Journal **19**, 6275 (2000).

[21] M. R. Koivunen, L. P. Sarin, and D. H. Bamford, *Structure–function insights into the rna-dependent rna polymerase of the dsrna bacteriophage φ6,* Segmented Double-stranded RNA Viruses: Structure and Molecular Biology , 239 (2008).

[22] S. Duffy, L. A. Shackelton, and E. C. Holmes, *Rates of evolutionary change in viruses: patterns and determinants,* Nature Reviews Genetics **9**, 267 (2008).

[23] S. Crotty, C. E. Cameron, and R. Andino, *Rna virus error catastrophe: Direct molecular test by using ribavirin,* Proceedings of the National Academy of Sciences **98**, 6895 (2001), http://www.pnas.org/content/98/12/6895.full.pdf .

[24] M. Vignuzzi, J. K. Stone, J. J. Arnold, C. E. Cameron, and R. Andino, *Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population,* Nature **439**, 344 (2006).

[25] A. Einstein, *Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen,* Annalen der physik **4** (1905).

[26] J. D. Van der Waals, *Over de Continuiteit van den Gas-en Vloeistoftoestand,* Vol. 1 (Sijthoff, 1873).

[27] S. Arrhenius, *Über die reaktionsgeschwindigkeit bei der inversion von rohrzucker durch säuren,* Zeitschrift für physikalische Chemie **4**, 226 (1889).

[28] H. A. Kramers, *Brownian motion in a field of force and the diffusion model of chemical reactions,* Physica **7**, 284 (1940).

[29] E. Abbe, *Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung,* Archiv für mikroskopische Anatomie **9**, 413 (1873).

[30] C. Bustamante, Z. Bryant, and S. B. Smith, *Ten years of tension: single-molecule DNA mechanics,* Nature **421**, 423 (2003).

[31] T. Lionnet, A. Dawid, S. Bigot, F.-X. Barre, O. A. Saleh, F. Heslot, J.-F. Allemand, D. Bensimon, and V. Croquette, *Dna mechanics as a tool to probe helicase and translocase activity,* Nucleic acids research **34**, 4232 (2006).

[32] I. Vilfan, J. Lipfert, D. Koster, S. Lemay, and N. Dekker, *Magnetic tweezers for single-molecule experiments,* in *Handbook of Single-Molecule Biophysics* (Springer, 2009) pp. 371–395.

[33] J. P. Cnossen, D. Dulin, and N. H. Dekker, *An optimized software framework for real-time, high-throughput tracking of spherical beads,* Review of Scientific Instruments **85**, 103712 (2014).

[34] C. Bouchiat, M. D. Wang, J.-F. Allemand, T. R. Strick, S. M. Block, and V. Croquette, *Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements,* Biophysical Journal **76**, 409 (1999).

[35] J. F. Marko and E. D. Siggia, *Stretching dna,* Macromolecules **28**, 8759 (1995).

[36] T. Odijk, *Stiff chains and filaments under tension,* Macromolecules **28**, 7016 (1995).

[37] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, *Stretching DNA with optical tweezers,* Biophysical Journal **72**, 1335 (1997).

[38] I. Tinoco Jr and C. Bustamante, *The effect of force on thermodynamics and kinetics of single molecule reactions,* Biophysical chemistry **101**, 513 (2002).

[39] D. T. Gillespie, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions,* Journal of computational physics **22**, 403 (1976).

[40] M. Depken, E. Galburt, and S. Grill, *The origin of short transcriptional pauses,* Biophysical Journal **96**, 2189 (2009), cited By 44.

[41] M. D. Mulcair, P. M. Schaeffer, A. J. Oakley, H. F. Cross, C. Neylon, T. M. Hill, and N. E. Dixon, *A molecular mousetrap determines polarity of termination of dna replication in e. coli,* Cell **125**, 1309 (2006).

[42] M. M. Elshenawy, S. Jergic, Z.-Q. Xu, M. A. Sobhy, M. Takahashi, A. J. Oakley, N. E. Dixon, and S. M. Hamdan, *Replisome speed determines the efficiency of the tus-ter replication termination barrier,* Nature (2015).

[43] M. Pandey, M. M. Elshenawy, S. Jergic, M. Takahashi, N. E. Dixon, S. M. Hamdan, and S. S. Patel, *Two mechanisms coordinate replication termination by the escherichia coli tus–ter complex,* Nucleic acids research , gkv527 (2015).

I

METHODS FOR MULTIPLEXED MAGNETIC TWEEZERS ASSAYS

2

UNTANGLING REACTION PATHWAYS THROUGH MODERN APPROACHES TO HIGH-THROUGHPUT SINGLE-MOLECULE FORCE-SPECTROSCOPY EXPERIMENTS

> *An invasion of armies can be resisted,*
> *but not an idea whose time has come.*
>
> Victor Hugo

Single-molecule experiments provide a unique means for real-time observation of the activity of individual biomolecular machines. Through such techniques, insights into the mechanics of for example, polymerases, helicases, and packaging motors have been gleaned. Here we describe the recent advances in single-molecule force spectroscopy instrumentation that have facilitated high-throughput acquisition at high spatiotemporal resolution. The large datasets attained by such methods can capture rare but important events, and contain information regarding stochastic behaviors covering many orders of magnitude in time. We further discuss analysis of such data sets, and with a special focus on the pause states described in the general literature on RNA polymerase pausing we compare and contrast the signatures of different reaction pathways.

This chapter has been published as: David Dulin, Bojk A Berghuis, Martin Depken & Nynke H Dekker (2015) Untangling reaction pathways through modern approaches to high-throughput single-molecule force-spectroscopy experiments. *Curr Op Struct Biol* 34, 116-122 (2015). [1].

2.1. INTRODUCTION

THE correct readout, maintenance, repair, and replication of genomic information involves a stunning variety of carefully coordinated and regulated molecular machines, including key players such as RNA polymerases (RNAP) and replisomes. The progression of these machines is highly dynamic, as their activity is frequently and stochastically interrupted by numerous co-factors or intrinsic catalytic events. For example, during transcription, RNAP progression is interrupted by various types of pauses, including backtracking pauses resulting from RNAP diffusing upstream on the DNA template[2, 3], and regulatory pauses allowing for transcription-translation synchronization and RNA co-transcriptional folding[4–6]. During replication, the replisome also frequently halts, for example, for the formation of primers on the Okazaki fragment[7], due to DNA polymerase exchange[8], and possibly also due to proofreading[9–11].

To obtain quantitative insight into the functioning of molecular machines, probing at the single-molecule level with nanometre-scale spatial and millisecond temporal resolution has proved to be very successful[12]. This success results from the ability to detect transient intermediates or rare events that are masked when ensemble techniques are used. Single-molecule methods may be broadly grouped into optical methods visualizing individual molecules using fluorescence-based microscopy[13–15], and force-based methods such as atomic force microscopy, flow-based stretching, and optical and magnetic tweezers[16]. While both approaches have provided invaluable insight into the functioning of molecular machines, we here focus on the latter. Motivated by the sub-nanometer elementary length scales of biological substrates such as DNA, much effort has gone into developing instruments with high spatiotemporal resolution[12]. Provided sufficiently low enzyme kinetic rates (~1 bp/s), optical tweezers have been developed with single DNA base pair (bp) resolution[17]. Such tweezers have allowed detailed mechanistic studies of, for example, RNAp[18], the $\phi29$ packaging motor[19], and the Hepatitis C viral helicase NS3[20].

Though biological systems that catalyze chemical reactions along a single pathway[19, 21–23] have successfully been characterized with high-resolution optical tweezers, approaches suitable for analysis of more complicated pathways displayed remain a challenge due to the low yield of such technique. We here review recent approaches to parallelized single-molecule force-spectroscopy methods using widefield (e.g. camera-based) detection to simultaneously track the action of a large number of independent molecules and substrates (**Fig. 2.1**). For the case of polymerizing enzymes with an internal pause dynamics, we further discuss an analysis approach that capitalizes on large data sets to reveal information about the enzyme translocation pathway, and review the signatures of various pause types described in the literature.

2.2. SINGLE-MOLECULE APPROACHES GO PARALLEL

The simplest high-throughput single-molecule approach, typically denoted flow-stretch experiments, involves the use of liquid flow to exert drag forces on tethered beads (**Fig. 2.1a**). Flow-stretch experiments have the advantage of being relatively simple while still

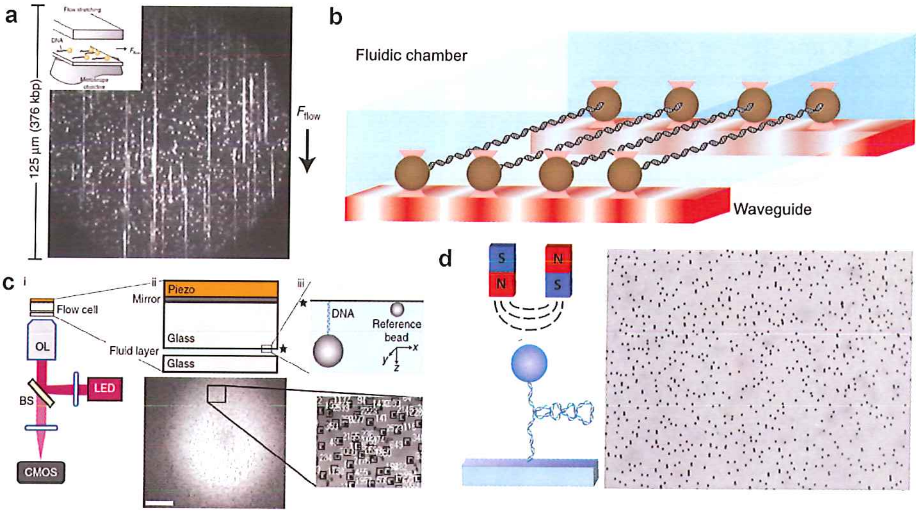


Figure 2.1: **Recent developments in high-throughput single-molecule force spectroscopy methods.** **(a)** A flow-stretch experiment where drag forces are applied to a DNA molecule by a sustained fluid flow, and the progression of leading strand replication is monitored through the length of the DNA tether between a bead and the surface. With this method it has been possible to observe, for example, primase induced pausing in lagging strand DNA synthesis[7] and coordination and exchange of DNA polymerases[8]. Figure panel adapted from [24] with permission. **(b)** In a nanophotonic standing-wave array trap assay, interference between two counter-propagating lasers beams produces a standing wave inside a wave guide and a linear array of optical traps in the juxtaposed flow chamber. By placing two waveguides in close proximity, it is possible to simultaneously trap and exert tension on an array of 'dumb-bell' geometries. **(c)** In the acoustic force spectroscopy assay, a piezo plate on the flow cell vibrates at MHz frequency, generating an acoustic wave that homogeneously pushes on beads tethered to the top surface of the flow-cell. The field of view is imaged on a CMOS camera, where thousands of beads can be followed with ~ 5 nm resolution. Figure panel adapted from [25] with permission. **(d)** A magnetic tweezers experiment that uses a pair of magnets to apply a force or torque to a tethered molecule. The magnetic tweezers assay is readily parallelized, as shown in the right image of hundreds of simultaneously tracked magnetic beads[26] in a homogenous force field[27–29].

allowing for the observation of enzymatic activity on long DNA templates (>40 kb) at about a 100 bp resolution. This approach has been extensively used to study the dynamics of replication in the bacteriophage T7 model system[24, 30]. A more recent approach, denoted nanophotonic standing-wave array trap (nSWAT), uses microfluidics combined with waveguiding to create an array of optical traps[31] (**Fig. 2.1b**). Being integrated into a micro-fabricated chip, the nSWAT assay substantially reduces the complexity of optical tweezers and the influence of mechanical noise. Another recent approach is acoustic force spectroscopy (AFS), which uses a flow-cell integrated piezo element to apply acoustic pressure to micron-sized beads[25] (**Fig. 2.1c**). AFS is ideal for high-throughput experiments, as it generates a homogenous force over distances much larger than the sizes of the force transducers. For the same reasons, magnetic tweezers (MT) are also inherently suitable for parallelization[26] (**Fig. 2.1d**). The capabilities of a widefield camera-based imaging approach combined with MT were first demonstrated by following either tens of molecules in real-time[32], or hundreds with post-processing image analysis[33]. The potential for multiplexing is further expanded by a recent increases in camera resolution (**Fig. 2.1d**), and the use of graphics processing units interfaced with the CUDA-based parallel computing framework[26, 34, 35]. It is now possible to simultaneously follow up to 800 beads in real-time, with a resolution of 1 nm at 25 Hz[26].

Further development of high-throughput experiments will greatly help the mapping of single-molecule events that are either very long-lived or rare. We now discuss how to visualize such data, and illustrate the two main scenario where multiplexing becomes important by discussing two recent studies utilizing high-throughput MT approaches to study DNA binding proteins and polymerase activity.

2.3. DWELL TIME DISTRIBUTIONS

Dwell time analysis has a long history of being applied to single-molecule experiments where it is possible to directly measure the duration (the dwell time) of the event of interest[23, 36]. With the increase in the type of events that can be studied with statistical significance in multiplexed force-spectroscopy experiments, it becomes important to be able to visualize the data over many orders of magnitude in time and probability. For such visualizations it is especially useful to collect dwell times in a weighted histogram, where the score in each bin is normalized with the bin width and total number of dwell times recorded. Such weighted histograms represent empirical dwell time distributions (DTD). The basic shape of an empirical DTD is insensitive to the binning used, which makes it possible to consistently visualize the data over several orders of magnitude by using various binning schemes (e.g. using log-scaled bins).

2.4. MAGNETIC TWEEZERS-BASED STUDY OF DNA-BINDING PROTEINS CAPITALIZES ON IMPROVED STATISTICS

High throughput proved essential in a recent study probing *Escherichia coli* replication fork arrest induced by the tight interaction between the DNA binding protein Tus and its cognate DNA-binding sequence *Ter*[37]. It had previously been argued that the tight interaction was dependent on specific interactions between Tus and the replisome[38]. By using a multiplexed MT DNA hairpin assay (**Fig. 2.2a**) to mimick fork progression without the proteins of the repliosome, it was shown that protein–protein interactions were not necessary for tight Tus–*Ter* interactions. Gathering the necessary dwell-time statistics would have been very demanding in time without multiplexing, as strand separation frequently remained blocked at the Tus–*Ter* site for hundreds of seconds (**Fig. 2.2b**).

2.5. MAGNETIC TWEEZERS-BASED STUDY OF POLYMERASE ACTIVITY CAPITALIZES ON IMPROVED STATISTICS

A recent study of viral mutagenesis used high-throughput magnetic tweezers to examine rare nucleotide misincorporation events during replication[39] (**Fig. 2.1d**). In this work, the progression of a viral RNA-dependent RNA polymerase (RdRP) was monitored at the single-molecule level (**Fig. 2.2c**). Different traces taken during the same experi-

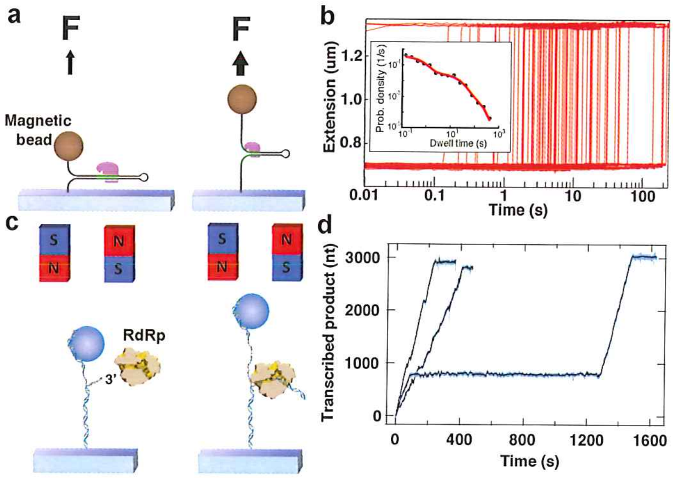


Figure 2.2: **High-throughput single-molecule assays and the resulting multiplexed datasets.** **(a)** In the Tus–*Ter* experimental assay, the DNA-protein lock is formed by force-induced DNA hairpin strand separation. The interaction strength is quantified by measuring the lifetime of the lock at a constant force. **(b)** This yields a dataset of dwell times until rupture events (i.e. full opening of the hairpin). The resulting dwell time distribution shows evidence of multiple exponential states, revealing the intermediate steps towards lock formation (inset). **(c)** A pair of magnets is used to apply a force to a double-stranded RNA molecule while an RdRP transcribes the RNA. As progression of the fork converts double-stranded to single-stranded RNA in the tether, transcription can be followed by monitoring the vertical position of the bead. Figure panel adapted from [39]. **(d)** Three traces of the transcription activity of an RdRP (blue points) acquired during the same experiment at an applied force of 20 pN and an NTP concentration of the order of mM, filtered with a 0.5 Hz low pass filter (black).

ment differ remarkably (**Fig. 2.2d**), and to characterize rare stable pauses large data sets were recorded at many experimental conditions. To analyze the data, a new approach based on the direct fitting of mechano-chemical models to complete data sets was used. We conclude by illustrating the basic ingredients of this approach, and comment on the signatures of the polymerase pause types found in the literature.

2.6. DWELL TIME BASED ANALYSIS FOR UNTANGLING MOLECULAR MOTOR TRANSLOCATION PATHWAYS

Standard approaches used to analyze single molecule traces require either extensive pre-processing of traces in the form of pause-picking algorithms[4, 18], or discards important information about pause durations through considering local velocities[6, 40]. Here we use simulations to outline a recent extension of dwell-time analysis that can be used to fit complete data sets directly to mechano-chemical models of translocating motors[39]. In this approach, the traces are not segmented with fixed time intervals as is done in velocity-based analysis methods, but are segmented with fixed translocation distances – the dwell time windows – to produce dwell time statistics. As one often lacks the resolution to gather dwell time statistics for single translocation steps, the data are instead collected over dwell time windows that can span multiple translocation steps[41]. We will here focus on pause analysis, and not discuss the delicate and largely unresolved influence of noise on the shorter time events, such as pause-free translocation. For illustrative purposes, we will assume that on the time scale of pauses it is possible to smooth the traces sufficiently to suppress the effect of noise.

In **Fig. 2.3a** we show a simulated example trace with a dwell-time window of 10 bp, and the series of scored dwell-times indicated as t_1, t_2, among others. Given the rates in any specific translocation scheme, one can in principle use first-passage time analysis[42] to calculate the DTD $P_1(t)$ of times t for taking one step. The DTD for taking n steps is then the n-fold convolution $P_n(t) = [P_1 * P_1 * \ldots * P_1](t)$ of the single-step distribution. The DTD can in principle be arbitrarily complicated, but we here focus on a few of the simplest important cases that result from translocation schemes suggested in the literature.

First order process. In the simplest scenario, each step along the elongation pathway is a first order process with a constant rate k (**Fig. 2.3b**, inset). The one-step DTD is then an exponential distribution, and the n step distribution is a Gamma distribution of order n and characteristic rate k. From **Fig. 2.3b** it is clear that the width of the Gamma distribution depends on the number of substeps, a fact that can be used to estimate n.

Off-pathway pauses. Polymerase elongation can be temporarily halted by thermally induced changes in enzyme structure[4, 18, 40], inducing what is referred to as off-pathway pauses. At its simplest, the structural change occurs with a first-order rate k_p, and is reversed with the rate k_e (**Fig. 2.3c**, inset). Though it is straightforward to calculate the DTD for arbitrary rates, again for simplicity, we focus on the case where pauses are entered infrequently in each dwell-time window, and typically last much longer than it takes to cross the dwell-time window without pausing. In this case, the first order process of the main elongation pathway captures most dwell-time windows, resulting in a Gamma distribution for short dwell times (**Fig. 2.3c**). For longer dwell times, infrequent entrance into the pause state results in exponentially distributed dwell-times with a lifetime $1/k_e$. The probabilistic weight under the short-time Gamma distribution and the long-time exponential distribution reports on the probability to move through a dwell-time window without pausing and the probability to enter at least one pause. Importantly, this type of pause will not have a lifetime that depends on nucleotide concentration, as it is exited through the reversal of the thermal fluctuation that triggered the pause. The probability of entering a pause will depend on nucleotide concentration though, as the pause entry directly competes with nucleotide addition.

On-pathway pauses. Another type of pause observed in the literature results from a stochastic and drastic reduction of the nucleotide addition speed, for example, after a non-cognate base has been inserted[39, 44, 45]. This leads to an on-pathway pause, which has a DTD similar to the off-pathway pause but with a pause-escape rate that is sensitive to nucleotide concentrations and an entrance probability that is independent on overall nucleotide concentration. Based on the different nucleotide dependencies of on-pathway and off-pathway pauses, it is possible to distinguish them using nucleotide

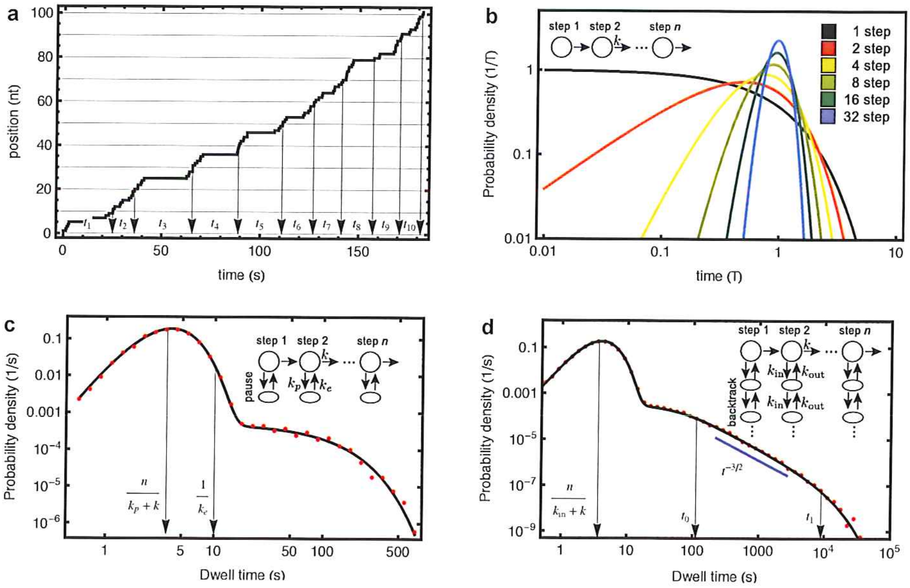


Figure 2.3: **Dwell-time distributions for translocating systems.** **(a)** Position versus time trace generated through a Gillespie simulation of a stochastic stepper with a simple pause (reaction scheme in inset of panel **(c)**). The times it takes to cross equal-sized dwell time windows $(t_1, t_2, t_3, \ldots)$ are scored as dwell times. **(b)** The distribution of dwell times generated from the translocation scheme sketched in the inset is given by the Gamma distribution. Keeping the average time τ it takes to cross the dwell time window constant, the width of the distribution decreases with the number of steps needed to cross the window. **(c)** For a translocation scheme with a rare and long off-pathway pauses (reaction scheme in the inset), the dwell time distribution will for short times follow a Gamma distribution centered on the average time it takes to translocate through a dwell-time window by n non-pausing steps $(n/(k+k_p))$. At long dwell times, the distribution will be dominated by the exponentially distributed pause-escape times (with the average $1/k_e$). The probabilistic weight of the Gamma and exponential-pause regions can be used to calculate the pause frequency. The basic shape of the dwell time distribution of on-pathway pauses is the same, while its dependence on, for example, nucleotide concentrations differ (see main text). **(d)** For a translocation scheme including rare backtracks (reaction scheme in the inset), we again have an initial pause-free region that follows a Gamma distribution. At later times, there is an algebraically distributed region decaying with the power $-3/2$, and bounded by the timescales $t_0 = (k_{\text{in}} k_{\text{out}})^{-1/2}$ and $t_1 = (k_{\text{in}}^{1/2} - k_{out}^{1/2})^{-2}$[43]. The red dots in panel **(c)** and **(d)** are generated by Gillespie simulations of the translocation schemes shown in the insets in the respective panels.

concentrations sweeps.

Hybrid pauses. A recent study[39] suggests that there exist off-pathway pauses, but where the paused state retains a degree of enzymatic activity. For such pauses, both pause probability and pause exit rate will depend on the overall nucleotide concentration.

Backtracking pauses. Another important pause is the backtracking pause, which results from the polymerase diffusively moving backwards and misaligning the product terminal with the active site. To reverse such a pause, the polymerase has to diffuse back to its original position[2, 3]. The diffusive return induces pauses with a distinct $-3/2$ power-law distributed dwell-times over a large time range[2, 43, 46] (**Fig. 2.3d**, inset). The width of this region is bounded by a lower corner time and upper cutoff time, both set by the hopping rates in the diffusive backtrack (**Fig. 2.3d**)[43].

Composite distributions. Experimental DTDs are generally more complicated than any of the above scenarios, both in that the timescales might not be well separated and that several types of pauses can be present at once. Still, the large data sets now becoming available through multiplexed experiments makes it possible to distinguish a large number of features (**Fig. 2.4**). Details regarding the precise reaction scheme can be extracted not only by noting that pause mechanisms can be differentiated by how their probabilities and lifetimes depend on nucleotide concentration, but also from force dependence, and from the introduction of non-native nucleotide analogs[39]. In **Fig. 2.4** we show the empirical DTD based on successively larger data sets collected from a viral RdRP[39] with several pause types. An increasingly significant and rich structure pertaining to long-time and rare events is seen as the data sets are enlarged. To fit models to

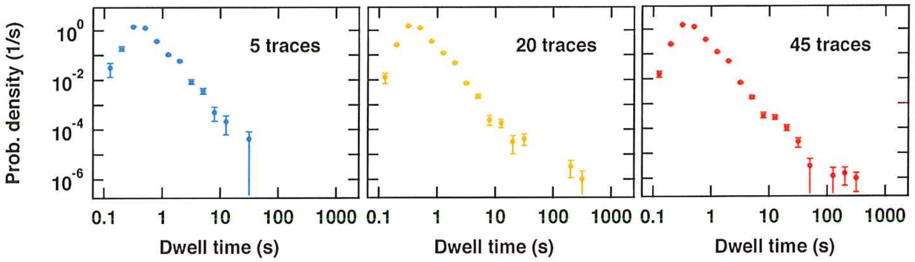


Figure 2.4: **Probability density distribution of the dwell times as a function of the number of P2 RdRp activity traces.** In the leftmost panel, the dwell-time distribution is composed of more than 1000 dwell times derived from 5 independent traces using a 10 bp dwell-time window. In this case, the short-time Gamma distribution, which represents around 95% of the events, can be clearly identified. In the center and right panels, the number of independent traces is increased to 20 (yellow data), and 45 (red data), respectively. Here, one can readily observe new features at long dwell times with increased statistics. These detailed probability density distributions can be described by a minimal composite model with two exponential pauses (dominating dwell times between 1 and 20 s), and a power law originating in a backtracked pause (dominating dwell times above 20 s)[39]. In all three panels, the dwell times are log binned, and the error bars on each bin represent the standard deviation as calculated from 1000 bootstrapped data sets.

DTD it is often convenient to determine the number of pause states through the Bayesian Information Criterion (BIC)[47], and fit out model parameters by Maximum Likelihood Estimation (MLE)[48] applied directly to the dwell times. At present, this approach requires at least a partial analytic solution for the first-passage time of the corresponding reaction scheme[39], though future developments of numerical approaches will likely eliminate this requirement.

2.7. OUTLOOK

THE approaches discussed here can be applied to many different types of molecular motors. For example, replication is regularly interrupted by rare and slow events, such as proofreading activity on the nascent DNA and polymerase exchange. A recent study has also demonstrated the utility of combining high-resolution single-molecule techniques with next generation sequencing technology[5] to report on the specific DNA sequence responsible for a regulatory transcriptional pause. With further progress in multiplexing and the ability to localize the absolute position of a molecular machine on its template, it should soon become possible to investigate sequence-dependent pausing patterns on a global scale.

2.8. ACKNOWLEDGEMENTS

We thank our current and former colleagues in the Nynke Dekker and Martin Depken groups of the Department of Bionanoscience, TU Delft for useful discussions, in particular Jelmer Cnossen, Behrouz Eslami-Mossallam, and Theo van Laar. We also acknowledge Elio Abbondanzieri for discussions on RNA polymerases, and Eric Snijder, Clara Posthuma, and Craig Cameron for discussions on RNA-dependent RNA polymerases. M.D. acknowledges early discussions with Stephan Grill and Eric Galburt regarding dwell-time distributions and polymerases. This work was financed by a TU Delft startup grant to M.D. and by VICI and TOP grants from the Netherlands Organisation for Scientific Research to N.H.D.

REFERENCES

[1] D. Dulin, B. A. Berghuis, M. Depken, and N. H. Dekker, *Untangling reaction pathways through modern approaches to high-throughput single-molecule force-spectroscopy experiments,* Current opinion in structural biology **34**, 116 (2015).

[2] E. Galburt, S. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, and C. Bustamante, *Backtracking determines the force sensitivity of rnap ii in a factor-dependent manner,* Nature **446**, 820 (2007), cited By 130.

[3] J. Shaevitz, E. Abbondanzieri, R. Landick, and S. Block, *Backtracking by single rna polymerase molecules observed at near-base-pair resolution,* Nature **426**, 684 (2003), cited By 215.

[4] K. M. Herbert, A. L. Porta, B. J. Wong, R. A. Mooney, K. C. Neuman, R. Landick, and S. M. Block, *Sequence-resolved detection of pausing by single {RNA} polymerase molecules,* Cell **125**, 1083 (2006).

[5] M. Larson, R. Mooney, J. Peters, T. Windgassen, D. Nayak, C. Gross, S. Block, W. Greenleaf, R. Landick, and J. Weissman, *A pause sequence enriched at translation start sites drives transcription dynamics in vivo,* Science **344**, 1042 (2014), cited By 18.

[6] B. Zamft, L. Bintu, T. Ishibashi, and C. Bustamante, *Nascent rna structure modulates the transcriptional dynamics of rna polymerases,* Proceedings of the National Academy of Sciences of the United States of America **109**, 8948 (2012), cited By 24.

[7] J.-B. Lee, R. Hite, S. Hamdan, X. Xie, C. Richardson, and A. Van Oijen, *Dna primase acts as a molecular brake in dna replication,* Nature **439**, 621 (2006), cited By 154.

[8] H. Geertsema, A. Kulczyk, C. Richardson, and A. Van Oijen, *Single-molecule studies of polymerase dynamics and stoichiometry at the bacteriophage t7 replication machinery,* Proceedings of the National Academy of Sciences of the United States of America **111**, 4073 (2014), cited By 4.

[9] B. Ibarra, Y. Chemla, S. Plyasunov, S. Smith, J. Lázaro, M. Salas, and C. Bustamante, *Proofreading dynamics of a processive dna polymerase,* EMBO Journal **28**, 2794 (2009), cited By 37.

[10] M. Manosas, M. Spiering, F. Ding, D. Bensimon, J.-F. Allemand, S. Benkovic, and V. Croquette, *Mechanism of strand displacement synthesis by dna replicative polymerases,* Nucleic Acids Research **40**, 6174 (2012), cited By 12.

[11] G. J. Wuite, S. B. Smith, M. Young, D. Keller, and C. Bustamante, *Single-molecule studies of the effect of template tension on t7 dna polymerase activity,* Nature **404**, 103 (2000).

[12] D. Dulin, J. Lipfert, M. C. Moolman, and N. H. Dekker, *Studying genomic processes at the single-molecule level: introducing the tools and applications,* Nat Rev Genet **14**, 9 (2013).

[13] B. Huang, M. Bates, and X. Zhuang, *Super-resolution fluorescence microscopy,* Annual Review of Biochemistry **78**, 993 (2009), cited By 457.

[14] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha, *Advances in single-molecule fluorescence methods for molecular biology,* Annual Review of Biochemistry **77**, 51 (2008), cited By 340.

[15] J. Lichtman and J.-A. Conchello, *Fluorescence microscopy,* Nature Methods **2**, 910 (2005), cited By 418.

[16] K. Neuman and A. Nagy, *Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy,* Nature Methods **5**, 491 (2008), cited By 603.

[17] J. Moffitt, Y. Chemla, S. Smith, and C. Bustamante, *Recent advances in optical tweezers,* Annual Review of Biochemistry **77**, 205 (2008), cited By 404.

[18] E. Abbondanzieri, W. Greenleaf, J. Shaevitz, R. Landick, and S. Block, *Direct observation of base-pair stepping by rna polymerase,* Nature **438**, 460 (2005), cited By 423.

[19] Y. Chemla, K. Aathavan, J. Michaelis, S. Grimes, P. Jardine, D. Anderson, and C. Bustamante, *Mechanism of force generation of a viral* DNA *packaging motor,* Cell **122**, 683 (2005), cited By 158.

[20] W. Cheng, S. Arunajadai, J. Moffitt, I. Tinoco Jr., and C. Bustamante, *Single-base pair unwinding and asynchronous* RNA *release by the hepatitis* C *virus* NS*3 helicase,* Science **333**, 1746 (2011), cited By 40.

[21] G. Chistol, S. Liu, C. Hetherington, J. Moffitt, S. Grimes, P. Jardine, and C. Bustamante, *High degree of coordination and division of labor among subunits in a homomeric ring* ATP*ase,* Cell **151**, 1017 (2012), cited By 26.

[22] J. Moffitt, Y. f. Chemla, K. Aathavan, S. Grimes, P. Jardine, D. Anderson, and C. Bustamante, *Intersubunit coordination in a homomeric ring* ATP*ase,* Nature **457**, 446 (2009), cited By 138.

[23] J. R. Moffitt, Y. R. Chemla, and C. Bustamante, *Chapter ten - methods in statistical kinetics,* in *Single Molecule Tools, Part B:Super-Resolution, Particle Tracking, Multiparameter, and Force Based Methods,* Methods in Enzymology, Vol. 475, edited by N. G. Walter (Academic Press, 2010) pp. 221 – 257.

[24] K. E. Duderstadt, R. Reyes-Lamothe, A. M. van Oijen, and D. J. Sherratt, *Replication-fork dynamics,* Cold Spring Harbor Perspectives in Biology **6** (2014), 10.1101/cshperspect.a010157, http://cshperspectives.cshlp.org/content/6/1/a010157.full.pdf+html .

[25] G. Sitters, D. Kamsma, G. Thalhammer, M. Ritsch-Marte, E. Peterman, and G. Wuite, *Acoustic force spectroscopy,* Nature Methods **12**, 47 (2015), cited By 2.

[26] J. P. Cnossen, D. Dulin, and N. H. Dekker, *An optimized software framework for real-time, high-throughput tracking of spherical beads,* Review of Scientific Instruments **85**, 103712 (2014).

[27] I. De Vlaminck, T. Henighan, M. T. J. van Loenhout, D. R. Burnham, and C. Dekker, *Magnetic forces and dna mechanics in multiplexed magnetic tweezers,* PLoS ONE **7**, e41432 (2012).

[28] J. Lipfert, X. Hao, and N. Dekker, *Quantitative modeling and optimization of magnetic tweezers,* Biophysical Journal **96**, 5040 (2009), cited By 43.

[29] Z. Yu, D. Dulin, J. Cnossen, M. Köber, M. M. van Oene, O. Ordu, B. A. Berghuis, T. Hensgens, J. Lipfert, and N. H. Dekker, *A force calibration standard for magnetic tweezers,* Review of Scientific Instruments **85**, 123114 (2014).

[30] H. Geertsema, K. Duderstadt, and A. Van Oijen, *Single-molecule observation of prokaryotic dna replication,* Methods in Molecular Biology **1300**, 219 (2015), cited By 0.

[31] M. Soltani, J. Lin, R. Forties, J. Inman, S. Saraf, R. Fulbright, M. Lipson, and M. Wang, *Nanophotonic trapping for precise manipulation of biomolecular arrays,* Nature Nanotechnology **9**, 448 (2014), cited By 8.

[32] N. Ribeck and O. Saleh, *Multiplexed single-molecule measurements with magnetic tweezers,* Review of Scientific Instruments **79** (2008), 10.1063/1.2981687, cited By 56.

[33] I. De Vlaminck, T. Henighan, M. Van Loenhout, I. Pfeiffer, J. Huijts, J. Kerssemakers, A. Katan, A. Van Langen-Suurling, E. Van Der Drift, C. Wyman, and C. Dekker, *Highly parallel magnetic tweezers by targeted* DNA *tethering,* Nano Letters **11**, 5489 (2011), cited By 25.

[34] A. Huhle, D. Klaue, H. Brutzer, P. Daldrop, S. Joo, O. Otto, U. F. Keyser, and R. Seidel, *Camera-based three-dimensional real-time particle tracking at khz rates and ångström accuracy,* Nature Communications **6** (2015), 10.1038/ncomms6885.

[35] B. M. Lansdorp, S. J. Tabrizi, A. Dittmore, and O. A. Saleh, *A high-speed magnetic tweezer beyond 10,000 frames per second,* Review of Scientific Instruments **84**, 044301 (2013).

[36] R. Roy, S. Hohng, and T. Ha, *A practical guide to single-molecule* FRET, Nature Methods **5**, 507 (2008), cited By 560.

[37] B. A. Berghuis, D. Dulin, Z.-Q. Xu, T. van Laar, B. Cross, R. Janissen, S. Jergic, N. E. Dixon, M. Depken, and N. H. Dekker, *Strand separation establishes a sustained lock at the tus-ter replication fork barrier,* Nat Chem Biol **11**, 579 (2015).

[38] S. Mulugu, A. Potnis, J. Taylor, K. Alexander, and D. Bastia, *Mechanism of termination of dna replication of* Escherichia coli *involves helicase-contrahelicase interaction,* Proceedings of the National Academy of Sciences of the United States of America **98**, 9569 (2001).

[39] D. Dulin, I. D. Vilfan, B. A. Berghuis, S. Hage, D. H. Bamford, M. M. Poranen, M. Depken, and N. H. Dekker, *Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase,* Cell Reports **10**, 983 (2015).

[40] K. Neuman, E. Abbondanzieri, R. Landick, J. Gelles, and S. Block, *Ubiquitous transcriptional pausing is independent of* RNA *polymerase backtracking,* Cell **115**, 437 (2003), cited By 185.

[41] E. Galburt, S. Grill, and C. Bustamante, *Single molecule transcription elongation,* Methods **48**, 323 (2009), cited By 18.

[42] S. Redner, *A guide to first-passage processes* (Cambridge University Press, 2007).

[43] M. Depken, E. Galburt, and S. Grill, *The origin of short transcriptional pauses,* Biophysical Journal **96**, 2189 (2009), cited By 44.

[44] Y.-C. Tsai and K. Johnson, *A new paradigm for* DNA *polymerase specificity,* Biochemistry **45**, 9675 (2006), cited By 145.

[45] X. Yang, E. Smidansky, K. Maksimchuk, D. Lum, J. Welch, J. Arnold, C. Cameron, and D. Boehr, *Motif d of viral* RNA*-dependent* RNA *polymerases determines efficiency and fidelity of nucleotide addition,* Structure **20**, 1519 (2012), cited By 19.

[46] M. Voliotis, N. Cohen, C. Molina-París, and T. Liverpool, *Fluctuations, pauses, and backtracking in* DNA *transcription,* Biophysical Journal **94**, 334 (2008), cited By 57.

[47] G. Schwarz, *Estimating the Dimension of a Model,* The Annals of Statistics **6**, 461 (1978), 10.1214/aos/1176344136.

[48] S. Cowan, *Statistical Data Analysis* (Oxford University Press, 1998).

3

COVALENT DNA ANCHORING STRATEGIES

> *See first, think later, then test. But always see first.*
> *Otherwise, you will only see what you were expecting.*
> *Most scientists forget that.*
>
> Douglas Adams

Magnetic tweezers are a powerful single-molecule technique that allows real-time quantitative investigation of biomolecular processes under applied force. High pulling forces exceeding tens of picoNewtons may be required, e.g. to probe the force range of proteins that actively transcribe or package the genome. Frequently, however, the application of such forces decreases the sample lifetime, hindering data acquisition. To provide experimentally viable sample lifetimes in the face of high pulling forces, we have designed a novel anchoring strategy for DNA in magnetic tweezers. Our approach, which exploits covalent functionalization based on heterobifunctional poly(ethylene glycol) crosslinkers, allows us to strongly tether DNA while simultaneously suppressing undesirable nonspecific adhesion. A complete force and lifetime characterization of these covalently anchored DNA-tethers demonstrates that, compared to more commonly employed anchoring strategies, they withstand 3-fold higher pulling forces (up to 150 pN) and exhibit up to 200-fold higher lifetimes (exceeding 24 h at a constant force of 150 pN). This advance makes it possible to apply the full range of biologically relevant force scales to biomolecular processes, and its straightforward implementation should extend its reach to a multitude of applications in the field of single-molecule force spectroscopy.

This chapter has been published as: Richard Janissen, Bojk A Berghuis, David Dulin, Max Wink, Theo van Laar & Nynke H Dekker. Invincible DNA tethers: covalent DNA anchoring for enhanced temporal and force stability in magnetic tweezers experiments *Nucl Acids Res* (2014). [1].

3.1. INTRODUCTION

SINGLE-MOLECULE techniques have become increasingly important in the last two decades, as they have permitted detailed insights to biological processes that are not readily apparent in ensemble measurements. They also provide access to novel parameters such as force, which plays a fundamental role in a variety of biological processes ranging from cellular motility to the replication, repair and segregation of DNA [2, 3]. Examples of single-molecule methods capable of applying force to biological systems include atomic force spectroscopy (AFM), optical tweezers and magnetic tweezers [3, 4]. The magnetic and optical tweezers methods have garnered particular attention as they provide high temporal, spatial and force sensitivity in a range from hundreds of femtoNewtons to tens of picoNewtons (pN), and, in contrast to AFM, they are also capable of measuring and applying torque [5–8]. More recent developments in magnetic tweezers also include the possibility of measuring multiple samples simultaneously (multiplexing) [9–11], allowing for a broad, statistically sound characterization of biomolecular machines at the single-molecule level.

Studies of biomolecular processes such as DNA unzipping [12], polymerase dynamics [13], nucleosome unwrapping [14], protein unfolding [15], conformational changes such as the overstretching transition of nucleic acid structures [16] and viral genome packaging motors [17] have particularly stringent requirements on the applied forces, which need to exceed several tens of pN. For example, the dsDNA B-S conformational transition is only observed at pulling forces of ~ 65 pN [18]. However, as the lifetime of the non-covalent bonding typically employed for the anchoring of biomolecules in magnetic or optical tweezers decreases with applied force [19–21], such studies are limited to short measurement times in the order of minutes. As a consequence, the low amount of data that can be collected from individual tethers can hamper reliable characterization of the biomolecular process of interest. To allow measurements over longer timeframes and broader force spectra, the anchoring of biomolecules in tweezers experiments should employ strong chemical bonding.

Despite significant developments in tweezers instrumentation and methodology in recent years [5–8, 22], the chemistry of the biomolecular anchoring techniques has remained largely unaltered [23–25]. Focusing more specifically on magnetic tweezers, typically DNA constructs containing digoxigenin (DIG)-modified nucleotides (incorporated by enzymatic polymerase chain reaction (PCR)) are bound onto surfaces to multiple non-specifically adsorbed anti-digoxigenin IgG antibodies (anti-DIG) [8]. The attachment of superparamagnetic beads to DNA tethers is mainly realized by biotin:streptavidin linkages [8, 10, 22]. While this methodology permits a rapid and reliable DNA attachment, allowing it to become widespread, known limitations exist. For example, a single DIG:anti-DIG interaction has a higher thermodynamic dissociation constant ($K_D = 1.2 \times 10^{-9}$ M) and significantly lower force stability ($F_{max} \sim 25$ pN) [26] compared to a biotin:streptavidin complex, for which $K_D = 10^{-14}$ M and $F_{max} \sim 200$ pN [27]. While multiple of these non-covalent ligand:receptor bondings may be employed to extend the force stability and lifetime of tethered DNA [8, 10, 22], the facilitated dissociation of bonds due to the application of external forces [28] continues to constrain both parameters.

In this work, we achieve the covalent anchoring of DNA tethers for use in magnetic tweezers (**Figs. 3.1** and **3.3a**). To do so, we introduce the application of ethanolamine and heterobifunctional poly(ethylene glycol) (PEG) linkers that allow the covalent binding of DNA molecules to either glass or magnetic beads [29, 30]. This development relies on the ability to selectively change the functional moieties of PEG-crosslinkers [29]. We quantitatively characterize both the PEG-based surface passivation and DNA-tether stability in the magnetic tweezers. Our studies of the attachment stabilities of DNA tethers, probed by dynamic force spectroscopy and lifetimes experiments, demonstrate that DNA molecules anchored via covalent coupling to PEG-coated surfaces and magnetic beads withstand pulling forces up to 150 pN and exhibit lifetimes exceeding 24 h under constant force loads of 45 and 150 pN. Our use of amphiphilic PEG polymers, widely used in single-molecule AFM force experiments [30–33], provides the additional benefit of excellent biocompatibility [34–37] and, as we experimentally demonstrate, increased suppression of non-specific adhesions [38, 39] over a range of ionic conditions compared to the commonly applied DNA-tethering method using nitrocellulose surfaces and streptavidin-coated beads. These advances overcome the current limitations on the applicable force and lifetime stabilities of DNA tethers, thereby vastly extending the impact of future magnetic and optical tweezers studies.

3.2. RESULTS AND DISCUSSION

We systematically characterize our three different chemical anchoring strategies for DNA (**Fig. 3.1**, Methods section) according to criteria designed to test their suitability in single-molecule magnetic tweezers experiments. First, we analyze the non-specific adhesion of functionalized magnetic beads to different surface coatings, as excessive non-specific adhesion may limit the multiplexing efficacy of magnetic tweezers experiments. The effect of different ionic concentrations is taken into account, as magnetic tweezers experiments may require adapted buffer conditions for the study of specific biological processes. Next, we examine the spatial stability of covalently surface-bound polystyrene reference beads to surfaces (**Fig. 3.2**). Such reference beads are typically employed in single-molecule experiments to correct for instrumental drift. To evaluate the force and lifetime stability of different chemically anchored DNA tethers, we perform force-dependent rupture experiments with pulling forces up to 150 pN and lifetime experiments with durations up to 24 h under constant force loads of 45 (**Fig. 3.3**) or 150 pN. A summary of the quantitative results is supplied in **Table 3.1**.

3.2.1. EFFECT OF PEG ON SURFACE PASSIVATION

We characterize the different combinations of functionalized flow cell surfaces and magnetic beads to investigate the presence of non-specific interactions. For quantitative evaluation, we count the number of non-specific adhered magnetic beads per surface area (examples shown in **Fig. 3.2a,b**) for each surface-bead combination in the presence of 1% (v/v) BSA, which is predominantly used as non-specific adhesion-decreasing adjunct [40]. To additionally investigate the effect of ionic strength on surface-bead interactions, we perform these experiments at two different monovalent ion concentra-

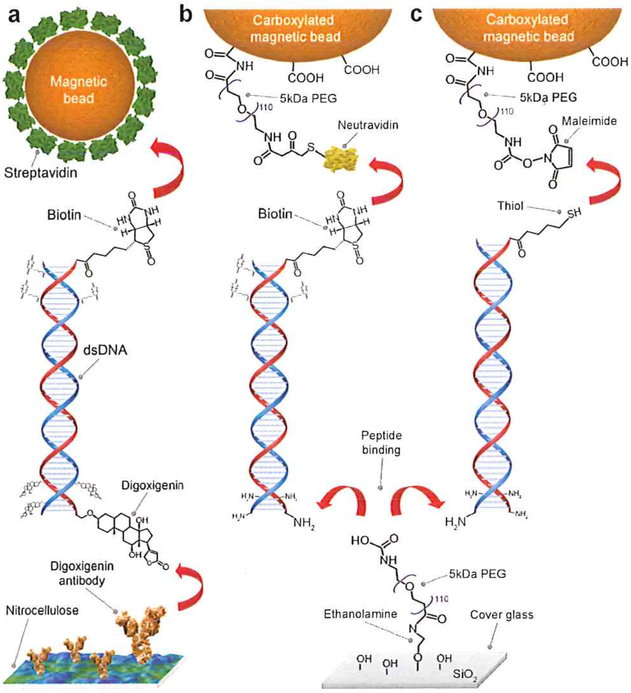


Figure 3.1: **Schematic overview of DNA anchoring strategies used in this study.** **(a)** Tethering of DNA in magnetic tweezers by coupling a multiply biotinylated handle on one extremity of the DNA to a streptavidin-coated magnetic bead, and by binding a handle containing multiple digoxigenin molecules on the other extremity of the DNA to anti-digoxigenin IgG antibodies adsorbed on nitrocellulose. **(b,c)** Tethering of one DNA extremity via covalent coupling NH_2-enriched handles on the DNA to a surface. The other DNA extremity can then be bound to either **(b)** neutravidin-coated magnetic beads using labeled DNA handles containing multiple biotins or to **(c)** maleimide-labeled magnetic beads using a single DNA 5' thiol-modification. In both cases **(b,c)** the magnetic bead coatings include PEG polymers that serve both as a passivation layer and as a covalent crosslinker.

tions (50 mM NaCl, 500 mM NaCl) and examine the effect of adding divalent magnesium cations (addition of 7 mM $MgCl_2$ to 500 mM NaCl) (**Fig. 3.2c**). To evaluate statistically our results, we apply an analysis of variances (ANOVA) followed by a Tukey's post-hoc test with a significance threshold level $P < 0.05$ [41].

Our experimental results (**Fig. 3.2**) indicate that when no or solely one PEG crosslinker is used in the different bead-surface combinations, similar degrees of non-specific interactions occur between different functionalized magnetic beads and surface compositions under all tested buffer conditions (**Fig. 3.2c**). For example, we observe that the standard method of using streptavidin-coated beads and nitrocellulose surfaces results in a relatively low density ($\sim 650 \pm 170$ beads/mm^2) of non-specifically adhered beads at low salt conditions (**Fig. 3.2a**; **Fig. 3.2c**, gray bar). Under these ionic conditions, PEGylated magnetic beads exposing neutravidin (**Fig. 3.2c**, red bar) and maleimide (**Fig. 3.2c**, blue bar) exhibit similar densities ($\sim 710 \pm 180$ beads/mm^2 and $\sim 500 \pm 230$ beads/mm^2, respectively) of adhered beads to nitrocellulose surfaces. Comparable densities of adhered magnetic beads are also observed when streptavidin-coated beads are tested on PEG surfaces (**Fig. 3.2c**, gray hatched bar, (~ 720± 280 beads/mm^2). The use of higher salt conditions (500 mM NaCl and 500 mM NaCl supplemented with 7 mM $MgCl_2$; **Fig. 3.2c**) leads to moderately increased bead adhesion for PEGylated beads that expose neutravidin and maleimide (**Fig. 3.2c**, red and blue bar, respectively), but the adhesion of streptavidin-coated beads to nitrocellulose and PEG surfaces (**Fig. 3.2c**, gray and gray hatched bars) remain mostly unaffected.

The density of non-specifically adhered beads can be reduced ~3-fold by switching to the combination of PEG-coated surfaces and PEGylated magnetic beads (**Fig. 3.2b**). For example, for neutravidin-exposing PEGylated beads tested on PEG surfaces, the bead density is reduced to ~ 200 ± 80 beads/mm^2 (**Fig. 3.2c**, red hatched bar). Similar results are obtained for maleimide-exposing PEGylated beads tested on PEG surfaces (**Fig. 3.2c**, blue hatched bars), and overall, the results appear independent of the salt conditions tested. In general, non-specific bead adhesions may result from either electrostatic forces – e.g. Coulomb forces or the van der Waals force – or entropic forces [38, 43, 44]. The use of non-charged PEG crosslinkers provides a spatial separation between the magnetic bead and neighboring surfaces on the order of ~2 nm per PEG layer (corresponding to the Flory radius of 5 kDa PEG [38]). The Debye length under the ionic conditions tested lies below the provided surface-bead distances of the PEG coatings, suggesting that the introduction of PEG layers does not affect electrostatic interactions except when bead and surface are already in very close proximity. Any reduction in adhesion due to the introduction of PEG layers, which is most pronounced in the case of PEG-coated beads combined with PEGylated surfaces, is therefore likely a result of entropy repulsion, as was also concluded in a previous report that utilized longer, more brush-like PEG layers [38]. Remaining adhesion between PEGylated beads and PEG surfaces could possibly result from defects within the PEG-coatings.

3.2.2. SPATIAL REFERENCE BEAD FIXATION STABILITY

We next verify the compatibility of our procedure for the covalent coupling of DNA tethers to surfaces with the stable fixation of so-called reference beads. Position-stable reference beads on sample surfaces are essential to the success of magnetic tweezers measurements [11, 45], as they permit the removal of artifacts related to instrumentation and sample drift. A common approach is to affix polystyrene reference beads to nitrocellulose surfaces by non-specific adhesion, as described in the 'Methods' section [3, 6, 8, 10, 11, 22, 46]. In the case of PEG-coated surfaces, however, this bead attachment method can no longer be applied, as result of adhesion suppression by the PEG polymer layer as described in the previous subsection. An alternative approach is to melt polystyrene beads onto glass surfaces prior to the organochemical surface coating. When surfaces are coated with nitrocellulose, this method guarantees a simple and reliable method for reference bead fixation. In the case of PEG-coated surfaces, however, the required high temperatures damage the ethanolamine and PEG layers, resulting in non-homogenous coverage. In areas of decreased PEG-linker densities this would consequently lead to increased non-specific binding of magnetic beads and biomolecules.

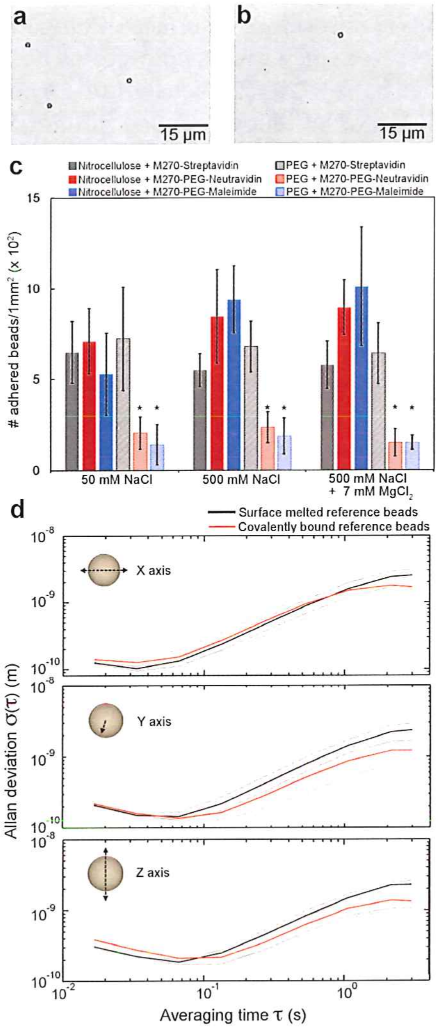


Figure 3.2: **Non-specific adhesion and spatial stability of beads.** **(a,b)** Images of a flow cell with non-specifically adhered magnetic beads. **(c)** The number of adhered magnetic beads is summarized in the histogram for different combinations of bead-surface for a variety of buffer conditions. The error bars represent the mean standard deviations and the asterisks denote the significance threshold level of the applied analysis of variance (1-way ANOVA; $*P < 0.05$) for each buffer condition. **(d)** The position stability of reference beads fixed via surface-melting (black) or covalent attachment (red) is analyzed by computing their respective Allan deviations [42] as a function of time in all three dimensions. For each attachment method, the median of $n = 25$ reference beads is presented as continuous lines and the 25th and 75th percentiles as dashed lines.

We have therefore developed a new anchoring procedure for reference beads that relies on the same accessible carboxylic moieties on surface-anchored PEG crosslinkers employed in the covalent surface attachment of amino-labeled DNA tethers (**Fig. 3.1b,c**). This method allows the fixation of amine-coated reference beads.

To test the fixation of covalently attached beads, we compare their spatial stability with that of surface-melted polystyrene beads. To do so, we tracked the motions of beads anchored via either approach in three dimensions over 2000 s at 60 Hz ($n = 25$ for both cases). The tracked motions of any individual bead were corrected for instrumental drift in all dimensions by subtracting the coordinates of a second, similarly anchored bead. We analyze the resulting drift-corrected bead movements by computing the corresponding overlapping Allan deviation [42] for averaging times between ~20 ms and 3 s (**Fig. 3.1d**). The experimental results for both anchoring strategies display deviations of bead motion in all three dimensions that fall below 13 Å (**Fig. 3.1d**). Additionally, the differences in the Allan deviations between two approaches for reference bead fixation are <4 Å over the entire averaging time scale (**Fig. 3.1d**, compare black and red dashed lines). For short averaging times between ~20 and 100 ms, all bead motions generally remain below 4 Å for all dimensions. Toward higher averaging times up to 3 s, the bead stability decreases globally, with excursions up to 13 Å that result from the increasing influence of instrumental drift. These results demonstrate that both approaches are equally suitable for high-resolution measurements [45, 47, 48] and, more importantly, that we have provided a protocol for reference bead fixation that is compatible with covalent tethering of DNA tethers.

3.2.3. FORCE-DEPENDENT EXTENSION OF LINEAR AND DNA HAIRPIN CONSTRUCTS

We now discuss the use of our functionalization process and the addition of PEG-polymers for tethering molecules such as DNA. Before characterizing the overall force- and lifetime characteristics of populations of DNA tethers (see 'Force stability and tether limetimes of linear dsDNA' in 'Results and Discussion' section), we have verified that our newly-introduced DNA anchoring methods do not affect the mechanical properties of individual DNA molecules [24, 25, 49, 50]. To do so, we apply force to both individual covalently surface-anchored linear 4.8 kbp dsDNA constructs (**Fig. 3.1b**) and to 1 kbp DNA hairpin containing constructs. The latter molecules are typically used to study DNA-binding proteins and serve as sensitive and reliable force sensors [48, 51, 52]. We perform the experimental evaluation conducting dynamic force spectroscopy (DFS) measurements (**Fig. 3.3a**) with a constant force loading of 10 pN/s up to 115 pN.

For both covalently tethered dsDNA (**Fig. 3.3b**) and hairpin-DNA (**Fig. 3.3c**), we find that their response to force exhibits the expected behavior. For example, the application of force to a covalently anchored linear dsDNA construct demonstrates regular entropic stretching behavior at pulling forces below 5 pN, followed by enthalpic stretching at higher forces and leading to the observation of the well-described force-induced overstretching transition at ~60 pN (**Fig. 3.3b**). The clearly observable overstretching transition is associated with a rapid ~1.7-fold length increase, in agreement with previous reports [24, 50]. As expected, the force-extension curve of dsDNA up to the overstretching transition can be well-described by the extensible worm-like chain (WLC) model [49, 53]. Application of this model to our dsDNA stretching data (**Fig. 3.3b**, dashed line) yields a

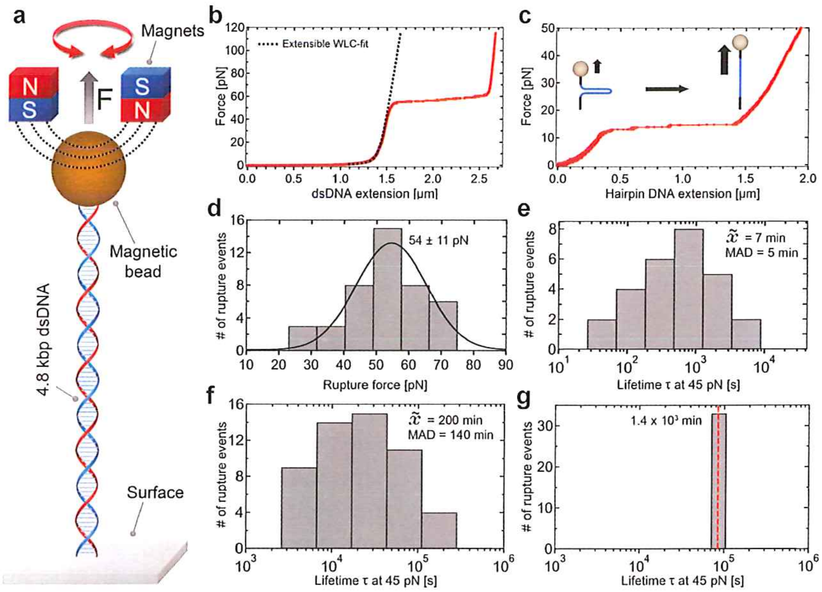


Figure 3.3: **DNA stability under high forces.** **(a)** DNA tethers were stretched by applying a magnetic field via a pair of superparamagnetic magnets. **(b)** Force-extension plot for a covalently anchored linear 4.8 kbp dsDNA tether. The black dashed line is a fit of the extensible worm-like-chain fit to the data. At ~60 pN, the overstretching transition of DNA is observed. **(c)** Force-extension plot for a covalently surface-bound single 1 kbp hairpin DNA construct. **(d)** Histogram of measured rupture forces ($n = 43$) via DFS with constant loading rate of 10 pN/s up to 150 pN and applied Gaussian fit (black curve). **(e)** Histogram of measured lifetimes at constant pulling force of 45 pN ($n = 37$) for dsDNA tethers anchored via DIG:anti-DIG to the surface and streptavidin:biotin to the magnetic bead. **(f,g)** Histograms of observed lifetimes at applied forces of 45 pN for DNA molecules anchored covalently to the surface alone ($n = 52$). **(f)** and anchored covalently to both surface and bead ($n = 32$) (g). The median of tether lifetimes in **(e,f)** is denoted within the histogram panels as $\tilde{x}$ with the corresponding median absolute deviation (MAD). The red dashed line in **(g)** is meant to indicate that the measured lifetimes are lower bounds, as the experiment itself was terminated after 24 h.

bending persistence length of ~40 nm and a stretching modulus of ~800 pN, in agreement with previous studies [25, 49, 54]. The application of force to DNA tethers containing a 1-kbp hairpin demonstrates the mechanical unzipping of the double-stranded hairpin structure (**Fig. 3.3c**). We observe full opening of the 1 kbp hairpin, as witnessed by a length increase of ~1 μm, at a force of ~13 pN with, also in agreement with previous studies [12, 55, 56].

3.2.4. FORCE STABILITY AND TETHER LIFETIMES OF LINEAR DSDNA

We now test the coupling stability of the three dsDNA tether anchoring methods by measuring the overall force- and life-time characteristics of populations of DNA tethers. To do so, dsDNA constructs anchored via the different strategies are subjected to increasing forces up to 150 pN. Separately, we conduct lifetime experiments at constant forces of 45 or 150 pN over 24 h. The results of the force-dependent tether stability and the construct lifetimes are summarized in **Table 3.1**.

Construct	Applicable force	Lifetime at 45 pN	Lifetime at 150 pN
Standard	54 ± 11 pN	~7 min	
Covalent bond at one extremity	150 pN	~3 h	~1 min
Covalent bond at both extremities	150 pN	>24 h	>24 h

Table 3.1: Maximum applicable pulling forces and median average tether lifetimes at a constant pulling forces of 45 or 150 pN for the three different DNA tethering methods.

Our experiments indicate that there are significant differences in the force stability and tether lifetimes for dsDNA tethers anchored via the three different approaches (**Fig. 3.3**), with the common approach of relying on DIG:anti-DIG and streptavidin:biotin interactions (**Fig. 3.1a**) providing the weakest linkages. The pulling of linear dsDNA tethers coupled using this approach results in average pulling force stabilities of F_{max} ~55 pN (**Fig. 3.3d**). We perform lifetime measurements at 45 pN, a force selected because it is ~20% lower than the average force at which the DIG:anti-DIG coupling fails. Under this force load, this dsDNA construct demonstrates widely distributed lifetimes ranging from a few seconds to ~40 min. The spread in lifetimes likely reflects the number of contributing DIG:anti-DIG complexes to the dsDNA anchorage [28]. On average, however, dsDNA anchored via this common approach last only ~7 min (**Fig. 3.3e**).

In stark contrast, the approaches introducing one or more covalent linkages, which in principle provide pulling force stabilities up to nanoNewtons [57], provide vastly-improved force stabilities and lifetimes. For example, dsDNA tethers anchored covalently onto PEGylated surfaces by multiple amide bonds and coupled to magnetic beads via either multiple neutravidin:biotin complexes or a single maleimide:thiol coupling (**Fig. 3.1b,c**) exhibited no force-dependent tether ruptures even for applied forces as high as 150 pN (**Table 3.1**). Similarly, atomic force spectroscopy studies reported pulling force stabilities ≥200 pN for both binding complexes [27, 58, 59]. Concretely, the fact that the maximal force is similarly high in these two cases demonstrates that the neutravidin:biotin coupling does not limit the maximal force that can be applied. Differences between the approaches relying on one versus two covalent linkages (**Fig. 3.1b,c**) nevertheless appear in the lifetime experiments. These were again performed at a constant force of 45 pN, to compare to the data in **Fig. 3.3e**. However, it was necessary to increase the duration of the lifetime experiments to 24 h. The dsDNA constructs covalently attached to the PEGylated surfaces and tethered to magnetic beads by several neutravidin:biotin complexes now exhibit a distribution of lifetimes that ranges from ~1 to 24 h, where the range likely reflects that the presence of differing numbers of neutravidin:biotin binding complexes. On average, their lifetime exhibits a dramatic 30-fold increase to ~3 h compared to the common attachment strategy (**Fig. 3.3f**). Impressively, the dsDNA constructs covalently anchored to the PEG-coated surface and covalently tethered to maleimide-modified magnetic beads (**Fig. 3.1c**) by a single thioether linkage remain stable at this force load over the entire duration of 24 h (**Fig. 3.3g**). More dramatic differences between these two constructs were observed at a constant force load of 150 pN: here, dsDNA constructs covalently tethered only to PEGylated surfaces demonstrate a lifetime distribution that range from few s to ~3 h lifetimes, with a median of ~1 min (**Table 3.1**, $n = 94$), while dsDNA constructs covalently tethered at both extremities continue to display lifetimes in excess of 24 h (**Table 3.1**, $n = 28$). This result demonstrates that covalent anchorage of both DNA extremities provides the ideal experimental scenario.

The results for the experiments on tether rupture force and lifetime results are summarized in **Table 3.1**. The statistical overview clearly demonstrate explicitly the disadvantage of using non-covalent interactions for DNA-anchoring, as the observed lifetimes at high forces are limited to the order of minutes, which drastically impacts the observation time for biophysical experiments. Strategies that employ multiple of such weak linkages can contribute to both force- and life-time stability and can prove useful in certain experiments, but, as our experiments show, eliminating DIG:anti-DIG coupling will enhance the force range while eliminating neutravidin:biotin coupling will enhance the lifetime stability. Indeed, using covalent tethering to both DNA extremities makes it possible to perform magnetic tweezers experiments on DNA molecules that can withstand pulling forces up to 150 pN with lifetimes over 24 h. In such cases, the applicable forces and lifetimes are solely limited by the intrinsic stability of the studied biomolecules or biomolecular complexes.

3.3. CONCLUSIONS

We have demonstrated a significant increase in pulling force stability and lifetimes under force load for DNA tethers by applying covalent chemical tether-anchoring using heterobifunctional PEG crosslinker for magnetic tweezers experiments. By applying this methodology, the limit of lifetime and applicable force will principally be defined by the stability of the biological sample itself. Conveniently, the implementation of biocompatible PEG crosslinkers also provides suppression against non-specific adhesion of magnetic beads and biomolecules on the surface. These advantages provide vastly enhanced high-throughput efficiency in magnetic tweezers experiments and can lead, together with advances in multiplexing capability [9–11] to improved statistics for the description of biomolecular processes. The anchoring method that we present is not limited only to magnetic tweezers: it can be easily applied to other single-molecule techniques, such as optical tweezers and fluorescence-based studies, as well as for e.g. biomaterial engineering and biosensors, where stable biomolecule attachment is required. It can also readily be extended to the binding of proteins, peptides and a wide range of other organic molecules. This development opens new possibilities for single-molecule experiments that probe the stability of biomolecules and interactions under high force, such as DNA-binding protein interactions, protein–protein complexes and protein unfolding studies, amongst others.

3.4. MATERIALS AND METHODS

3.4.1. BEADS

Streptavidin-coated and carboxylated superparamagnetic beads (M-270 DynaBeads, Life-Technologies, USA) with a diameter of 2.8 μm were used within this study. Commercially available amine-coated polystyrene beads (Polysciences GmbH, Germany) with a diameter of 3 μm were used as reference beads fixed onto the functionalized glass surfaces.

3.4.2. DNA CONSTRUCTS

Linear dsDNA and single DNA hairpin constructs, each with an enzymatically ligated dsDNA handle at both ends, were synthesized as samples.

For the chemical anchoring stability characterization and tether lifetime survey, linear double-stranded DNA constructs with a total length of 4.8 kbp were synthesized based on a 3.6 kbp fragment of the pRL-SV40 plasmid (Promega, USA), digested with BamHI and XbaI and ligated in a second step to the 600 bp dsDNA handles, as previously described [8, 60]. To create handles, a 1.2 kb fragment from pBluescript (Stratagene, USA) was amplified by PCR in the presence of Biotin-16-dUTP (Roche Diagnostics, Switzerland), Digoxigenin-11-dUTP (Roche Diagnostics, Switzerland) or Aminoallyl-dUTP (Thermo Scientific, USA) using forward primer 5'-GACCGAGATAGGGTTGAGTG and reversed primer 5'-CAGGGTCGGAACAGGAGAGC. Prior to enzymatic ligation via T4 DNA ligase (New England Biolabs Ltd., UK), the handles were digested with either *BamHI* or *XbaI*. Thiol-labeled DNA handles were produced by PCR using the same template and primers with the same sequence as mentioned above, but containing a 5' thiol-labeled nucleotide (Biolegio B.V, The Netherlands). Three different full-length construct combinations used in this study were assembled as shown (**Fig. 3.1**).

To prepare a DNA hairpin construct, a 1 kb Lambda phage dsDNA fragment (LifeTechnologies, USA) was amplified by PCR using the forward primer 5'-CTGCGGTCTCGTTGCTTACCGTCACCAGAAATTACCGTCAC and reversed primer 5'-CCATCTTGGTCTCCTAGGTTTTTAGCAGCGAAGCGTTTGATAAG, digested with *BsaI* and ligated at one DNA terminus with a 42 bp oligonucleotide to form a U-turn (**Table 3.2**). Biotin-16-dUTP and Aminoallyl-11-dUTP enriched handles (see above) were ligated to double-stranded spacer oligonucleotides (Biolegio B.V., The Netherlands; **Table 3.2**). These handle/spacer constructs were allowed to anneal in order to form a short (50 bp) double-stranded stem with a *BsaI* overhang, which were ligated to the free accessible terminals of the 1 kb dsDNA-hairpin fragment.

3.4.3. POLY(ETHYLENE GLYCOL) CROSSLINKER

For surface passivation and covalent DNA coupling procedures, a mixture of mono- and heterobifunctional PEG linker of 5 kDa (~110 monomers) were used. For covalent amino-labeled DNA anchoring, NHS-PEG-COOH and NHS-mPEG (LaysanBio, USA) were selected as surface linker; for PEGylation of the superparamagnetic beads NH2-PEG-maleimide (Nanocs, USA), NH$_2$-PEG-SH and NH$_2$-mPEG (IRIS Biotech, Germany) were used.

oligonucleotide	sequence
U-turn	5'-CCTAAGCTCGCCGAGGCGAGCGAAAGCTCGCCTCGGCGAGCT-3'
upper handle, forward	5'-GGCCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAG-TTGCTCTTGCCATGCTCTTTACAACCGGTTGACTGCTTCAGGGG-TCGATCCCGCTTTGTAC-3'
upper handle, reversed	5'-GGCAAGAGCAACTCGGTCGCCGC-ATACACTATTCTCAGAATGACTTGGTT-3'
lower handle, forward	5'-GCAAGTACAAAGCGGGATCGACCCCTGAAGCAGTCAACCGGT-TGTAAAGAGCATCGATCGTTGTCAGAAGTAAGTTGGCCGCAGT-GTTATCACTCATGGTTATGCCAGATGGTAAGCCCTCCCGTATCGTAG-TTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGA3'
lower handle, reversed	5'-GATCTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTG-TAGATAACTACGATACGGGAGGGCTTACCATCTGGC3'

Table 3.2: Used oligonucleotides.

3.4.4. FUNCTIONALIZATION OF SUPERPARAMAGNETIC BEADS

For the functionalization of magnetic beads, we started with commercially available carboxylated superparamagnetic beads. For a final volume of 1 ml with a bead concentration of 3 mg/ml (~ 2 × 108beads), 100 μl of the stock solution was used for any functionalization strategy (**Fig. 3.1**).

The vials with the stock solution were shaken for 30 min using a tilting mixer to avoid colloidal aggregates prior to the functionalization steps. Afterward, 100 μl of the carboxylated bead stock solution was transferred to a 1 ml Eppendorf tube and placed in a magnetic holder (DynaMag-5, LifeTechnologies, USA) for 2 min, after which the supernatant was discarded. The beads were washed twice with 1 ml 100 mM MES (2-(N-morpholino)ethane sulfonic acid, Sigma, USA) buffer (pH 4.7) for 10 min with a tube shaker at 1200 rpm (Thermomixer, Eppendorf, USA), and the supernatant was discarded after each washing step which includes a 2 min incubation in the magnetic holder. The beads were resuspended in 100 μl MES buffer and mixed with a 100 μl MES solution containing 4 mg of NH_2-PEG-SH (for binding neutravidin in a second step) or NH_2-PEG-maleimide (depending on the DNA anchoring strategy, see Figure 1), 8 mg NH_2-mPEG and 20 mg EDC (1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide, Sigma, USA). The components were mixed thoroughly for 2 min with a vortex, and afterward the bead solution was sonicated for 2 h in an ultrasonic bath (Bransonic B1510, Branson, USA) with an additional 30 s mixing step every 20 min. Following the PEGylation reaction, the functionalized beads were washed three times with 1 ml PBS (phosphate buffered saline, Sigma, USA) buffer (pH 7.4) containing 2% BSA (bovine serum albumin, Sigma, USA). Each washing step consists of 10 min of shaking at 1200 rpm followed by placement of the tube in the magnetic holder for 2 min, after which the supernatant was discarded. The functionalized magnetic beads were resuspended in 1 ml PBS buffer (2% BSA, pH 7.4).

For the coating of magnetic beads with neutravidin, PEGylated beads with NH_2-PEG-SH were resuspended in 100 μl sodium phosphate (SP, pH 7) buffer (Sigma, USA) and mixed with 100 μl SP buffer containing 100 μg of maleimide-labeled neutravidin (Pierce, USA). After vortexing the mixture for 1 min, the coupling reaction was carried out in the tube shaker at 1200 rpm for 2 h. Afterward, the neutravidin-coated beads were washed three times with 1 ml PBS buffer (2% BSA, pH 7.4) for 10 min at 1200 rpm, constantly discarding the supernatant after placing the tube in the magnetic holder for 2 min and after each mixing instance. The magnetic beads were resuspended in 1 ml PBS buffer (2% BSA, pH 7.4).

All functionalized magnetic beads were stored at 4°C and used within 30 days following bead functionalization. No noticeable decreases in the DNA-coupling efficiency were observed during this period.

3.4.5. FUNCTIONALIZATION OF FLOW CELL SURFACES

Borosilicate cover glasses (#1, Menzel GmbH, Germany) with dimensions of 24 × 60 mm serve as the flow cell surface for DNA anchoring. The glasses were placed into a Teflon holder and incubated in a 5% (V/V) aqueous Hellmanex III (Hellma GmbH, Germany) solution and sonicated for 20 min at 40°C. After washing thoroughly with water, the glasses were covered in water and sonicated for further 20 min. After another washing step with water, the glass surfaces were dried in a nitrogen flow [29, 30]. The preactivation of the glass surfaces to generate a high density of silanole groups—which allows a homogeneous surface functionalization—was realized by applying oxygen plasma (Plasma-PREEN I, Plasmatic Systems Inc., USA) for 2 min with an oxygen flow rate of 2.5 SCFH and a power of 200 W.

For the coating of glass surfaces with nitrocellulose, nitrocellulose membrane paper (Invitrogen, USA) was dissolved (1% m/V) in acetic acid pentyl ester (Sigma, USA) by mixing the components for 1 h at 250 rpm in a tube shaker at 35°C. Prior to polymer coating, polystyrene beads with a diameter of 3 μm were melted to the previously cleaned cover glasses to serve as reference beads for the magnetic tweezers experiments. For that purpose, 3 μl of a 250× dilution of the polystyrene bead stock solution in ethanol (Sigma, USA) was added to the cover glasses and spread evenly on the surface with the lateral side of the pipette tip. For bead fixation, the cover glasses were heated to 150°C for 3 min on a ceramic heat plate (PC-420D, Corning, USA). Afterward, the prepared nitrocellulose solution was diluted 10×, and 3 μl of it was spread evenly onto each surface and cured by evaporation at 90°C for 1 min.

The glass surfaces for covalent DNA attachment were aminated by esterification of ethanolamine with surface silanole groups [29, 30]. The clean cover glasses were incubated overnight in a 5 M ethanolamine hydrochloride (EA, Sigma, USA) solution in anhydrous dimethyl sulfoxide (DMSO, Sigma, USA). The dissociation of EA in DMSO was performed at 65°C for ~1 h, accompanied by some occasional shaking to ensure complete dissolution. After the surface amination, the cover glasses were washed thoroughly with water and dried with a nitrogen stream.

The functionalized glass surfaces were afterward used for flow cell assembly as previously described [8]. In brief, a double layer of parafilm spacer was placed onto the functionalized surfaces, and the flow cell was closed by a second coverslip on top containing inlet and outlet apertures for fluidic couplings. The assembly was performed by melting the parts together at 90°C for 30 s.

3.4.6. EVALUATION OF NON-SPECIFIC MAGNETIC BEAD ADHESION

To quantify the non-specific interactions between functionalized magnetic beads and coated surfaces, we counted the number of non-specific adhered beads per field of view ($\sim 260 \times 190 \mu m$) of the complementary metal-oxide-semiconductor (CMOS) camera for each different bead-surface configuration. Experimentally, chemically functionalized magnetic beads with a concentration of 6×10^7 beads/ml were incubated on the functionalized surfaces for 30 min in PBS/BSA buffer (1% BSA, pH 7.4), followed by washing with 3 ml of the same buffer until no more mobile beads were observed during flow cell flushing. For each bead-surface configuration, 20 of such areas were investigated to calculate the average number and standard deviation of non-specifically adhered beads. The data were additionally evaluated by a Kruskal–Wallis one-way analysis of variance (ANOVA), followed by a Tukey's post-hoc test [41].

3.4.7. DNA ANCHORING AND MAGNETIC BEAD TETHERING

The DNA constructs with different functional moieties were anchored to the functionalized surfaces according to different chemical approaches (**Fig. 3.3**).

As a standard method, DNA constructs with biotin and digoxigenin (DIG) labeled handles were surface-anchored via anti-DIG antibodies and coupled to streptavidin-coated superparamagnetic beads as previously described [8, 10, 22]. In detail, first a volume of 50 μl PBS buffer (pH 7.4) containing 100 $\mu g/ml$ of anti-DIG antibodies (Roche, USA) was flushed into the flow cell and incubated for 1 h. Afterward, the flow cell was washed with 1 ml PBS/BSA buffer (1% BSA, pH 7.4) and incubated with this solution for another hour to passivate the surface, decreasing the non-specific adhesion of biomolecules and magnetic beads. After another washing step with 500 μl PBS/BSA buffer, 16 pM of the DNA construct in 100 μl PBS/BSA buffer was incubated within the flow cell for 1 h for surface anchoring, followed by a washing step using 1 ml PBS/BSA buffer. A volume of 100 μl PBS/BSA buffer containing 100 μg of streptavidin-coated magnetic beads was then incubated for another hour within the flow cell to tether the DNA. A final washing step with PBS/BSA buffer was applied until all non-tethered beads were removed.

To covalently anchor amino-labeled DNA to the surface via peptide binding (**Fig. 3.1b,c**), the previously aminated surfaces were PEGylated. To achieve this, a mixture of 3.5 mg NHS-PEG-COOH and 6.5 mg NHS-mPEG in 500 μl MES/EDC buffer (100 mM MES, 50 mM EDC, pH 4.7) was prepared and incubated in the flow cell for 1 h. After washing with 3 ml water, a concentration of 16 pM amino-labeled DNA constructs in 100 μl MES/EDC buffer was mixed with 1.5×10^4 amino-labeled polystyrene reference beads and added to the flow cell for 1 h. After covalent anchoring of DNA and reference beads to the PEG surface, the flow cells were washed with 1 ml water, followed by incubation in PBS/BSA buffer for another hour as an additional passivation step. The flow cells were then washed with 500 μl PBS/BSA buffer and bead tethering was performed in one of two possible ways, depending on the functional moieties included on the free accessible end of the DNA constructs.

Thus, to bind DNA constructs that contain biotinylated handles to magnetic beads, 100 μg of neutravidin-coated magnetic beads in 100 μl PBS/BSA buffer was incubated within the flow cell for 1 h. Alternatively, to covalently tether sulfhydryl-labeled DNA to magnetic beads, the thiol groups were reduced prior by incubation with 2 mM DTT (Dithiothreitol) in PBS buffer (pH 7) for 1 h, followed by a washing step using 1 ml PBS buffer. Afterward, 100 μg of maleimide-modified magnetic beads in 100 μl PBS/BSA buffer was added to the flow cell and incubated for 1 h. In both cases, a final washing step was applied using PBS/BSA buffer until all non-tethered beads were removed.

3.4.8. MAGNETIC TWEEZERS EXPERIMENTAL CONFIGURATION

The magnetic tweezers implementation used in this study has been described previously [10, 22]. Briefly, light transmitted through the sample was collected by a 100× oil-immersion objective (Apochromat 100×, NA = 1.25, Olympus, USA) and projected onto a 4.1 MP CMOS camera (Falcon 4M60, Teledyne Dalsa, Canada) with a sampling frequency of 60 Hz. The applied magnetic field was generated by a pair of vertically aligned permanent neodymium-iron-boron magnets (SuperMagnete, Switzerland) separated by a distance of 300 μm, suspended on a motorized stage (M-126.PD2, Physik Instrumente, Germany) above the flow cell. Additionally, the magnet pair could be rotated about the illumination axis by an applied DC servo step motor (C-149.PD, Physik Instrumente, Germany).

Image processing of the collected light allowed us to track the real-time position of both surface-attached reference beads and superparamagnetic beads coupled to DNA tethers in three dimensions. This was achieved using a cross-correlation algorithm realized with custom-written software in LabView (2011, National Instruments Corporation, USA). After subtraction of the relative reference bead position to correct for instrumental drift, the x, y and z position of the DNA-tethered beads were determined with a spatial accuracy of < 3 nm. The upward stretching forces on the DNA tethers by the superparamagnetic beads were calibrated from the bead Brownian motions, whereby spectral corrections were employed to correct for camera blur and aliasing [47].

3.4.9. TETHER RUPTURE FORCE AND LIFETIME MEASUREMENT

To statistically compare the mechanical stability between DNA molecules tethered using the three different chemical anchoring strategies (**Fig. 3.1**), we employed two approaches: we applied dynamic force spectroscopy to evaluate the DNA tether rupture force, and we performed force clamp experiments at 45 and 150 pN to evaluate DNA tether lifetimes.

Prior to the characterization of tether stability, we performed a control measurement designed to detect magnetic beads tethered to multiple DNA molecules. This involved applying a pulling force of ~10 pN and comparing the response of the beads upon clockwise versus counterclockwise magnet rotation of 60 turns. When a single magnetic bead is tethered to the surface via multiple DNA molecules, a significant decrease of the DNA tether extension should be observed in both cases, an effect caused by the wrapping of the molecules about one another [61]. Such multiply tethered magnetic beads were excluded from analysis. To evaluate the force-dependent rupture stability of the different types of DNA tethers, we performed dynamic force spectroscopy up to a maximum force of 150 pN, using a constant force loading rate of 10 pN/s. The lifetime analysis for each DNA construct was performed by applying a constant force of 45 pN over the duration of 24 h and detecting the time of tether rupture. For the covalently coupled DNA constructs, this experiment was also repeated at a constant force of 150 pN. All dynamic force spectroscopy and lifetime experiments were carried out in 10 mM TRIS/HCl buffer (pH 7.5, Sigma, USA) containing 100 mM NaCl.

REFERENCES

[1] R. Janissen, B. A. Berghuis, D. Dulin, M. Wink, T. van Laar, and N. H. Dekker, *Invincible DNA tethers: covalent DNA anchoring for enhanced temporal and force stability in magnetic tweezers experiments,* Nucleic Acids Research **42**, e137 (2014).

[2] W. J. Greenleaf, M. T. Woodside, and S. M. Block, *High-resolution, single-molecule measurements of biomolecular motion,* Annual review of biophysics and biomolecular structure **36**, 171 (2007).

[3] D. Dulin, J. Lipfert, M. C. Moolman, and N. H. Dekker, *Studying genomic processes at the single-molecule level: introducing the tools and applications,* Nat Rev Genet **14**, 9 (2013).

[4] K. Neuman and A. Nagy, *Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy,* Nature Methods **5**, 491 (2008), cited By 603.

[5] F. M. Fazal and S. M. Block, *Optical tweezers study life under tension,* Nature photonics **5**, 318 (2011).

[6] X. J. A. Janssen, J. Lipfert, T. Jager, R. Daudey, J. Beekman, and N. H. Dekker, *Electromagnetic Torque Tweezers: A Versatile Approach for Measurement of Single-Molecule Twist and Torque,* Nano Letters **12**, 3634 (2012).

[7] A. La Porta and M. D. Wang, *Optical torque wrench: angular trapping, rotation, and torque detection of quartz microparticles,* Physical review letters **92**, 190801 (2004).

[8] J. Lipfert, J. W. J. Kerssemakers, T. Jager, and N. H. Dekker, *Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments,* Nature Methods **7**, 977 (2010).

[9] I. De Vlaminck, T. Henighan, M. T. J. van Loenhout, D. R. Burnham, and C. Dekker, *Magnetic forces and dna mechanics in multiplexed magnetic tweezers,* PLoS ONE **7**, e41432 (2012).

[10] J. Lipfert, D. A. Koster, I. D. Vilfan, S. Hage, and N. H. Dekker, *Single-Molecule Magnetic Tweezers Studies of Type IB Topoisomerases,* in *DNA Topoisomerases,* Methods in Molecular Biology, Vol. 582 (Humana Press, 2009) pp. 71–89.

[11] N. Ribeck and O. Saleh, *Multiplexed single-molecule measurements with magnetic tweezers,* Review of Scientific Instruments **79** (2008), 10.1063/1.2981687, cited By 56.

[12] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, *Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins,* PNAS **103**, 6190 (2006).

[13] J. Zhou, V. Schweikhard, and S. M. Block, *Single-molecule studies of rnapii elongation,* Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms **1829**, 29 (2013).

[14] I. V. Dobrovolskaia and G. Arya, *Dynamics of forced nucleosome unraveling and role of nonuniform histone-dna interactions,* Biophysical journal **103**, 989 (2012).

[15] F. Rico, L. Gonzalez, I. Casuso, M. Puig-Vidal, and S. Scheuring, *High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations,* Science **342**, 741 (2013).

[16] E. Herrero-Galán, M. E. Fuentes-Perez, C. Carrasco, J. M. Valpuesta, J. L. Carrascosa, F. Moreno-Herrero, and J. R. Arias-Gonzalez, *Mechanical identities of rna and dna double helices unveiled at the single-molecule level,* Journal of the American Chemical Society **135**, 122 (2012).

[17] D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante, *The bacteriophage &phis; 29 portal motor can package dna against a large internal force,* Nature **413**, 748 (2001).

[18] L. Bongini, L. Melli, V. Lombardi, and P. Bianco, *Transient kinetics measured with force steps discriminate between double-stranded dna elongation and melting and define the reaction energetics,* Nucleic acids research , gkt1297 (2013).

[19] G. I. Bell, *Models for the specific adhesion of cells to cells,* Science **200**, 618 (1978).

[20] E. Evans and K. Ritchie, *Dynamic strength of molecular adhesion bonds,* Biophysical journal **72**, 1541 (1997).

[21] E. Evans, *Probing the relation between force-lifetime-and chemistry in single molecular bonds,* Annual review of biophysics and biomolecular structure **30**, 105 (2001).

[22] J. Lipfert, M. J. Wiggin, J. W. J. Kerssemakers, F. Pedaci, and N. H. Dekker, *Freely orbiting magnetic tweezers to directly monitor changes in the twist of nucleic acids,* Nature Communications **2**, 439 (2011).

[23] T. R. Strick, J.-F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, *The Elasticity of a Single Supercoiled DNA Molecule,* Science **271**, 1835 (1996).

[24] S. B. Smith, Y. Cui, and C. Bustamante, *Overstretching b-dna: the elastic response of individual double-stranded and single-stranded dna molecules,* Science **271**, 795 (1996).

[25] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, *Ionic effects on the elasticity of single DNA molecules,* Proceedings of the National Academy of Sciences **94**, 6185 (1997).

[26] G. Neuert, C. Albrecht, E. Pamir, and H. Gaub, *Dynamic force spectroscopy of the digoxigenin–antibody complex,* FEBS letters **580**, 505 (2006).

[27] M. de Odrowąż Piramowicz, P. Czuba, M. Targosz, K. Burda, and M. Szymoński, *Dynamic force measurements of avidin–biotin and streptavdin–biotin interactions using afm,* Acta Biochim. Pol **53**, 93 (2006).

[28] R. W. Friddle, A. Noy, and J. J. De Yoreo, *Interpreting the widespread nonlinear force spectra of intermolecular bonds,* Proceedings of the National Academy of Sciences **109**, 13573 (2012).

[29] R. Janissen, L. Oberbarnscheidt, and F. Oesterhelt, *Optimized straight forward procedure for covalent surface immobilization of different biomolecules for single molecule applications,* Colloids and Surfaces B: Biointerfaces **71**, 200 (2009).

[30] R. Janissen and F. Oesterhelt, *Directed deposition of single molecules on surfaces,* Journal of nanoscience and nanotechnology **10**, 5328 (2010).

[31] E. L. Bentzen, I. D. Tomlinson, J. Mason, P. Gresch, M. R. Warnement, D. Wright, E. Sanders-Bush, R. Blakely, and S. J. Rosenthal, *Surface modification to reduce non-specific binding of quantum dots in live cell assays,* Bioconjugate chemistry **16**, 1488 (2005).

[32] E. M. Puchner and H. E. Gaub, *Force and function: probing proteins with afm-based force spectroscopy,* Current opinion in structural biology **19**, 605 (2009).

[33] D. Ho, K. Falter, P. Severin, and H. E. Gaub, *Dna as a force sensor in an aptamer-based biochip for adenosine,* Analytical chemistry **81**, 3159 (2009).

[34] Z. Yang, J. A. Galloway, and H. Yu, *Protein interactions with poly (ethylene glycol) self-assembled monolayers on glass substrates: diffusion and adsorption,* Langmuir **15**, 8405 (1999).

[35] C. Rodriguez-Emmenegger, E. Brynda, T. Riedel, M. Houska, V. Šubr, A. B. Alles, E. Hasan, J. E. Gautrot, and W. T. Huck, *Polymer brushes showing non-fouling in blood plasma challenge the currently accepted design of protein resistant surfaces,* Macromolecular rapid communications **32**, 952 (2011).

[36] R. T. Chan, H. Marçal, R. A. Russell, P. J. Holden, and L. J. R. Foster, *Application of polyethylene glycol to promote cellular biocompatibility of polyhydroxybutyrate films,* International Journal of Polymer Science **2011** (2011).

[37] N. A. Alcantar, E. S. Aydil, and J. N. Israelachvili, *Polyethylene glycol-coated biocompatible surfaces,* Journal of biomedical materials research **51**, 343 (2000).

[38] S. Upadhyayula, T. Quinata, S. Bishop, S. Gupta, N. R. Johnson, B. Bahmani, K. Bozhilov, J. Stubbs, P. Jreij, P. Nallagatla, *et al.*, *Coatings of polyethylene glycol for suppressing adhesion between solid microspheres and flat surfaces,* Langmuir **28**, 5059 (2012).

[39] A. Jain, R. Liu, Y. K. Xiang, and T. Ha, *Single-molecule pull-down for studying protein interactions,* Nature protocols **7**, 445 (2012).

[40] B. Sweryda-Krawiec, H. Devaraj, G. Jacob, and J. J. Hickman, *A new interpretation of serum albumin surface passivation,* Langmuir **20**, 2054 (2004).

[41] P. Games, H. Keselman, and J. Rogan, *A review of simultaneous painvise multiple comparisons,* Statistica Neerlandica **37**, 53 (1983).

[42] F. Czerwinski, A. C. Richardson, and L. B. Oddershede, *Quantifying noise in optical tweezers by allan variance,* Optics express **17**, 13255 (2009).

[43] J. Israelachvili, *Differences between non-specific and bio-specific, and between equilibrium and non-equilibrium, interactions in biological systems,* Quarterly reviews of biophysics **38**, 331 (2005).

[44] D. Leckband and J. Israelachvili, *Intermolecular forces in biology,* Quarterly reviews of biophysics **34**, 105 (2001).

[45] I. Vilfan, J. Lipfert, D. Koster, S. Lemay, and N. Dekker, *Magnetic tweezers for single-molecule experiments,* in *Handbook of Single-Molecule Biophysics* (Springer, 2009) pp. 371–395.

[46] M. Kruithof, F. Chien, M. de Jager, and J. van Noort, *Subpiconewton Dynamic Force Spectroscopy Using Magnetic Tweezers,* Biophysical Journal **94**, 2343 (2008).

[47] A. J. W. te Velthuis, J. W. J. Kerssemakers, J. Lipfert, and N. H. Dekker, *Quantitative Guidelines for Force Calibration through Spectral Analysis of Magnetic Tweezers Data,* Biophysical Journal **99**, 1292 (2010).

[48] K. Kim and O. A. Saleh, *A high-resolution magnetic tweezer for single-molecule measurements,* Nucleic acids research **37**, e136 (2009).

[49] C. Bouchiat, M. D. Wang, J.-F. Allemand, T. R. Strick, S. M. Block, and V. Croquette, *Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements,* Biophysical Journal **76**, 409 (1999).

[50] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay, and F. Caron, *Dna: an extensible molecule,* Science **271**, 792 (1996).

[51] M. Manosas, A. Meglio, M. M. Spiering, F. Ding, S. J. Benkovic, F.-X. Barre, O. a. Saleh, J.-F. Allemand, D. Bensimon, and V. Croquette, *Magnetic Tweezers for the Study of DNA Tracking Motors,* in *Single Molecule Tools, Part B: Super-Resolution, Particle Tracking, Multiparameter, and Force Based Methods*, Methods in Enzymology, Vol. 475 (Elsevier, 2010) pp. 297–320.

[52] P. B. Tarsa, R. R. Brau, M. Barch, J. M. Ferrer, Y. Freyzon, P. Matsudaira, and M. J. Lang, *Detecting force-induced molecular transitions with fluorescence resonant energy transfer,* Angewandte Chemie International Edition **46**, 1999 (2007).

[53] T. Odijk, *Polyelectrolytes near the rod limit,* Journal of Polymer Science: Polymer Physics Edition **15**, 477 (1977).

[54] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, *Stretching DNA with optical tweezers,* Biophysical Journal **72**, 1335 (1997).

[55] A. Bosco, J. Camunas-Soler, and F. Ritort, *Elastic properties and secondary structure formation of single-stranded dna at monovalent and divalent salt conditions,* Nucleic acids research **42**, 2064 (2014).

[56] M. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, *Probing single-stranded dna conformational flexibility using fluorescence spectroscopy,* Biophysical Journal **86**, 2530 (2004).

[57] M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann, and H. E. Gaub, *How strong is a covalent bond?* Science **283**, 1727 (1999).

[58] C. Yuan, A. Chen, P. Kolb, and V. T. Moy, *Energy landscape of streptavidin-biotin complexes measured by atomic force microscopy,* Biochemistry **39**, 10219 (2000).

[59] P. Zheng, Y. Cao, and H. Li, *Facile method of constructing polyproteins for single-molecule force spectroscopy studies,* Langmuir **27**, 5713 (2011).

[60] A. Crut, P. A. Nair, D. A. Koster, S. Shuman, and N. H. Dekker, *Dynamics of phosphodiester synthesis by dna ligase,* Proceedings of the National Academy of Sciences **105**, 6894 (2008).

[61] G. Charvin, A. Vologodskii, D. Bensimon, and V. Croquette, *Braiding dna: experiments, simulations, and models,* Biophysical journal **88**, 4124 (2005).

4

HIGH-THROUGHPUT, HIGH-FORCE PROBING OF DNA-PROTEIN INTERACTIONS WITH MAGNETIC TWEEZERS

Science is nothing, but trained and organized common sense.

Thomas Huxley

Recent advances in high-throughput single-molecule magnetic tweezers have paved the way for obtaining information on individual molecules as well as ensemble-averaged behavior in a single assay. Here we describe how to design robust high-throughput magnetic tweezers assays that specifically require application of high forces (> 20 pN) for prolonged periods of time (> 1000 s). We elaborate on the strengths and limitations of the typical construct types that can be used and provide a step-by-step guide towards a high tether yield assay based on two examples. Firstly, we discuss a DNA hairpin assay where force-induced strand separation triggers a tight interaction between DNA-binding protein Tus and its binding site *Ter*, where forces up to 90 pN for hundreds of seconds were required to dissociate Tus from *Ter*. Secondly, we show how the LTag helicase of Simian virus 40 unwinds dsDNA, where a load of 36 pN optimizes the assay readout. The approaches detailed here provide guidelines for the high-throughput, quantitative study of a wide range of DNA-protein interactions.

This chapter has been submitted as: Bojk A. Berghuis, Mariana Koeber, Theo van Laar & Nynke H. Dekker. High-throughput, high-force probing of DNA-protein interactions with magnetic tweezers.

4.1. INTRODUCTION

IN recent decades, single-molecule techniques have become a valuable addition to existing bulk assays to study biological processes. The direct access to reaction kinetics has proven to provide crucial insights into the stochastic behavior and transient dynamics of individual molecules.[1–6] An additional challenge is to efficiently build up statistics in order to correctly place the occurrence of individual events in perspective of the total distribution of events. A single-molecule technique well equipped for this challenge is magnetic tweezers (MT), especially since recent developments in hardware and data acquisition software have paved the way to perform hundreds of single-molecule experiments simultaneously.[7, 8]

In any magnetic tweezers assay, single molecules are used to tether micrometer-sized magnetic beads to the surface of a fluid chamber. The beads are visualized through a microscope objective, and the resulting image is recorded with a camera (**Fig. 4.1a**). The properties of the individual molecules are inferred through the movement of the beads: continuously in motion due to Brownian motion of the surrounding water molecules, but limited in their movement as a result of the anchoring via the molecule of interest. An increased upward pulling force induced by lowering a pair of magnets towards the fluid chamber stretches the tethers and limits bead movement even further. It is this force-extension relationship that is molecule-specific and can be used to e.g. infer the rate at which a helicase unwinds a double-stranded (ds) DNA helix.[9, 10]

Choosing the best possible manner through which to read out enzyme activity or the binding of the protein of interest is arguably the single most fundamental and creative part of a single-molecule MT experiment. Most commonly, a DNA tether is used as a means to detect enzyme activity or protein binding in MT experiments. RNA tethers have also been used with great success [8, 11–16], but as these form a smaller subset of experiments, for simplicity we focus exclusively on DNA in this article. Given the ease with which DNA can be manipulated using tools of molecular biology, the possibilities for tether design are rich. This has been exemplified by the many creative designs used over the past years.[9, 10, 17–19] While summing up all designs used to date is beyond the scope of this paper, we identify roughly three classes of DNA constructs: the linear – forked or nicked – dsDNA construct, the rotationally constrained dsDNA construct that allows the introduction of supercoils [20], and the DNA hairpin (**Fig. 4.1b,c**).

A linear dsDNA construct is the archetypal manner to tether a bead to a surface as a means to infer the properties of this construct (or the proteins/enzymes interacting with it) through tracking the bead movement through conventional light microscopy (**Fig. 4.2a**). The behavior of dsDNA and single-stranded (ss) DNA under applied tension is well studied [9] and can be modeled using the worm-like chain (WLC) model or extensions thereof.[22, 23] In detecting enzyme activity or protein binding using a dsDNA construct, the difference in force-extension characteristics between dsDNA and ssDNA is essential (**Fig. 4.2d**), as the conversion of the ds helical structure into ssDNA (or protein-covered ssDNA) under force can then be observed as a change in the end-to-end extension of the construct/tether (~ 0.17 nm/bp at 25 pN). For a nicked or forked construct, the applica-

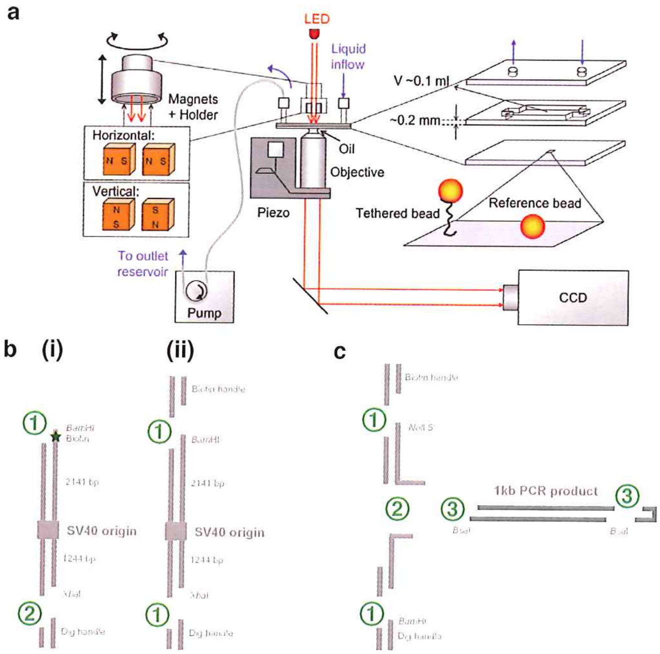


Figure 4.1: **The magnetic tweezers setup and construct design & assembly.** **(a)** The magnetic tweezers setup: an inverted microscope stage that is used to project the image of surface-tethered beads onto a CCD/CMOS camera chip.[21] **(b)** Typical design of a torsionally unconstrained (i) and a torsionally constrained (ii) dsDNA construct, different sequential or parallel ligation steps shown (green numbered circles) **(c)** Design and assembly strategy of the *Ter* DNA hairpin.

tion of force is on one strand and has a destabilizing effect on the double helix, culminating in the overstretching transition to ssDNA at forces around 65 pN. At lower forces (up to ~ 35 pN), pulling force is known to assist enzymes in plowing through the double helix, leading to less and shorter stochastic polymerase pausing, for instance.[8]

While a forked dsDNA assay suffices in cases where processive enzyme activity or extensive DNA-binding is measured, the readout of e.g. a single protein-binding event or a short burst of activity might be somewhat limited. Here the use of a rotationally constrained dsDNA construct might be more apt (**Fig. 4.2b**). At low forces (< 1 pN), a reduction of the linking number through the application of negative turns leads to the formation of plectonemic supercoils, and thereby to a decrease in tether extension (**Fig. 4.2e**).[24] At higher forces, it results in a reduction of twist and concomitant denaturation [25, 26] with initially little change in extension (**Fig. 4.2e**) until an increase is observed upon significant underwinding (**Fig. 4.4d**). Such a rotationally constrained dsDNA construct can be used e.g. to probe for removal of supercoils by topoisomerases [27–30] or to probe for protein- or enzyme-induced opening up of the double helix, as such increased unwinding leads to a compensatory change in dsDNA extension at low force.[3, 18] In the latter case, monitoring changes in the state of DNA supercoiling provides a more sensitive measure of enzymatic activity than simple ds to ss conversion, as also shown by experiments that probe progression of RNA polymerases by monitoring the positive torsional strain upstream of their location as they move along the DNA.[31]

A third class of DNA constructs is the DNA hairpin (**Fig. 4.2c**). The use of DNA hairpins has several important advantages compared to dsDNA constructs. First of all, as a DNA hairpin implies exertion of force on both strands of the double helix equally, this becomes a very direct way to measure unwinding activity of the myriad of enzymes equipped to do so. Applying a force lowers the energy barrier the enzyme has to overcome in the most direct way possible, namely by pulling apart the double helix. Secondly, DNA hairpins have a higher resolution in the readout compared to linear dsDNA constructs, as each disruption of a Watson-Crick base pair leads to a jump of two single-stranded inter-base distances (~ 0.6 nm/bp, as opposed to ~ 0.12 nm/bp at 12 pN for ds to ssDNA conversion on a linear construct).[14, 31] Thirdly, DNA hairpins allow for a direct control over the rehybridization process as the work required to unzip a DNA hairpin is equivalent to the work performed by an enzyme performing the unzipping (**Fig. 4.2f**). The work needed to melt a linear dsDNA construct lies significantly higher. However, the use of DNA hairpins also comes with certain drawbacks that must be taken into consideration. For example, the unwinding activity of a helicase on a hairpin under low force may not be readily detectable, as the newly exposed ssDNA will form a random coil.[32] Furthermore, the use of hairpins offers no possibility for torsional control (unless there is a third anchoring point [33]) and lastly, there is always tension on an initiation site at or downstream of the fork.

As the DNA construct leaves the drawing board, an often-recurring problem arising in making the DNA construct is low yield of the final product. As there are usually several intermediate products (e.g. biotin- or digoxygenin-labelled handles, hairpin loops, mismatch regions, primers, etc.) that need to be ligated together in one or several steps, the yield is highly dependent on the order through which this happens as well as the choice of the restriction sites/enzymes.

Over recent years, we have refined and improved the method of DNA construct production and assembly, and have come up with a robust method to produce a wide variety of construct designs based on a few fundamental steps. We will elaborate on two different dsDNA constructs as well as a DNA hairpin. Furthermore, as construct yield is the initial of many steps towards tether yield in the experiment, we discuss the subsequent experimental setup and present several strategic choices to be made leading to a successful high-throughput single-molecule experiment, exemplified through a number of typical enzymatic and protein-nucleic acid interaction assays. We first touch upon the assay-specific calibrations necessary to perform the experiments and then elaborate on two protein-DNA interaction assays: an assay where the specific binding and locking of Tus onto a 23 bp *Ter* site is investigated with a DNA hairpin [34], followed by tracking of the enzymatic unwinding activity of the LTag helicase as it unwinds dsDNA.

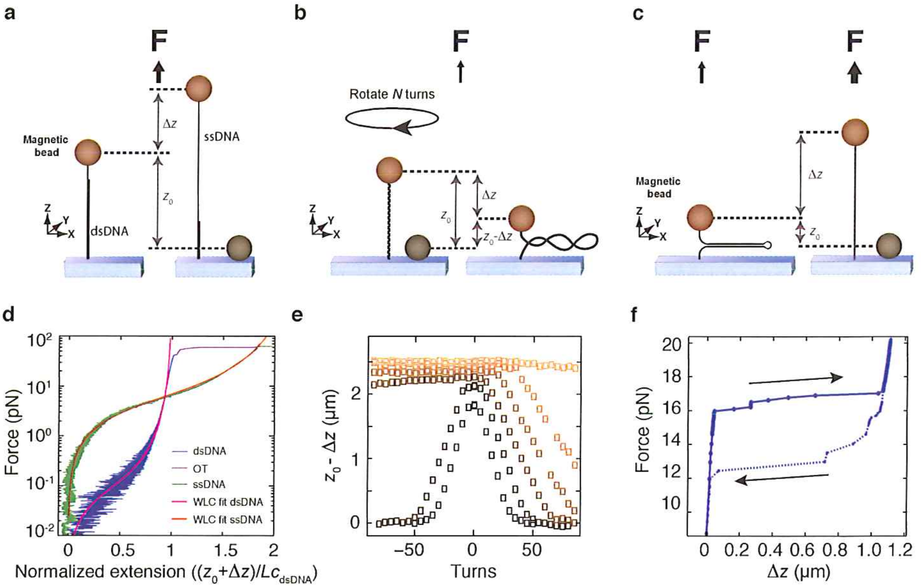


Figure 4.2: **The magnetic tweezers assays and examples of the achievable readouts.** **(a)** A linear dsDNA construct where only one of the DNA strands is attached to the bead renders it rotationally unconstrained and implies the pulling force is applied on one of the ssDNA strands only. We calibrate the extension change that would result from ds to ssDNA conversion through enzymatic activity by measuring the force-extension relationships for dsDNA (blue data) and ssDNA (green data) in separate experiments, as shown in **(d)**. At the overstretching transition (OT, magenta), the force is sufficiently high to allow sudden peeling of the dsDNA helix, such that the tethered strand becomes single-stranded. The difference in electrostatic properties of ds and ssDNA imply that at forces above ~ 8 pN ds-to-ss conversion yields tether lengthening, while at forces lower than ~ 8 pN tether shortening would be observed. **(b)** The application of turns to a torsionally constrained dsDNA. At low forces, plectonemic supercoils (Wr) can be introduced by applying turns to a torsionally constrained dsDNA as illustrated schematically. The response of dsDNA to applied rotation at different forces is indicated (e, from dark to light, $F = 0.25, 0.5, 1.1, 2, 3.5$, and 6.5 pN). **(c)** A DNA hairpin is closed at low forces (< 13 pN), so the end-to-end extension $\langle z \rangle$ is zero. Increasing the force (> 16 pN) will open the hairpin: for every broken base pair, two ssDNA base lengths are added to the tether length. This force extension relationship is also characterized in a force-extension experiment **(f)**.

4.2. METHODS

4.2.1. PREPARATION OF DNA HAIRPINS

The 1.1 kb *Ter* hairpins were generated by PCR using plasmids pTER and pTER-Rev as a template, containing the *TerB* site in either the nonpermissive or permissive orientation, respectively, and flanked by phage λ sequences, were obtained from Invitrogen (Invitrogen/Life Technologies, Carlsbad, CA). The hairpins were constructed in a multistep process, where the specific order of assembly described here contributes significantly to the final product yield (**Fig. 4.1c**). First, a 1-kb fragment containing the *TerB* site was amplified from the pTER plasmid using primers 1 and 2 (**Table 4.1**). This fragment was digested with the nonpalindromic restriction enzyme BsaI (New England BioLabs Inc., Ipswich, MA) and ligated at one end to a 42-bp oligonucleotide forming a U-turn (oligonu-

Primer no.	Oligonucleotide sequence
1	5’-CTGCGGTCTCGTTGCTTACCGTCACCAGAAATTACCGTCAC-3’
2	5’-CCATCTTGGTCTCCTAGGTTTTTAGCAGCGAAGCGTTTGATAAG-3’
3	5’-CCTAAGCTCGCCGAGGCGAGCGAAAGCTCGCCTCGGCGAGCT-3’
4	5’-GACCGAGATAGGGTTGAGTG-3’
5	5’-CAGGGTCGGAACAGGAGAGC-3’
6	5’-GGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTT-3’
7	5’-GGCCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTG-CCATGCTCTTTACAACCGGTTGACTGCTTCAGGGGTCGATCCCGCTTTGTAC-3’
8	5’-GATCTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAG-ATAACTACGATACGGGAGGGCTTACCATCTGGC-3’
9	5’-GCAAGTACAAAGCGGGATCGACCCCTGAAGCAGTCAACCGGTTGTAAAGA-GCATCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGCC-AGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAAC-TATGGATGAACGA-3’
10	5’-CCATCTTGGTCTCCGACATTATAGCACAGTCGTGGTGAC-3’
11	5’-CTGCGGTCTCGAGGCGGTTAATATTATGGCGCGTTG-3’
12	5’-P-GCCTACTTTAGTTACAACATACTTATT-3’
13	5’-P-TGTCAAACCTCATGTTGTAACTAAAGT-3’

Table 4.1: Used oligonucleotide sequences

cleotide 3). The hairpin handles were created by PCR amplification of a 1.2-kb pBluescript SK+ (Stratagene–Agilent Technologies, Santa Clara, CA) fragment using primers 4 and 5 in the presence of either biotin-16-dUTP or digoxigenin-11-dUTP (Roche Diagnostics, Basel, Switzerland). Prior to ligation to spacer oligonucleotides, handles were digested with either *BamHI* or *NotI*. The upper spacer of the hairpin was generated by annealing 5’-phosphorylated oligonucleotides 6 and 7 and ligating this double-stranded DNA fragment to the *NotI*-digested biotin-labeled handle. The lower spacer was made by annealing 5’-phosphorylated primers 8 and 9 and ligating them to the *BamHI*-digested digoxigenin-labeled handle. Finally, the overhangs of these handle-spacer constructs were allowed to anneal to form a short (50-bp) stem with a 5’-GCAA overhang that was ligated to the complementary *BsaI* site of the 1-kb *TerB* fragment. Oligonucleotides were obtained from Biolegio B.V., Nijmegen, the Netherlands and from Ella Biotech GmbH, Martinsried, Germany.

4.2.2. PREPARATION OF A DNA HAIRPIN CONTAINING A MISMATCH REGION

To create a 1 kb fragment containing a 5-base mismatch between bases 3–7 in the *Ter* site, two fragments of 500 bp were generated by PCR using pTER as template and primer combinations 1 and 10, and 2 and 11 (**Table 4.1**), respectively. These fragments were digested with BsaI and ligated to each end of the annealed primer pair 12 and 13 containing the mismatch.

4.2.3. PREPARATION OF 3.4 KB DSDNA CONSTRUCT CONTAINING THE SV40 ORIGIN OF REPLICATION

To monitor enzyme activity, we engineered a torsionally unconstrained 3.4 kb dsDNA construct containing the SV40 origin of replication close to the center of the construct (**Fig. 4.1b (i)**). For calibration purposes, where we mechanically unwind the DNA, we created a torsionally constrained construct (**Fig. 4.1b (ii)**). The design of these constructs is based on the plasmid pRL-SV40 (Promega Corporation, Madison, WI), with handles containing biotin on one end and digoxigenin on the other end. Using restriction sites for *BamHI* and *XbaI* in the plasmid, a fragment was created with 2141 bp on one side of the SV40 origin and 1244 bp on the other side of the origin. For the torsionally constrained constructs, 600 bp handles containing either biotin or digoxigenin are ligated on the ends. These handles are amplified by PCR from pBluescriptSKII+ using forward primer 5'- gaccgagatagggttgagtg and reverse primer 5'- cagggtcggaacaggagagc in the presence of biotin-16-dUTP or digoxigenin-11-dUTP (Roche Diagnostics, Basel, Switzerland) in a ratio of 1:5 with dTTP. Therefore, the expected number of labelled nucleotides is approximately 60 per 600-bp handle. For the torsionally unconstrained construct, pRL-SV40 (Promega Corporation, Madison, WI) was digested with BamHI and the 5'-overhang was filled with Klenow (New England BioLabs Inc., Ipswich, MA) in the presence of biotin-labelled dATP (Invitrogen/Life Technologies, Carlsbad, CA). Subsequently, the DNA was digested with XbaI and ligated to a digoxigenin-labelled handle that was digested with XbaI as described above.

4.2.4. FUNCTIONALIZATION OF FLOW CELL SURFACES

Glass microscope cover slips (#1, 24 × 60 mm, Menzel GmbH, Germany) serve as the flow cell surface for DNA anchoring as well as the top for sealing off the flow cell. Inlet/outlet holes were drilled in the top cover slip. Then all cover slips were placed into a Teflon holder and sonicated in acetone for 30 minutes, followed by another sonication step in 2-propanol for another 30 minutes, after which they were allowed to air dry. Then the bottom glass surfaces were coated with nitrocellulose. For this, nitrocellulose membrane paper (Invitrogen, USA) was dissolved (1% m/V) in ethanol as a stock by mixing the components in a tube shaker at 35°C. Prior to use, the nitrocellulose solution was diluted to 0.2% m/V. At this stage polystyrene (3 μm diameter) reference beads can be added to the nitrocellulose solution, the concentration of beads depends on the MT field of view size (typically a $\sim 1000\times$ dilution is used). Another option would be to add the polystyrene reference beads dissolved in buffer to the flow cell at a later stage. The melting allows for more thorough surface attachment with the risk of having clusters of molten beads on your surface, the addition at a later stage allows for better control, with the risk of losing more beads when flushing. Irrespective of the choice made, 5 min sonication of the polystyrene beads prior to addition is recommended. Functionalization of the flow cell surface is performed by adding 3 μl of the nitrocellulose solution and spreading it out evenly using the lateral side of the pipette tip. If this solution contains reference beads, the cover glasses should be heated to 150 °C for 3 min on hot plate.

4.2.5. FLOW CELL ASSEMBLY AND PREPARATION

A double layer of parafilm spacer was placed onto the functionalized surfaces, and the flow cell was closed by a second coverslip on top containing inlet and outlet holes. The cover slip parafilm sandwich was sealed by melting the parts together at 90°C for ~ 30 s. Prior to the addition of DNA-linked magnetic beads, the bottom surface was functionalized by incubation with 1 mg/ml anti-digoxigenin (Roche) in PBS for 30 min to provide for DNA attachment. The surface was passivated by incubating blocking aid (an undisclosed mix of proteins from Sigma) for 10 min followed by a 10 min incubation of high salt buffer (e.g. 700 mM KCl).

4.2.6. BEAD TETHERING AND POST-INCUBATION CLEANUP

DNA constructs (final concentration ~50 pg/μl) were mixed and incubated for 2 min with 20 μl streptavidin-coated paramagnetic beads (M270 Dynabeads) at room temperature in Tris buffer (50 mM Tris-HCl pH 7.9, 50 mM KCl, 0.1 mM EDTA, 0.01% Triton X-100). Triton X-100 is used to avoid bead clustering. The supernatant was replaced by 50 μl Tris buffer followed by a 15 min incubation of the bead-DNA solution in the flow cell containing an anti-digoxigenin–coated nitrocellulose surface. Non-tethered beads were removed by flushing with 1 ml Tris buffer, applying a high (30–40 pN) force while rotating the magnets (10 turns and back and forth repeatedly at 10 Hz), and followed by flushing with more buffer (at low force (0.1 pN) to keep the removed beads from settling back onto the tethers) until all non-tethered beads had been flushed out.

4.2.7. SPECIFICATIONS OF THE MAGNETIC TWEEZERS

The magnetic tweezers implementation used in this study has been previously described.[7, 8, 35] Briefly, light transmitted through the sample was collected by an oil-immersion objective (Olympus UPLSAPO60XO 60×, numerical aperture (NA) = 1.35, Olympus, USA) and projected onto a 12-megapixel CMOS camera (Falcon FA-80-12M1H, Teledyne Dalsa, Canada) with a sampling frequency of 58 Hz at full field of view, or higher when cropped. A 2-inch 200-mm tube lens between objective and camera resulted in an effective magnification of 67×. As a result of the aforementioned camera and magnification specifications, the field of view (fov) size is approximately $300 \times 400 \mu$m, allowing for the possibility of tracking hundreds of beads simultaneously. The applied magnetic field was generated by a pair of vertically aligned permanent neodymium-iron-boron magnets (SuperMagnete, Switzerland) separated by a distance of 1.0 or 0.5 mm and suspended on a motorized stage (M-126.PD2, Physik Instrumente, Germany) above the flow cell. Additionally, the magnet pair could be rotated about the illumination axis by an applied DC servo step motor (C-150.PD, Physik Instrumente, Germany).

4.2.8. PRE-EXPERIMENT HAIRPIN CALIBRATION AND TESTING

Prior to measuring the dwell times of *Ter*-bound Tus on the DNA hairpins, we characterized the two DNA hairpins described above in force-extension experiments. The purpose of the calibration experiments is to check the constructs for their length, the opening force of the bare hairpin and verify the location of the *Ter* site upon the addition of Tus. To obtain the force extension curves we start at low force (~8 pN) and lower the magnet position in a linear fashion until a full opening of the DNA hairpin is observed, and back (**Fig. 4.2c**). The linear magnet movement will lead to an exponential force ramp, which for the purpose of this experiment yields a rapid and straightforward qualitative test to check whether the hairpins produce the expected unfolding and folding pattern (**Fig. 4.2f**).

4.2.9. CANONICAL PRE-MEASUREMENT TEST OF DSDNA TETHERS

For each experiment, we select for DNA tethers that have the expected length by measuring the difference in tether extension between high (> 10 pN) and low (< 100 fN) forces. At zero force the bead bounces on the flow cell surface, so the lowest measured extension (after filtering appropriately) will define zero extension. At forces above ~ 10 pN the DNA is fully stretched such that the average extension at this force will yield the tether contour length.[8, 36]

4.2.10. DATA ACQUISITION

Image processing of the collected bead diffraction patterns was used to track the real-time position of both surface-attached reference beads and superparamagnetic beads coupled to DNA tethers in three dimensions. We implemented custom written software in C++, CUDA and LabView (2011, National Instruments Corporation, USA) that is suited for high-throughput tracking in magnetic tweezers.[7] Tracking of the x,y coordinates is performed using center-of-mass computation followed by a further refinement using the quadrant interpolation algorithm. Localization of the bead's z-coordinate is achieved by creating a radial profile using the refined x,y coordinates and comparing this profile to a prerecorded look-up table of radial profiles. After subtraction of the reference bead position to correct for instrumental drift, the x, y and z positions of the DNA-tethered beads were determined with a spatial accuracy of < 3 nm. The upward stretching forces on the DNA tethers by the superparamagnetic beads were calibrated from analysis of the extent of their Brownian motion, whereby spectral corrections were employed to correct for camera blur and aliasing.[37, 38]

4.3. MEASUREMENT TYPES

We now combine the DNA constructs described above with several measurement approaches, of which all exploit the multiplexing capacity of MT to gain statistics on biologically relevant processes that either have a low yield, or take a long time to acquire. All measurements described here are taken at high forces (> 20 pN) for prolonged periods of time (up to ~ 2000 s). These types of measurements require the combination of M270 beads and vertically aligned magnets with a gap size of 0.3-1 mm.[38] These measurements also strongly benefit from the high anti-digoxigenin concentration used here (1 mg/ml) and tether handles containing multiple digoxigenin- and biotin-dUTP moieties.

4.3.1. TUS MEASUREMENTS ON HPDNA — MEASURING A SINGLE DWELL TIME MANY TIMES

Recently, we investigated the tight interaction between the DNA binding protein Tus and its cognate DNA-binding sequence *Ter* -- the system known to be involved in the termination of DNA replication in *Escherichia coli*.[34] It had previously been argued that the tight interaction was dependent on specific interactions between Tus and the replisome.[39] By using a multiplexed MT DNA hairpin assay (**Fig. 4.2c**) to mimic fork progression in the absence of replisome proteins, it was shown that protein–protein interactions were not necessary for tight Tus–*Ter* interactions. Gathering the necessary dwell-time statistics would have been very time-consuming without the ability to multiplex, as strand separation frequently remained blocked at the Tus–*Ter* site for hundreds of seconds.

4.3.2. HAIRPIN DWELL TIME MEASUREMENTS

For the purpose of clarity, we here discuss the measurements performed on the hairpin construct that contains the *Ter* site in the blocking (nonpermissive) orientation only. Constant-force dwell time experiments were obtained by lowering the magnets in a linear fashion (10 mm/s) to the desired distance. The dwell time is the time measured between arrival of the magnets at their final position and the further opening of the hairpin from the locked to the fully opened state. Rupture of the Tus–*Ter* lock results in a sudden opening of the DNA hairpin (**Fig. 4.3a**): rupture points were easily identified as a sharp peak in the derivative of the z-trace.

The magnetic tweezers multiplexing capability allows for rapid collection of Tus-*Ter* dwell time distributions (**Fig. 4.3a**), as a typical experiment starts out with ~ 80 DNA hairpins. The dwell times showed to be multi-exponentially distributed and typically three single exponentials were needed to fit a dwell time distribution of Tus-*Ter*. This implies that there are three states in which the Tus-*Ter* lock can find itself trapped, each state with its own characteristic lifetime and probability (**Fig. 4.3b**). The longest lived state was found to be the signature of full lock formation, and this state was longest-lived for wt Tus-*Ter*. Mutations in the lock domain led to a decrease in the lifetime and/or probability of this state, while mutations in other Tus domains led to wt-like behavior (**Fig. 4.3c,d**).

4.3.3. ENZYMATIC ACTIVITY — LTAG UNWINDING DSDNA — MEASURING MANY DWELL TIMES IN A SINGLE BURST OF ACTIVITY

Using our assays, we have also studied the unwinding of duplex DNA by large tumor antigen (LTag) [40], the helicase of the Simian Virus 40 (SV40), which serves as a model system to understand eukaryotic DNA replication.[41] Two LTag hexamers assemble head-to-head at the origin of replication, and subsequently unwind dsDNA bidirectionally. While electron microscopy suggested that the double hexamer may function as a stationary unit that draws in the parental DNA [42], single-molecule fluorescence studies found efficient unwinding by two separate hexamers travelling apart.[43] We have probed whether the two LTag hexamers function as a single unit or not by using magnetic tweezers to investigate LTag-based unwinding with nanometer resolution at high forces (~ 35 pN). At this force, the extension of ssDNA exceeds that of dsDNA. If the two hexamers separate to unwind the DNA, the creation of ssDNA in an unwinding bubble should (providing absence of rehybridization) result in an increase in tether extension (**Fig. 4.4a**). If, on the other hand, DNA is being unwound by a stationary double hexamer that draws in parental DNA, an extension decrease is expected (**Fig. 4.4b**). Similarly to the experiments probing Tus binding described above, here we also benefit from the

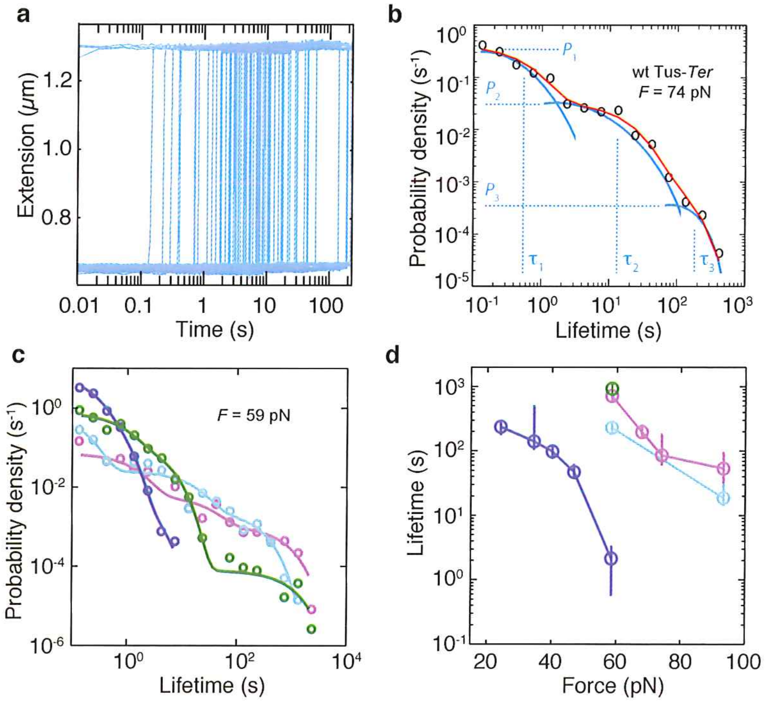


Figure 4.3: **Experimental readout as well as the resulting dwell time and lifetime distributions of Tus-*Ter* DNA hairpin experiments.** **(a)** Tus-*Ter* rupture events on the DNA hairpin construct are marked by a sudden, single jump in the extension, when the hairpin opens from the *Ter* site (extension = ~ 0.6μm) to the fully single-stranded form (ext. = 1.3μm). Data taken at 74 pN with wt Tus-*Ter*. **(b)** The resulting dwell time distribution of the dataset represented in **(a)**, which contains multiple single exponential distributions (global fit in red, guide to the eye in cyan). **(c)** The dwell-time distributions are altered by mutations in (H144A, blue) or near (E49K, green) the lock domain, but less affected by a mutation elsewhere (Q250A, cyan) when compared to wt Tus-*Ter* (purple, circles are data, solid lines are fits). **(d)** The lifetime extracted for the longest-lived state shows that Tus-*Ter* lock lifetimes depend on force and that a mutation in the lock domain of Tus is what mainly causes a deviation from wt-like behavior (same color scheme as **(c)**, the error bars indicate the $1-\sigma$ confidence intervals).

use of high-throughput magnetic tweezers, as they allow us to obtain multiple unwinding traces in a single experiment, limited only by the efficiency of helicase assembly on dsDNA in the flow cell.

4.3.4. CALIBRATING THE EXTENSION CHANGE INVOKED BY THE FORMATION OF AN UNWINDING BUBBLE

To quantify the number of basepairs unwound by the LTag helicase, we need to determine the extension change that results from the unwinding of a dsDNA base pair at a given force. To this end, we mechanically unwind an initially relaxed, but torsionally constrained, DNA (**Fig. 4.4c**). This reduces its linking number *Lk*, which is the sum of twist (Tw) and writhe (Wr). Changes in *Lk* are conveniently expressed using

$\sigma = (Lk - Lk_0)/Lk_0 = (\Delta Tw + Wr)/Lk_0$, where Lk_0 is the linking number of torsionally relaxed DNA.[44] When we reduce σ from 0 to -2 at low forces (< 2.5 pN), this primarily results in the formation of plectonemic supercoils (negative Wr) and thereby to a decrease in tether extension (**Fig. 4.4d**, e.g. blue data). A similar decrease of σ at higher forces yields a decrease in Tw and concomitant denaturation [26, 45], resulting in an extension increase (**Fig. 4.4d**, e.g. red data). At the highest forces probed (36 pN), the unwinding of duplex DNA leads to a significant increase in DNA extension. Fitting the expression $N_u = (\Delta l - a)/b$ to this dataset (for $-1.7 < \sigma < 0$), where N_u is the number of unwound basepairs, allows us to calculate the degree of unwinding of this 3.4 kb construct from the change in DNA extension Δl, and provides the basis for probing LTag unwinding.

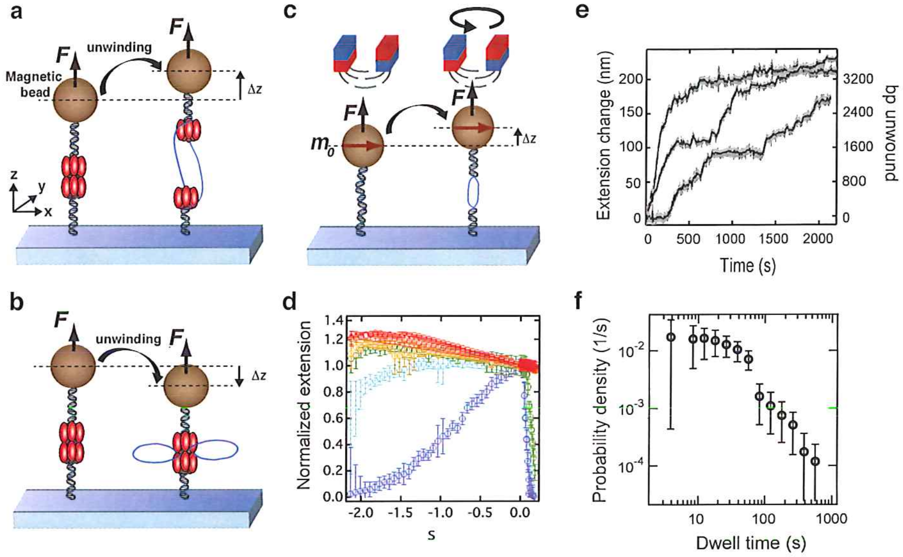


Figure 4.4: **Experimental readout of LTag unwinding dsDNA containing an origin of replication. (a), (b)** Unwinding of dsDNA either by two separate hexamers **(a)** or by a stationary double hexamer **(b)**. The unwinding of dsDNA by two separating hexamers will create an unwinding bubble. At high forces (> 6 pN), this will lead to an extension increase. A stationary double hexamer, on the other hand, will draw in parental dsDNA, and the extension will decrease. **(c)** To calibrate the extension change that results from the unwinding of a base pair at a given force, we unwind a torsionally constrained DNA molecule mechanically. **(d)** A reduction of the linking number through the application of negative turns leads to an extension decrease at low force, as a result of the formation of plectonemic supercoils, and to an extension increase at higher forces, due to a reduction of twist and concomitant denaturation (from blue to red, $F = 1.2, 2.5, 6, 12, 24$, and 36 pN; measurements were performed in 10 mM Tris-HCl, pH 7.4, and 7 mM $MgCl_2$). Fitting the expression $N_u = (\Delta l - a)/b$ to the dataset at 36 pN, we find $a = 0.3$ nm and $b = 0.0625$ nm/bp unwound. **(e)** Typical traces of LTag unwinding different dsDNA molecules. The unwinding yields an extension increase, indicating that under the tested conditions, dsDNA is likely unwound by two separate hexamers as depicted in **(a)**. The traces shown here were acquired in the presence of 17 nM LTag hexamer and 3 nM RPA. Full-bandwidth traces are shown along with 1 Hz low-pass filtered data. **(f)** From the median of the dwell time distribution of the dataset shown in **(e)**, we determine an unwinding rate of 4.4 bp/s (the dwell times are normalized to one bp unwound).

4.3.5. DNA UNWINDING BY LTAG HELICASE

To monitor DNA unwinding by LTag helicase, we approached the magnets to the flow cell to apply the desired force. 17 nM LTag hexamer was introduced into the flow cell in a buffer containing 10 mM Tris-HCl, pH 7.4, and 7 mM $MgCl_2$, supplemented with 6 mM adenosine triphosphate (ATP), 0.1 mg/mL bovine serum albumin (BSA), 1 mM dithiothreitol (DTT), and 6.8 nM competitor DNA (linearized plasmid pUC19, 2.7 kbp) to impede origin-independent unwinding.[46] After 45 min, 3 nM human replication protein A (RPA) was added in the same buffer to stabilize newly formed ssDNA, supplemented with 6 mM ATP, 0.1 mg/mL BSA, and 1 mM DTT.

Under these conditions, we observed activity associated increase in the extension for ~ 1/3 of the DNA tethers present in the flow cell (**Fig. 4.4e**), with remaining tethers showing no activity, likely due to absent or inefficient helicase loading. LTag processivity is low in the absence of RPA, which is most likely due to a poor stabilization of ssDNA by LTag alone. In the presence of RPA however, we mostly observe complete LTag mediated construct unwinding (**Fig. 4.4e**). For all acquired traces showing activity, we constructed a dwell-time distribution. From the median of this dwell-time distribution, we determine an unwinding rate for LTag helicase of 4.4 bp/s (**Fig. 4.4f**), in agreement with [43, 47]. Thus, this establishes an assay with which we can probe the unwinding of dsDNA by LTag starting from an origin of replication. Notably, in our high force experiments, we never observed a decrease in the extension of the DNA tethers. This suggests that, under the probed conditions, two spatially separated LTag hexamers unwind the dsDNA (**Fig. 4.4a**). Future experiments will be required to determine whether this unwinding mechanism is similarly applied at lower forces.

4.3.6. TIPS

- If the yield of the DNA construct is low, it is advisable to reconsider the order in which different fragments are ligated.
- The use of non-palindromic restriction enzymes (e.g. *BsaI*) avoids self-ligation of the PCR products or primers, resulting in an increased construct yield.
- Two easy to change factors contribute significantly to DNA tether strength: the Anti-dig concentration (a higher concentration implies stronger tethering) and the DNA-digoxygenin/biotin handle length (the longer, the more anchoring points).
- Pre-experiment cleanup: avoid being too careful with the tethers (weakly bound tethers will add to mid-experiment failures)
- Post-experiment cleanup: after performing the experiment, flush through a low-salt buffer through in high-speed bursts (if needed combined with tapping the tubing connected to the flow cell): this will remove all the tethers and stuck beads, the flow cell is ready for another round of experiments.
- Keep the occasions where air bubbles enter the flow cell chamber to a minimum: this will increase the lifetime of the flow cell.

- Add detergent (e.g. Triton X-100, 0.01% v/v) to buffers to avoid (reference) bead clustering. After the post incubation cleanup buffers lacking detergent (proteins might be sensitive to detergent) can be used.

- Keep in mind the reference bead size (bigger is better).

4.4. CONCLUDING REMARKS

Here we have described detailed protocols and methods for performing high force, multiplexed MT measurements. Improvements on the design of DNA constructs allow us to producing the DNA constructs with a high yield, which directly translates into a high DNA tether yield in the MT assay. We have drastically reduced tether loss caused by prolonged exposure to high forces (20–95 pN) by using long (600 bp) dsDNA handles labeled with multiple digoxygenin or biotin labels and by increasing the anti-digoxygenin concentration on the flow cell surface. By performing multiplexed measurements on DNA hairpins containing a single nonpermissive *Ter* site we were able to capture the distribution of dwell times that arises from the rupture of the wt Tus-*Ter* lock and various Tus mutants. Using a dsDNA molecule containing the origin of replication for the LTag helicase, we show that the dsDNA is unwound by two separate hexamers. These experiments demonstrate that the MT instrumentation and protocols described here have reached a level where performing multiplexed, high force single-molecule assays has become routine to such an extent that bridging the gap between the dynamics of a single molecule and the ensemble-averaged behavior is now within the realm of possibility.

REFERENCES

[1] A. M. van Oijen and J. J. Loparo, *Single-molecule studies of the replisome,* Annual review of biophysics **39**, 429 (2010).

[2] A. N. Kapanidis, E. Margeat, S. O. Ho, E. Kortkhonjia, S. Weiss, and R. H. Ebright, *Initial transcription by rna polymerase proceeds through a dna-scrunching mechanism,* Science **314**, 1144 (2006).

[3] A. Revyakin, C. Liu, R. H. Ebright, and T. R. Strick, *Abortive initiation and productive initiation by rna polymerase involve dna scrunching,* Science **314**, 1139 (2006).

[4] M. H. Larson, R. Landick, and S. M. Block, *Single-molecule studies of rna polymerase: one singular sensation, every little step it takes,* Molecular cell **41**, 249 (2011).

[5] C. Bustamante, W. Cheng, and Y. X. Mejia, *Revisiting the central dogma one molecule at a time,* Cell **144**, 480 (2011).

[6] C. E. Aitken, A. Petrov, and J. D. Puglisi, *Single ribosome dynamics and the mechanism of translation,* Annual review of biophysics **39**, 491 (2010).

[7] J. P. Cnossen, D. Dulin, and N. H. Dekker, *An optimized software framework for real-time, high-throughput tracking of spherical beads,* Review of Scientific Instruments **85**, 103712 (2014).

[8] D. Dulin, I. D. Vilfan, B. A. Berghuis, S. Hage, D. H. Bamford, M. M. Poranen, M. Depken, and N. H. Dekker, *Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase,* Cell Reports **10**, 983 (2015).

[9] C. Bustamante, Z. Bryant, and S. B. Smith, *Ten years of tension: single-molecule DNA mechanics,* Nature **421**, 423 (2003).

[10] T. Lionnet, A. Dawid, S. Bigot, F.-X. Barre, O. A. Saleh, F. Heslot, J.-F. Allemand, D. Bensimon, and V. Croquette, *Dna mechanics as a tool to probe helicase and translocase activity,* Nucleic acids research **34**, 4232 (2006).

[11] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco Jr, and C. Bustamente, *Reversible Unfolding of Single RNA Molecules by Mechanical Force,* Science **292**, 733 (2001).

[12] J. Abels, F. Moreno-Herrero, T. Van der Heijden, C. Dekker, and N. Dekker, *Single-molecule measurements of the persistence length of double-stranded rna,* Biophysical journal **88**, 2737 (2005).

[13] S. Dumont, W. Cheng, V. Serebrov, R. K. Beran, I. Tinoco, A. M. Pyle, and C. Bustamante, *Rna translocation and unwinding mechanism of hcv ns3 helicase and its coordination by atp,* Nature **439**, 105 (2006).

[14] W. Cheng, S. Arunajadai, J. Moffitt, I. Tinoco Jr., and C. Bustamante, *Single-base pair unwinding and asynchronous* RNA *release by the hepatitis* C *virus* NS*3 helicase,* Science **333**, 1746 (2011), cited By 40.

[15] J.-D. Wen, L. Lancaster, C. Hodges, A.-C. Zeri, S. H. Yoshimura, H. F. Noller, C. Bustamante, and I. Tinoco, *Following translation by single ribosomes one codon at a time,* Nature **452**, 598 (2008).

[16] J. Lipfert, G. M. Skinner, J. M. Keegstra, T. Hensgens, T. Jager, D. Dulin, M. Köber, Z. Yu, S. P. Donkers, F.-C. Chou, *et al.*, *Double-stranded rna under force and torque: Similarities to and striking differences from double-stranded dna,* Proceedings of the National Academy of Sciences **111**, 15408 (2014).

[17] J. Gore, Z. Bryant, M. D. Stone, M. Nöllmann, N. R. Cozzarelli, and C. Bustamante, *Mechanochemical analysis of dna gyrase using rotor bead tracking,* Nature **439**, 100 (2006).

[18] A. Revyakin, R. H. Ebright, and T. R. Strick, *Promoter unwinding and promoter clearance by rna polymerase: detection by single-molecule dna nanomanipulation,* Proceedings of the National Academy of Sciences of the United States of America **101**, 4776 (2004).

[19] Z. Bryant, F. C. Oberstrass, and A. Basu, *Recent developments in single-molecule dna mechanics,* Current opinion in structural biology **22**, 304 (2012).

[20] D. Dulin, J. Lipfert, M. C. Moolman, and N. H. Dekker, *Studying genomic processes at the single-molecule level: introducing the tools and applications,* Nat Rev Genet **14**, 9 (2013).

[21] J. Lipfert, X. Hao, and N. Dekker, *Quantitative modeling and optimization of magnetic tweezers,* Biophysical Journal **96**, 5040 (2009), cited By 43.

[22] P. Gross, N. Laurens, L. B. Oddershede, U. Bockelmann, E. J. Peterman, and G. J. Wuite, *Quantifying how dna stretches, melts and changes twist under tension,* Nature Physics **7**, 731 (2011).

[23] C. Bouchiat, M. D. Wang, J.-F. Allemand, T. R. Strick, S. M. Block, and V. Croquette, *Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements,* Biophysical Journal **76**, 409 (1999).

[24] J. Lipfert, J. W. J. Kerssemakers, T. Jager, and N. H. Dekker, *Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments,* Nature Methods **7**, 977 (2010).

[25] J. Allemand, D. Bensimon, R. Lavery, and V. Croquette, *Stretched and overwound dna forms a pauling-like structure with exposed bases,* Proceedings of the National Academy of Sciences **95**, 14152 (1998).

[26] M. Y. Sheinin, S. Forth, J. F. Marko, and M. D. Wang, *Underwound dna under tension: structure, elasticity, and sequence-dependent behaviors,* Physical review letters **107**, 108102 (2011).

[27] T. R. Strick, V. Croquette, and D. Bensimon, *Single-molecule analysis of dna uncoiling by a type ii topoisomerase,* Nature **404**, 901 (2000).

[28] N. Dekker, V. Rybenkov, M. Duguet, N. Crisona, N. Cozzarelli, D. Bensimon, and V. Croquette, *The mechanism of type ia topoisomerases,* Proceedings of the National Academy of Sciences **99**, 12126 (2002).

[29] D. A. Koster, V. Croquette, C. Dekker, S. Shuman, and N. H. Dekker, *Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB,* Nature **434**, 671 (2005).

[30] D. A. Koster, K. Palle, E. S. Bot, M.-A. Bjornsti, and N. H. Dekker, *Antitumour drugs impede dna uncoiling by topoisomerase i,* Nature **448**, 213 (2007).

[31] Y. Harada, O. Ohara, A. Takatsuki, H. Itoh, N. Shimamoto, and K. Kinosita, *Direct observation of dna rotation during transcription by escherichia coli rna polymerase,* Nature **409**, 113 (2001).

[32] N. Ribeck, D. L. Kaplan, I. Bruck, and O. A. Saleh, *Dnab helicase activity is modulated by dna geometry and force,* Biophysical journal **99**, 2170 (2010).

[33] J. T. Inman, B. Y. Smith, M. A. Hall, R. A. Forties, J. Jin, J. P. Sethna, and M. D. Wang, *Dna y structure: a versatile, multidimensional single molecule assay,* Nano letters **14**, 6475 (2014).

[34] B. A. Berghuis, D. Dulin, Z.-Q. Xu, T. van Laar, B. Cross, R. Janissen, S. Jergic, N. E. Dixon, M. Depken, and N. H. Dekker, *Strand separation establishes a sustained lock at the tus-ter replication fork barrier,* Nat Chem Biol **11**, 579 (2015).

[35] R. Janissen, B. A. Berghuis, D. Dulin, M. Wink, T. van Laar, and N. H. Dekker, *Invincible DNA tethers: covalent DNA anchoring for enhanced temporal and force stability in magnetic tweezers experiments,* Nucleic Acids Research **42**, e137 (2014).

[36] D. Klaue and R. Seidel, *Torsional Stiffness of Single Superparamagnetic Microspheres in an External Magnetic Field,* Physical Review Letters **102**, 028302 (2009).

[37] A. J. W. te Velthuis, J. W. J. Kerssemakers, J. Lipfert, and N. H. Dekker, *Quantitative Guidelines for Force Calibration through Spectral Analysis of Magnetic Tweezers Data,* Biophysical Journal **99**, 1292 (2010).

[38] Z. Yu, D. Dulin, J. Cnossen, M. Köber, M. M. van Oene, O. Ordu, B. A. Berghuis, T. Hensgens, J. Lipfert, and N. H. Dekker, *A force calibration standard for magnetic tweezers,* Review of Scientific Instruments **85**, 123114 (2014).

[39] D. Bastia, S. Zzaman, G. Krings, M. Saxena, X. Peng, and M. M. Greenberg, *Replication termination mechanism as revealed by Tus-mediated polar arrest of a sliding helicase,* Proceedings of the National Academy of Sciences **105**, 12831 (2008).

[40] D. Li, R. Zhao, W. Lilyestrom, D. Gai, R. Zhang, J. A. DeCaprio, E. Fanning, A. Jochimiak, G. Szakonyi, and X. S. Chen, *Structure of the replicative helicase of the oncoprotein sv40 large tumour antigen,* Nature **423**, 512 (2003).

[41] E. Fanning and K. Zhao, *Sv40 dna replication: from the a gene to a nanomachine,* Virology **384**, 352 (2009).

[42] R. Wessel, J. Schweizer, and H. Stahl, *Simian virus 40 t-antigen dna helicase is a hexamer which forms a binary complex during bidirectional unwinding from the viral origin of dna replication.* Journal of Virology **66**, 804 (1992).

[43] H. Yardimci, X. Wang, A. B. Loveland, I. Tappin, D. Z. Rudner, J. Hurwitz, A. M. van Oijen, and J. C. Walter, *Bypass of a protein barrier by a replicative dna helicase,* Nature **492**, 205 (2012).

[44] T. C. Boles, J. H. White, and N. R. Cozzarelli, *Structure of plectonemically supercoiled dna,* Journal of molecular biology **213**, 931 (1990).

[45] R. Vlijm, A. Mashaghi, S. Bernard, M. Modesti, and C. Dekker, *Experimental phase diagram of negatively supercoiled dna measured by magnetic tweezers and fluorescence,* Nanoscale **7**, 3205 (2015).

[46] G. Goetz, F. Dean, J. Hurwitz, and S. Matson, *The unwinding of duplex regions in dna by the simian virus 40 large tumor antigen-associated dna helicase activity.* Journal of Biological Chemistry **263**, 383 (1988).

[47] Y. Murakami and J. Hurwitz, *Functional interactions between sv40 t antigen and other replication proteins at the replication fork.* Journal of Biological Chemistry **268**, 11008 (1993).

5

THE EFFECT OF TEMPERATURE ON THE REVERSIBLE UNFOLDING OF DNA HAIRPINS

Yesterday's sensation is today's calibration,
and tomorrow's background.

Richard Feynman

Single-molecule experiments like magnetic tweezers assays are aiming to tackle increasingly more complex biological systems, and doing so requires more stringent control of parameters. Temperature is one of these parameters, as for instance the multi-enzyme complexes that carry out *E. coli* DNA replication are known to form only above 30°C. To this end, we have built an easy-to-implement, home-built, Arduino-based temperature control system for magnetic tweezers setups, which allows for precise temperature control as well as the ability to perform multiplexed measurements. As a demonstration, we extract the temperature-dependent energetics of reversible unfolding of a short DNA hairpin, such as the force and temperature dependence of the equilibrium constant for reversible unfolding. Our measurements show that the hairpin energetics are proportionally sensitive to changes in absolute temperature and suggest that life has evolved to leverage the narrow absolute temperature range at which it operates.

5.1. INTRODUCTION

FROM the onset of the first single-molecule experiments, the field of single-molecule biophysics has made gigantic leaps forward in terms of the extent of detail with which biological phenomena can be observed. In addition to this, advances in high-throughput methods have made it possible to perform hundreds of single-molecule experiments simultaneously (multiplexing) [1], thereby creating the opportunity to gain direct insight into how single-molecule events, dynamics, and kinetics build up the classically more familiar ensemble-averaged observations. Magnetic tweezers (MT) is a useful technique in the field of single-molecule, high-throughput manipulation as recent studies show [2, 3], and current state-of-the art setups are steadily heading towards the multiplexing capacity of lower-resolution single-molecule techniques such as flow stretching assays.[4] This allows for MT to be used as a tool to study complex and low initiation rate events such as DNA replication at the single-molecule level with unprecedented spatiotemporal resolution.

The increased biological complexity of in vitro single-molecule experiments, however, often demands a more stringent control over relevant parameters as the assay yield (or more generally: the fitness of the system of interest) across parameter space tends to become more peaked as biological complexity increases. This is demonstrated by the fact that e.g. individual *E. coli* replisome components such as the DnaB helicase or the DNA polymerase display activity at room temperature [5], yet the replisome as a whole is reported to not display any in vitro (single-molecule) activity below 30°C (Van Oijen assays, personal communication [4]). As temperature is an obvious and relatively easy-to-tweak parameter in MT experiments [6], we have designed an easy to adopt and control MT temperature control (TC) system. We demonstrate the workings of this TC system by inferring the temperature-dependent thermodynamic properties of reversible DNA hairpin folding.

Because (reversible) hairpin unfolding is a well-studied and well-understood phenomenon [7–9], this provides a good test system to study the effect of temperature on the equilibrium dynamics of hairpin unfolding – an aspect that to our knowledge has not been investigated thus far. Characterizing thermally induced fluctuations during single-molecule thermodynamic transformations of small systems (i.e. single molecules or small protein complexes) with MT is readily possible.[10] As others have readily shown, a short nucleic acid hairpin that behaves as a simple, two-state system in a single-molecule force spectroscopy assay provides an excellent sample to directly measure physical parameters such as the equilibrium free-energy difference, the enthalpy difference, and the entropy difference.[11–14] Free-energy landscape formalisms are the standard conceptual framework for describing (DNA hairpin) folding, and DNA hairpin folding in turn provides an excellent model system for investigating energy landscapes and understanding the effects of structure on folding.[15, 16] Reversible unfolding with a simple, single transition energy barrier at a constant load requires a short DNA hairpin that folds as a two-state system.[9]

Here we report on the effect of temperature on the reversible unfolding of a 20-bp hairpin

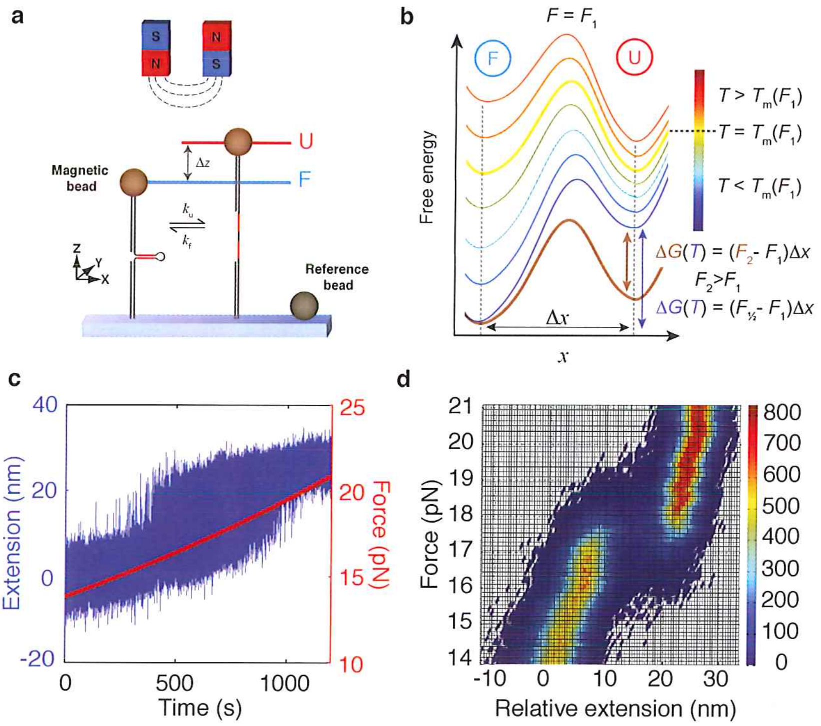


Figure 5.1: **The specifics and readout of reversible DNA hairpin unfolding in a temperature controlled magnetic tweezers assay. a)** The DNA hairpin (20 bp stem, 75% GC and a 4 base loop) is flanked by dsDNA handles of 1 kb each. At specific force-temperature combinations, we observe rapid changes in the extension (Δz), corresponding to the opening and closing of the hairpin from the folded (F) to the unfolded (U) state and back, respectively. b) The effect of temperature and force on the hairpin free-energy landscape along the reaction coordinate Δx. At a constant force ($F = F_1$), an increase in temperature can shift the equilibrium from mainly folded (F) to unfolded (U), thereby passing the (force-dependent) melting temperature ($T_m(F_1)$) (From low to high T: blue to red). An increase in force ($F = F_2$) tilts the free-energy landscape by lowering the free-energy difference ΔG between the folded and the unfolded state (shown in brown, for the lowest temperature (blue)). The free energy at zero force for a given temperature $\Delta G_0(T)$ is obtained by multiplying the force at which the states are equally populated $F_{1/2}(T)$ with the distance Δx between the states (Here $F_1 = F_{1/2}$ for $T = T_m(F_1)$). **c)** The extension of the 20 bp DNA hairpin over time shows a transition from the folded to the unfolded extension as the force is incrementally increased following a 20 s measurement at constant force, covering a range of ~7 pN in total (600 Hz acquisition rate). **d)** The force-extension relationship obtained by the measurement shown in (**c**). Each row represents a histogram of the extension at constant force for 20 s (color code shows the data points per bin).

under a constant load. By examining the population of closed versus open hairpin states as a function of force and temperature, we extract the probability of opening $P_U(F,T)$ and the equilibrium constant $K(F,T)$. The probability of the unfolded state as a function of force fits well to the Boltzmann relation $P_U(F) = (1+\exp[(F_{1/2}-F)\cdot\Delta x/(k_B T)])^{-1}$, where $F_{1/2}$ is the force at which the closed and open hairpin states are equally populated, Δx is the distance between the closed and the open state, and k_B is the Boltzmann constant. This allows us to infer parameters such the unfolding force $F_{1/2}(T)$, the difference in free energy between closed and open states at zero force $\Delta G_0(T)$, as well as the reaction enthalpy ΔH and entropy ΔS. Our room temperature observations are in good agreement with observations by others for similar hairpin structures [9], as well as the values predicted using mfold. The unfolding force and free-energy decreases linearly with temperature and a temperature jump of 10°C decreases the unfolding force by ~2 pN, and the free energy difference by ~ 15 kJ.mol^{-1}, all other measurement conditions being equal. This implies that a ~3.7% change in absolute temperature leads to a 10% change in the unfolding force and a 11% change in the reaction free-energy.

This observation, and given the importance of temperature in determining the fate of myriads of (bio)chemical reactions, suggests that performing single-molecule biophysics experiments at temperatures relevant to the system of interest is of paramount importance — a parameter best no longer ignored now that TC implementation has become readily feasible. Past studies have shown for instance that a temperature difference of only a few degrees Celsius has a dramatic effect on force-induced phase transitions of DNA.[17, 18] We therefore envisage that temperature control will become standard in the context of (multiplexed) MT assays, opening up many possible experimental possibilities of characterizing increasingly complex DNA-enzyme interactions. The proportionally large percentage change in reaction energetics demonstrate as a response to temperature change indicates that the molecules that form the basis of life have evolved to employ temperature as a leveraged product, given the narrow temperature bandwidth that is known to sustain life as we know it.

5.2. RESULTS

Here we measure the extension of a short DNA hairpin with a 20 base pair (bp) 75% GC hairpin stem and 4 base loop that reversibly unfolds at a constant applied force in our magnetic tweezers assay (**Fig. 5.1a**). The hairpin resides in either a fully closed, or a fully opened state, indicating that a single transition energy barrier lies between the two states, indicative of a simple, two-state system. A typical and classical record of extension under increasing load and constant temperature shows that the extension of this molecule is stable well below the transition force (< 13 pN) (**Fig. 5.1c**). As the transition force is approached, jumps in extension between the lower, folded hairpin extension and the extension of the unfolded hairpin become increasingly more frequent, eventually at the transition or unfolding force ($F_{1/2}$) both extensions are equally probable (**Fig. 5.1d**, ~ 17.0 pN at 24°C). Increasing the force further leads to a decrease in the number of transitions to the lower extension until the unfolded state remains stable at forces well beyond the transition force (**Fig. 5.2a**).

Whether two complementary strands of DNA adopt or maintain their helical form is strongly dependent on temperature, and is specified by the free-energy difference $\Delta G_0(T)$ between the paired and unpaired states. The stability of the double helix is reduced as the temperature increases, until at a certain point the melting temperature (T_m) is reached – the temperature at which 50% of the DNA helices are unfolded. This implies that temperature, like force, is a reaction coordinate that helps to determine the free-energy difference between the folded and unfolded state: thus, a similar state transi-

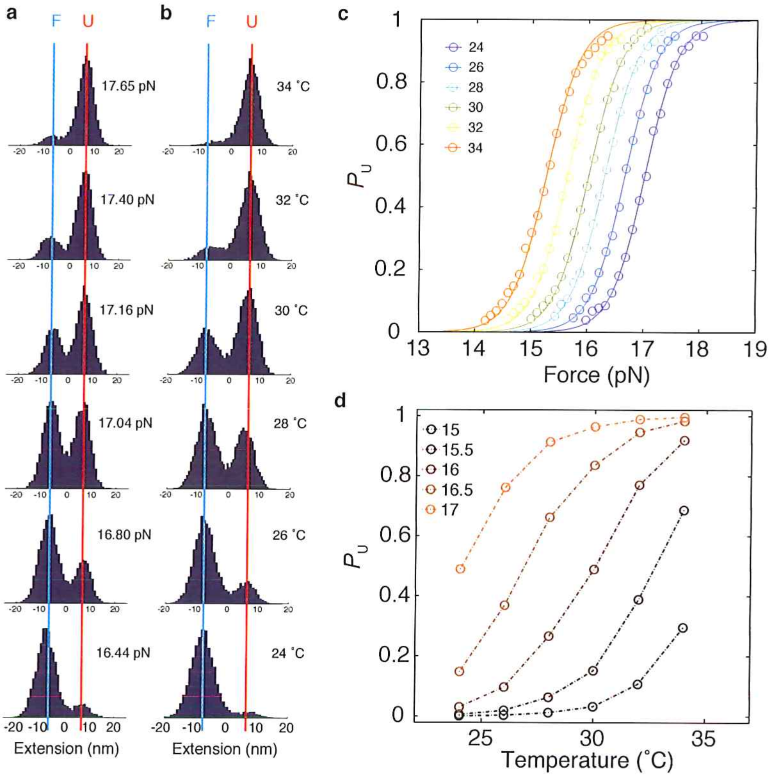


Figure 5.2: **Bi-stable nature of the DNA hairpin is force and temperature dependent.** **a)** A histogram of the hairpin extension at a constant temperature (24°C, rows in **Fig. 1d**) reveals two distinct populations (folded F, and unfolded U) co-exist at forces around the unfolding force ($F_{1/2}$). **b)** Histograms of the hairpin extension at a constant load (16.21 pN) show a similar force-induced melting transition at the measured temperature range. **c)** Fitting each population observed in (**a**) with a single Gaussian, we obtain the probability of the unfolded state (P_U) versus force for the measured temperature range (from blue to orange: 24, 26, 28, 30, 32, 34 °C resp. Circles are data). Solid lines show that P_U as a function of force fits well to the Boltzmann relation (i.e. the probability of a two-state system) $P_\mathrm{U}(F) = (1+\exp[(F_{1/2}-F)\cdot\Delta x/(k_B T)])^{-1}$. **d)** The probability of the unfolded state (P_U) versus temperature visualizes the force-induced melting transition for different forces (from dark to light: 15, 15.5, 16, 16.5, and 17 pN resp.).

tion pattern should be observable with temperature. To this end, we recorded force-extension curves as described above at 24, 26, 28, 30, 32, and 34°C using the same DNA hairpin in all experiments. At a force close to the transition force at low temperature, a shift from a mainly folded hairpin to a mainly opened hairpin can be observed as the temperature is increased over the measured range (**Fig. 5.2b**). The unfolding force decreases from 17.0 pN at 24°C to 15.3 pN at 34°C. Fitting the obtained histograms for the entire force and temperature range with either one or two Gaussians yields the probability of the unfolded state (P_u) versus force for each temperature (**Fig. 5.2c**). This probability fits well to the Boltzmann relation $P_\mathrm{U}(F) = (1+\exp[(F_{1/2}-F)\cdot\Delta x/(k_B T)])^{-1}$, where Δx is the distance between the two states along the reaction coordinate, which in this case is the difference in extension between closed and open hairpin states. A similar sigmoidal transition characteristic for a two-state system is observed at constant force over the temperature range (**Fig. 5.2d**). A global fit yields $\Delta x = 13.16$ nm, which is substantially lower than we would expect, as others find $\sim$ 18 nm for similarly sized hairpins, and a $F_{1/2}$ that decreases linearly with temperature (**Fig. 5.3a**). This linear relationship between the unfolding force and temperature results from the contribution of entropy to the free energy of double helix formation. This entropic contribution destabilizes the double helix as the temperature increases. The change in entropy upon hairpin unfolding can be extracted from the force-temperature unfolding relationship via the Clausius-Clapeyron relationship [19] $\partial F_{1/2}/\partial T = -\Delta S/\Delta x$, yielding $\Delta S = 1.38$ kJ.mol^{-1}.K^{-1}, significantly lower than the value predicted using mfold for this hairpin (2.01 kJ.mol^{-1}.K^{-1}).

From our dataset, we also determine the reaction equilibrium constant $\ln K(F) = 1/P_\mathrm{u} - 1 = -\Delta G/RT$, which increases exponentially with force for all measured temperatures (**Fig. 5.3b**). An expression analogous to the van‘t Hoff equation holds $\partial \ln K(F)/\partial F = \Delta x(F)/(k_B T)$, which implies that the slopes in **Fig. 5.3b**, when multiplied with $k_B T$, yield Δx at the transition force, which at 12.95 nm is in good agreement with the value obtained above. A global fit of $\ln K = (F(T)\cdot\Delta x)/(RT) = -\Delta G(T)/(RT)$ yields the free-energy difference at zero force $\Delta G_0(T)$ at every measured temperature (**Fig. 5.3c**). $\Delta G_0(T)$ decreases linearly over the measured temperature range, starting with a value of $\sim$ 142 kJ.mol^{-1} at 24°C, which is in good agreement with observations by others [9], though higher than predicted by mfold ($\sim$ 128 kJ.mol^{-1} at 24°C). At 34°C the free energy is reduced by 14 kJ.mol^{-1} to $\sim$119 kJ.mol^{-1}. Extrapolation of this linear trend implies that a spontaneous (zero force) transition to the unfolded state ($\Delta G \leqslant 0$) will start occurring at $\sim$ 83.8°C, which is in reasonable agreement with the value predicted by mfold (87.5°C). The entropy increase ΔS due to the increase in temperature can then be obtained from the slope of ΔG_0 as a function of temperature, as—$\partial \Delta G_0/\partial T = \Delta S_0 - \Delta S_\mathrm{stretch} \approx \Delta S = 2.38$ kJ.mol^{-1}.K^{-1}. To compare the entropy change with that of bulk calorimetric measurements (which measures ΔS_0), the entropy change should in principle be corrected to take into account the force-dependent elastic term $\Delta S_\mathrm{stretch}$.[15] However, as the stretching contribution of SSDNA varies slowly with temperature and the ratio of $k_B T$ and the linking number Lk is roughly constant, this term can be neglected.[20]

From the fit of the equilibrium constant versus force, we can obtain values for the equilibrium constant as a function of temperature, which also shows to be exponential, as expected (**Fig. 5.3d**). The positive linear slope with respect to temperature represents the shift of the reaction equilibrium to the unfolded state as the temperature increases. The spectrum of the equilibrium constant in the force temperature parameter space (**Fig. 5.3f**) summarizes the separate observations of **Figs. 5.3b** and **5.3d**.

The reaction enthalpy ΔH can be obtained through $\Delta G = \Delta H - T\Delta S$, as well as from the expression $\partial(F_{1/2}.\Delta x/T)/\partial(1/T) = \Delta H - \Delta H_\mathrm{stretch}$, i.e. the slope of $\Delta G/T$ versus $1/T$ is the enthalpy (**Fig. 5.3e**), which is 544 kJ.mol^{-1}, whereas mfold predicts an enthalpy of 727 kJ.mol^{-1} for this hairpin.

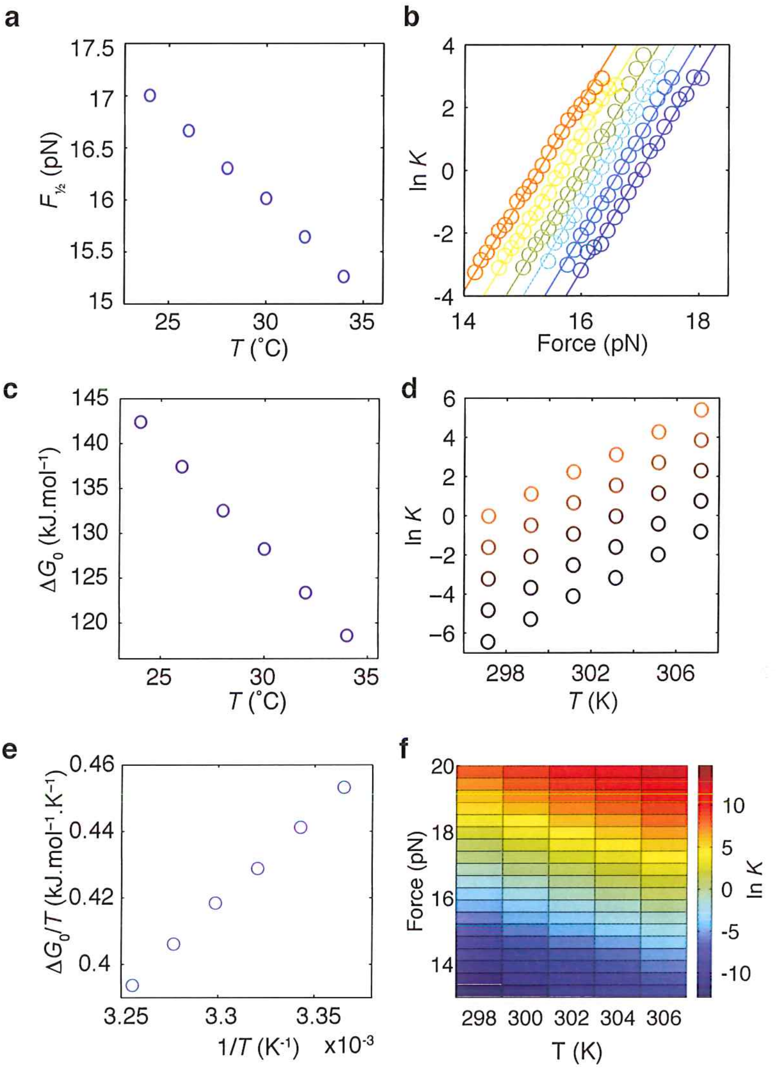


Figure 5.3: **The force and temperature dependence of the DNA hairpin free-energy and equilibrium constant. a)** The unfolding force, the force at which states F and U are equally populated ($F_{1/2}$) decreases linearly over the measured temperature range. **b)** The equilibrium constant ($\ln K(F) = 1/P_U - 1 = -\Delta G/(RT)$) decreases exponentially with force at a given temperature (from blue to orange: 24, 26, 28, 30, 32, 34 °C resp. Circles are data). Solid lines are the result of a global fit of $\ln K(F, T)$. The slope $\partial \ln K(F)/\partial F = \Delta x(F)/(k_B T)$ is mainly influenced by the distance Δx between the states (see **Figs. 5.9** and **5.10**). **c)** The free energy at zero force ΔG_0 decreases linearly over the measured temperature range. **d)** From the global fit in (**b**) we extract values for $\ln K(F)$ versus temperature (same forces as **Fig. 2d**, from dark to light: 15, 15.5, 16, 16.5, and 17 pN resp.). **e)** The slope of $\Delta G_0/T$ as a function of $1/T$ yields the hairpin enthalpy. **f)** The landscape of the (logarithm of the) equilibrium constant versus force and temperature ($\ln K(F, T)$). A positive $\ln K$ implies the ratio of unfolded versus folded hairpin state occupancy is larger than 1, $\ln K = 0$ implies equal state occupancy. Note that only $\ln K$ values in the range $\sim \{-5, 5\}$ can be obtained experimentally.

5.3. DISCUSSION

Our measurements show that temperature has a substantial effect on the force at which the hairpin unfolds, and thus shows that temperature has a similarly substantial effect on the free energy difference between the closed and the open hairpin states. This has all been inferred from applying Boltzmann statistics on the time-averaged occupation of the states at a given force and temperature. The clear exponential dependence of the equilibrium constant shows that the time-averaged state occupancy can be resolved sufficiently at the rate at which we acquire our datasets (600 Hz). The linear trends per degree Kelvin observed in the unfolding force and free energy difference at zero force ΔG_0 over this temperature range indicate that the temperature steps relative to each other implemented by the control system are correctly calibrated (see **Supplementary Information** for calibration details).

The distance Δx between the closed and open states however, is significantly smaller (13-14 nm) than that expected for a hairpin of this size ($\sim$ 18 nm) over this force range, as $\Delta x \sim 1$ nm/bp around $F_{1/2} = 16$ pN. Quite likely, this is due to an unexpected feature in the DNA construct assembly that led to a shorter than expected DNA hairpin stem. While this is currently being verified by subjecting a different DNA hairpin to the same tests described above, the trends in the energetics extracted are expected to be quite similar, though a change in Δx will change the slope of $\ln K$ versus force, thus change the slope of ΔG_0 versus force (see **Figs. 5.9 and 5.10** for more information). Nevertheless, we can still quantify the effect that a relatively small change in absolute temperature has on both the unfolding force and the free energy difference between the folded and unfolded hairpin states. If we assume that the unfolding force is not affected by the hairpin length -- or that temperature affects the unfolding force in an identical manner, we see that a 3.7% increase in absolute temperature (297.15 K to 307.15 K) leads to a ~10% decrease in the unfolding force (17.0 pN to 15.3 pN) as well as a 10% decrease in the reaction free energy ΔG_0 (134 to 119 kJ/mol). This means a change in temperature leads to an amplified change in other energetic parameters.

In summary, we have designed an easy to implement and control temperature control system for magnetic tweezers setups, and tested the system by measuring the welknown properties of force-induced reversible unfolding of a short DNA hairpin, yet now with temperature as an additional parameter. This allows us to obtain the temperature-dependence of quantities like the free-energy difference between the folded and the unfolded state, the enthalpy and the entropy of a single DNA hairpin. These experiments show that small changes in absolute temperature lead to proportionally large effects on the mechanochemistry and thermodynamics of this biological information carrier. As the reaction kinetics of not only DNA, but also arguably all biological enzymes and proteins have evolved to be sensitive to small changes in absolute temperature, we argue that temperature control will become a standard and necessary addition to single-molecule magnetic tweezers assays in the near future.

Primer no.	Oligonucleotide sequence
1	5'-CCATCTTGGTCTCCTAGGGGTTTAACAAGAGCAATTAGGGCC-3'
2	5'-CCATCTTGGTCTCCGAGCCATGCCGGTTGCCGCTGTTACC-3'
3	5'-CCATCTTGGTCTCCGCCTACCGTGACGCCTGCCAGCACCTC-3'
4	5'-CCATCTTGGTCTCCTCAACCGGGGAAGCGGGGATAACTTCA-3'
5	5'-P-GCTCGGCCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGA- GTTGCTCTGCTAGCTCGCCGAGGCGAGCGAAAGCTCGCCTCGG- CGAGCTAGCTGCCATGCTCTTTACAACCGGTTGACTGCTTCAGGG- GTCGATCCCGCTTT-3'
6	5'-AGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGGCC-3'
7	5'-P-AGGCAAAGCGGGATCGACCCCTGAAGCAGTCAACCGGTTGTAAAGAGCATGGCA-3'
8	5'-GACCGAGATAGGGTTGAGTG-3'
9	5'-TACGCATGGTCGTGCGTTTTAGCTGCGGAACAGG- GCGAGAAACGCCGCTCTGTGTAGGCTGGAGCTGCTTC3'
10	5'-CCATCTTGGTCTCCTTGACGCTCTAGAACTAGTGGATCCCCC-3'

Table 5.1: Used oligonucleotide sequences

5.4. METHODS

5.4.1. DNA HAIRPIN CONSTRUCT DESIGN AND ASSEMBLY

The hairpin DNA was cloned in a multi-step process. First, two fragments of 1 kb were amplified by PCR with primers 1-2 and 3-4 respectively (see **Table 5.1** for oligo sequences) using plasmid pbluescr1,2,4+pSFvI (kindly provided by Susanne Hager) as template. These fragments were digested with *BsaI.* To create the hairpin, 5'-phosphorylated oligo 5 was annealed to oligo 6 and to 5'-phosphorylated oligo 7, resulting in double-stranded DNA with a 20 bp hairpin with 4 bases of single stranded DNA and with a 4 bp 5'-overhang on both sites, complementary to the 5'-overhang of the *BsaI*-digested 1-kb fragments. This hairpin-containing fragment was ligated to the *BsaI*-digested 1-kb fragments. To create handles for the hairpin, a 643 bp fragment from pBluescript SK+ (Stratagene/ Agilent Technologies, Inc, Santa Clara, CA) was amplified by PCR in the presence of either Biotin-16-dUTP (Roche Diagnostics, Basel, Switserland) or Digoxigenin-11-dUTP (Roche Diagnostics, Basel, Switserland) using primers 8 and 9 or 8 and 10 respectively. Handles were digested with *BsaI* and ligated to the hairpin-containing 2-kb fragment from the first ligation step. All oligonucleotides were obtained from from Ella Biotech GmbH, Martinsried, Germany.

5.4.2. TEMPERATURE CONTROL SYSTEM HARDWARE

The temperature control system built here is based on a design implemented by Gollnick *et al.* [6], yet has a MT software integrated PID control loop, physically implemented by an Arduino microprocessor. For a list of parts see **Table 5.2**.

5.4.3. TEMPERATURE CONTROL LOOP

For elaborate information regarding the temperature control system we refer to the Supplementary Information. In short, precise control of the flowcell (sample) temperature is obtained through direct heating and monitoring of the baseplate that holds the flowcell, as well as the objective below the flowcell. A PID feedback loop (Supplementary Fig. 5.5a) ensures a stable temperature ($\sigma_{\max}(T < 40°\mathrm{C}) = 20$ mK) is reached within ~ 10 minutes through using the difference $e(t)$ between the current temperature and the set temperature value to calculate the voltage output $u(t)$ for a power supply unit (PSU):

$$u(t) = K_p\Big(e(t) + \frac{1}{T_i}\int_0^t e(\tau d\tau + T_d\frac{d}{dt}e(t)\Big) \tag{5.1}$$

5.4.4. MAGNETIC TWEEZERS INSTRUMENTATION

The magnetic tweezers implementation used in this study has been previously described [2, 3, 21]. Briefly, light transmitted through the sample was collected by an oil-immersion objective (Olympus UPLSAPO60XO 60×, numerical aperture (NA) = 1.35, Olympus, USA) and projected onto a 12-megapixel CMOS camera (Falcon FA-80-12M1H, Teledyne Dalsa, Canada) with a sampling frequency of 58 Hz at full field of view, or higher when cropped. A 2-inch 200-mm tube lens between objective and camera resulted in an effective magnification of 67×. As a result of the aforementioned camera and magnification specifications, the field of view (fov) size is approximately 300×400 μm, allowing for the possibility of tracking hundreds of beads simultaneously. The applied magnetic field was generated by a pair of vertically aligned permanent neodymium-iron-boron magnets (SuperMagnete, Switzerland) separated by a distance of 1.0 mm and suspended on a motorized stage (M-126.PD2, Physik Instrumente, Germany) above the flow cell. Additionally, the magnet pair could be rotated about the illumination axis by an applied DC servo step motor (C-150.PD, Physik Instrumente, Germany).

5.4.5. DATA ACQUISITION

Image processing of the collected bead diffraction patterns was used to track the real-time position of both surface-attached reference beads and superparamagnetic beads coupled to DNA tethers in three dimensions. We implemented custom written software in C++, CUDA and LabView (2011, National Instruments Corporation, USA) that is suited for high-throughput tracking in magnetic tweezers.[1] Tracking of the x, y coordinates is performed using center-of-mass computation followed by a further refinement using the quadrant interpolation algorithm. Localization of the bead's z-coordinate is achieved by creating a radial profile using the refined x, y coordinates and comparing this profile to a prerecorded look-up table of radial profiles. After subtraction of the reference bead position to correct for instrumental drift, the x, y and z positions of the DNA-tethered beads were determined with a spatial accuracy of < 3 nm. The upward stretching forces on the DNA tethers by the superparamagnetic beads were calibrated from analysis of the extent of their Brownian motion, whereby spectral corrections were employed to correct for camera blur and aliasing.[22, 23]

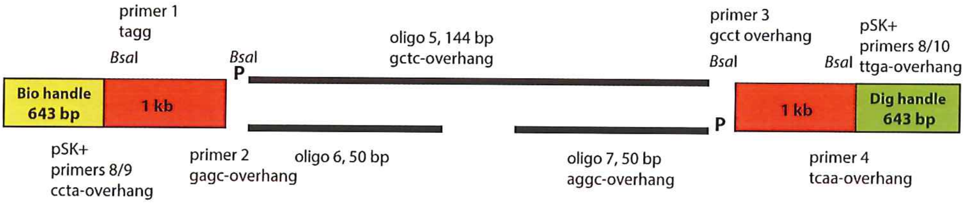


Figure 5.4: **DNA hairpin design and assembly.** The hairpin DNA was cloned in a multi-step process. First, two fragments of 1 kb were amplified by PCR with primers 1-2 and 3-4 respectively (see **Table 5.1** for oligo sequences) using plasmid pbluescr1,2,4+pSFvI (kindly provided by Susanne Hage) as template. These fragments were digested with *BsaI*. To create the hairpin, 5'-phosphorylated oligo 5 was annealed to oligo 6 and to 5'-phosphorylated oligo 7, resulting in double-stranded DNA with a 20 bp hairpin with 4 bases of single stranded DNA and with a 4 bp 5'-overhang on both sites, complementary to the 5'-overhang of the *BsaI*-digested 1-kb fragments. This hairpin-containing fragment was ligated to the BsaI-digested 1-kb fragments. To create handles for the hairpin, a 643 bp fragment from pBluescript SK+ (Stratagene/ Agilent Technologies, Inc, Santa Clara, CA) was amplified by PCR in the presence of either Biotin-16-dUTP (Roche Diagnostics, Basel, Switserland) or Digoxigenin-11-dUTP (Roche Diagnostics, Basel, Switserland) using primers 8 and 9 or 8 and 10 respectively. Handles were digested with *BsaI* and ligated to the hairpin-containing 2-kb fragment from the first ligation step. All oligonucleotides were obtained from from Ella Biotech GmbH, Martinsried, Germany.

Component	Description
Pt-100 RTD	Temperature sensors
PlayingWithFusion Dual MAX31865 PT-100 RTD to Digital Breakout board	Integrated circuit (IC) for digital temperature readout
SparkFun Inventor's Kit V3.1	Arduino for temperature readout
Computer with LabView	Software implementation
TTi CPX200DP	LabView controllable power supply
Minco HR5160R44.0L12E	Baseplate heaters
Minco HK5210E16.3L12A	Objective heater
Minco stretch tape (11m)	Heat resistant tape
Minco epoxy	Heat resistant glue

Table 5.2: List of the hardware used in the implementation of the temperature control system.

Sensor	T_{melt} (°C)	T_{boil} (°C)
1	0.25 ± 0.16	100.27 ± 0.08
2	0.23 ± 0.17	100.13 ± 0.08
3	0.38 ± 0.13	100.34 ± 0.07

Table 5.3: Average measured temperature and standard deviations for all three sensors at the melting and boiling points of milliQ water.

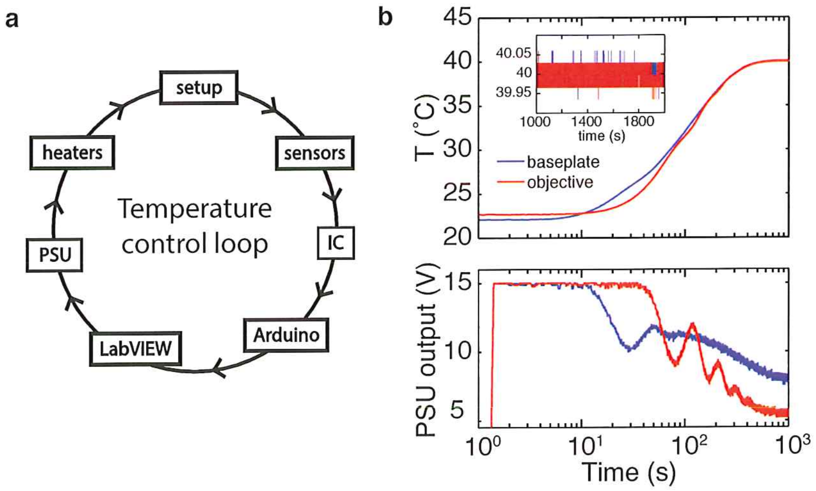


Figure 5.5: **The MT temperature control loop. a)** Using a home-built, Arduino-based temperature control system, we can monitor the experimental temperature and set it to a fixed value by using a PID feedback loop. The feedback loop controls the amount of heat added to the system via resistive heaters attached to the microscope objective as well as to the base plate that holds the flow cell. b) A stable objective and base plate temperature is reached within 10 mins, with typical standard deviations over ~ 10 min not surpassing 20 mK (see **Fig. 5.6b**). As there will be a lag between the time at which the objective and base plate stabilize, and the flow cell interior, we typically allow temperatures to stabilize for >30 min. Buffers are pre-heated before flushing in.

5.5. SUPPLEMENTARY INFORMATION

Table 5.4 shows the analysed errors in different components of the temperature measurements, resulting in a total theoretical maximum error of ±0.73°C, assuming the errors are additive. Our calibration showed a smaller uncertainty than this upper limit however. From the offsets between the measured and expected values of up to 0.4 °C (Table 5.3) and fluctuations between the sensors of ±0.1°C, we consider it fair to use ±0.5°C as absolute temperature readout error of the system.

Source of error	Maximum error (°C)
Pt-100 class A	±0.23
MAX31865	±0.5
Sensor wiring	0
Conversion	0
Circuitry	Negligible if no read errors

Table 5.4: Table of considered errors.

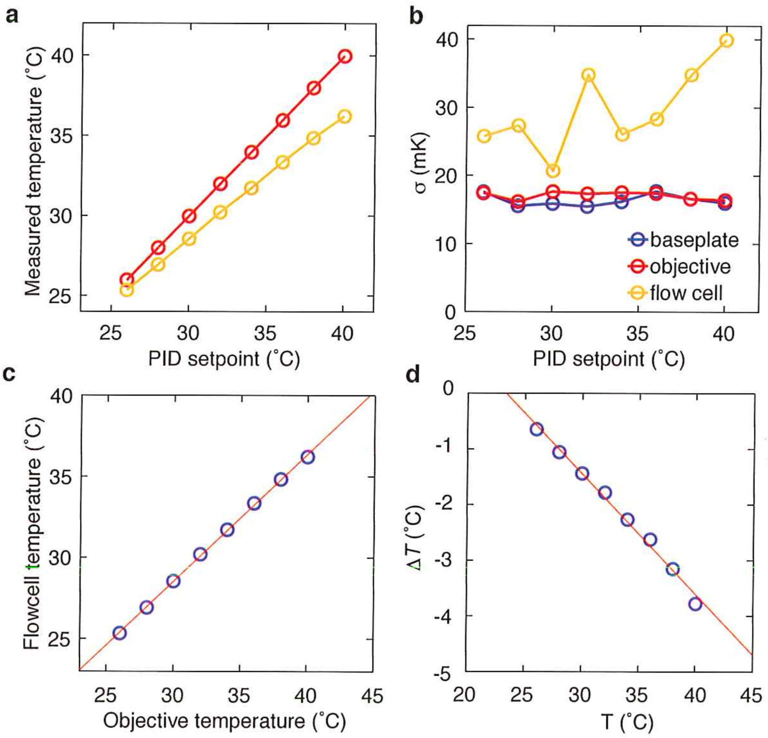


Figure 5.6: **Flow cell temperature calibration. a)** A series of temperature jumps allows us to measure the drop in temperature between setup (red, blue) and flow cell (orange). b) Standard deviations (σ, 10 min measurements) of the objective and flow cell remain below 20 mK over the measured temperature range, the flow cell σ is higher and shows an increase over the measured temperature range, with $\sigma_{\max} = 40$ mK at 40°C. **c)** The slope of objective versus flow cell temperature is linear (data in blue, linear fit in red; $y = 0.7834x + 5.0621$). d) The temperature difference ΔT between the PID setpoint and the flow cell ($y = -0.2166x + 5.0621$). The temperature differences between objective and flow cell agree well with observations by others.[6]

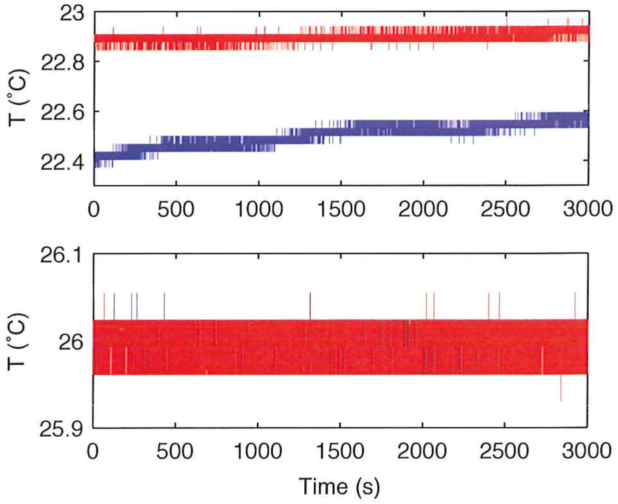


Figure 5.7: **The temperature stability of the MT temperature control system.** The objective (red) and base plate (blue) at room temperature (RT, top), and with the TC PID loop actively maintaining a temperature of 26°C (bottom). Over the 50 min measurement, the standard deviation (σ) of the base plate (bp) at RT, $\sigma_{\text{bp}}(RT) = 44.5$ mK, the objective $\sigma_{\text{obj}}(RT) = 18.9$ mK. At 26°C, $\sigma_{\text{bp}}(26) = 14.5$ mK and $\sigma_{\text{obj}}(26) = 14.8$ mK.

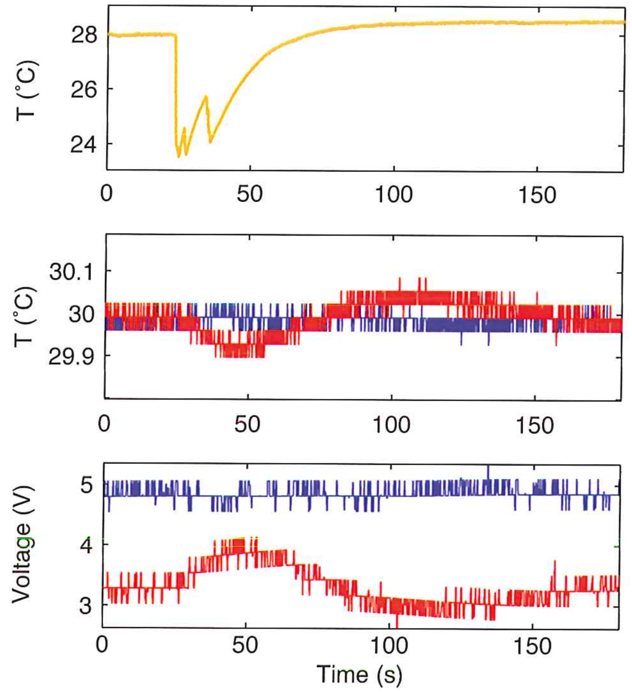


Figure 5.8: **System response to changes in the flow cell temperature.** Flushing room temperature buffer into the flow cell leads to a sudden drop in the flow cell temperature (top). This causes a drop in the temperature of the objective with a small time delay (middle, red), as the objective is in direct contact with the flow cell channel through the objective immersion oil. The PID control loop compensates for the perturbation of the objective temperature (bottom, red). The base plate temperature is not perturbed by the temperature of the buffer (middle and bottom, blue), though a stable base plate temperature reduces the drift significantly (data not shown).

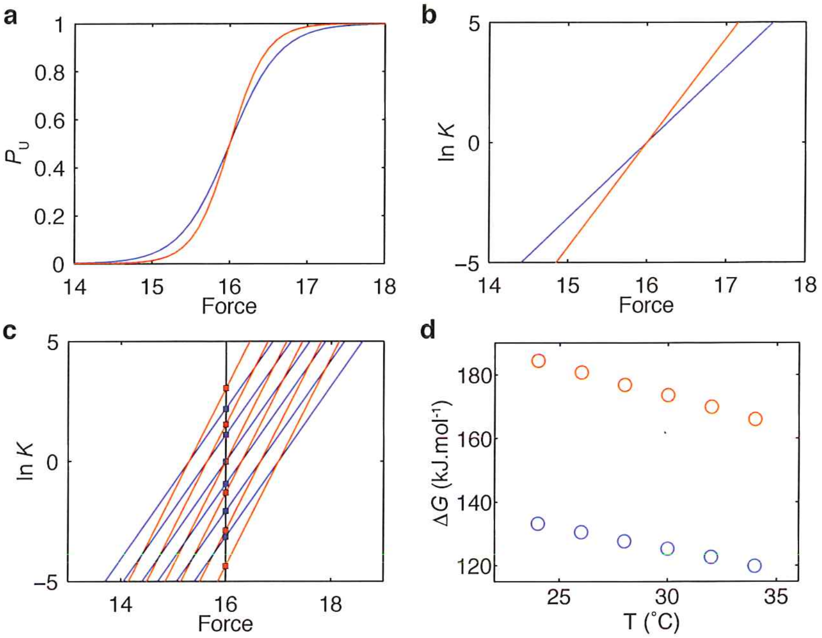


Figure 5.9: **The effect of hairpin length ($\Delta x_1 = 13$ nm (blue), $\Delta x_2 = 18$ nm (red)) on the kinetics and energetics of equilibrated reversible unfolding.** **a)** The force-dependent transition from the folded to the unfolded state is described by $P_{\mathrm{U}} = [1 + \exp[(F_{1/2} - F)\Delta x/(k_B T)]]^{-1}$, such that $\lim(x \to 0) P_{\mathrm{U}} = 0.5$ and $\lim(x \to \infty)\{P_{\mathrm{U}}(F < F_{1/2}) = 0, P_{\mathrm{U}} = 1 \mathrm{otherwise}\}$. The effect of a length change can be seen for the length observed in current experiments ($\Delta x = 13$ nm, blue) and the expected length of a 20 bp stem 4 base loop hairpin ($\Delta x = 18$ nm, red). b) The equilibrium constant $K = P_{\mathrm{U}}^{-1}{-}1$, depends exponentially on the applied force F, the slope of $\ln K$ increases with increasing Δx. c) The y-intercept of $\partial \ln K/\partial F$ at $F = 0$ multiplied by $-RT$ yields the free energy ΔG (y-intercept at $F = 16$ pN is shown for the unfolding forces $F_{1/2}(T)$ measured), such that: d) Both the magnitude as the slope of ΔG versus T are influenced by a change in Δx (assuming a constant GC-content of the hairpin and length-independent unfolding forces).

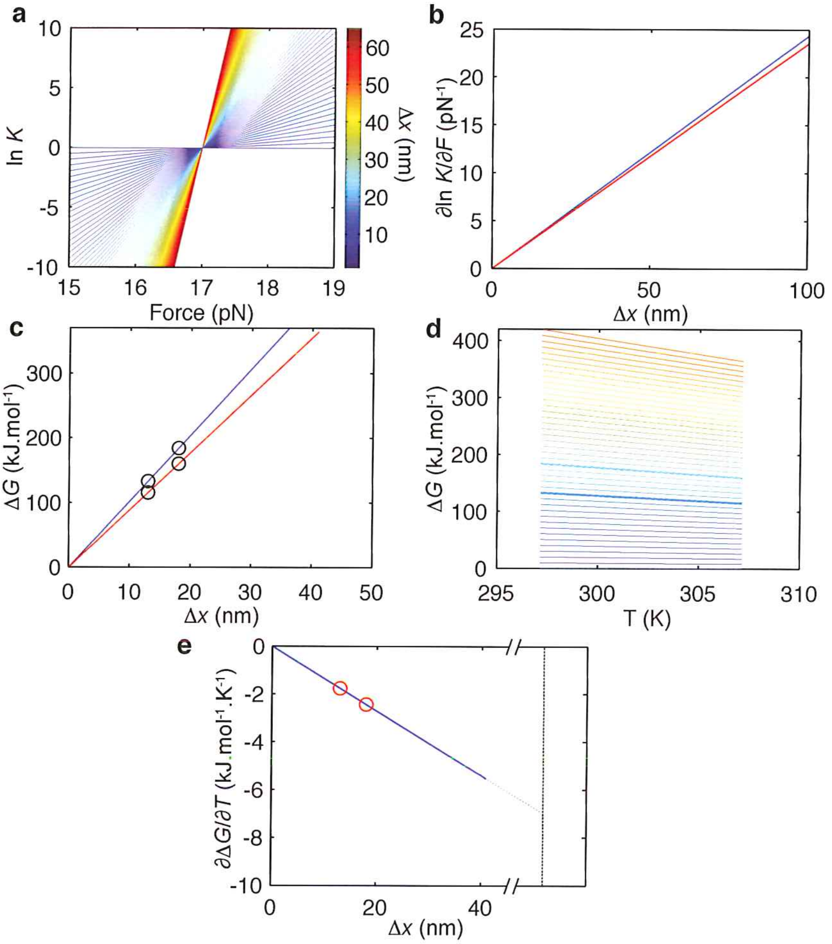


Figure 5.10: **The effect of the distance between the states Δx of an equilibrated 2-state (hairpin) system with a free energy difference at zero force of ΔG_0.** **a)** As shown in Fig. S2a,b; the slope of $\ln K$ depends on the distance Δx between the two states. **b)** The change in slope of $\ln K$ versus F increases linearly with increasing Δx, and has a (small) dependence on $F_{1/2}(T)$, as a result of a temperature difference (24°C blue, 34°C red) In our measured temperature range, we will not be able to detect this difference based on the ratio of folded versus unfolded species. What it physically means is that with increasing T the transition from one state to the other will have a slightly larger force range. **c)** This implies a linear increase of the free energy ΔG_0 versus Δx, with a slope that is set by the temperature-dependent unfolding force $F_{1/2}(T)$ (24°C blue, 34°C red). **d)** As seen in **Fig. 5.9d**, the magnitude of $\Delta G_0(T)$ increases with a slope that becomes more negative for increasing Δx (assuming Δx does not affect $F_{1/2}(T)$, and a constant GC-content). Each line represents a constant Δx (13 nm (blue) and 18 nm (cyan) highlighted). e) $\partial\Delta G_0/\partial T$ (the slope of iso-Δx lines in (**d**)) decreases linearly with Δx, this holds true as long as the system can be considered an equilibrium system (thus has an equilibrium constant that is nonzero or not infinite) such that Boltzmann statistics apply (dashed vertical line shows boundary, red circles show $\partial\Delta G_0/\partial T$ at Δx = 13 and 18 nm, resp.).

REFERENCES

[1] J. P. Cnossen, D. Dulin, and N. H. Dekker, *An optimized software framework for real-time, high-throughput tracking of spherical beads,* Review of Scientific Instruments **85**, 103712 (2014).

[2] D. Dulin, I. D. Vilfan, B. A. Berghuis, S. Hage, D. H. Bamford, M. M. Poranen, M. Depken, and N. H. Dekker, *Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase,* Cell Reports **10**, 983 (2015).

[3] B. A. Berghuis, D. Dulin, Z.-Q. Xu, T. van Laar, B. Cross, R. Janissen, S. Jergic, N. E. Dixon, M. Depken, and N. H. Dekker, *Strand separation establishes a sustained lock at the tus-ter replication fork barrier,* Nat Chem Biol **11**, 579 (2015).

[4] A. M. van Oijen and J. J. Loparo, *Single-molecule studies of the replisome,* Annual review of biophysics **39**, 429 (2010).

[5] N. Ribeck, D. L. Kaplan, I. Bruck, and O. A. Saleh, *Dnab helicase activity is modulated by dna geometry and force,* Biophysical journal **99**, 2170 (2010).

[6] B. Gollnick, C. Carrasco, F. Zuttion, N. S. Gilhooly, M. S. Dillingham, and F. Moreno-Herrero, *Probing dna helicase kinetics with temperature-controlled magnetic tweezers,* Small **11**, 1273 (2015).

[7] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, *Mechanical separation of the complementary strands of dna,* Proceedings of the National Academy of Sciences **94**, 11935 (1997).

[8] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco Jr, and C. Bustamente, *Reversible Unfolding of Single RNA Molecules by Mechanical Force,* Science **292**, 733 (2001).

[9] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, *Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins,* PNAS **103**, 6190 (2006).

[10] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, *Single-molecule studies of dna mechanics,* Current opinion in structural biology **10**, 279 (2000).

[11] M. C. Williams, J. R. Wenner, I. Rouzina, and V. A. Bloomfield, *Entropy and heat capacity of dna melting from temperature dependence of single molecule stretching,* Biophysical Journal **80**, 1932 (2001).

[12] C. Hyeon and D. Thirumalai, *Can energy landscape roughness of proteins and rna be measured by using mechanical unfolding experiments?* Proceedings of the National Academy of Sciences **100**, 10249 (2003).

[13] M. Palassini and F. Ritort, *Improving free-energy estimates from unidirectional work measurements: theory and experiment,* Physical Review Letters **107**, 060601 (2011).

[14] M. Ribezzi-Crivellari and F. Ritort, *Free-energy inference from partial work measurements in small systems,* Proceedings of the National Academy of Sciences **111**, E3386 (2014).

[15] I. Tinoco Jr and C. Bustamante, *The effect of force on thermodynamics and kinetics of single molecule reactions,* Biophysical chemistry **101**, 513 (2002).

[16] F. Ritort, C. Bustamante, and I. Tinoco, *A two-state kinetic model for the unfolding of single molecules by mechanical force,* Proceedings of the National Academy of Sciences **99**, 13544 (2002).

[17] J. van Mameren, P. Gross, G. Farge, P. Hooijman, M. Modesti, M. Falkenberg, G. J. Wuite, and E. J. Peterman, *Unraveling the structure of dna during overstretching by using multicolor, single-molecule fluorescence imaging,* Proceedings of the National Academy of Sciences **106**, 18231 (2009).

[18] H. Fu, H. Chen, J. F. Marko, and J. Yan, *Two distinct overstretched dna states,* Nucleic acids research , gkq309 (2010).

[19] S. B. Smith, Y. Cui, and C. Bustamante, *Overstretching b-dna: the elastic response of individual double-stranded and single-stranded dna molecules,* Science **271**, 795 (1996).

[20] S. de Lorenzo, M. Ribezzi-Crivellari, J. R. Arias-Gonzalez, S. B. Smith, and F. Ritort, *A temperature-jump optical trap for single-molecule manipulation,* Biophysical journal **108**, 2854 (2015).

[21] R. Janissen, B. A. Berghuis, D. Dulin, M. Wink, T. van Laar, and N. H. Dekker, *Invincible DNA tethers: covalent DNA anchoring for enhanced temporal and force stability in magnetic tweezers experiments,* Nucleic Acids Research **42**, e137 (2014).

[22] Z. Yu, D. Dulin, J. Cnossen, M. Köber, M. M. van Oene, O. Ordu, B. A. Berghuis, T. Hensgens, J. Lipfert, and N. H. Dekker, *A force calibration standard for magnetic tweezers,* Review of Scientific Instruments **85**, 123114 (2014).

[23] A. J. W. te Velthuis, J. W. J. Kerssemakers, J. Lipfert, and N. H. Dekker, *Quantitative Guidelines for Force Calibration through Spectral Analysis of Magnetic Tweezers Data,* Biophysical Journal **99**, 1292 (2010).

II

RNA POLYMERASES

6

ELONGATION-COMPETENT PAUSES GOVERN THE FIDELITY OF A VIRAL RNA-DEPENDENT RNA POLYMERASE

Knowledge is an unending adventure at the edge of uncertainty.

Jacob Boronowski

RNA viruses have specific mutation rates that balance the conflicting needs of an evolutionary response to host antiviral defenses and avoidance of the error catastrophe. While most mutations are known to originate in replication errors, difficulties of capturing the underlying dynamics have left the mechanochemical basis of viral mutagenesis unresolved. Here, we use multiplexed magnetic tweezers to investigate error incorporation by the bacteriophage $\phi 6$ RNA-dependent RNA polymerase. We extract large datasets fingerprinting real-time polymerase dynamics over four magnitudes in time, in the presence of nucleotide analogs, and under varying NTP and divalent cation concentrations and fork stability. Quantitative analysis reveals a new pause state that modulates polymerase fidelity and so ties viral polymerase pausing to the biological function of optimizing virulence. Adjusting the frequency of such pauses offers a target for therapeutics and may also reflect an evolutionary strategy for virus populations to track the gradual evolution of their hosts.

This chapter has been published as: David Dulin, Igor D. Vilfan, Bojk A. Berghuis, Susanne Hage, Dennis H. Bamford, Minna M. Poranen, Martin Depken & Nynke H. Dekker. Elongation-competent pauses govern the fidelity of a viral RNA-dependent RNA polymerase *Cell Reports* 10, 983-992 (2015). [1].

6.1. GRAPHICAL ABSTRACT

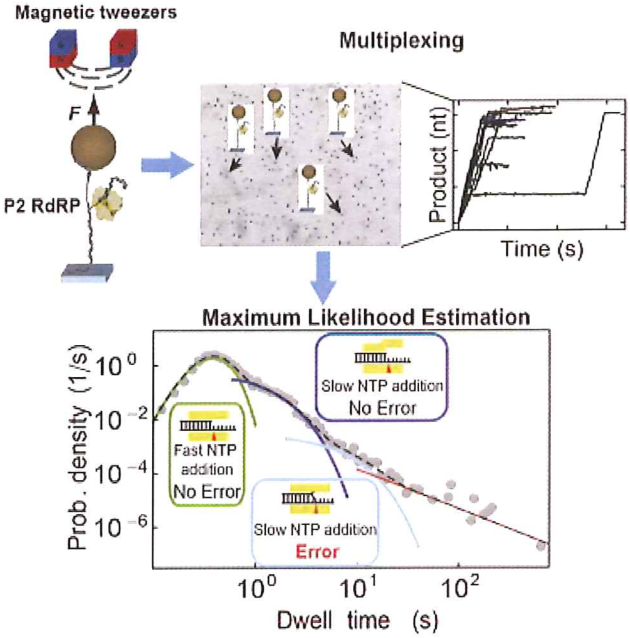


Figure 6.1

6.2. HIGHLIGHTS

- The RNAP elongation dynamics of a dsRNA virus are captured by magnetic tweezers
- Large sets of high-resolution data allow quantification of polymerase fidelity
- Maximum likelihood estimation allows extraction of critical model parameters
- Errors are predominantly incorporated via new error-prone catalytic pathway

6.3. INTRODUCTION

RNA viruses are responsible for many human pandemics, including hepatitis C, polio, influenza, and dengue fever. Because of their high mutation rates, RNA viruses evolve rapidly and are difficult to target with vaccines[2]. On the molecular level, the dominant source of mutations is the error-prone RNA-dependent RNA polymerases (RdRPs) responsible for replicating the viral genomes[3]. A high mutation rate increases evolvability, but also induces many deleterious mutations, and a delicate balance needs to be struck to ensure the pathogenicity of the viral population. Given the strict demands on the precision of mutation rates, RdRPs have become an important target for antiviral therapies that seek to either decrease[4, 5] or increase[6] RdRP replication fidelity. Understanding how RdRPs influence viral mutation rates therefore carries direct implications for human health and the development of antiviral therapies, but is also of fundamental importance for our comprehension of viral evolution. Despite this, little is known about the dynamics of RNA elongation by RdRPs[7] and in particular of the nucleotide selection process—the origin of most mutations.

Direct probing of error incorporation is challenging, as errors are infrequent random events easily masked in bulk measurements. Stop-flow and quench-flow experiments have greatly elucidated the dynamics of nucleotide addition, but such experiments often rely on nucleotide starvation conditions to induce otherwise rare error-incorporation events[7–10]; as the polymerization dynamics are severely perturbed under such conditions, it is not a priori evident that the error probabilities estimated in this way represent error rates under physiological conditions. Single-molecule experiments[11–13] do in principle have the potential to detect error incorporation in the presence of all nucleotides, but the limited throughput of these experiments has so far precluded a detailed statistical study of such rare events.

To gain a deeper insight into the origin of viral mutagenesis, we here study error incorporation of a model RdRP at the single-molecule level. Specifically, as all the structurally characterized RNA-dependent viral polymerases share a high degree of structural conservation [15, 16], we use the well characterized P2 (**Fig. 6.2a**)—the RdRP of the double-stranded RNA (dsRNA) bacteriophage ϕ6—as a model system for viral RdRPs and RTs[17, 18]. To overcome limitations induced by bulk averaging and limited statistics, we have developed a stable and massively multiplexed magnetic-tweezers assay (**Fig. 6.8b**) that retains the single-molecule resolution while producing the statistics necessary to study rare error-incorporation events. We collected $\sim$ 1000 elongation traces (RdRP position versus time over a maximum of 2,800 nucleotides [nt]) at varying nucleotide (NTP) and divalent cation concentrations in the presence of nucleotide analog and under varying fork stability (**Figs. 6.8b** and **Fig. 6.8e**). Our elongation traces are highly stochastic (**Fig. 6.2c**), with force-insensitive and rapid elongation ($\sim$ 20 nt/s) interrupted by pauses of duration 1–10 s and a small population of pauses lasting 10–1000 s. Based on a maximum-likelihood analysis tailored to our large single-molecule force-spectroscopy data sets, we uncover evidence that the ϕ6 RdRP, and by extension other RdRPs and RTs, utilizes two distinct but coupled pathways for error incorporation: one high-fidelity pathway without significant pausing and one low-fidelity pathway acces-

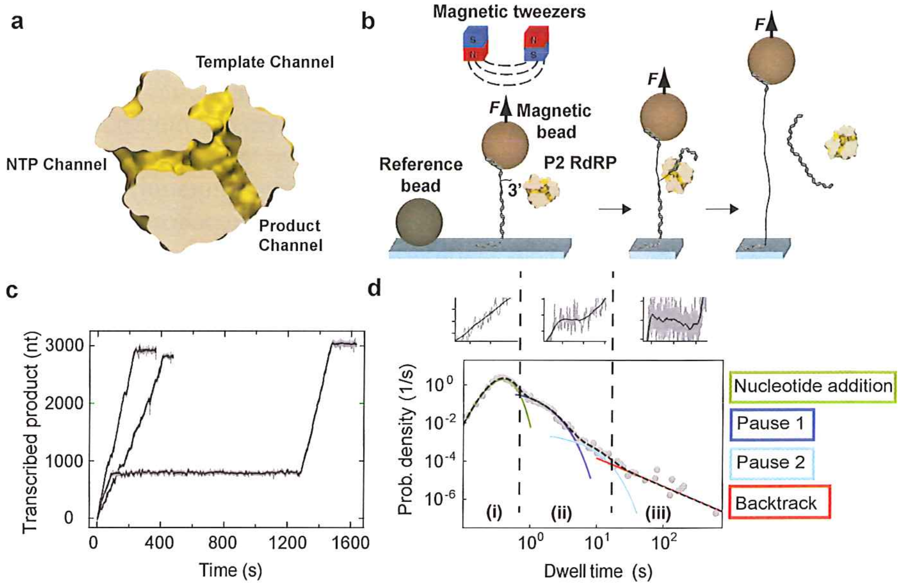


Figure 6.2: **Detecting P2 activity at the single-molecule level.** **(a)** Illustration of a cross-section of P2 in elongation, displaying the NTP and template tunnels (according to structure of the initiation complex, PDB number: 1UVI; [14]) together with the product channel, which has been drawn in manually for purposes of illustration. **(b)** Schematic of the experimental setup for monitoring P2 transcription on magnetic tweezers. In all three panels, *F* refers to the force applied by the magnetic tweezers. A primarily duplex RNA is tethered to a magnetic bead at one extremity and to a surface at the other. This RNA construct is built on a 4.2-kb plus strand to which a 2.8-kb template strand for P2 transcription is hybridized. The template strand is fully complementary to the 4.2-kb plus-strand apart from a 15-nt overhang at the 3' end to facilitate enzyme initiation (**Fig. 6.8**). **(c)** Three typical traces of P2-catalyzed transcription representing the increase in product length versus time at 20 pN force and $[\text{NTP}]_{\text{opt}}$. All three traces were acquired at 25 Hz (gray) and low-pass filtered at 0.5 Hz (black). These three traces present different dynamic behavior that is attributable to three unique P2 enzymes. The leftmost trace includes almost no pauses. The middle trace is interrupted by few short pauses, and the rightmost trace includes two distinct regions of fast transcription activity separated by a very long pause. **(d)** The dwell time distribution is extracted from 52 traces of P2 transcription activity acquired at 20 pN and $[\text{NTP}]_{\text{opt}}$ (gray dots). We fit this distribution to a stochastic-pausing model (Supplemental Information) by using MLE (dashed black line). For clarity, we individually plot each contribution to the dwell-time distribution: in green, the Gamma distribution; in dark blue, the first short pause (Pause 1); in cyan, the second short pause (Pause 2); and in red, the power law distribution of pause times originating in backtracking. On top of the histogram are plotted representative events in P2 activity: from left to right (1) fast incorporation without pause, described by the Gamma distribution; (2) short pauses, described by the two exponential distribution; and (3) long backtracked pauses, described by a power law distribution. The distinct appearance of the nucleotide-addition peak and the fact that the fitted Gamma distribution captures both its width and average (using a single free parameter, assuming that the polymerase takes 1-nt steps) show that our approach is not appreciably affected by experimental noise and avoids convolving elongation with pausing (Supplemental Information; **Fig. 6.10a**).

sible through a novel pause state. Direct mutations of the active site that have a large effect on mutation rates are generally deleterious[19]. Therefore, the ability to gradually tune between two pathways with different inherent fidelities offers an attractive alternative evolutionary strategy to achieve the precise mutation rates needed to track the evolution of the host and ensure continued viral pathogenicity. Detailed information regarding how error probabilities are dynamically set could also facilitate the development and refinement of drug therapies specifically targeting viral mutagenesis[4–6].

6.4. RESULTS

6.4.1. MULTIPLEXED MAGNETIC TWEEZERS DETAIL RDRP DYNAMICS ON SUB-SECOND TO HOUR TIMESCALES

In our magnetic tweezers assay, dozens of magnetic beads (**Fig. 6.8b**) are tethered to a surface via dsRNA (**Fig. 6.2b**). When P2 is added to the reaction buffer, it initiates at a 3' overhang present on each tether (**Figs. 6.2b** and **6.8a**)[17, 20–22]. By processively elongating the RNA product, P2 converts the tethered dsRNA into single-stranded RNA (ssRNA) while producing a free dsRNA strand (**Fig. 6.2b**). Because of the length difference between ssRNA and dsRNA (**Fig. 6.8c**), the motion of each elongating P2 molecule can be determined from the change in vertical position of the magnetic bead (**Fig. 6.2b**) with an experimental spatial resolution of ~5 nt or better, depending on the applied force and the length of the exposed ssRNA (with the instrumental noise, measured to equal 0.3 nm at 0.5 Hz on a fixed bead, providing the ultimate limit; **Fig. 6.8b**). The absolute bead position at initiation is determined with a 10- to 20-nt precision, limited by occurrences of P2 pausing shortly after initiation. Each trace originates in the activity of a single P2, as the initiation site is removed once P2 has converted the 3' overhang into duplex RNA (**Fig. 6.2b**). Different traces taken during the same experiment differ remarkably (**Figs. 6.2c** and **6.8d**) and demonstrate the highly stochastic nature of P2 elongation dynamics; for example, in **Fig. 6.2c**, the fastest enzyme shown spent at most a few seconds in a paused state, while the slowest enzyme paused for more than 1000 s.

6.4.2. MAXIMUM LIKELIHOOD ESTIMATION BASED ON THE GENERAL SEQUENCE INDEPENDENT ELONGATION MODEL WITH TWO PAUSES AND BACKTRACKS

To analyze the stochastic P2 elongation dynamics, we perform a probabilistic analysis of the times it takes the polymerase to transcribe through consecutive windows of 10 nt along the trace. This window size is set to equal at least twice the experimental spatial resolution of the noisiest configuration (lowest force, ssRNA tethers) in order to avoid noise-triggered crossings in and out of the dwell-time window. The recorded times are referred to as dwell times. By collecting a large number of dwell times (for the exact total number of dwell time in each condition presented here, see **Fig. 6.8e**), we can construct empirical dwell-time distributions (**Fig. 6.2d**) that capture the probabilistic nature of P2 dynamics with high precision (see typical error bars indicated in **Fig. 6.9c**). A few characteristic features become immediately apparent (**Fig. 6.2d**; Supplemental Information): (1) a short-time peak—the nucleotide-addition peak—originating from P2 elongating through the dwell-time window without pausing, with the position of the apex of this peak reflecting the typical time taken to cross the dwell-time window without pausing, and its width is set by the number of steps taken in doing so; (2) a shoulder situated at intermediate times, corresponding to one or more short pauses with exponentially distributed lifetimes; and (3) a long-lived pause with a broad distribution of lifetimes that is consistent with a polymerase paused by backtracking[23–27]. As described in the Supplemental Information, we use the Bayes-Schwartz information criterion[28] to objectively determine that the data support the existence of two separate short pauses—Pauses 1 and 2—represented in the shoulder (2) of our measured dwell-time distributions (**Fig. 6.2d**). We use maximum-likelihood estimation (MLE) (Supplemental Information) to fit a general stochastic-pausing model to each measured dwell-time distribution and extract the nucleotide addition rate (k_{na}), the probability of nucleotide addition without pausing (P_{na}), as well as the pause entrance probabilities (P1 and P2) and the exit rates (k_1 and k_2) for the two short pauses. The fitting parameters extracted from the MLE are not altered by the noise encountered in our experiments (**Fig. 6.10a**). As the exponential-pause shoulders in the dwell-time distributions hide the features needed to determine the amount of backtracking, we merely include this pause type to avoid introducing a late-time cutoff that might bias the analysis. Under any single realization of experimental conditions, the fitted dwell-time distribution is consistent with a multitude of reaction schemes including the specified pauses. In the Discussion, we argue that the data from our concentration sweeps suggest a particular minimal reaction scheme. As we cannot accurately determine the absolute position of the RdRP, we do not include sequence effects in our model. In the Supplemental Information, we also argue why sequence effects are not likely to be significant. Even so, if there is strong sequence dependence, our parameter estimates can be seen as representing averages over sequence.

6.4.3. THE NTP CONCENTRATION DEPENDENCE OF P2 CATALYZED RNA ELONGATION DYNAMICS

To probe the dynamics of nucleotide incorporation, we next vary the nucleotide concentration in our assay in the μM to mM range, all at a constant force (30 pN) and nucleotide stoichiometry (**Figs. 6.3** and **6.9a**). We find that the average time P2 required to transcribe the template increases with a decrease in the NTP concentration: from 130 s at the highest concentration $[\text{NTP}]_{\text{opt}}$ (a mM concentration optimized for P2 initiation; see Supplemental Information and [29, 30] to 1750 s at $[\text{NTP}]_{\text{opt}}/500$ Comparing the dwell-time distributions obtained at $[\text{NTP}]_{\text{opt}}$ and $[\text{NTP}]_{\text{opt}}/100$ (**Fig. 6.3a**, black circles and red triangles, respectively), we see that a hundred-fold decrease in the nucleotide concentration only results in a slight shift in the nucleotide-addition rate (as judged from the position of the apex of the nucleotide-addition peak in **Fig. 6.3a**). At the same time, the time spent in short pauses is greatly increased (as judged from the weight under the exponential shoulders in **Fig. 6.3a**). These results suggest that in the concentration range examined, the major shift in dynamics originates in a change of pausing behavior, and not in a slowdown of the nucleotide-addition process. The empirical dwell-time distribution is well described by our stochastic-pausing model over the concentration range examined (**Fig. 6.3b,c**). Importantly, since we can directly judge from **Fig. 6.9a** if the elongation peak is discernable against the background of the pausing shoulder, we can determine the concentration range over which it is possible to estimate the bare nucleotide addition rate without the risk of inadvertently including the effect of pauses. This constitutes a clear advantage over standard attempts of fitting out pause-free velocities as detailed in the Supplemental Information and in **Fig. 6.10b,c**.

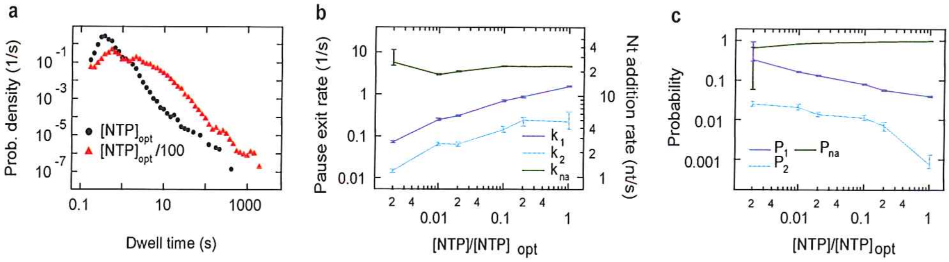


Figure 6.3: **The nucleotide concentration dependence of P2 transcription dynamics (a)** The empirical dwell-time distributions for two different NTP concentrations at the same force (30 pN): black dots, $[NTP]_{opt}$; red triangles, $[NTP]_{opt}/100$. **(b)** The pause exit and nucleotide addition rates as a function of NTP concentration. The error bars represent the SD of a MLE procedure applied to 200 bootstrapped data sets. The nucleotide addition rate (right axis) is represented by green dots connected by green lines as guides for the eye. The value of the nucleotide addition rate at $[NTP]_{opt}/500$ is not well estimated because of the disappearance of the Gamma distribution behind Pause 1 distribution (**Fig. 6.9a**). The pause exit rates (left axis) are represented by dots connected by lines as guides for the eye (dark blue, k_1; cyan, k_2). (c) Probability for P2 to be in one of the three different states (green, nucleotide addition; dark blue, Pause 1; cyan, Pause 2) as a function of relative NTP concentration.

6.4.4. ERROR INCORPORATION IS ASSOCIATED WITH P2 PAUSING

To probe how error incorporations affect RdRP dynamics, we introduced the nucleotide analog inosine 5'-triphosphate (ITP) to the reaction buffer. ITP has previously been used to probe error incorporation by eukaryotic RNA polymerase II (RNAPII)[31] and *Escherichia coli* (*E. coli*) RNAP[26]. As P2 is neither structurally nor evolutionary related to the multi-subunitsRNAPs of cellular organisms[15, 32], we first analyzed how P2 RdRP incorporates ITP in bulk experiments. We used ethidium bromide (EtBr) staining (**Fig. 6.4a**, upper panel) and radiolabeling ([α-32P]ITP) (**Fig. 6.4a**, lower panel) to quantify the product synthesis and the ITP incorporation when one of the canonical NTPs was replaced with ITP. We observed that ITP could replace GTP, as the reaction successfully continues to completion (**Fig. 6.4a**, lane "no GTP") with only a slight reduction in efficiency (**Fig. 6.4a**, upper panel, compare lanes "no GTP" and "all NTPs"): this implies that v will frequently replace GTP in competitive nucleotide addition on a template cytosine. The intermediate band observed in the "no GTP" lane is likely due to secondary structures in the ssRNA template, which transiently stall P2 during the replication process. When replacing other nucleotides with ITP, no replication products were observed (**Fig. 6.4a**): this implies that ITP acts as a mismatched base when pairing with G, U, and A in the template strand. The behavior of P2 in the presence of ITP is directly comparable to the behavior of RNAPII under similar conditions[31], which implies that ITP can also be used to probe error incorporation by P2.

We next studied P2 transcription in the presence of ITP in our single-molecule assay. At 1-mM ITP, we observed a marked increase in the time spent in short pauses (**Fig. 6.4b**), strongly suggesting that these pauses are connected to error incorporation. The empirical dwell-time distribution is well described by our stochastic-pausing model in the presence of ITP (**Fig. 6.4c,d**).

6.4.5. MANGANESE INCREASES THE NUCLEOTIDE ADDITION RATE AND LOWERS THE PROBABILITY OF P2 PAUSING

Multiple studies on both DNA polymerases and RdRPs have shown that their fidelity is decreased when there is only manganese (i.e., no magnesium) present in the reaction buffer[33–35]. To investigate the effect of manganese on P2 elongation dynamics, we titrated the manganese concentration between 0 to 2 mM once the P2 has entered elongation (Supplemental Information). Varying the manganese concentration resulted in significant changes in the dwell-time distributions (**Fig. 6.5a**): as the manganese concentration is increased, Pause 1 and Pause 2 probabilities decrease by 30%–40% and 60%–70%, respectively, and the Pause 1 exit rate increases by 30%–40%, while the Pause 2 exit rate shows no significant change. Additionally, the nucleotide addition rate increases by 10%–20% (**Fig. 6.5b,c**). Taken together, these results are consistent with an increase in overall elongation rate with increased manganese concentrations, as reported in an earlier biochemical bulk study[22].

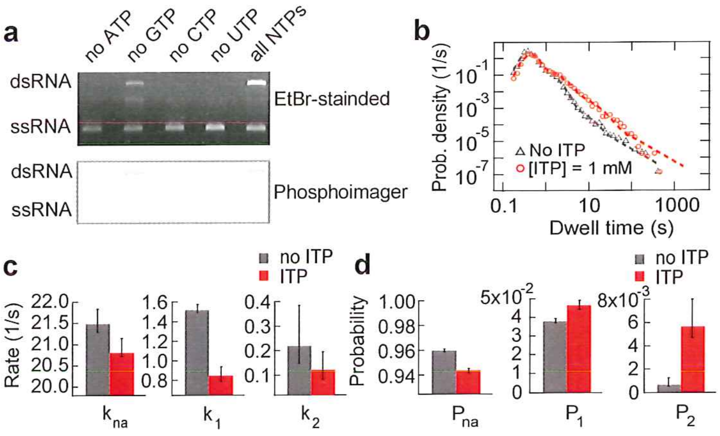


Figure 6.4: **ITP Incorporation by P2 slows catalytic activity and increases pause 2 probability.** **a)** Replacement of GTP by ITP in P2-catalyzed replication reaction. Initiation reactions were car- ried out in the presence of 1-mM GTP and ATP. The formed elongation complexes were purified, and the formation of new initiation complexes was prevented by heparin. The subsequent elongation reactions were performed in [NTP]opt, supplemented with 35-nM [α–^{32}P]ITP (lane all NTPs) or in conditions where one of the canonical rNTPs was replaced with 1-mM ITP as indicated. The reaction products were analyzed by standard agarose gel electrophoresis using EtBr staining (upper) or visualization by phosphorimaging (lower). The mobility of the 2948-nt-long ssRNA template and the dsRNA product is indicated on the left. The radioactively labeled ssRNA bands observed in the no-ATP and no-CTP lanes (lower) likely correspond elongation complexes that were stalled at early stage of the elongation phase, close to the 3' end of the ssRNA template. **b)** Empirical dwell-time distribution from single molecule experiments with (red circles) and without (gray circles) ITP, at 30 pN force and [NTP]$_{opt}$. The dashed lines represent the MLE fits. **c)** Rates extracted from the MLE fits of the distributions in (**b**). (**d**) Probabilities extracted from the MLE fits of the distributions in (**b**). The errors bars in (**c**) and (**d**) represent the standard deviation of a MLE procedure applied to 200 bootstrapped data sets.

6.4.6. THE FORCE DEPENDENCE OF P2-CATALYZED RNA ELONGATION

P2 performs both replication and transcription on ssRNA templates. Here we observe the force dependence of P2 enzymatic activity when P2 has to displace the complementary strand of the template. In the applied force range (16–35 pN), the base pairs at the ds-ssRNA junction (**Fig. 6.2b**) are destabilized by 0.6-2k_BT due to the applied force. We acquired approximately 45 traces per applied force at [NTP]$_{\mathrm{opt}}$ (**Figs. 6.6a** and **6.9b**) and found that an increase in the force leaves the nucleotide-addition peak unaffected, but decreases the amount of time spent pausing (**Fig. 6.6a**). The increased time spent pausing is quantitatively confirmed by the rate and probability estimates produced from fitting the dwell-time distributions (**Fig. 6.6b,c**). In **Fig. 6.10d**, we show how these results contrast with those that result from the use of the more standard approach of attempting to extract pause-free velocities from velocity histograms.

6.5. DISCUSSION

6.5.1. BASIS FOR A KINETIC MODEL OF P2-CATALYZED RNA ELONGATION

Our multiplexed magnetic-tweezers setup enables us to collect very large dwell-time data sets detailing the dynamics of the ϕ6 RdRP P2 over four orders of magnitude in time. The size of the data sets allows us to appreciably constrain the type of models consistent with our data. We use MLE to analyze our dwell time distributions. Similar stochastic analysis has previously been used in the study of ion channels[36] and single-molecule fluorescence microscopy[37], but not in single-molecule force spectroscopy experiments – in part due to the relative paucity of data that could be collected. Using the Bayesian information criterion (Supplemental Information), we determine that the inclusion of at least two exponentially distributed short pauses is supported by the data. In constructing a kinetic model[38] of P2-catalyzed RNA elongation, we therefore choose to include the two simplest and most widely accepted pathways for polymerase pausing: pausing due to nucleotide misalignment arising from error incorporations[10] (**Fig. 6.13a**) and pausing that originates in a thermally driven large-scale conformational change that renders the enzyme catalytically inactive[39] (**Fig. 6.13c**).

6.5.2. EVIDENCE FOR TWO DISTINCT NUCLEOTIDE ADDITION PATHWAYS

A key observation is that the pauses we identify fail to satisfy the nucleotide concentration dependencies expected for the two above-mentioned categories of off- and on-pathway pauses (for the full argument, see Supplemental Information, especially **Fig. 6.13a-d**). In order to capture the observed decrease in pause lifetimes with increased nucleotide concentration (**Fig. 6.3b**), we need to relax the assumption of complete catalytic inactivity[39, 40] in the RdRP conformation responsible for the standard off-pathway pause (**Fig. 6.7a**). To capture the concomitant decrease in pause probability, we further need to assume that nucleotide addition from the off-pathway configuration forces its reversal, bringing the RdRP back to its original on-pathway configuration (**Fig. 6.7a**). With this, we can rationalize the concentration dependencies of **Fig. 6.3b** in terms of Michaelis-Menten kinetics of nucleotide addition from different catalytically active configurations of the RdRP, with and without a mismatched terminal base (see **Fig. 6.7b** and the discussion in next subsection). Interestingly, a recent fluorescence resonance en-

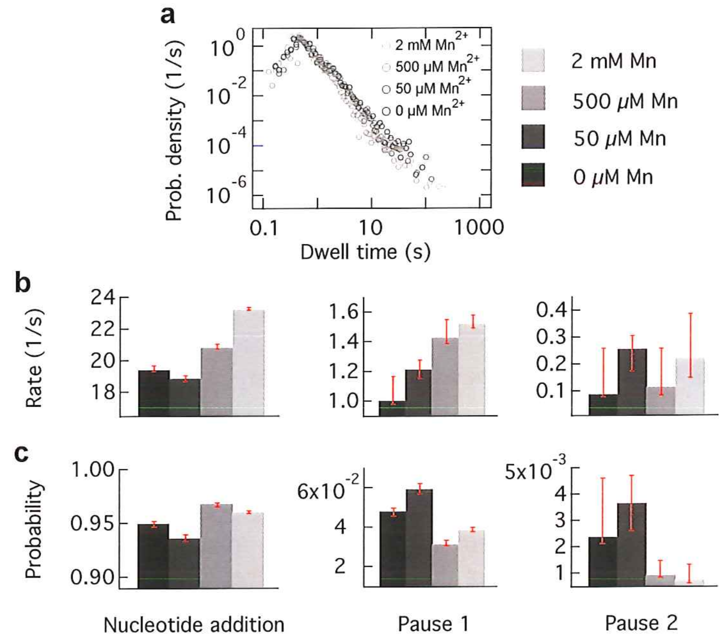


Figure 6.5: **Mn^{2+} increases the nucleotide addition rate and reduces the probability of pausing.** **(a)** The empirical dwell-time distribution for four different concentrations of Mn2+: 0 mM, 50 mM, 500 mM, and 2 mM. The concentration of Mg^{2+} is maintained constant at 5 mM, as is the nucleotide concentration (fixed at [NTP]$_{\text{opt}}$). **(b)** The nucleotide addition rate, the Pause 1 exit rate, and the Pause 2 exit rate, plotted for the different manganese concentrations. **(c)** The probabilities for P2 to be present in nucleotide addition, Pause 1, or Pause 2, plotted for the different manganese concentrations.

ergy transfer study[41] of an error-prone family Y DNA polymerase, related to RdRPs and RTs[15], also provides evidence of an alternative pathway to increased error incorporation.

6.5.3. LONG PAUSES DIRECTLY REPORT ON SINGLE-ERROR INCORPORATION EVENTS

We now expect one of our two observed pauses to correspond to a slowdown after an error has been incorporated; the probability of this pause should be a direct readout of the error rate of the process. Assuming Pause 1 is the pause reporting on errors unrealistically leads to an error estimate of one error every 26 ± 1 bp at [NTP]$_{\text{opt}}$ (**Fig. 6.3c**), while assuming it is Pause 2 leads to an error estimate of one error every 1350 ± 410 (**Fig. 6.3c**), in broad agreement with error-estimates for related enzymes[3]. This suggests that Pause 1 corresponds to the entrance into a long-lived elongation-competent state (**Fig. 6.3b**), with a reduced – albeit nonzero – catalytic activity; Pause 2 arises from a catalytic slowdown (**Fig. 6.3b**) when an error has just been incorporated, as is reported for other polymerases upon error incorporation[7, 10]. Since the probability of entering Pause 2 is largely proportional to the probability of entering Pause 1 (**Fig. 6.3c**), we infer that most errors are incorporated through the slow catalytic state that accompanies Pause 1, rendering this a fidelity-controlling pause. From here on we refer to the fast active state as the high-fidelity catalytic (HFC) state, the state corresponding to Pause 1 as the low-fidelity catalytic (LFC) state, and the state corresponding to Pause 2 as the terminal-base mismatched catalytic (TMC) state (**Fig. 6.7a**). By considering **Figs. 6.3b** and **6.7a**, we can infer which path dominates the exit out of each state. Correct nucleotide addition dominates the exit from the HFC state, as most catalytic steps occur without pausing (p_{na} is close to one). The same holds true for LFC state, as the total exit rate from the LFC state depends on NTP concentration, and only a small fraction of transitions originating in the LFC state enters into the TMC state. The fact that the nucleotide addition rate in the TMC state is proportional to but lower than that of the LFC state (**Fig. 6.3b**) indicates that the two states have a similar nucleotide dissociation constants, but that there is an increased barrier to nucleotide addition in the TMC.

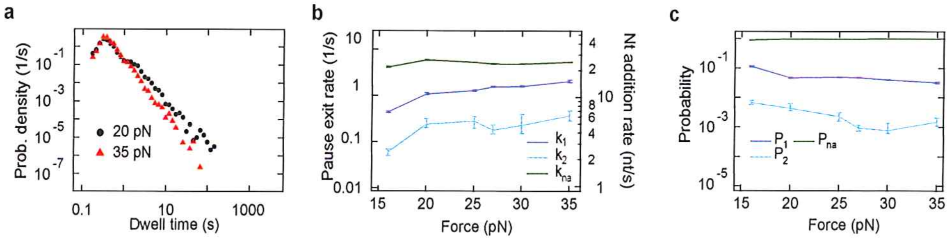


Figure 6.6: **The force dependence of P2 transcription.** **(a)** The dwell time distribution for two different applied forces at [NTP]opt: black dots, 20 pN; red triangles, 35 pN. **(b)** The pause exit and nucleotide addition rates as a function of force at $[\text{NTP}]_{\text{opt}}$. The error bars represent the SD of a MLE procedure applied to 200 bootstrapped data sets. The nucleotide addition rate (right axis) is represented by green dots connected by green lines as guides for the eye. The pause exit rates (left axis) are represented by dots connected by lines as guides for the eye (dark blue, k_1; cyan, k_2). **(c)** Probability for P2 to be in one of the three different states (green, nucleotide addition; dark blue, Pause 1; cyan, Pause 2).

6.5.4. NUCLEOTIDE ANALOG AND MANGANESE TITRATION CONFIRM THE LINK BETWEEN LONG PAUSES AND ERROR INCORPORATION

To confirm that the long pauses observed are a direct result of error incorporation, we examined the effect of introducing ITP into the reaction buffer (**Fig. 6.4b-d**). Inosine has previously been shown to act like an error after incorporation by other polymerases[42, 43], but to bind more strongly than a standard mismatch to the template base in the active site. Consequently, assays utilizing inosine will allow us to detect the effects of increased error incorporation already at modest levels of inosine. As the effect of ITP has not been established for P2, we first performed bulk experiments showing that ITP influences P2-directed RNA elongation in a manner consistent with the effect in other polymerases: it acts like an error when incorporated instead of all canonical bases apart from G, which it replaces with only a slight reduction in efficiency (**Fig. 6.4a**). With the introduction of ITP in our buffer, we expect to see a rise in the probability of entering the TMC state. We also expect a modest reduction of nucleotide addition rate because

ITP replaces GTP with a slight reduction in efficiency. As the entrance into the LFC state is assumed to be through a thermally driven transition in competition with nucleotide addition, we also expect a small rise in the probability of Pause 1 as the nucleotide addition rate is moderately decreased. The exit rate out of the LFC state (Pause 1) is also expected to decrease, as correct nucleotide addition is the dominant escape route and ITP is incorporated instead of GTP with a reduced efficiency. All of these effects are observed experimentally (**Fig. 6.4c,d**). Additionally, and in accord with a run-off replication assay[22], our manganese titration experiments show that the nucleotide addition rate increases with manganese concentration (**Fig. 6.5b**). With an increased nucleotide addition rate, we would expect a decreased entry to the LFC state and consequently a fall of the probabilities of both Pause 1 and Pause 2; this is also observed in our data (**Fig. 6.5c**), adding further support to our model.

Interestingly, our manganese experiments lead to a different conclusion for polymerase fidelity from what has been proposed based on stop-flow experiment on the poliovirus RdRP. Replacing magnesium with manganese in the reaction buffer has previously been suggested to decrease RdRP fidelity[33] not increase it, as suggested by our data (**Fig. 6.5c**). Possible explanations are that there is a significant enzymatic differences between the polio and the $\phi6$ RdRPs or that we did not completely replace magnesium with manganese in the reaction buffer. An interesting alternative explanation for the observed differences is that for enzymes with dual incorporation pathways, stop/quench flow assays are liable to skew the balance between the pathways used by allowing the HFC and LFC states to equilibrate between nucleotide additions. For our system, starving the polymerase of nucleotides would lead to an overestimate of error incorporation, an effect that is enhanced by the addition of manganese (**Fig. 6.5b**), as it increases the catalytic rates.

6.5.5. FORCE DEPENDENCE OF PAUSE PROBABILITIES SUGGESTS TRANSLOCATION BEFORE CHEMISTRY IN THE SLOW CATALYTIC STATES

We observe no force dependence in the nucleotide addition rate from the HFC state, implying that either the translocation step is not rate limiting or that P2 translocates through a power stroke that renders it insensitive to the forces we apply. However, the nucleotide addition rates in the LFC and TMC states are still sensitive to force (**Fig. 6.6b**). Fitting an Arrhenius law to the exit rates (**Fig. 6.11**), we determine that the transition states are situated 0.17 ±0.02 and 0.12 ± 0.09 nm from the pretranslocated position, respectively; these distances are consistent with the decrease in tether length due to the opening of one base pair in the tethered dsRNA (**Fig. 6.8c**) and strongly suggest that the chemistry of bond formation does not provoke translocation in the LFC and TMC states, but rather that translocation of P2 is here thermally driven.

6.6. CONCLUSIONS

HERE, we have extensively probed the mechanochemistry of viral mutagenesis. Specifically, we studied the elongation dynamics of P2 from bacteriophage $\phi6$, an established model system for RdRPs and RTs, by employing high-throughput magnetic tweez-

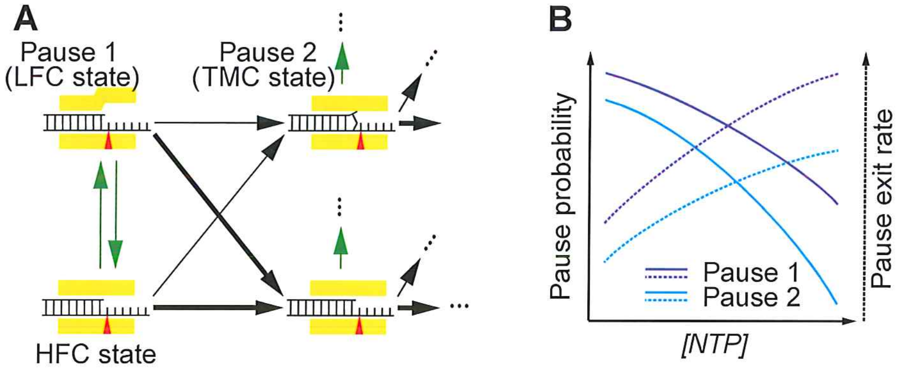


Figure 6.7: **Model for on-pathway polymerase pauses.** **(a)** In our model, the polymerase has two catalytically active configurations, indicated by a narrow or wide space around the active site in the schematic polymerase structure (yellow). This, together with the state of the nascent RNA – having a terminal mismatch or not – gives rise to our two observed pause types. The resulting states for the RdRP and RNA are an HFC state, which is fast, without terminal mismatch, and with a low error rate; a LFC state, which is slow, without terminal mismatch, error prone, and entered through a thermal transition from the HFC state; and a TMC state, entered upon error incorporation, where nucleotide misalignment results in very slow nucleotide addition (Pause 2). In the reaction scheme, black arrows represent transitions dependent on nucleotide concentration, while green arrows represent thermal transitions. The thick arrow out of each state represents the dominant path to leave that state, as deduced from our experimental observations (see main text). **(b)** The expected trends in pause probabilities and pause exit rates as a function of NTP concentration for the pauses in **(a)**. These rates are dominated by the transitions indicated by thick arrows in **(a)**, and from **Fig. 6.3b**, we see that these exit rates are consistent with Michaelis-Menten kinetics.

ers combined with maximum-likelihood analysis. This approach has allowed us to capture the P2-catalyzed RNA elongation process for both correct and incorrect nucleotides with unprecedented detail. By analyzing how the replication dynamics of P2 RdRP are affected by nucleotide concentrations, destabilization of the fork, and the presence of both nucleotide analog and manganese, we have found strong evidence for a stochastic partitioning between two catalytically active configurations with different catalysis rates and substrate selectivity. Importantly, bulk techniques such as stop/quench flow assays would not be able to discriminate between the two catalytic pathways we have described and would always under estimate the fidelity of the polymerase.

Our observations are of both potential therapeutic and fundamental importance: identifying compounds that affect the probability of entering the fidelity controlling pause we describe here (Pause 1 in the above) might lead to ways of modifying mutation rates in RNA viruses and consequently decrease their pathogenicity; the gradual tuning of probabilities of entering fidelity controlling pauses might similarly be a general evolutionary strategy that allows viral populations to continuously adapt to the evolution of the host's antiviral defenses. Such an evolutionary strategy is especially appealing since mutations at the active site of viral polymerases induce large fidelity changes[3] that may inactivate the virus[19]. Using our novel approach for single-molecule error quantification at physiological nucleotide stoichiometry, future studies will examine whether the conclusions we draw for P2 hold true also for other polymerases.

Finally, and importantly, our high-throughput single-molecule approach is able to detail the dynamic behavior of enzymes over four orders of magnitude in time; used together with the accompanying theoretical modeling, this should help to usher in a mode of single-molecule approaches where a vast array of kinetic effects can be separated with statistical significance, increasing the biological impact of single-molecule studies. This is especially exciting in the context of understanding the molecular effect of drugs, as our approach has the potential to yield precise dynamic fingerprints of any antiviral drug[4, 44] affecting the elongation process.

6.7. EXPERIMENTAL PROCEDURES

6.7.1. SINGLE-MOLECULE TRANSCRIPTION EXPERIMENTS

Once the RNA construct length is calibrated inside a flow cell containing P2 reaction buffer (50-mM HEPES [pH 7.9], 20-mM ammonium acetate, 3% w/v polyethylene glycol 4000 Da, 0.1-mM EDTA [pH 8.0], 5-mM $MgCl_2$, 2-mM $MnCl_2$, 0.01% Triton X-100, 5% Superase RNase inhibitor [Life Technologies], and 20-mg/ml BSA) (see Supplemental Information for P2 and RNA production and **Fig. 6.8a** for the RNA construct), 9 nM of P2 is flushed in. We perform experiments at 21oC for 1 hr at constant force and fixed NTP concentration while recording images of the magnetic beads at 25 Hz. Because of the large field of view of our camera we are able to follow up to 200 tethers at a time[45, 46]. The recorded images of the beads are converted into (x, y, z) positions using custom-written routines in Labview[47, 48]. Distinct traces are low-pass filtered at 0.5 Hz, providing an optical resolution of 0.3 nm along the optical axis (z axis) (**Fig. 6.9b**) and synchronized with regards to starting position. The changes in extension are converted into the numbers of transcribed nucleotides using the force-extension relationships for dsRNA and ssRNA constructs obtained in P2 reaction buffer (**Fig. 6.3b**)[49].

6.7.2. STALLING REACTION WITH P2

Following a 30-min incubation with the P2 initiation buffer (P2 reaction buffer supplemented with 1-mM rATP, 1-mM rGTP, 0.2-mM rCTP), we rinse the reaction chamber with an excess amount of P2 reaction buffer containing 0.5% Superase RNase inhibitor (Life Technologies) (no NTPs). We then trigger elongation by adding P2 reaction buffer supplemented with rNTPs (and ITP) as indicated. The dynamic of P2 is not affected by the stalling reaction (**Fig. 6.12**). In the case of the Mn^{2+} titration experiments, no EDTA was added into the P2 reaction buffer for all experiments containing less than 2-mM Mn^{2+}.

6.7.3. BULK REPLICATION EXPERIMENTS

RNA replication reactions were carried out in 50-mM HEPES (pH 7.5), 20-mM ammonium acetate, 6% (w/v) polyethylene glycol 4000, 5-mM $MgCl_2$, 2-mM $MnCl_2$, 0.1-mM EDTA, 0.1% Triton X-100, and 0.8-U/ml RNasin Ribo Lock (Thermo Scientific). Initiation reactions were carried out (125-nM ssRNA and 500-nM P2) in the presence of 1-mM rGTP and rATP (to allow the synthesis of the first 10 nts complementary to the 3' end of the ssRNA template) at 30 °C for 20 min, the formed elongation complexes were purified using Sephacryl S-300 matrix (Pharmacia) and the reaction buffer with no NTPs for elution. The formation of new initiation complexes was prevented by heparin (5 mg/ml), and the incubation was continued at 30 °C for 5 min. The subsequent elongation reactions were supplemented with 0.2 mCi/ml of $[\alpha-^{32}\text{P}]$ITP (6000 Ci/mmol; Cambro Scientific GmbH) and the indicated nucleotides, and the incubation at 30 °C was continued for 2 h.

6.7.4. GENERATION OF EMPIRICAL DWELL TIME DISTRIBUTIONS

When constructing empirical dwell-time distributions, dwell-times from different polymerases operating at identical conditions were grouped together (**Fig. 6.8e**). Looking at dwell-time distribution for individual polymerases, a small percentage consistently spend a much longer than average time pausing; to prevent these pause-prone polymerases from biasing the data, we remove the 5% with the highest and lowest dwell-time densities at 3 s.

6.8. SUPPORTING INFORMATION

6.8.1. PLASMIDS TEMPLATES FOR SSRNA PRODUCTION

Plasmid pBB10 (Petrushenko et al., 2006) was used for the production of AB, CD, SP, BIO and DIG SSRNAS and plasmid pLM659 (Gottlieb et al., 1992) for the production of S+ SSRNA.

6.8.2. PRODUCTION OF SSRNA MOLECULES BY T7 TRANSCRIPTION

Single-stranded RNA molecules are produced by in vitro run-off transcriptions using T7 RNA polymerase from the Ribomax large-scale RNA production systems (Promega). Transcription reactions for AB, CD and SP SSRNA fragments (**Fig. 6.8a**) contain 500 ng DNA template (purified PCR fragment that includes a T7 promoter), 10 μl T7 buffer, 1 μl CTP (100 mM), 1 μl ATP (100 mM), 1 μl GTP (100 mM), 1 μl UTP (100 mM) and 5 μl T7 polymerase in a reaction volume of 50 μl. For the production of BIO SSRNA the UTP amount is reduced to half and the reaction is supplemented with 4.7 μl biotin-16-UTP (10 mM, Roche), while for DIG SSRNA production the UTP amount is reduced to 0.63 μl and the reaction is supplemented with 3.7 μl of digoxigenin-11-UTP (10 mM, Roche).

Transcription products are purified with the RNeasy MinElute cleanup kit (Qiagen). Samples are eluted with 1 mM sodium citrate (pH 6.4) (Life Technologies) and concentrations of SSRNA molecules are determined using a Nanodrop (NanoDrop, Wilmington, DE). The S+ SSRNA for bulk experiments is produced and purified as previously described[50].

6.8.3. DESCRIPTION OF HYBRIDIZATION PROTOCOL AND TETHERING CONDITIONS

The different SSRNA molecules that make up the complete RNA construct for single-molecule studies of P2 transcription (**Fig. 6.8a**) are combined at equimolar ratios (taking into account the differences in their lengths), with the exception of the short handles (BIO and DIG) which are added in a four times molar excess compared to the AB strand (1 μg AB strand, 1.4 μg CD strand, 360 ng SP strand, 450 ng BIO strand, and 516 ng DIG strand). The reaction is carried out in 0.5× saline-sodium citrate (Promega) and 1 mM sodium citrate in a reaction volume of 200 μl.

Hybridization is performed in a PCR machine that is allowed to slowly cool down starting from 65 °C. The temperature is kept constant at 65 °C for an hour and subsequently reduced in incremental steps of 1.2 °C until a final temperature of 25 °C is reached. At each intermediate step the temperature is maintained constant for 5 min. dsRNA is purified after hybridization with the RNeasy MinElute cleanup kit (Qiagen). Hybridization products are eluted with 1 mM sodium citrate. Concentrations are determined using a Nanodrop (NanoDrop, Wilmington, DE).Tethering of the RNA construct was carried out in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 mM EDTA, 15 mM sodium azide, 0.01% Triton X-100, and 0.5% Superase RNase inhibitor (Life Technologies). This buffer was used in all the steps prior to the addition of P2 enzyme to the reaction chamber.

6.8.4. PURIFICATION OF ϕ6 P2

Recombinant N-terminally histidine-tagged ϕ6 RdRP was expressed from plasmid pAA5 [30] in *E. coli* BL21(DE3) [51] and purified using Ni-NTA affinity column (Qiagen), HiTrap$_{\mathrm{TM}}$ Heparin HP column and HiTrap$_{\mathrm{TM}}$ Q HP column (GE Healthcare) as previously described [29, 30]. The non-tagged RdRP used for bulk experiments was expressed from plasmid pEMG2[52] in *E. coli* BL21(DE3) (Studier and Moffatt, 1986) and purified as previously described[29, 52]. The purified proteins were stored in 50% glycerol, 140 mM NaCl, 50 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 0.1% Triton-X 100 at -20 °C.

6.8.5. SINGLE-MOLECULE TRANSCRIPTION EXPERIMENTS

Once the RNA construct length is calibrated inside a flow cell containing P2 reaction buffer [50 mM HEPES pH 7.9, 20 mM ammonium acetate, 3% w/v polyethylene glycol 4000 Da, 0.1 mM EDTA pH 8.0, 5 mM $MgCl_2$, 2 mM $MnCl_2$, 0.01% Triton X-100, 5% Superase RNase inhibitor (Life Technologies), and 20 μg/ml BSA] (see Supplemental Experimental Procedures for P2 and RNA production and **Fig. 6.8a** for the RNA construct), 9 nM of P2 is flushed in. We perform experiments at 21 °C for 1 h at constant force and fixed NTP concentration while recording images of the magnetic beads at 25 Hz. Due to the large field of view of our camera we are able to follow up to 200 tethers at a time[46]. The recorded images of the beads are converted into (x, y, z) positions using custom-written routines in Labview[47, 48]. Distinct traces are low-pass filtered at 0.5 Hz and synchronized, providing an optical resolution of 0.3 nm along the optical axis (z axis) (**Fig. 6.8b**). The changes in extension are converted into the numbers of transcribed nucleotides using the force-extension relationships for dsRNA and ssRNA constructs obtained in P2 reaction buffer (**Fig. 6.8c**)[49].

6.8.6. STALLING REACTION WITH P2

Following a 30 min incubation with the P2 initiation buffer (P2 reaction buffer supplemented with 1 mM rATP, 1 mM rGTP, 0.2 mM rCTP), we rinse the reaction chamber with an excess amount of P2 reaction buffer containing 0.5% Superase RNase inhibitor (Life Technologies)(no NTPs). We then trigger elongation by adding P2 reaction buffer supplemented with rNTPs (and ITP) as indicated. The dynamic of P2 is not affected by the stalling reaction (**Fig. 6.12**). In the case of the Mn^{2+} titration experiments, no EDTA was added into the P2 reaction buffer for all experiments containing less than 2 mM Mn^{2+}.

6.8.7. BULK REPLICATION EXPERIMENTS

RNA replication reactions were carried out in 50 mM HEPES pH 7.5, 20 mM ammonium acetate, 6% (wt/vol) polyethylene glycol 4000, 5 mM $MgCl_2$, 2 mM $MnCl_2$, 0.1 mM EDTA, 0.1% Triton X-100 and 0.8 unit/μl RNasin Ribo Lock (Thermo Scientific Ltd.). Initiation reactions were carried out (125 nM ssRNA and 500 nM P2) in the presence of 1 mM rGTP and rATP (to allow the synthesis of the first 10 nts complementary to the 3'-end of the ssRNA template) at 30 °C for 20 min, the formed elongation complexes were purified using Sephacryl S-300 matrix (Pharmacia) and the reaction buffer with no NTPs for elution. The formation of new initiation complexes was prevented by heparin (5 mg/ml) and the incubation was continued at 30 °C for 5 min. The subsequent elongation reactions were supplemented with 0.2 mCi/ml of $[\alpha-^{32}\mathrm{P}]$ITP (6000 Ci/mmol; Cambro Scientific GmbH) and the indicated nucleotides, and the incubation at 30 °C was continued for 2h.

6.8.8. GENERATION OF EMPIRICAL DWELL-TIME DISTRIBUTIONS

When constructing empirical dwell-time distributions, dwell-times from different polymerases operating at identical conditions were grouped together. Looking at dwell-time distribution for individual polymerases (data not shown), it is clear that a small percentage consistently spend a much longer-than-average time pausing; to prevent these pause-prone polymerases from biasing the data, we remove the 5% with the highest and lowest dwell-time densities at 3 s.

6.8.9. FULL DESCRIPTION OF NTP CONDITIONS EMPLOYED IN P2 STUDIES

$[\mathrm{NTP}]_{\mathrm{opt}}$	1 mM rATP, 1 mM rGTP, 0.2 mM rCTP, 0.2 mM rUTP
$[\mathrm{NTP}]_{\mathrm{opt}}/5$	0.2 mM rATP, 0.2 mM rGTP, 0.04 mM rCTP, 0.04 mM rUTP
$[\mathrm{NTP}]_{\mathrm{opt}}/10$	0.1 mM rATP, 0.1 mM rGTP, 0.02 mM rCTP, 0.02 mM rUTP
$[\mathrm{NTP}]_{\mathrm{opt}}/50$	20 μM rATP, 20 μM rGTP, 4 μM rCTP, 4 μM rUTP
$[\mathrm{NTP}]_{\mathrm{opt}}/100$	10 μM rATP, 10 μM rGTP, 2 μM rCTP, 2 μM rUTP
$[\mathrm{NTP}]_{\mathrm{opt}}/500$	2 μM rATP, 2 μM rGTP, 0.4 μM rCTP, 0.4 μM rUTP

Table 6.1: rNTP concentrations used:

6.8.10. CONVERTING BEAD POSITION TO P2 POSITION

By recording the vertical position of the magnetic bead, we can observe the change in extension of our construct due to P2 converting dsRNA in the tether to ssRNA (**Fig. 6.2b**). To convert the position of the bead to the number of transcribed nucleotides $N_T(F)$, we use the linear-interpolation equation:

$$N_T(F) = N\frac{L(F) - L_{ds}(F)}{L_{ss}(F) - L_{ds}(F)} \tag{6.1}$$

For a given applied force F, N is the number of nucleotides in the template (2820 nt), $L(F)$ is the measured extension our construct, $L_{ds}(F)$ is the extension of a fully dsRNA tether, and $L_{ss}(F)$ is the length of a fully ssRNA tether. The functions $L_{ss}(F)$ and $L_{ds}(F)$ are extracted fromthe force-extension measurement of ssRNA and dsRNA shown in **Fig. 6.8c**. The assumption of linear conversion is validated by the force-extension measurement of tether with equal parts dsRNA and ssRNA, whose values agree well with the prediction of $[L_{ss}(F)+L_{ds}(F)]/2$ (**Fig. 6.8c**).

6.8.11. STOCHASTIC-PAUSING MODEL

There are many kinetic models that are consistent with the empirical dwell-time distributions we observe. We first try to identify the fraction of dwell-times p_n, for which the exit rate k_n from pause n is the rate-limiting step for crossing the dwell-time window. As the short pauses are broadly distributed and all far longer than the typical time between nucleotide additions, the dwell-time will be strongly dominated by the longest pause it contains; unfortunately, this washes out most details of the kinetic scheme that connects pauses and nucleotide addition. Due to this though, we can give the general form of the dwell-time distribution without specifying how the pauses are connected to the nucleotide addition pathway1 :

$$p_{dw}(t) = p_{na}\Gamma(t; N_{dw}, 1/k_{na}) + \sum_{n=1}^{N_{sp}} p_n k_n e^{-k_n t} + \frac{a_{bt}}{2t^{3/2}} \tag{6.2}$$

In the above expression (eqn.6.2), the first term contributes the portion p_{na} of dwell-times that originate in P2 crossing the dwell-time window of size N_{dw} base pairs without pausing; the second term is a sum of contributions originating in pause-dominated transitions, each contributing a fraction p_n of dwell-times; the third term captures the asymptotic power-law decay – with amplitude a_{bt} – of the probability of dwell-times dominated by a backtrack.

6.8.12. THE DIMENSIONALITY OF THE MODEL (CHOOSING N_{sp})

The number of short pauses N_{sp} justified by the data can be estimated using Bayes-Schwartz information criteria (BSIC)[28]. We consistently find that two or more short pauses are needed to describe the data (**Fig. 6.9b,c**). The difference in estimated exit rates between using a two and three exponentially distributed pauses falls within the error bars when using only two pauses. Therefore, we focus on the two pauses we can clearly detect at all conditions, and choose $N_{sp} = 2$.

6.8.13. MAXIMUM LIKELIHOOD ESTIMATION

The dwell-time distribution (eqn.6.2) is fit to the experimentally collected dwell-times $\{t_i\}_i$ by minimizing the likelihood function[53]:

$$L = -\sum_i \ln p_{dw}(t_i) \tag{6.3}$$

with respect to rates and probabilistic weights. We calculate the errors in our parameter estimates by bootstrapping [54] the system 200 times, and report the one-sigma confidence intervals among the bootstrapped data sets.

6.8.14. DOMINATING IN A DWELL-TIME WINDOW VS. DOMINATING IN ONE STEP

The fractions P_n represent the probability that a particular rate k_n dominates the dwell-time. We want to relate this to the probability p_n that a specific exit rate dominates within a one-nt transcription window. Assuming we have labelled the pauses so that $k_{n-1} > k_n$, we can relate the probability of having rate n dominating in N_{dw} steps to the probability of having it dominate in one step:

$$p_n = (\sum_{m=0}^{n} P_m)^{N_{dw}} - (\sum_{m=0}^{n-1} P_m)^{N_{dw}}, \qquad p_0 = p_{na} = p_{na}^{N_{dw}} = p_0^{N_{dw}} \tag{6.4}$$

The first term in eqn.6.4 represents the probability of having no pauses longer than the nth pause in the dwell-time window, and the second term represents the probability of having no pauses longer than the $n-1$th pause. The difference between the two terms is the probability that the nth pause will dominate. This can be inverted to yield:

$$P_n = (\sum_{m=0}^{n} p_m)^{1/N_{dw}} - (\sum_{m=0}^{n-1} p_m)^{1/N_{dw}}, \qquad P_0 = p_0^{1/N_{dw}} \tag{6.5}$$

In **Fig. 6.3c** and **Fig. 6.6c** of the main text we plot the P_n's (derived from the p_n's fitted from the data) as a function of the NTP concentration and of the applied force.

6.8.15. COMPARISON OF OUR ANALYSIS METHOD WITH STANDARD ANALYSIS METHOD

Crucially, our MLE based on dwell times circumvents some of the pitfalls associated with the more common approach of segmenting traces based on thresholding a smoothed polymerase velocity (**Fig. 6.10**): in such approaches many pauses might escape detection, falling below the smoothing window size and the velocity threshold. The distinct appearance of the nucleotide-addition peak (see e.g. **Fig. 6.2d**) and the fact that the fitted Gamma distribution (Supplemental Experimental Procedures) captures both its width and average (using a single free parameter, assuming that the polymerase takes 1 nt steps) shows that the present approach largely avoids such issues.

6.8.16. BUILDING A MODEL DESCRIBING THE P2 RDRP DATA

The construction of detailed kinetic models of nucleic-acid elongation has a long history [55], and much has been learned about transcription kinetics along the dominant reaction paths through analyzing quench-flow and stop-flow experiments[7, 56]. Single-molecule experiments have captured biologically relevant and rare events[25, 39], but limited statistics has so far hampered detailed comparisons to theoretical models. With our large single-molecule datasets we are able to study the dynamics of both rapid elongation and infrequent pause states. By considering how the dwell-time distribution changes under different experimental conditions we can appreciably constrain the set of models capable of describing the full data set.

In the next section we comment on alternative models including sequence-dependence, but conclude that introduction of such details is not warranted by the data. Our final kinetic model, **Fig. 6.7a** and **Fig. 6.13g**, offers a simple and economical explanation of our full data set, but does not comment on the veracity of more detailed theories focusing on the sub- steps of the nucleotide addition cycle.

6.8.17. ON- AND OFF-PATHWAY PAUSES

When considering transcription and replication, pauses are often conveniently grouped as either on- or off-pathway events, respectively, referring to whether the long-lived state is traversed as a step in the nucleotide addition cycle, or visited only as a side step (**Fig. 6.13a,c**). With this definition, the slow incorporation following a mismatched base at the termini of the nascent RNA strand[10], would constitute a simple on-pathway pause (**Fig. 6.13a**). A simple off-pathway pause would arise from a thermally activated and long-lived conformational change in the polymerase structure that renders it inactive (**Fig. 6.13c**), such as the ubiquitous pauses reported for bacterial RNAP[39, 40, 57]. To rationalize the pauses detected in our data (e.g. **Fig. 6.2c,d**), we here consider the possible NTP dependence of on- and off-pathway pauses.

On-pathway pause: If the selection of nucleotides occurs from a single state through competitive binding and differential catalysis, then the probability of incorporating an error (giving an on-pathway pause) will be independent of the overall nucleotide concentration (**Fig. 6.13a,b**). The exit rate from the same on-pathway pause increases with increasing NTP concentration, as by definition these pauses are escaped by the addition of nucleotides (**Fig. 6.13a,b**).

Off-pathway pause: The situation is very different for the off-pathway pause, as here the entrance into the pause occurs in competition with nucleotide-dependent elongation (**Fig. 6.13c**), and the pause probability should decrease with NTP concentration (**Fig. 6.13d**). If the chance of repeated pause entry and exit between nucleotide additions is negligible, the escape rate from off-pathway pauses is independent of NTP concentration (**Fig. 6.13c,d**). In our experiments, significant repeated entry into off-pathway pauses between nucleotide additions is negligible since the probability of pausing $(1-P_{\text{na}})$ is small (**Fig. 6.3c**), and hence the probability of entering the pause n times, with $n \geq 2, (1-P_{\text{na}})^n$, is negligible. However, it is possible that the pause can be entered infrequently overall, but entered repeatedly only on specific sequence motifs and so display a concentration-dependent effective escape rate. Such sequence dependence could result from structural changes triggered by the newly formed dsRNA within the RdRP (**Fig. 6.2b**), but this seems unlikely given that the shape of the RNA duplex is largely independent of sequence[58, 59]. Alternatively, the sequence of the duplex RNA immediately ahead of the RdRP (**Fig. 6.2b**) could influence the pausing process, but this also seems unlikely given that the pause probabilities remain largely unperturbed as we vary the base paring energies (with up to $2k_BT/\text{bp}$) at the junction by applying a force to the tether (**Fig. 6.6c**). It is worth noting that even if off-pathway pauses would be entered repeatedly between nucleotide incorporations, they would still only give rise to a single exponential shoulder in the dwell time distribution.

A mix of off- and on-pathway pauses: Including both on- and off- pathway pauses in the reaction scheme (**Fig. 6.13e**) does not introduce any new concentration dependence

(**Fig. 6.13f**). A straight mix of pauses is therefore not sufficient to explain the concentration dependence of our data (**Fig. 6.3b,c**).

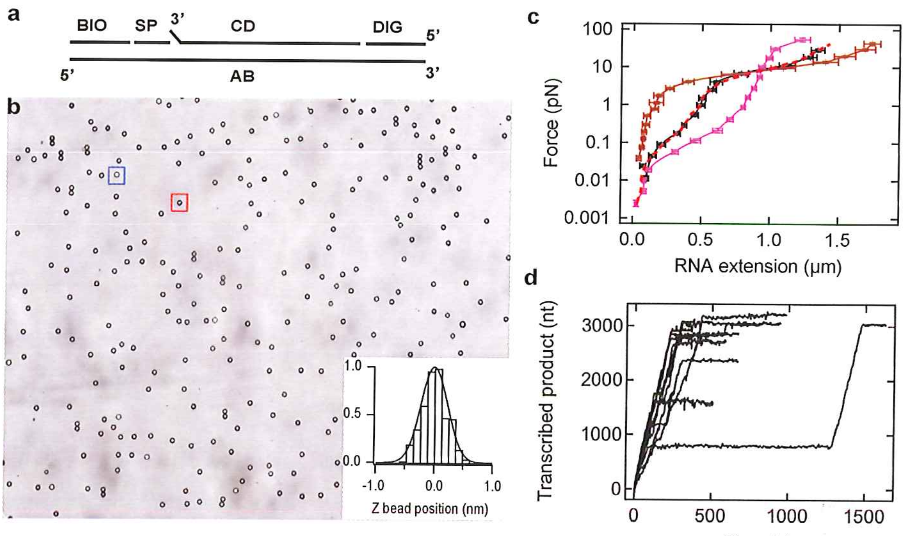
A novel type of pause: In order to capture the observed P2 activity (Fig. 2b, c) we therefore need to revise the assumption that the thermally activated off-pathway state must be completely enzymatically inactive (**Fig. 6.7a**). This gives us a new pause type. By this simple adjustment to the scheme in **Fig. 6.13e**, which yields the scheme of **Fig. 6.13g**, we obtain the concentration trends of **Fig. 6.3b,c**, as depicted in **Fig. 6.13h**. Note that the thermal transition that triggered this pause is assumed to be reversed upon nucleotide addition (see also Discussion in the main text).

6.9. SUPPLEMENTAL DATA

Fig. 6.8 shows the assembly of the RNA construct for single-molecule study of P2 transcription dynamics, related to **Figs. 6.2-6.6.** The RNA construct is built upon a 4.2 kb plus-strand (denoted AB, length of 4169 nucleotides) to which we hybridize a 2.8 kb template strand (denoted CD, length of 2820 nucleotides) for P2 transcription (Supplemental Experimental Procedures). This template strand is fully complementary to a 2.8 kb long section of the 4.2 kb plus-strand, with the exception of a 15 nt overhang that provides a free 3' end (sequence: CCCUUUUUUUUUUUU. The overhang facilitates template entry into the template channel of P2 during initiation (**Fig. 6.2b**) [18]. To tether the RNA construct to the bead, we further hybridize a shorter complementary strands (labelled BIO, and of length of 471 nucleotides) to one extremity of the 4.2 kb plus-strand. The strand labelled DIG (length of 538 nucleotides) is hybridized to the other extremity of the 4.2 kb plus-strand to tether the RNA construct to the surface. The strand labelled SP serves to provide a duplex spacer region between the site of P2 initiation and the magnetic bead following hybridization. All five separate single-strands of RNA [60] are synthesized via run-off in vitro transcription by T7 RNA polymerase (Promega) (Supplemental Experimental Procedures). Upon hybridization (Supplemental Experimental Procedures), these RNAs combine to yield a duplex RNA shown here that can be tethered in the magnetic tweezers (**Fig. 6.2b**). **(b)** Typical field of view in 50× configuration. The beads that appear to have a thick dark ring as their circumference are tethered 2.8 μm diameter magnetic beads (see e.g. the bead inside the red square). The other, brighter beads are 3 μm diameter reference beads that are fixed to the surface (see e.g. the bead inside the blue square). Inset: we monitor the z-positions of two reference beads (3 μm diameter, polystyrene) that are fixed to the bottom surface of the flow cell over the course of 134 s at 25 Hz and 50× magnification. Subsequently, we subtract their respective z-positions to eliminate mechanical drift in the instrument, which provides a spatial resolution of 1 nm (standard deviation). Following low-pass filtering at 0.5 Hz, we obtain a spatial resolution along the optical axis of 0.3 nm (standard deviation). Bins: experimental data; solid line: Gaussian fit to the filtered data with a σ of 0.3 nm. **(c)** Force versus extension of dsRNA, ssRNA, half dsRNA /half ssRNA constructs. Force-extension curves for RNA constructs with different length CD strand A (**Fig. 6.8a**). The pink curve represents force-extension curve measurements of molecules that have a full-length CD strand and are thus fully duplex (pink, $N = 6$). These dsRNA molecules are identical to the RNA construct that is present at the start of the reaction in the P2 experiments. The black curve represents force-extension curve measurements of molecules that have a half-length CD strand (1464 nucleotides) and thus consist of half ssRNA/half dsRNA in the relevant 2.8 kb region (black, $N = 8$). This reflects the mechanical situation that is present when P2 has transcribed half of its template. The brown curve represents force-extension curve measurements of molecules that lack a CD strand entirely and thus contain a 2787 nucleotide stretch of ssRNA (brown, $N = 6$). These molecules contain exactly the same number of ssRNA nucleotides as the RNA construct that is present at the end of the reaction in the P2 experiments. In all of these curves, the points correspond to experimental data and the solid lines correspond to a linear interpolation. The dashed red line corresponds to the linear combination of the interpolated data for dsRNA and ssRNA (Supplemental Experimental Procedures). We observe that there is excellent agreement between the experimental data on the half ssRNA/half ssRNA construct (black points) and the linear combination of the interpolated data for the ssRNA and dsRNA constructs. This indicates that we can calibrate the P2 experiments by employing the linear combination between ssRNA and dsRNA to convert the experimentally-measured tether length to the number of nucleotides transcribed. **(d)** Use of multi-bead tracking to simultaneously acquire multiple P2 activity traces at the single-molecule level. By using multi-bead tracking, we can simultaneously observe the activity of multiple P2. Shown here are 16 simultaneously acquired traces. In a small fraction of the traces, we are not able to follow P2 transcription until the end of the template was reached – a consequence of tether loss due to surface unbinding or RNA degradation. In all cases, traces are analyzed to the last observed nucleotide addition. The data shown here have been low-pass filtered at 0.5 Hz (sampling rate: 25 Hz). Data were taken at a nucleotide concentration equal to [NTP]opt (Supplemental Experimental Procedures) and an applied force of 20 pN. **(e)** Total number of dwell times for each experimental condition presented in the study. This represented for the force condition, NTPs concentration (as defined in Supplemental Experimental Procedures), and Mn^{2+} concentration. (*) represents the data set of **Fig. 6.12** for stalled and re-initiated P2 at 16 pN. As we use a dwell-time window of 10nt, the total number of nucleotide (nt) transcribed by P2 is 10 times the total number of dwell in any particular condition.

6.10. ACKNOWLEDGEMENTS

We thank Wiecher Kamping and Quinten Esajas for preliminary contributions to this work; Riitta Tarkiainen for technical assistance; Iwijn de Vlaminck for help with the experimental configuration; Jelmer Cnossen for assistance with software; Eric Snijder, Clara Posthuma, Jan Lipfert, and Craig Cameron for fruitful discussions; and Christophe Danelon, Clara Posthuma, and Jan Lipfert for critically reading the manuscript. This work was financially supported by the Academy of Finland grants 272507, 250113, and 256069 to M.M.P., grants 255342 and 256518 to D.H.B., and grants from the Sigrid Jusélius Foundation to D.H.B. and M.M.P. jointly, as well as by a TU Delft startup grant to M.D. and by a EURYI grant from the European Science Foundation and a TOP grant from the Netherlands Organisation for Scientific Research to N.H.D.

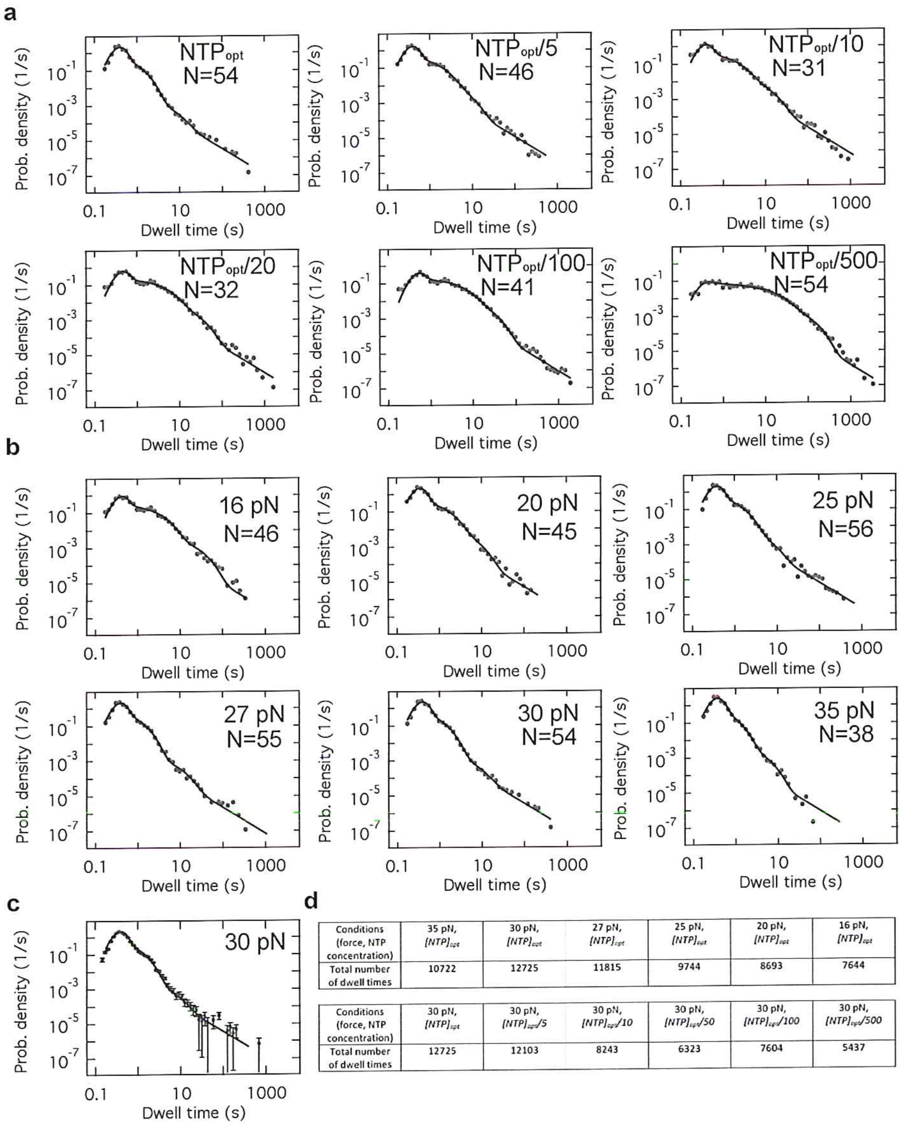


e Force [NTPs] [Mn^{2+}]	35 pN *$[NTP]_{opt}$* 2 mM Mn^{2+}	30 pN *$[NTP]_{opt}$* 2 mM Mn^{2+}	27 pN *$[NTP]_{opt}$* 2 mM Mn^{2+}	25 pN *$[NTP]_{opt}$* 2 mM Mn^{2+}	20 pN *$[NTP]_{opt}$* 2 mM Mn^{2+}	16 pN *$[NTP]_{opt}$* 2 mM Mn^{2+}
Total number of dwell times (nt = x 10)	**10722**	**12725**	**11815**	**9744**	**8693**	**7644**

Force [NTPs] [Mn^{2+}]	30 pN *$[NTP]_{opt}$* 2 mM Mn^{2+}	30 pN *$[NTP]_{opt}/5$* 2 mM Mn^{2+}	30 pN *$[NTP]_{opt}/10$* 2 mM Mn^{2+}	30 pN *$[NTP]_{opt}/50$* 2 mM Mn^{2+}	30 pN *$[NTP]_{opt}/100$* 2 mM Mn^{2+}	30 pN *$[NTP]_{opt}/500$* 2 mM Mn^{2+}
Total number of dwell times (nt = x 10)	**12725**	**12103**	**8243**	**6323**	**7604**	**5437**

Force [NTPs] [Mn^{2+}]	30 pN *$[NTP]_{opt}$* ,1 mM ITP 2 mM Mn^{2+}	30 pN *$[NTP]_{opt}$* 0 mM Mn^{2+}	30 pN *$[NTP]_{opt}$* 50 µM Mn^{2+}	30 pN *$[NTP]_{opt}$* 500 µM Mn^{2+}	(*) 16 pN *$[NTP]_{opt}$* 2 mM Mn^{2+}
Total number of dwell times (nt = x 10)	**7366**	**9957**	**6162**	**11264**	**8189**

Figure 6.8: **Aspects of the multiplexed MT P2 assay.** **(a)**The assembly of the RNA construct for single-molecule study of P2 transcription dynamics, related to **Figs. 6.2-6.6.** **(b)** A typical field of view in the 50× configuration used in our assay. **(c)**Force versus extension of dsRNA, ssRNA, half dsRNA /half ssRNA constructs. **(d)** Use of multi-bead tracking to simultaneously acquire multiple P2 activity traces at the single-molecule level. **(e)** The otal number of dwell times for each experimental condition presented in the study. See text for detailed explanation.



Conditions (force, NTP concentration)	35 pN, $[NTP]_{opt}$	30 pN, $[NTP]_{opt}$	27 pN, $[NTP]_{opt}$	25 pN, $[NTP]_{opt}$	20 pN, $[NTP]_{opt}$	16 pN, $[NTP]_{opt}$
Total number of dwell times	10722	12725	11815	9744	8693	7644

Conditions (force, NTP concentration)	30 pN, $[NTP]_{opt}$	30 pN, $[NTP]_{opt}/5$	30 pN, $[NTP]_{opt}/10$	30 pN, $[NTP]_{opt}/50$	30 pN, $[NTP]_{opt}/100$	30 pN, $[NTP]_{opt}/500$
Total number of dwell times	12725	12103	8243	6323	7604	5437

Figure 6.9: **(a) Dwell time distributions obtained at different [NTP]s**, related to **Figs. 6.3 and 6.5.** The full set of P2 transcription activity dwell-time distributions acquired as a function of NTP concentration at an applied force of 30 pN. The dwell-time distributions are fitted with eqn.6.2 (Supplemental Experimental Procedures). In a log-log representation, the power law appears as a straight line. As noted in the main text, at $[\text{NTP}]_{\text{opt}}/500$ the probability of traversing the 10 nt transcription window without entering a pause state is so low that the nucleotide addition peak nearly disappears. The nucleotide concentration and the number of traces per data set (*N*) are indicated in the top right corner of each panel. **(b)Dwell-time distributions obtained at different forces.** The full set of P2 transcription activity dwell-time distributions acquired as a function of force at [NTP]opt. The dwell-time distributions are fitted with eqn.6.2 (Supplemental Experimental Procedures). The applied force and the number of traces per data (*N*) set are indicated in the top right corner of each panel. (c) Dwell-time distribution obtained at 30 pN force and $[\text{NTP}]_{\text{opt}}$. The error bars represent 1 standard deviation and are generated from 10000 bootstraps.

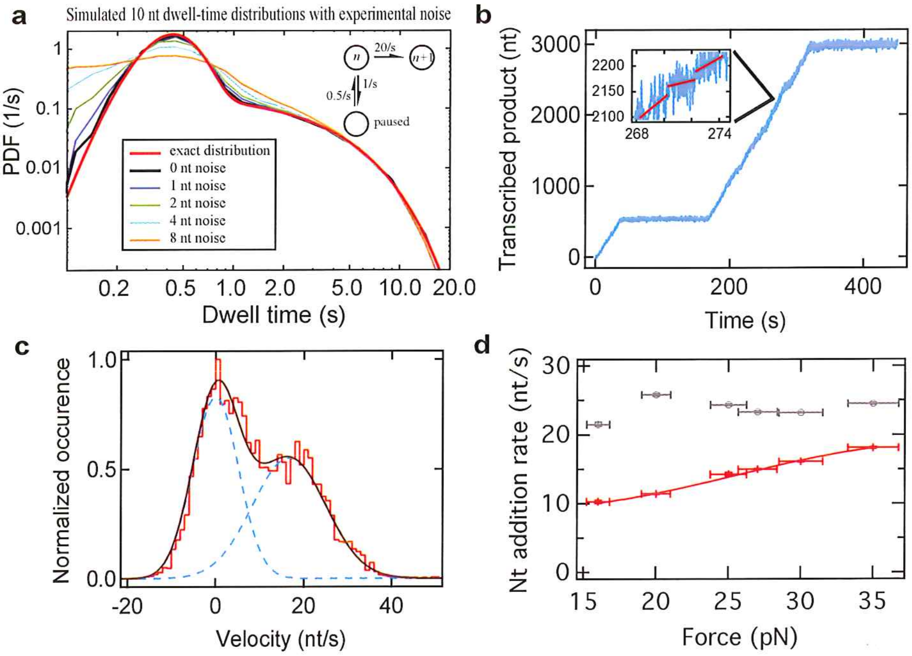


Figure 6.10: **a) The influence of experimental noise on dwell-time distributions, related to Figs. 6.2-6.6. (a)** To examine the influence of experimental noise on our empirical dwell-time distributions, we simulate traces for the simple model shown in the inset: elongating from product length n to $n+1$ with rate 20/s, pausing with rate 1/s, and escaping pauses with rate 0.5/s. Results from simulations with Gaussian noise added at 0 (black), 1 (blue), 2 (green), 4 (cyan), and 8 (orange) nt standard deviation and 25 Hz are shown. The simulation results should be compared to the exact dwell-time distribution shown in red. As expected, experimental noise predominantly affects short dwell times. Importantly, as soon as experimental noise starts influencing the results, the elongation peak is broadened, making it impossible to get a good fit of the elongation peak using a Gamma distribution with shape parameter 10. For a specific dwell time window, the width of the elongation peak can therefore be used to check that the influence of experimental noise is negligible. This should be contrasted to the standard method of extracting pause- free velocities, where the effect of noise is more obscure. **(b) Determination of pause-free velocities and comparison to the dwell times analysis.** A trace of P2 transcription activity (blue points) acquired at an applied force of 30 pN and $[\text{NTP}]_{\text{opt}}$. The data have not been low-pass filtered. The instantaneous velocity is measured by extracting the local slope of the trace within a 50 time-point interval (2 s) using a least mean squared linear fit (solid red lines). **(c)** For each trace, the velocities extracted in this manner are binned into a histogram with a bin size of 1 nt/s. The resulting distribution is fitted with a double Gaussian function (black line). The leftmost Gaussian is centered at 0 nt/s (pauses), the rightmost Gaussian is centered on the pause-free velocity value of 16.7 nt/s. **(d)** The velocity of P2 as a function of force as determined using pause-free analysis (red) and dwell time analysis (grey). The fitted pause-free velocities, averaged over several traces acquired under similar conditions, as a function of the applied force. We observe that the pause-free velocity acquired in this manner demonstrates significant force dependence. Note that this contrasts sharply with the nearly constant value for k_{na} that results from our overall analysis of the dwell-time distributions. Error bars represent the standard error of the mean for the velocity axis and the standard deviation (±5%, see [61]) of the applied force on the abscissa. The solid red curve is a fit according to the functional form of $v(f) = v_{\text{max}}/(1+Ae^{-\delta f/k_B T}$ [62], with $v_{\text{max}} = 20.7(18.9-24.5)$ nt/s, $A = 4.7(4.2-6.3)$, and $\delta = 3.8(2.7-4.7)$ nm. Here, the one-sigma confidence interval (reported within parentheses) is calculated over 200 bootstrapped data sets.

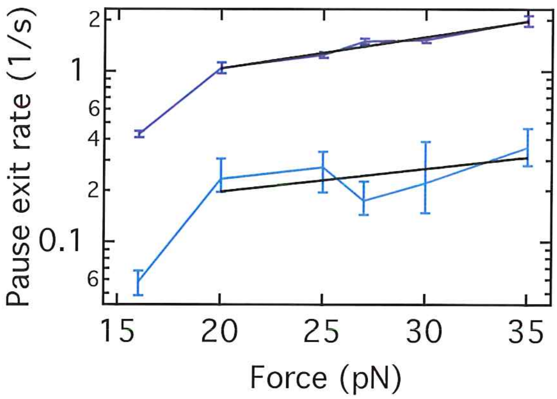


Figure 6.11: **Transition state fit of the pause lifetimes, related to Fig. 6.5.** To determine the distance to the transition state we use the form $k(F) = k_0 \exp(F\Delta x / k_B T)$. The fits (black lines) give $\Delta x = 0.17 \pm 0.02$ nm for Pause 1 (dark blue) and $\Delta x = 0.12 \pm 0.09$nm for Pause 2 (cyan). Note that the data points at 16 pN are excluded from the fits, as they appear to deviate from the trend exhibited by the data obtained at higher forces. We hypothesize that this change in behavior at 16 pN results from an influence of the longest-lived pauses (the backtracked pauses) on the fits of the shorter-lived Pause 1 and Pause 2 states.

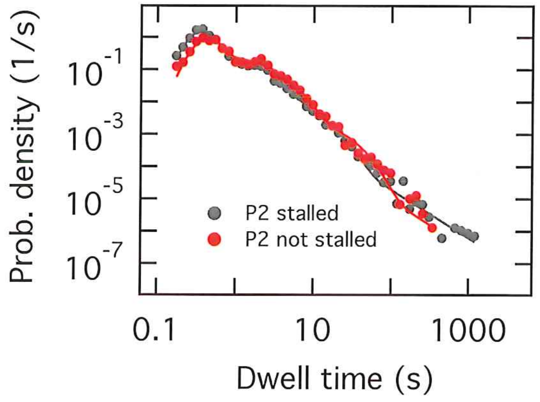


Figure 6.12: **P2 transcription behavior is unaffected by stalling, related to Experimental Procedures.** A dwell-time distribution of P2 transcription in the absence of stalling (red) and a dwell-time distribution of P2 transcription obtained when P2 was reactivated by the addition of a fourth NTP following stalling (grey). Both datasets were collected with an applied force of 16 pN and the nucleotide concentration $[\mathrm{NTP}]_{\mathrm{opt}}$.

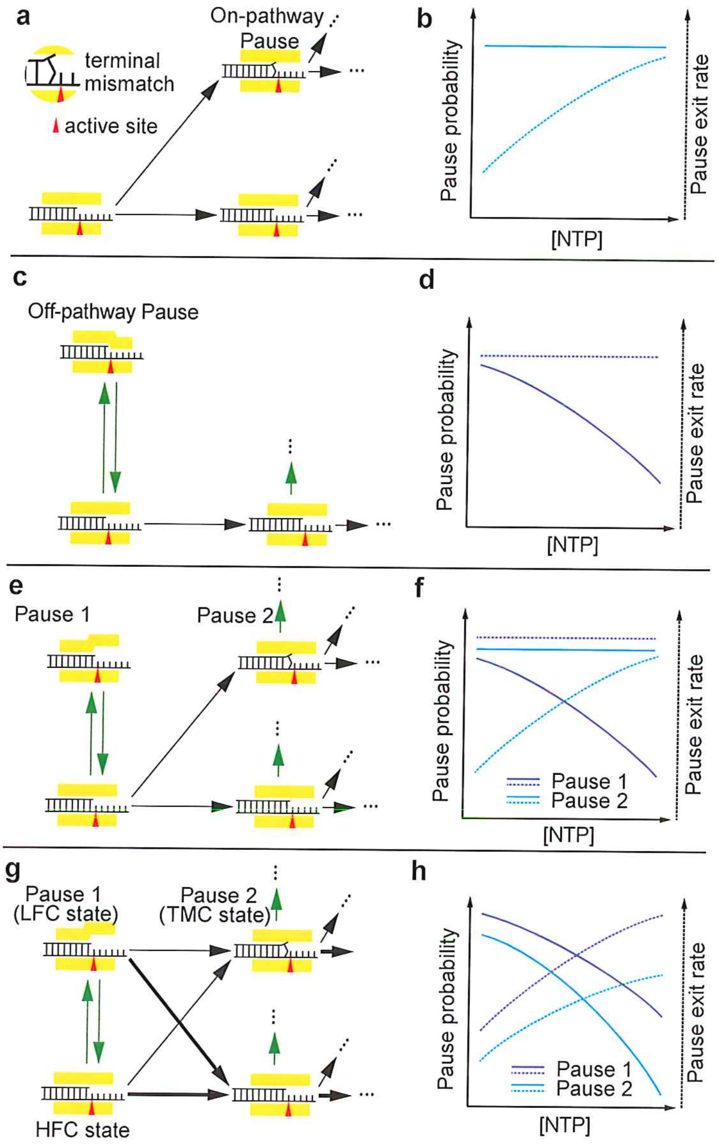


Figure 6.13: **Models for on- and off-pathway polymerase pausing, related to 6.7.** **a)** Schematic of an on-pathway pause capturing the slowdown in the nucleotide incorporation after a mismatch in the 3'-end of the nascent RNA molecule. **b)** The expected pause probabilities and pause exit rates as a function of NTP concentration for the pause in (**a**). **c)** Schematic of an off-pathway pause capturing the slowdown in nucleotide incorporation due to e.g. a structural rearrangement rendering the catalytic site inactive. **d)** The expected pause probabilities and pause exit rates as a function of NTP concentration for the pause in (**c**). **e)** A combination of on-pathway pauses and catalytically inactive off-pathway pauses. **f)** The expected pause probabilities and pause exit rates as a function of NTP concentration for the pauses in (**e**). **g)** A modification of the model depicted in (**e**) with the off-pathway pause displaying a low but non-zero catalytic activity. **h)** The expected pause probabilities and pause exit rates as a function of NTP concentration for the pauses in (**g**). In all panels, black arrows represent nucleotide dependent transitions, while green arrows represent transitions independent of nucleotides. In (**g**) the thick arrow out of each state represents the dominant path to leave that state, as deduced from our experimental observations (see main text); these exit rates can be inferred to obey Michaelis-Menten kinetics from **Fig. 6.3b**.

REFERENCES

[1] D. Dulin, T. J. Cui, J. Cnossen, M. W. Docter, J. Lipfert, and N. H. Dekker, *High spatiotemporal-resolution magnetic tweezers: Calibration and applications for dna dynamics,* Biophysical journal **109**, 2113 (2015).

[2] A. S. Lauring, J. Frydman, and R. Andino, *The role of mutational robustness in rna virus evolution,* Nat Rev Micro **11**, 327 (2013).

[3] M. Vignuzzi, J. K. Stone, J. J. Arnold, C. E. Cameron, and R. Andino, *Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population,* Nature **439**, 344 (2006).

[4] S. Crotty, D. Maag, J. J. Arnold, W. Zhong, J. Y. N. Lau, Z. Hong, R. Andino, and C. E. Cameron, *The broad-spectrum antiviral ribonucleoside ribavirin is an rna virus mutagen,* Nat Med **6**, 1375 (2000).

[5] M. Vignuzzi, J. K. Stone, and R. Andino, *Ribavirin and lethal mutagenesis of poliovirus: molecular mechanisms, resistance and biological implications,* Virus Research **107**, 173 (2005), virus Entry into Error Catastrophe as a New Antiviral Strategy.

[6] M. Vignuzzi, E. Wendt, and R. Andino, *Engineering attenuated virus vaccines by controlling replication fidelity,* Nat Med **14**, 154 (2008).

[7] X. Yang, E. Smidansky, K. Maksimchuk, D. Lum, J. Welch, J. Arnold, C. Cameron, and D. Boehr, *Motif d of viral* RNA*-dependent* RNA *polymerases determines efficiency and fidelity of nucleotide addition,* Structure **20**, 1519 (2012), cited By 19.

[8] Z. Jin, J. Deval, K. A. Johnson, and D. C. Swinney, *Characterization of the elongation complex of dengue virus rna polymerase: Assembly, kinetics of nucleotide incorporation, and fidelity,* Journal of Biological Chemistry **286**, 2067 (2011), http://www.jbc.org/content/286/3/2067.full.pdf+html .

[9] Z. Jin, V. Leveque, H. Ma, K. A. Johnson, and K. Klumpp, *Assembly, purification, and pre-steady-state kinetic analysis of active rna-dependent rna polymerase elongation complex,* Journal of Biological Chemistry **287**, 10674 (2012), http://www.jbc.org/content/287/13/10674.full.pdf+html .

[10] K. A. Johnson, *Role of induced fit in enzyme specificity: A molecular forward/reverse switch,* Journal of Biological Chemistry **283**, 26297 (2008), http://www.jbc.org/content/283/39/26297.full.pdf+html .

[11] D. Dulin, J. Lipfert, M. C. Moolman, and N. H. Dekker, *Studying genomic processes at the single-molecule level: introducing the tools and applications,* Nat Rev Genet **14**, 9 (2013).

[12] H. J. Geertsema and A. M. van Oijen, *A single-molecule view of* DNA *replication: the dynamic nature of multi-protein complexes revealed,* Current Opinion in Structural Biology **23**, 788 (2013), protein-carbohydrate interactions / Biophysical methods.

[13] M. H. Larson, R. Landick, and S. M. Block, *Single-molecule studies of rna polymerase: one singular sensation, every little step it takes,* Molecular cell **41**, 249 (2011).

[14] P. S. Salgado, E. V. Makeyev, S. J. Butcher, D. H. Bamford, D. I. Stuart, and J. M. Grimes, *The structural basis for* RNA *specificity and ca2+ inhibition of an* RNA-*dependent* RNA *polymerase,* Structure **12**, 307 (2004).

[15] H. A. M. Mönttinen, J. J. Ravantti, D. I. Stuart, and M. M. Poranen, *Automated structural comparisons clarify the phylogeny of the right-hand-shaped polymerases,* Molecular Biology and Evolution **31**, 2741 (2014), http://mbe.oxfordjournals.org/content/31/10/2741.full.pdf+html .

[16] K. K.-S. Ng, J. Arnold, and C. Cameron, *Structure-function relationships among rna-dependent rna polymerases,* in *RNA Interference,* Current Topics in Microbiology and Immunology, Vol. 320, edited by P. Paddison and P. Vogt (Springer Berlin Heidelberg, 2008) pp. 137–156.

[17] S. J. Butcher, J. M. Grimes, E. V. Makeyev, D. H. Bamford, and D. I. Stuart, *A mechanism for initiating rna-dependent rna polymerization,* Nature **410**, 235 (2001).

[18] E. V. Makeyev and J. M. Grimes, *Rna-dependent* RNA *polymerases of ds*RNA *bacteriophages,* Virus Research **101**, 45 (2004), viral molecular machines: Replication systems within the inner cores of dsRNA viruses.

[19] A. Acevedo, L. Brodsky, and R. Andino, *Mutational and fitness landscapes of an rna virus revealed through population sequencing,* Nature **505**, 686 (2014).

[20] E. V. Makeyev and D. H. Bamford, *The polymerase subunit of a dsrna virus plays a central role in the regulation of viral rna metabolism,* The EMBO Journal **19**, 6275 (2000).

[21] L. P. Sarin, M. M. Poranen, N. M. Lehti, J. J. Ravantti, M. R. L. Koivunen, A. P. Aalto, A. A. van Dijk, D. I. Stuart, J. M. Grimes, and D. H. Bamford, *Insights into the pre-initiation events of bacteriophage $\phi6$ rna-dependent rna polymerase: towards the assembly of a productive binary complex,* Nucleic Acids Research **37**, 1182 (2009), http://nar.oxfordjournals.org/content/37/4/1182.full.pdf+html .

[22] S. Wright, M. M. Poranen, D. H. Bamford, D. I. Stuart, and J. M. Grimes, *Noncatalytic ions direct the* RNA-*dependent* RNA *polymerase of bacterial double-stranded* RNA *virus $\phi6$ from de novo initiation to elongation,* Journal of Virology **86**, 2837 (2012), http://jvi.asm.org/content/86/5/2837.full.pdf+html .

[23] A. C. M. Cheung and P. Cramer, *Structural basis of rna polymerase ii backtracking, arrest and reactivation,* Nature **471**, 249 (2011).

[24] M. Depken, E. Galburt, and S. Grill, *The origin of short transcriptional pauses,* Biophysical Journal **96**, 2189 (2009), cited By 44.

[25] E. Galburt, S. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, and C. Bustamante, *Backtracking determines the force sensitivity of rnap ii in a factor-dependent manner,* Nature **446**, 820 (2007), cited By 130.

[26] J. Shaevitz, E. Abbondanzieri, R. Landick, and S. Block, *Backtracking by single rna polymerase molecules observed at near-base-pair resolution,* Nature **426**, 684 (2003), cited By 215.

[27] M. Voliotis, N. Cohen, C. Molina-París, and T. Liverpool, *Fluctuations, pauses, and backtracking in* DNA *transcription,* Biophysical Journal **94**, 334 (2008), cited By 57.

[28] G. Schwarz, *Estimating the Dimension of a Model,* The Annals of Statistics **6**, 461 (1978), 10.1214/aos/1176344136.

[29] E. V. Makeyev and D. H. Bamford, *Replicase activity of purified recombinant protein p2 of double-stranded rna bacteriophage* $\phi 6$, The EMBO Journal **19**, 124 (2000).

[30] I. D. Vilfan, A. Candelli, S. Hage, A. P. Aalto, M. M. Poranen, D. H. Bamford, and N. H. Dekker, *Reinitiated viral rna-dependent rna polymerase resumes replication at a reduced rate,* Nucleic Acids Research **36**, 7059 (2008), http://nar.oxfordjournals.org/content/36/22/7059.full.pdf+html .

[31] M. J. Thomas, A. A. Platas, and D. K. Hawley, *Transcriptional fidelity and proofreading by {RNA} polymerase {II},* Cell **93**, 627 (1998).

[32] T. A. Steitz, *Dna polymerases: Structural diversity and common mechanisms,* Journal of Biological Chemistry **274**, 17395 (1999), http://www.jbc.org/content/274/25/17395.full.pdf+html .

[33] J. J. Arnold, , and C. E. Cameron, *Poliovirus rna-dependent rna polymerase (3dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of mg2+†,* Biochemistry **43**, 5126 (2004), pMID: 15122878, http://dx.doi.org/10.1021/bi035212y .

[34] R. A. Beckman, A. S. Mildvan, and L. A. Loeb, *On the fidelity of dna replication: manganese mutagenesis in vitro,* Biochemistry **24**, 5810 (1985), http://dx.doi.org/10.1021/bi00342a019 .

[35] M. F. Goodman, S. Keener, S. Guidotti, and E. W. Branscomb, *On the enzymatic basis for mutagenesis by manganese.* Journal of Biological Chemistry **258**, 3469 (1983), http://www.jbc.org/content/258/6/3469.full.pdf+html .

[36] D. Colquhoun and A. Hawkes, *The principles of the stochastic interpretation of ion-channel mechanisms,* in *Single-Channel Recording,* edited by B. Sakmann and E. Neher (Springer US, 1995) pp. 397–482.

[37] X. Xie and H. Lu, *Single-molecule enzymology,* in *Single Molecule Spectroscopy,* Springer Series in Chemical Physics, Vol. 67 (Springer Berlin Heidelberg, 2001) pp. 227–240.

[38] S. J. Greive and P. H. von Hippel, *Thinking quantitatively about transcriptional regulation,* Nat Rev Mol Cell Biol **6**, 221 (2005).

[39] K. Neuman, E. Abbondanzieri, R. Landick, J. Gelles, and S. Block, *Ubiquitous transcriptional pausing is independent of* RNA *polymerase backtracking,* Cell **115**, 437 (2003), cited By 185.

[40] L. Donehower, F. Wong-Staal, and D. Gillespie, *Divergence of baboon endogenous type c virogenes in primates: genomic viral rna in molecular hybridization experiments.* Journal of Virology **21**, 932 (1977), http://jvi.asm.org/content/21/3/932.full.pdf+html .

[41] A. Brenlla, R. P. Markiewicz, D. Rueda, and L. J. Romano, *Nucleotide selection by the y-family dna polymerase dpo4 involves template translocation and misalignment,* Nucleic Acids Research **42**, 2555 (2014), http://nar.oxfordjournals.org/content/42/4/2555.full.pdf+html .

[42] F. H. Martin, M. M. Castro, F. Aboul-ela, and I. Tinoco, *Base pairing involving deoxyinosine: implications for probe design,* Nucleic Acids Research **13**, 8927 (1985), http://nar.oxfordjournals.org/content/13/24/8927.full.pdf+html .

[43] H. Matsuzaki, G. A. Kassavetis, and E. Geiduschek, *Analysis of {RNA} chain elongation and termination by saccharomyces cerevisiae {RNA} polymerase {III},* Journal of Molecular Biology **235**, 1173 (1994).

[44] S. Crotty, C. E. Cameron, and R. Andino, *Rna virus error catastrophe: Direct molecular test by using ribavirin,* Proceedings of the National Academy of Sciences **98**, 6895 (2001), http://www.pnas.org/content/98/12/6895.full.pdf .

[45] J. P. Cnossen, D. Dulin, and N. H. Dekker, *An optimized software framework for real-time, high-throughput tracking of spherical beads,* Review of Scientific Instruments **85**, 103712 (2014).

[46] I. De Vlaminck, T. Henighan, M. Van Loenhout, I. Pfeiffer, J. Huijts, J. Kerssemakers, A. Katan, A. Van Langen-Suurling, E. Van Der Drift, C. Wyman, and C. Dekker, *Highly parallel magnetic tweezers by targeted* DNA *tethering,* Nano Letters **11**, 5489 (2011), cited By 25.

[47] J. Lipfert, J. J. W. Kerssemakers, M. Rojer, and N. H. Dekker, *A method to track rotational motion for use in single-molecule biophysics,* Review of Scientific Instruments **82**, 103707 (2011).

[48] M. T. vanÂ Loenhout, J. W. Kerssemakers, I. D. Vlaminck, and C. Dekker, *Non-bias-limited tracking of spherical particles, enabling nanometer resolution at low magnification,* Biophysical Journal **102**, 2362 (2012).

[49] B. Maier, D. Bensimon, and V. Croquette, *Replication by a single dna polymerase of a stretched single-stranded dna,* Proceedings of the National Academy of Sciences **97**, 12002 (2000), http://www.pnas.org/content/97/22/12002.full.pdf .

[50] M. M. Poranen, P. S. Salgado, M. R. L. Koivunen, S. Wright, D. H. Bamford, D. I. Stuart, and J. M. Grimes, *Structural explanation for the role of mn2+ in the activity of ϕ6 rna-dependent rna polymerase,* Nucleic Acids Research **36**, 6633 (2008), http://nar.oxfordjournals.org/content/36/20/6633.full.pdf+html .

[51] F. Studier and B. A. Moffatt, *Use of bacteriophage* T7 RNA *polymerase to direct selective high-level expression of cloned genes,* Journal of Molecular Biology **189**, 113 (1986).

[52] M. M. Poranen, M. R. L. Koivunen, and D. H. Bamford, *Nontemplated terminal nucleotidyltransferase activity of double-stranded rna bacteriophage ϕ6 rna-dependent rna polymerase,* Journal of Virology **82**, 9254 (2008), http://jvi.asm.org/content/82/18/9254.full.pdf+html .

[53] S. Cowan, *Statistical Data Analysis* (Oxford University Press, 1998).

[54] W. Press, B. Flannery, S. Teukolsky, and W. T. Vetterling, *NUMERICAL RECIPES IN C: THE ART OF SCIENTIFIC COMPUTING* (Cambridge University Press, 1992).

[55] L. M. Okada and D. G. Gillespie, *The impact of family planning programs on unplanned pregnancies,* Family planning perspectives **9**, 173 (1977).

[56] Y.-C. Tsai and K. Johnson, *A new paradigm for* DNA *polymerase specificity,* Biochemistry **45**, 9675 (2006), cited By 145.

[57] K. M. Herbert, A. L. Porta, B. J. Wong, R. A. Mooney, K. C. Neuman, R. Landick, and S. M. Block, *Sequence-resolved detection of pausing by single {RNA} polymerase molecules,* Cell **125**, 1083 (2006).

[58] F. U. Gast and P. J. Hagerman, *Electrophoretic and hydrodynamic properties of duplex ribonucleic acid molecules transcribed in vitro: evidence that a-tracts do not generate curvature in rna,* Biochemistry **30**, 4268 (1991), http://dx.doi.org/10.1021/bi00231a024 .

[59] Y. H. Wang, M. T. Howard, and J. D. Griffith, *Phased adenine tracts in double-stranded rna do not induce sequence-directed bending,* Biochemistry **30**, 5443 (1991), http://dx.doi.org/10.1021/bi00236a017 .

[60] I. D. Vilfan, W. Kamping, M. van den Hout, A. Candelli, S. Hage, and N. H. Dekker, *An rna toolbox for single-molecule force spectroscopy studies,* Nucleic Acids Research **35**, 6625 (2007), http://nar.oxfordjournals.org/content/35/19/6625.full.pdf+html .

[61] J. Lipfert, X. Hao, and N. Dekker, *Quantitative modeling and optimization of magnetic tweezers,* Biophysical Journal **96**, 5040 (2009), cited By 43.

[62] E. Abbondanzieri, W. Greenleaf, J. Shaevitz, R. Landick, and S. Block, *Direct observation of base-pair stepping by rna polymerase,* Nature **438**, 460 (2005), cited By 423.

7

BACKTRACKING BEHAVIOR IN VIRAL RNA-DEPENDENT RNA POLYMERASE PROVIDES THE BASIS FOR A SECOND INITIATION SITE

Transcription in RNA viruses is highly dynamic, with a variety of pauses interrupting nucleotide addition by RNA-dependent RNA polymerase (RdRp). For example, rare but lengthy pauses (>20 s) have been linked to backtracking for viral single-subunit RdRps. However, while such backtracking has been well characterized for multi-subunit RNA polymerases (RNAPs) from bacteria and yeast, little is known about the details of viral RdRp backtracking and its biological roles. Using high-throughput magnetic tweezers, we quantify the backtracking by RdRp from the double-stranded (ds) RNA bacteriophage $\phi6$, a model system for RdRps. We characterize the probability of entering long backtracks as a function of force and propose a model in which the bias toward backtracking is determined by the base paring at the dsRNA fork. We further discover that extensive backtracking provides access to a new 3'-end that allows for the de novo initiation of a second RdRp. This previously unidentified behavior provides a new mechanism for rapid RNA synthesis using coupled RdRps and hints at a possible regulatory pathway for gene expression during viral RNA transcription.

This chapter has been published as: David Dulin, Igor D. Vilfan, Bojk A. Berghuis, Minna M. Poranen, Martin Depken & Nynke H. Dekker. Backtracking behavior in viral RNA-dependent RNA polymerase provides the basis for a second initiation site *Nucl Acids Res* (2015). [1].

7.1. INTRODUCTION

ORGANISMS in all domains of life share a need to replicate and transcribe their genome.[2] Replication and transcription are highly dynamic processes in which rapid nucleotide incorporation is frequently interrupted by a variety of pauses that differ in nature depending on the molecular machinery involved[3–5]. For example, the T7 replisome pauses during primer synthesis[6], Escherichia coli RNAP pauses in a sequence-dependent manner during elongation[7, 8], and the incorporation of incorrect nucleotides causes the dsRNA bacteriophage ϕ6 RdRp to pause[9]. In both eukaryotic and bacterial multi-subunit RNAPs, long pauses (>10 s) have been assigned to backtracked pauses, where the RNAP diffuses upstream on the DNA template[10, 11]. During a backtracked pause, the nascent RNA and the template are held in register, and the nascent 3'-end exits the core RNAP through the NTP channel. Backtracking ends either through spontaneous realignment of the 3'-end and the catalytic site, or through the creation of a new 3'-end at the catalytic site via protein-dependent cleavage[10–13]. As the backtracked state is a stochastic and rare event, it is particularly well suited to study by single-molecule approaches that avoid the inherent averaging of bulk experiments. Though backtracking and its attendant dynamics have been studied in bacteria and eukaryotes using single-molecule force spectroscopy[10, 11], the phenomenon has only recently been observed for an RdRp of viral origin[9].

RdRps are responsible for both replication and transcription in RNA viruses, which remain at the forefront of many human and animal epidemics. The genomes of RNA viruses may be single-stranded with either sense ((+)RNA) (e.g. poliovirus, hepatitis C virus) or anti-sense ((-)RNA) (e.g. influenza virus, ebola virus), or they can be double-stranded (e.g. rotavirus, bacteriophage ϕ6). Due to the high error rate of their RdRps[14], RNA viruses evolve very rapidly[15, 16] making it challenging to design effective vaccines. RdRps also play a central role in viral RNA recombination—a separate pathway for viral evolution[17]. RdRps of RNA viruses share a conserved catalytic site that is also found among the reverse transcriptases (e.g. human immunodeficiency virus, HIV)[18–20]. Given the relative inefficiency of vaccines, RdRps have become direct targets for drug development. For example, ribavirin, a mutagen nucleotide analogue that is incorporated by RdRp, has been introduced as therapy for hepatitis C and poliovirus infections[21].

We present a study of P2, the RdRp of the bacteriophage ϕ6[22] and a general model for RNA virus RdRps. P2 initiates transcription and replication de novo on a free 3' RNA end[23–26]. Following initiation, P2 enters the RNA elongation phase, which we have recently characterized at the single-molecule level using a highly parallelized magnetic tweezers assay. This approach, which is particularly powerful in studying rare events, has revealed evidence for a previously unknown catalytically competent and error-prone pathway[9]. Presently, we utilize the power of this approach to study RdRp backtracking. We demonstrate that the probability for the polymerase to enter long backtracks state is force dependent, and we provide a mechanochemical model explaining this force dependence in terms of re-hybridization of the RNA strands at the fork. Interestingly, following long backtracks we occasionally observe a stable and processive reversal of the direction of transcription. We propose that the 3' RNA that protrudes as a result of backtracking provides a new initiation site for a second P2. Following successful initiation, such a second P2 displaces the first P2 in the upstream direction along its template. Strikingly, at low tether forces the apparent nucleotide addition rate of the second P2 RdRp is twice faster than that of the first, highlighting a possible pathway for rapid RNA synthesis during viral transcription. Given the structural similarity amongst different viral RdRps[19], it is likely that such a mechanism would be conserved amongst the viral RdRps that rely on *de novo* initiation (e.g. hepatitis C virus)[27].

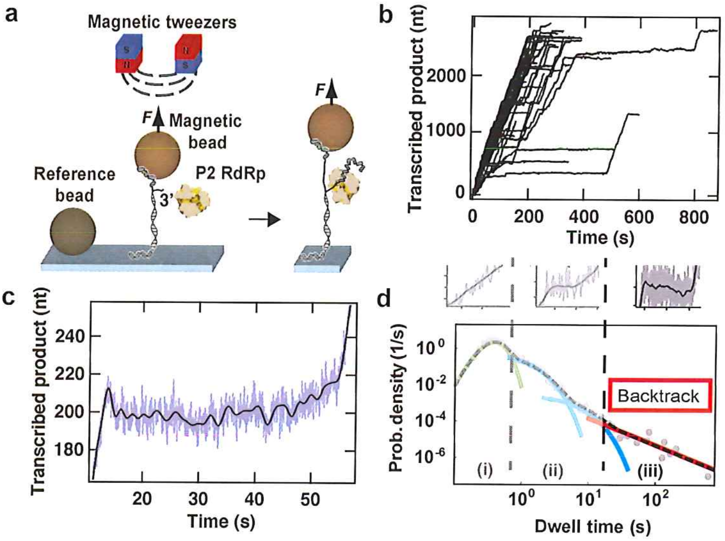


Figure 7.1: **Experimental configuration of the magnetic tweezers assay and identification of RdRp backtracking. (a)** Schematic representation of the experimental assay (not to scale). A predominantly dsRNA construct, attached to the surface by one end and to a micron-sized magnetic bead by the other end, experiences a constant force applied to a magnetic bead (light brown sphere) via a pair of permanent magnets (red and blue cubes) placed above the flow cell (not shown). One of the strands of the construct has a free 3' end for initiation by a single P2 RdRp. A reference bead (dark brown sphere) is non-specifically adsorbed to the surface to correct for the influence of mechanical drift. During elongation, P2 RdRp synthesizes a complementary copy to the template strand, and in doing so exposes a ssRNA strand between the flow cell surface and the magnetic bead. The resulting change in extension is monitored and converted into a number of incorporated nucleotides. **(b)** Transcription activity by P2 RdRp versus time monitored using parallelized detection, yielding 52 traces in a single experiment. The experimental conditions include an applied force of 30 pN, an acquisition frequency of 25 Hz (low-pass filtered at 0.5 Hz) and 1 mM ATP/GTP and 0.2 mM CTP/UTP. **(c)** A sample trace illustrating backtracking behavior by P2 RdRp. A decrease in the length of the transcribed product by ≈15 nts is observed at ≈15 s. Subsequently, P2 RdRp pauses for ≈40 s before resuming elongation. For this trace, the applied force is 35 pN (Materials and Methods). The raw data (blue) are acquired at 25 Hz and filtered at 0.5 Hz (black line). **(d)** A dwell-time distribution extracted from 52 traces of P2 transcription activity acquired at 20 pN and $[\text{NTP}]_{\text{opt}}$ (gray dots). We fit this distribution to a stochastic-pausing model by using MLE (dashed black line)[9]. For clarity, we individually plot each contribution to the dwell-time distribution: in green, a Gamma distribution capturing the elongation peak; in blue, the first short exponential pause (Pause 1) and the second short exponential pause (Pause 2); and in red, the power law distribution of pause times originating from backtracking. Above the dwell-time distribution, representative P2 activity events are plotted: from left to right, fast incorporation without pause (light gray), short pauses (light gray) and long backtracked pauses (dark gray). The fit parameters for the Gamma distribution and the exponential distributions extracted from the MLE are available in[9].

7.2. RESULTS

7.2.1. OBSERVATION OF ϕ6 P2 RDRP ELONGATION KINETICS THROUGH MULTIPLEXING

To observe the activity of ϕ6 P2 RdRp during the elongation phase, we use our previously described assay[9]. Briefly, a dsRNA construct is tethered to a magnetic bead and kept at a constant tension applied through a pair of permanent magnets positioned above the flow chamber. The dsRNA construct has one strand partially non-hybridized at its 3'-end, hence providing a template for P2 initiation. Once a P2 RdRp has initiated, it displaces the template strand from the complementary strand that tethers the bead to the surface (**Fig. 7.1a**). Because of the length difference between ssRNA and dsRNA, we can monitor the motion of each elongating P2 with a resolution ≈5 nt at 0.5 Hz[9]. The massive parallelization of our instrument and computational approach provides the possibility of real-time acquisition of up to 52 polymerase activity traces simultaneously (**Fig. 7.1b**)[28].

When the RdRp backtracks, the template strand rehybridizes to the tether strand, resulting in a shortening of the construct, hence in an apparent backward motion of the polymerase (**Fig. 7.1c**). When we provide an applied force as high as 30–35 pN, the spatiotemporal resolution of our apparatus is sufficient to observe backward motion of 10 nt or longer. In **Fig. 7.1c**, we show how pauses of ≈40 s follow an initial backward motion of ≈10 bases. Due to a decreased spatiotemporal resolution at lower forces, we cannot directly observe the backward motion of P2 for most of the long pauses (>20 s), but instead infer their existence by their duration, knowing that the shorter pauses (1–10 s) derive from a different origin[9].

We performed a dwell-time analysis for each experimental condition by measuring the time it takes the polymerase to transcribe through consecutive windows of 10 nt along the trace[9]. The obtained dwell-time distribution contains the features of different probability distribution functions (**Fig. 7.1d**, gray dots). We determined three different trends in the dwell-time distribution: a short-time gamma distributed peak ((i) in **Fig. 7.1d**) originating from the P2 polymerase incorporating 10 successive nt without a pause, an exponential shoulder situated at the intermediate times ((ii) in **Fig. 7.1d**), and a long-lived pause with a broad distribution ((iii) in **Fig. 7.1d**) consistent with an algebraic decay generated by backtracking pauses[9, 11, 29, 30]. The parameters of each distribution can be extracted via Maximum Likelihood Estimation (MLE) (black dashed line in **Fig. 7.1d**)[9].

7.2.2. THE NATURE OF THE P2 RDRP BACKTRACKED STATE IS FORCE DEPENDENT

To examine the force dependence of backtracking, we acquire data at forces of 16 pN, 20 pN, 25 pN, 30 pN and 35 pN (blue, cyan, green, yellow and red respectively, **Fig. 7.2a**) using the buffer conditions described in **Materials and Methods**. In this range, the applied force destabilizes the ds-ssRNA junction by 0.6-2 $k_B T$ [9]. We perform a dwell-time analysis for each force (**Fig. 7.2a,b**) and find that the recorded dwell-times span almost five orders of magnitude in time. Each distribution contains between ≈12000 (35 pN) and ≈30000 (20 pN) dwell-times. From **Fig. 7.2b**, we observe that the long-lived pauses representing the backtracked state (**Fig. 7.1d**) are populated to a degree that depends on the applied force: higher force provides lower probability density distribution of the dwell-times. We observe a force dependent hierarchy from 16 pN to 35 pN, with the error bars (one standard deviation confidence interval extracted from 1000 bootstrapped data sets) of the different distributions being well separated. The probability density distribution of the dwell-times does not depend on the presence of P2 RdRp in the reaction buffer, as shown in a previous study (Figure S5 of ([9])). In the same study, we also observed that the lifetime of the exponentially distributed pauses (**Fig. 7.1d**) does not exceed 17 s at $[\text{NTP}]_{\text{opt}}$[9], and to avoid the influence of shorter non-backtracked pauses (region (ii) in **Fig 7.1d**), we similarly score only pauses longer than 20 s as in[9]. In this way, we measure the probability of P2 to pause longer than 20 s within 10 nt incorporation for each force (**Fig. 7.2c**). We observe a strong force dependence, where for example the probability of being in a pause longer than 20 s during a 10 nt incorporation cycle is 0.0228 ± 0.0012 at 16 pN and decreases nearly 20-fold to 0.0013 ± 0.0003 at 35 pN (**Fig. 7.2c**).

7.2.3. THE P2 RDRP BACKTRACKED STATE IS INFLUENCED BY TEMPLATE STRAND REHYBRIDIZATION

The force dependence of the backtracking behavior of P2 RdRp can be understood in terms of the model illustrated in **Fig. 7.2d**. In this model, thermally induced fraying of the duplex region formed between the newly synthesized and the template strand will liberate the 3'-end of the nascent RNA strand. Further exposure of the nascent RNA strand, and concomitant backtracking by P2, is facilitated by the rehybridization of the downstream template strand with the 4.2 kb (+) strand. Tension in the template strand lowers the energetic gain of hybridization with the plus strand, and thus the rehybridization process is favored at lower forces[31] (**Fig. 7.2d**). Provided that the exposed 3'-end can be accommodated (the NTP channel being a likely candidate, see e.g. PDB 1UVI[32], backward diffusion of P2 RdRp along the template strand will be favored at lower forces, consistent with the observed increased probability for long backtracked pauses (**Fig. 7.2c**). As an applied force changes the energy gain for hybridization between the tether and plus strand, our model suggests a possible sequence dependence of the backtracked state as determined by base-pairing energies.

7.2.4. P2 RDRP REVERSES DIRECTION DURING TRANSCRIPTION

In our single-molecule experiments, we occasionally observe events in which P2 appears to 'reverse' its direction during transcription (**Fig. 7.3a**). Prior to such events, P2 transcript elongation starts normally (**Fig. 7.3a**, data up to ≈600 s). However, following a long pause, we occasionally observe a decrease in the extension of the RNA tether (**Fig. 7.3a**, black arrow). This decrease typically continues until the tether length of the RNA construct reverts to its original value measured prior to P2 initiation. Such 'reversal' events occur in a small percentage of all traces. The occurrence of reversals in a trace is, however, significantly more likely at applied forces below 20 pN (**Fig. 7.3b**), with a probability of 0.15 ± 0.07 and 0.027 ± 0.026 at applied forces of 16 pN and 35 pN, respectively. In very

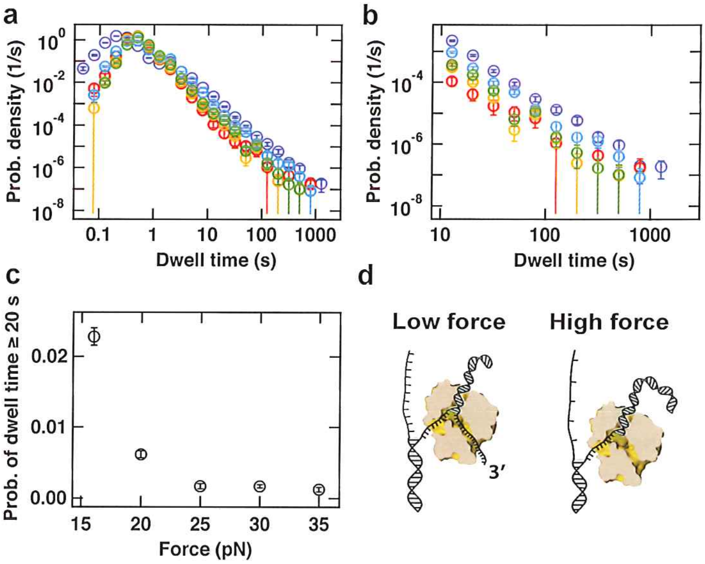


Figure 7.2: **Force dependence of the probability of entering into a backtracked state for P2 RdRp.** **(a)** Probability density distributions for P2 RdRp transcriptional activity acquired at 16 pN (dark blue, 102 traces), 20 pN (light blue, 184 traces), 25 pN (green, 200 traces), 30 pN (yellow, 210 traces) and 35 pN (red, 76 traces). The error bars correspond to one standard deviation estimated from 1000 bootstraps. **(b)** A zoom-in of **(a)** for dwell-times longer than 10 s. **(c)** Probability that a dwell-time exceeds 20 s as a function of the applied force. The error bars are the standard deviation of the distribution extracted from 1000 bootstraps. **(d)** Proposed model that accounts for the force-dependence of the probability of finding P2 RdRp in a backtracked state. At low force (small distance between single nucleotides on the non-template strand), the tension at the dsRNA fork is small enough to allow the rehybridization of the template strand to the non-template strand, compensating the melting of the dsRNA product, leading to the backtracking of the RdRp. At high force (which results in a larger distance between single nucleotides on the non-template strand), the increase in tension impairs the rebridization of the template to the non-template strand, preventing backtracking to happen.

rare instances (0.2% of P2 traces collected), we also observe a renewed ‘forward’ motion upon the completion of a reversal (**Fig. 7.3c,d**).

To understand the origin of the reversals, we perform two additional experiments. First, we test whether the reversal events could be attributed to polymerization by P2 on the single-stranded regions exposed on the RNA tether as transcription progresses. We test this by monitoring the extension of an RNA construct that lacks the template strand (**Fig. 7.1a**) in the presence of P2 at 16 pN applied force. From the observation of 46 tethers, no active polymerization events are detected. Second, we test whether the reversals require the presence of free P2 in the flow cell buffer. To this end, we stall P2 on the template strand by depriving it of one NTP, rinse the flow cell and flush in the reaction buffer containing all four NTPs but devoid of free P2[9]. Under these conditions, 50 data traces displayed continued transcription by P2, yet not one of these contained a reversal, further demonstrating the need of at least a second RdRp for reversal to occur. As the probability of observing a reversal event at this force (16 pN) in the presence of free P2 is 0.15 (**Fig.**

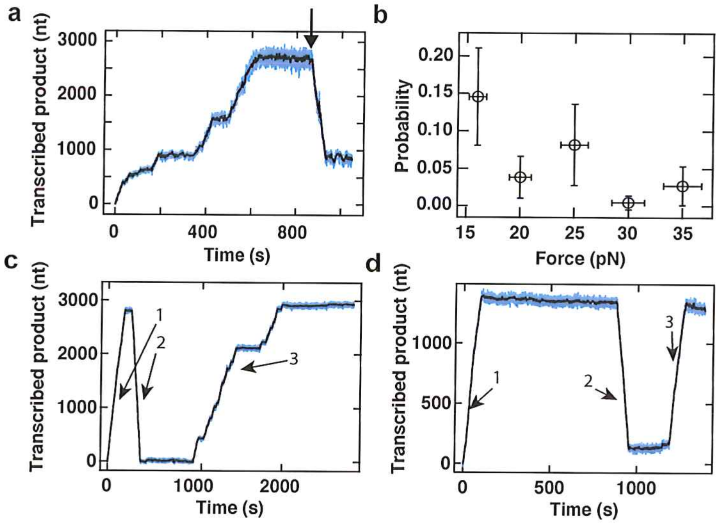


Figure 7.3: **Long pauses in the elongation dynamics are rarely followed by processive changes in the direction of transcription.** **(a)** The extension of the transcribed product as a function of time. At 800 s (indicated by arrow), P2 RdRp exhibits a reversal behavior. The experimental conditions include an applied force of 16 pN force and an acquisition frequency of 25 Hz. **(b)** The probability of observing a reversal event as a function of the applied force (see panel **(a)**). The error bars represent the 95% confidence interval for a binomial distribution for the ordinate and the standard deviation of the applied force (±5%, see ([33])) for the abscissa. **(c and d)** P2 RdRp transcription activity traces that exhibit very rare behavior in which multiple switches in the apparent directionality of the signal can be observed. In both panels, the extension of the RNA construct (1) increases, then (2) decreases (as in a reversal event) and (3) increases again. Data are acquired at an applied force of 25 pN using buffer conditions described in Materials and Methods. For panels **(a)**, **(c)** and **(d)**, raw data are shown in blue and data low-pass filtered at 0.5 Hz are shown in black.

7.3b), the probability of not observing a single reversal in 50 traces can be estimated to be less than 0.0003.

Lastly, reversals appear to be associated with a higher rate of nucleotide addition and a decrease in pauses frequency in comparison to initial forward transcription (**Fig. 7.3a**). To quantify this, we utilized the largest data set of observed reversals (acquired at 16 pN) to construct a dwell-time distribution derived only from the reversed part of the traces (pink circles, **Fig. 7.4a**). We can determine the maximum of the distribution, which reports on the apparent catalytic rate. Notwithstanding the lower statistics for the reversals data set at 16 pN, it is clear that the peak position shifts to shorter time in comparison with the peak position of the data set at 35 pN (red circles, **Fig. 7.4a**), the largest assisting force. The corresponding apparent nucleotide addition rate can be determined as 53.4 ± 5.4 nt/s (**Fig. 7.4a,b**). This is nearly 2-fold higher than the rate of nucleotide addition extracted from the forward transcriptions at the same NTP concentration and 35 pN (25.6 ± 0.2 nt/s) (**Fig. 7.3b**)[9].

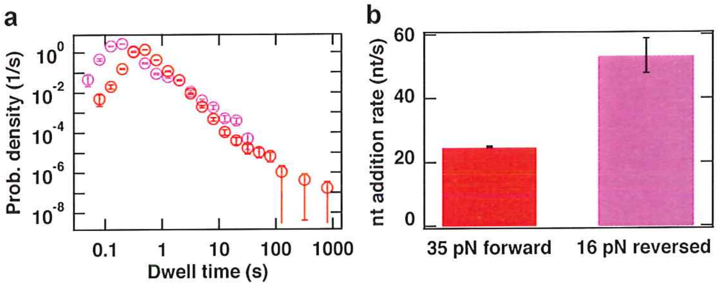


Figure 7.4: **P2 RdRp exhibits different kinetic properties in standard elongation compared to reversal events.** **(a)** The dwell-time distribution for 16 reversal traces obtained at 16 pN (pink circles) and for the forward traces obtained at 35 pN (red circles). Error bars are one standard deviation confidence intervals derived from 1000 bootstraps. **(b)** The nucleotide addition rates extracted from the maxima of the dwell-time distributions in **(a)** (identical color codes employed). Error bars are one standard deviation confidence intervals derived from 10000 bootstraps.

7.3. DISCUSSION

RdRp is involved in the replication and the transcription of the viral genome and is therefore essential for virus survival. While the catalytic activity of RdRps has been intensively studied in bulk[34–36] and, more recently, in single-molecule experiments[9], non-catalytic pauses such as backtracked pauses have not previously been directly observed in RdRp elongation dynamics. For the first time, we show direct evidence for the backward motion characteristic of backtracking (**Fig. 7.1c**) for a viral RdRp. The statistics provided by our experimental assay have been key in quantifying the long-lived pauses that result from a polymerase backtracking. Considering pause dynamics over more than two orders of magnitude in time, we show how the energy needed to propagate the fork influences the probability of backtracking.

The proposed model for backtracking (**Fig. 7.2d**) also accounts for the observed reversals. Backtracking can introduce a newly exposed 3'-end for a new polymerase to initiate on, and we hence interpret the reversals as a sign that a new polymerase has initiated and drives the ssRNA/dsRNA junction backward (**Fig. 7.5a**). This association between backtracks and reversal events is supported both by the similar force dependence of their probabilities (compare **Figs. 7.2c** and **7.3b**) and the fact that free P2 in solution is required for the observation of reversals.

Alternative scenarios that are consistent with the observed decrease in extension can be eliminated. These include exonuclease activity by P2, as this is not supported by structural evidence[23]; and nucleotide excision by P2 via pyrophosphorolysis, as results on the structurally-related bovine viral diarrhea virus RdRp[37] and human immunodeficiency virus (HIV) reverse transcriptase[38] indicate that pyrophosphate concentrations as high as 100 μM would be required, and these are neither present in our fresh HPLC-purified NTP stock nor produced by the polymerization activity in our single-molecule assay given the low number of active elongation complexes. A recent study has shown

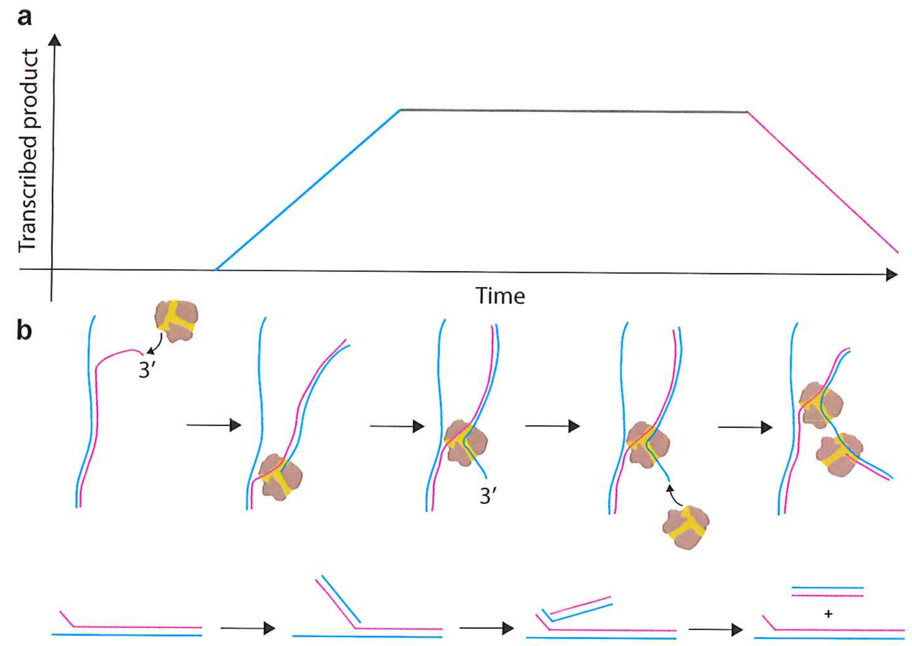


Figure 7.5: **Mechanism accounting for the observation of P2 RdRp reversal events.** **(a)** (Top) The mechanochemical model and (bottom) a schematic trace for a reversal event (**Fig. 7.3a**). Reversal events are proposed to originate from a second P2 RdRp that initiates on the 3' RNA product end that extrudes from the NTP channel of a first, backtracked P2 RdRp. The second P2 RdRp is free to initiate template-dependent RNA polymerization on the exposed RNA 3'-terminus. Subsequent elongation will cause the first P2 RdRp to be pushed back along the template strand. (+)RNA is shown in blue and (-)RNA is shown in pink. The background shading represents the different stages of RNA synthesis: gray for the absence of catalytic activity, blue for (+)RNA synthesis and pink for (-)RNA synthesis. **(b)** Schematic representing RNA production first by the forward transcribing RdRp and then by the second RdRp upon reversal (color code identical to **(a)**). Strands aligned in parallel fashion are fully hybridized. The net product of a complete reversal event is the original dsRNA construct plus a shorter dsRNA.

that the hepatitis C virus RdRp is able to use an NTP to excise the nucleotide at the 3' end[36]. However, the catalytic rate observed for such a reaction is three orders of magnitude slower than what the catalytic rate we measured during a reversal with P2 RdRp (**Fig. 7.3b**). Furthermore, this nucleotide excision mechanism has only been demonstrated with a C terminated 3' RNA product[36], whereas our template does not contain long stretches of C residues. We can also discard a scenario in which dissociation by an initial P2 from its template strand is followed by subsequent initiation of a second P2 on the exposed 3' end of the newly synthesized complement. While such a scenario would account for the presence of a second P2, it contradicts observations that P2 forms a stable elongation complex (as demonstrated in our previous work in which we are able to reactivate P2 following a significant period of stalling[9]) and fails to explain the observed force-dependence of the probability of reversals (**Fig. 7.3b**). It remains intriguing that the nucleotide addition rate is found to be 2-fold higher during reversals than during forward transcription activity (**Fig. 7.4a,b**). Two P2 enzymes working in close proximity thus seem to favor a more rapid translocation, but elucidating the reasons behind this will require further in-depth study. Interestingly, from the dwell-time distribution of the reversals at 16 pN applied force (**Fig. 7.4a**), we observe that the second polymerase does not enter into long pauses. One can explain this observation using the backtrack model of **Fig. 7.2d**: at low applied force, the small energy difference between RNA tether-RNA template base pairs and the RNA product-RNA template base pairs facilitates the re-hydridization of the RNA template strand to the RNA tether. This favored hybridization prevents backtracking of the second enzyme, in a manner similar to that of the first enzyme in the high force situation.

7.3.1. IMPLICATIONS OF OUR MODEL

A link can be made between the observed force dependence of backtracking and the behavior of P2 within the $\phi6$ bacteriophage. Within the bacteriophage, the hexameric NTPase (denoted P4) operates as a conduit for the plus-strand RNA release[39–41]. By analogy with the role of the applied force in our magnetic tweezers experiments, we speculate that P4 may, by limiting the excursions of the plus-strand RNA, prevent P2 from becoming trapped in a backtracked state. Its anchoring capability could, amongst its other known roles[39], serve to maintain the overall P2 transcription rate and directionality. However, we cannot exclude the occurrence of backtracking within the viral capsid. Furthermore, given the presence of $\approx$10 P2 RdRps in the viral capsid[42], a number of them may be free of template (the viral genome being tri-segmented[43]). Occurrences of P2 RdRp backtracking under these conditions allow for the possibility of reversal events to occur within the viral capsid.

The biological implications of the ability of a second P2 RdRp being able to initiate de novo on an exposed 3'-end remains to be further investigated. A possible role for such reversal events could be the production of shorter RNA molecules (**7.5b**) that would be subsequently used for homologous[44] or heterologous recombination[45–47]. Indeed, it has been shown that two genomic segments of $\phi6$ can recombine to replace the damaged 3'-end of one of the segments[46]. One might speculate that the second polymerase could rescue the template on which the first one is stalled, either for intrinsic biochemical reasons (e.g. catalytically inactive) or due to the presence of a roadblock on the dsRNA. The nearly 2-fold higher rate of reversal events compared to forward transcription events could potentially be exploited by the virus to enhance the production of viral RNA in the host cell. Another interesting observation is that the sequence of events shown in **Figs. 7.3d** and **7.5a** would result in the amplification of a viral RNA that contains an incomplete part of the genomic template (**Fig. 7.5b**). Future work will be necessary to unravel any sequence dependence of the reversals. The degree of structural similarity between RdRps[19] suggests that this mechanism could be conserved amongst RNA viruses that initiate replication or transcription de novo (e.g. hepatitis C virus[27]).

7.4. MATERIALS AND METHODS

Magnetic tweezers apparatus The magnetic tweezers assay and the flow-cell preparation used here have been previously described in ([9, 28, 48].

7.4.1. DSRNA CONSTRUCT

The construct employed here is made of four single-stranded (ss) RNA strands annealed to one complementary strand of 4 kb in length as described in detail previously[9]. To link the construct to the bead, a ssRNA strand containing biotins is annealed at one end of the complementary strand. To attach the construct to the surface, the other end of the complementary strand is annealed to a ssRNA strand containing digoxygenin. The template strand, ≈2.9 kb in length, contains a free 3'-end starting with 3 C residues followed by 15 U residues. A ssRNA strand is annealed as a spacer between the biotinylated strand and the template strand.

7.4.2. REACTION BUFFER FOR P2

The reaction buffer is composed of 50 mM HEPES pH 7.9, 20 mM ammonium acetate, 3% w/v polyethylene glycol 4000, 0.1 mM EDTA (pH 8.0), 5 mM $MgCl_2$, 2 mM $MnCl_2$, 0.01% Triton X-100, 5% Superase RNase inhibitor (Life Technologies), 20 μg/ml bovine serum albumin, and an optimal concentration of nucleotides, $[\mathrm{NTP}]_{\mathrm{opt}}$, of 1 mM ATP and GTP, 0.2 mM CTP and UTP, coming from HPLC-purified stock solutions. This concentration has been shown to be optimal for initiation of the RdRp on the 3'-end of an RNA template starting with 3 C residues followed by 15 U residues[9, 49, 50].

7.4.3. SINGLE-MOLECULE TRANSCRIPTION EXPERIMENTS

Once the RNA construct is calibrated inside a flow cell containing P2 reaction buffer[9], 9 nM of P2 is flushed in with the reaction buffer. We perform experiments at 21°C for 1 h at constant force and fixed NTP concentration while recording images of the magnetic beads at 25 Hz. Subsequently, the images are analyzed in real time using custom-written routines in Labview and CUDA nVidia to determine their (x, y, z) position[28]. Up to 800 beads can be followed in real time at 25 Hz with this assay. Distinct traces are low-pass filtered at 0.5 Hz and synchronized. The changes in tether extension are converted into numbers of transcribed nucleotides using the force-extension relationships for dsRNA and ssRNA constructs obtained in P2 reaction buffer[9], as previously described[51].

7.4.4. GENERATION OF DWELL TIME DISTRIBUTIONS

We use a dwell-time analysis[9] to analyze the low-pass filtered data obtained as described above (Fig. 7.1c, inset). Thus, we scan each transcription trace with a 10 nt transcription window, and for each such window we record the time needed for the polymerase to transcribe through it, and record this as the dwell-time. We then slide the transcription window forward by 10 nt and record the subsequent dwell-time. This procedure is repeated until the end of the trace is reached, and the dwell-times that result from an individual trace are binned with those resulting from identical conditions (e.g., same force) to yield a dwell-time distribution. Looking at dwell-time distribution for individual polymerases, it is clear that a small percentage spend a much longer-than-average time pausing. To prevent these pause-prone polymerases from biasing the data, we remove the 5% with the most extreme pause densities.

REFERENCES

[1] David Dulin, Igor D Vilfan, Bojk A Berghuis, Minna M Poranen, Martin Depken, and Nynke H Dekker. Backtracking behavior in viral rna-dependent rna polymerase provides the basis for a second initiation site. *Nucleic acids research*, page gkv1098, 2015.

[2] Johnson A. Lewis J. Raff M. Roberts K. Walter P. Alberts, B. *Molecular Biology of the Cell (Garland Science, New York, 2002).*

[3] David Dulin, Jan Lipfert, M. Charl Moolman, and Nynke H. Dekker. Studying genomic processes at the single-molecule level: introducing the tools and applications. *Nat Rev Genet*, 14(1):9–22, 01 2013.

[4] Hylkje J Geertsema and Antoine M van Oijen. A single-molecule view of DNA replication: the dynamic nature of multi-protein complexes revealed. *Current Opinion in Structural Biology*, 23(5):788 – 793, 2013. Protein-carbohydrate interactions / Biophysical methods.

[5] Matthew H Larson, Robert Landick, and Steven M Block. Single-molecule studies of rna polymerase: one singular sensation, every little step it takes. *Molecular cell*, 41(3):249–262, 2011.

[6] J.-B. Lee, R.K. Hite, S.M. Hamdan, X.S. Xie, C.C. Richardson, and A.M. Van Oijen. Dna primase acts as a molecular brake in dna replication. *Nature*, 439(7076):621–624, 2006. cited By 154.

[7] Kristina M. Herbert, Arthur La Porta, Becky J. Wong, Rachel A. Mooney, Keir C. Neuman, Robert Landick, and Steven M. Block. Sequence-resolved detection of pausing by single {RNA} polymerase molecules. *Cell*, 125(6):1083 – 1094, 2006.

[8] M.H. Larson, R.A. Mooney, J.M. Peters, T. Windgassen, D. Nayak, C.A. Gross, S.M. Block, W.J. Greenleaf, R. Landick, and J.S. Weissman. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science*, 344(6187):1042–1047, 2014. cited By 18.

[9] David Dulin, Igor D Vilfan, Bojk A Berghuis, Susanne Hage, Dennis H Bamford, Minna M Poranen, Martin Depken, and Nynke H Dekker. Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase. *Cell Reports*, 10(6):983–992, February 2015.

[10] J.W. Shaevitz, E.A. Abbondanzieri, R. Landick, and S.M. Block. Backtracking by single rna polymerase molecules observed at near-base-pair resolution. *Nature*, 426(6967):684–687, 2003. cited By 215.

[11] E.A. Galburt, S.W. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, and C. Bustamante. Backtracking determines the force sensitivity of rnap ii in a factor-dependent manner. *Nature*, 446(7137):820–823, 2007. cited By 130.

[12] Martin Depken, JuanÂ M.R. Parrondo, and StephanÂ W. Grill. Intermittent transcription dynamics for the rapid production of long transcripts of high fidelity. *Cell Reports*, 5(2):521 – 530, 2013.

[13] Oleg Laptenko, Jookyung Lee, Ivan Lomakin, and Sergei Borukhov. Transcript cleavage factors grea and greb act as transient catalytic components of rna polymerase. *The EMBO Journal*, 22(23):6322–6334, 2003.

[14] Kym Lowry, Andrew Woodman, Jonathan Cook, and David J. Evans. Recombination in enteroviruses is a biphasic replicative process involving the generation of greater-than genome length 'imprecise' intermediates. *PLoS Pathog*, 10(6):e1004191, 06 2014.

[15] Adam S. Lauring, Judith Frydman, and Raul Andino. The role of mutational robustness in rna virus evolution. *Nat Rev Micro*, 11(5):327–336, 05 2013.

[16] Marco Vignuzzi, Jeffrey K. Stone, Jamie J. Arnold, Craig E. Cameron, and Raul Andino. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*, 439(7074):344–348, 01 2006.

[17] Etienne Simon-Loriere and Edward C. Holmes. Why do rna viruses recombine? *Nat Rev Micro*, 9(8):617–626, 08 2011.

[18] Kenneth K.-S. Ng, JamieJ. Arnold, and CraigE. Cameron. Structure-function relationships among rna-dependent rna polymerases. In PatrickJ. Paddison and PeterK. Vogt, editors, *RNA Interference*, volume 320 of *Current Topics in Microbiology and Immunology*, pages 137–156. Springer Berlin Heidelberg, 2008.

[19] AartjanJ.W. te Velthuis. Common and unique features of viral rna-dependent polymerases. *Cellular and Molecular Life Sciences*, 71(22):4403–4420, 2014.

[20] Heli A. M. Mönttinen, Janne J. Ravantti, David I. Stuart, and Minna M. Poranen. Automated structural comparisons clarify the phylogeny of the right-hand-shaped polymerases. *Molecular Biology and Evolution*, 31(10):2741–2752, 2014.

[21] Shane Crotty, David Maag, Jamie J. Arnold, Weidong Zhong, Johnson Y. N. Lau, Zhi Hong, Raul Andino, and Craig E. Cameron. The broad-spectrum antiviral ribonucleoside ribavirin is an rna virus mutagen. *Nat Med*, 6(12):1375–1379, 12 2000.

[22] Eugene V. Makeyev and Jonathan M. Grimes. Rna-dependent RNA polymerases of dsRNA bacteriophages. *Virus Research*, 101(1):45 – 55, 2004. Viral molecular machines: Replication systems within the inner cores of dsRNA viruses.

[23] Sarah J. Butcher, Jonathan M. Grimes, Eugeny V. Makeyev, Dennis H. Bamford, and David I. Stuart. A mechanism for initiating rna-dependent rna polymerization. *Nature*, 410(6825):235–240, 03 2001.

[24] Minna M. Poranen, Paula S. Salgado, Minni R. L. Koivunen, Sam Wright, Dennis H. Bamford, David I. Stuart, and Jonathan M. Grimes. Structural explanation for the role of mn2+ in the activity of $\phi 6$ rna-dependent rna polymerase. *Nucleic Acids Research*, 36(20):6633–6644, 2008.

[25] L. Peter Sarin, Sam Wright, Qing Chen, Linda H. Degerth, David I. Stuart, Jonathan M. Grimes, Dennis H. Bamford, and Minna M. Poranen. The c-terminal priming domain is strongly associated with the main body of bacteriophage ï*6 rna-dependent {RNA} polymerase. *Virology*, 432(1):184 – 193, 2012.

[26] Sam Wright, Minna M. Poranen, Dennis H. Bamford, David I. Stuart, and Jonathan M. Grimes. Noncatalytic ions direct the RNA-dependent RNA polymerase of bacterial double-stranded RNA virus ϕ6 from de novo initiation to elongation. *Journal of Virology*, 86(5):2837–2849, 2012.

[27] Todd C. Appleby, Jason K. Perry, Eisuke Murakami, Ona Barauskas, Joy Feng, Aesop Cho, David Fox, Diana R. Wetmore, Mary E. McGrath, Adrian S. Ray, Michael J. Sofia, S. Swaminathan, and Thomas E. Edwards. Structural basis for rna replication by the hepatitis c virus polymerase. *Science*, 347(6223):771–775, 2015.

[28] J P Cnossen, D Dulin, and N H Dekker. An optimized software framework for real-time, high-throughput tracking of spherical beads. *Review of Scientific Instruments*, 85(10):103712, October 2014.

[29] M. Depken, E.A. Galburt, and S.W. Grill. The origin of short transcriptional pauses. *Biophysical Journal*, 96(6):2189–2193, 2009. cited By 44.

[30] E.A. Galburt, S.W. Grill, and C. Bustamante. Single molecule transcription elongation. *Methods*, 48(4):323–332, 2009. cited By 18.

[31] Marie-Noëlle Dessinges, Timothée Lionnet, Xu Guang Xi, David Bensimon, and Vincent Croquette. Single-molecule assay reveals strand switching and enhanced processivity of uvrd. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17):6439–6444, 2004.

[32] Paula S. Salgado, Eugene V. Makeyev, Sarah J. Butcher, Dennis H. Bamford, David I. Stuart, and Jonathan M. Grimes. The structural basis for RNA specificity and ca2+ inhibition of an RNA-dependent RNA polymerase. *Structure*, 12(2):307 – 316, 2004.

[33] J. Lipfert, X. Hao, and N.H. Dekker. Quantitative modeling and optimization of magnetic tweezers. *Biophysical Journal*, 96(12):5040–5049, 2009. cited By 43.

[34] Jamie J. Arnold, , and Craig E. Cameron. Poliovirus rna-dependent rna polymerase (3dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of mg2+†. *Biochemistry*, 43(18):5126–5137, 2004. PMID: 15122878.

[35] Jamie J. Arnold and and Craig E. Cameron David W. Gohara. Poliovirus rna-dependent rna polymerase (3dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of mn2+. *Biochemistry*, 43(18):5138–5148, 2004. PMID: 15122879.

[36] Zhinan Jin, Vincent Leveque, Han Ma, Kenneth A. Johnson, and Klaus Klumpp. Ntp-mediated nucleotide excision activity of hepatitis c virus rna-dependent rna polymerase. *Proceedings of the National Academy of Sciences*, 110(5):E348–E357, 2013.

[37] Claudia M. D'Abramo, Luciano Cellai, and Matthias GÃ¶tte. Excision of incorporated nucleotide analogue chain-terminators can diminish their inhibitory effects on viral rna-dependent {RNA} polymerases. *Journal of Molecular Biology*, 337(1):1 – 14, 2004.

[38] Dominique Arion, Neerja Kaushik, Suzanne McCormick, Gadi Borkow, , and Michael A. Parniak*. Phenotypic mechanism of hiv-1 resistance to 3'-azido-3'-deoxythymidine (azt): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase. *Biochemistry*, 37(45):15908–15917, 1998. PMID: 9843396.

[39] Markus J. Pirttimaa, Anja O. Paatero, Mikko J. Frilander, and Dennis H. Bamford. Nonspecific nucleoside triphosphatase p4 of double-stranded rna bacteriophage ϕ6 is required for single-stranded rna packaging and transcription. *Journal of Virology*, 76(20):10122–10127, 2002.

[40] Denis E. Kainov, Markus Pirttimaa, Roman Tuma, Sarah J. Butcher, George J. Thomas, Dennis H. Bamford, and Eugene V. Makeyev. Rna packaging device of double-stranded rna bacteriophages, possibly as simple as hexamer of p4 protein. *Journal of Biological Chemistry*, 278(48):48084–48091, 2003.

[41] Xiaoyu Sun, Markus J. Pirttimaa, Dennis H. Bamford, and Minna M. Poranen. Rescue of maturation off-pathway products in the assembly of pseudomonas phage ϕ6. *Journal of Virology*, 87(24):13279–13286, 2013.

[42] Xiaoyu Sun, Dennis H. Bamford, and Minna M. Poranen. Probing, by self-assembly, the number of potential binding sites for minor protein subunits in the procapsid of double-stranded rna bacteriophage ϕ6. *Journal of Virology*, 86(22):12208–12216, 2012.

[43] Minna M Poranen and Roman Tuma. Self-assembly of double-stranded {RNA} bacteriophages. *Virus Research*, 101(1):93 – 100, 2004. Viral molecular machines: Replication systems within the inner cores of dsRNA viruses.

[44] Shiroh Onodera, Yang Sun, and Leonard Mindich. Reverse genetics and recombination in ϕ8, a dsrna bacteriophage. *Virology*, 286(1):113 – 118, 2001.

[45] S Onodera, X Qiao, P Gottlieb, J Strassman, M Frilander, and L Mindich. Rna structure and heterologous recombination in the double-stranded rna bacteriophage phi 6. *Journal of Virology*, 67(8):4914–4922, 1993.

[46] L Mindich, X Qiao, S Onodera, P Gottlieb, and J Strassman. Heterologous recombination in the double-stranded rna bacteriophage phi 6. *Journal of Virology*, 66(5):2605–2610, 1992.

[47] Xueying Qiao, Jian Qiao, and Leonard Mindich. An in vitro system for the investigation of heterologous rna recombination. *Virology*, 227(1):103 – 110, 1997.

[48] R Janissen, Bojk A Berghuis, D Dulin, M Wink, T van Laar, and N H Dekker. Invincible DNA tethers: covalent DNA anchoring for enhanced temporal and force stability in magnetic tweezers experiments. *Nucleic Acids Research*, 42(18):e137–e137, October 2014.

[49] Eugeny V. Makeyev and Dennis H. Bamford. Replicase activity of purified recombinant protein p2 of double-stranded rna bacteriophage ϕ6. *The EMBO Journal*, 19(1):124–133, 2000.

[50] Eugeny V. Makeyev and Dennis H. Bamford. The polymerase subunit of a dsrna virus plays a central role in the regulation of viral rna metabolism. *The EMBO Journal*, 19(22):6275–6284, 2000.

[51] Berenike Maier, David Bensimon, and Vincent Croquette. Replication by a single dna polymerase of a stretched single-stranded dna. *Proceedings of the National Academy of Sciences*, 97(22):12002–12007, 2000.

III

DNA ROADBLOCKS

8

STRAND SEPARATION ESTABLISHES A SUSTAINED LOCK AT THE TUS–*Ter* REPLICATION FORK BARRIER

Hic sudavit, sed non frustra

The bidirectional replication of a circular chromosome by many bacteria necessitates proper termination to avoid the head-on collision of the opposing replisomes. In *Escherichia coli*, replisome progression beyond the termination site is prevented by Tus proteins bound to asymmetric *Ter* sites. Structural evidence indicates that strand separation on the blocking (nonpermissive) side of Tus–*Ter* triggers roadblock formation, but biochemical evidence also suggests roles for protein-protein interactions. Here DNA unzipping experiments demonstrate that nonpermissively oriented Tus–*Ter* forms a tight lock in the absence of replicative proteins, whereas permissively oriented Tus–*Ter* allows nearly unhindered strand separation. Quantifying the lock strength reveals the existence of several intermediate lock states that are impacted by mutations in the lock domain but not by mutations in the DNA-binding domain. Lock formation is highly specific and exceeds reported *in vivo* efficiencies. We postulate that protein-protein interactions may actually hinder, rather than promote, proper lock formation.

This chapter has been published as: Bojk A Berghuis, David Dulin, Zhi-Qiang Xu, Theo van Laar, Bronwen Cross, Richard Janissen, Slobodan Jergic, Nicholas E Dixon, Martin Depken & Nynke H Dekker. Strand separation establishes a sustained lock at the Tus–*Ter* replication fork barrier. *Nature Chemical Biology* 11, 579-585 (2015). [1].

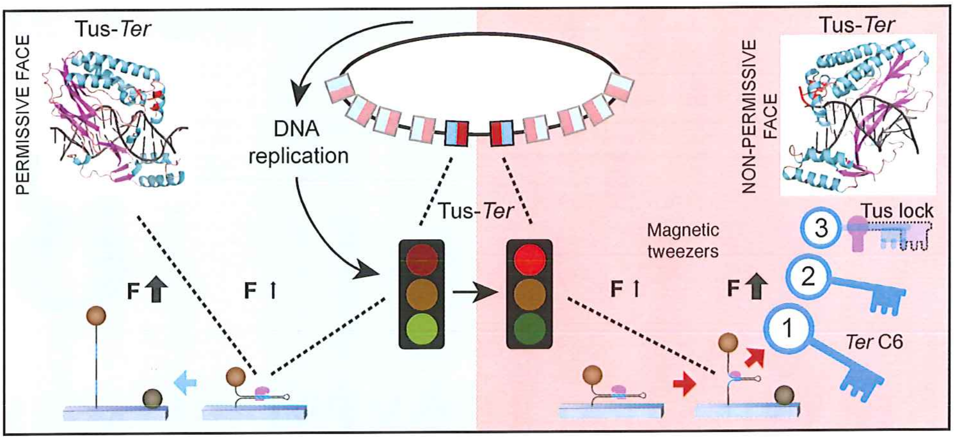


Figure 8.1: **Graphical abstract**

8.1. INTRODUCTION

DNA replication in *E. coli* initiates bidirectionally at *oriC*, creating two replication forks that proceed around the circular 4.6-Mbp chromosome in opposite directions. The forks progress at an average speed of 1 kbp/s until they meet again at the terminus region. As the replication forks approach the terminus, each encounters five 23-bp *Ter* DNA sites (out a total of ten, denoted *TerA-J*) bound in a specific orientation by a 36-kDa DNA-binding protein called Tus[2–5] and proceeds unhindered. However, when a replication fork continues beyond the terminus, Tus–*Ter* is approached from the opposite direction (**Fig. 8.2a**), triggering it to form a tightly locked complex and bringing the replication fork to a halt[2, 6–8]. Each *Ter* site is nonpalindromic, does not contain any direct repeats and has a strictly conserved G-C6 base pair followed by a highly conserved 13-bp core region. Tus is a monomeric protein that forms a simple 1:1 complex with *TerB*[9] (**Fig. 8.2b**). The structure of the Tus–*TerA* complex shows that many of the conserved residues among the *Ter* sites make base-specific contacts with the protein[5, 10]. The Tus–*TerB* complex has a reported dissociation constant (K_d) of 44 pM in 50 mM NaCl[11]. This renders it the most stable complex known between a monomeric sequence-specific DNA-binding protein and a duplex-DNA recognition sequence.

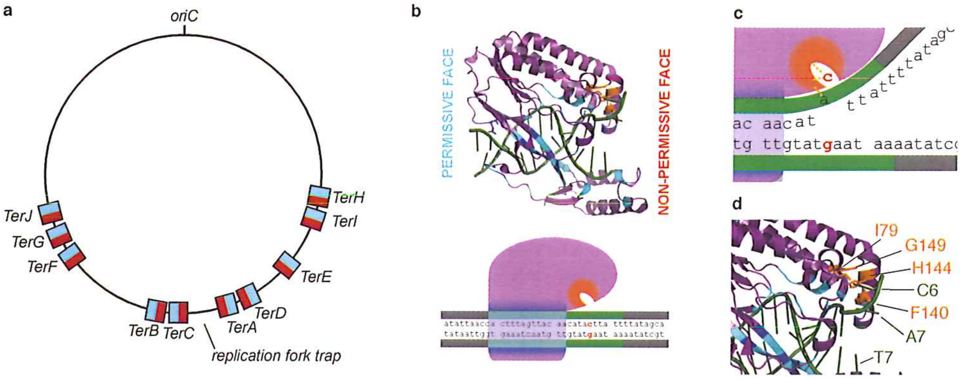


Figure 8.2: **The Tus-*Ter* complex structure and domains.** **(a)** Location and orientation (turquoise for permissive face, red for nonpermissive face) of the 23-bp *Ter* sites in the *E. coli* chromosome. **(b)** The crystal structure (top) of the locked Tus–*Ter* complex (PDB ID: 2EWJ) with a schematic representation (bottom) showing the protein has a DNA-binding domain located mainly in the two antiparallel β-strands interacting with the major groove of *Ter* DNA (dark blue for base-specific interactions, light blue for nonspecific interactions). **(c,d)** Schematic representation **(c)** and crystal structure **(d)** showing the lock domain and amino acid residues (in orange) in Tus that interact specifically with C6 upon strand separation. In **d**, *Ter* nucleotides in green.

A long-standing issue regards the manner in which the asymmetric blockage at Tus–*Ter* comes about. Does Tus itself function as a molecular roadblock, locking itself onto the *Ter* DNA as the DNA replication machinery approaches, or are there specific protein-protein interactions that lead to the polar arrest of the replisome? On one hand, various studies imply specific protein-protein interactions between Tus and the replicative helicase DnaB. Tus–*Ter* is much more effective in its natural host, for instance, while

8

the functionally similar but structurally unrelated *Bacillus subtilis* replication termination system works well in *E. coli*[12, 13]. Tus–*Ter* blocks DnaB, but not the Rep helicase, *in vitro*[14], and evidence from yeast two-hybrid analysis shows specific interactions between DnaB and Tus[15]. On the other hand, ample evidence suggests a protein-independent polar blocking mechanism. For example, RNA chain elongation catalyzed by T7, SP6 and *E. coli* RNA polymerases is impeded by Tus–*Ter* in a polar manner[16, 17]. Tus–*TerB* also blocks the actions of the UvrD, Rep, PriA and SV40 large T antigen helicases, indicating low specificity for DnaB alone[18–21].

In 2006 light was shed on this molecular roadblock through a surface plasmon resonance (SPR) study of the dissociation of Tus from forked *TerB* oligonucleotides, which was supported by a crystal structure of a forked, 'locked' Tus–*Ter* complex[22]. This locked complex exhibits significant structural differences at the fork-blocking (nonpermissive) face in comparison with the ds*TerA*-bound, but not locked, Tus structure elucidated a decade earlier[5]. The locked conformation reveals that, of the single-stranded (ss) DNA bases in the forked *Ter* region (*Ter* bases 1–7), the highly conserved C6 base is flipped out of the helical DNA axis and into the protein (**Fig. 8.2c**). In this conformation the C6 base undergoes tight interactions with several amino acids (**Fig. 8.2d**). These Tus lock domain residues are distinctly different from those involved in sequence recognition and binding affinity[23]. It was therefore proposed that the Tus–*Ter* system is the molecular analog of a mousetrap: the trap is set by Tus binding to *Ter* in an oriented fashion and triggered by strand separation invoked by the approaching replication machinery[22].

The mousetrap model has two major implications. First, it suggests that binding and lock formation are two different mechanisms that can be ascribed to different domains of Tus. Second, lock formation through strand separation could occur independently of any specific protein-protein interactions. Nevertheless, convincing evidence arose that translocation of DnaB on double-stranded (ds) DNA in the absence of unwinding is sufficient to provoke polar arrest[24]. Although this result did not require it, these authors proposed an alternative model in which the DnaB helicase binds specifically to Tus. They argued that formation of the locked complex may act as a backup mechanism when protein-protein interaction fails but may not be sufficient on its own.

In this study, we used the quantitative power of high-throughput single-molecule approaches to address both of the key implications of the Tus–*Ter* mousetrap model and to dissect the overall mechanism of lock formation. By applying a mechanical force to unwind a DNA hairpin containing a single *TerB* site, we mimicked replisome-mediated DNA unwinding and directly showed that strand separation alone could trigger the nonpermissively oriented Tus–*Ter* to form a strong and long-lived lock. Remarkably, the Tus–*Ter* lock formed in 100% of our hairpin-opening attempts. This was in contrast to Tus–*Ter* in the permissive orientation, where strand separation proceeded virtually unhindered. We quantified lock strength by measuring the lifetimes of the Tus–*Ter* complex at different forces. These experiments revealed that at high forces Tus dissociation occurred on three (or more) characteristic timescales, suggesting that strand separation at high forces partitioned the Tus–*Ter* structure into thermodynamically trapped substructures. We argue that the shorter-lived substructures correspond to intermediates in the process of full lock formation during replisomal strand separation and that the longest-lived structure is the full lock.

Our results strongly validated the molecular mousetrap model[22] by showing that Tus–*Ter* caused polar arrest of strand separation in the absence of any replication-related proteins. We showed that the interaction was efficient and was not limited by the rate of C6 flipping and finding the lock pocket. Using specific mutants we were able to discriminate DNA-binding and locking domains in this system. Residue H144, located deep in the Tus lock domain, determines the strength of interaction of the Tus–*Ter* lock: force-dependent lifetimes of H144A decreased more profoundly than those of any of the other single-site mutations tested. F140, located at the side of the lock pocket, was found to be involved in the specificity of the lock pocket for a C base. Notably, a mutation in residue E49, which is located outside the lock domain and thought to play a pivotal role in the specific interaction of Tus with DnaB, displayed a marked decrease in the probability of lock formation even though the lock lifetime was identical to that of wild-type (WT) Tus–*Ter*. This showed that E49 plays a crucial role in guiding C6 to the lock domain and that interfering with specific residues surrounding the Tus lock modulates the probability of forming a tightly locked Tus–*Ter* complex. Conversely, a mutation in the DNA-binding domain at the permissive face of the complex did not affect locking behavior. As *in vivo* experiments point toward probabilities of molecular-motor arrest substantially below those found here[13], we hypothesize that, instead of forming the basis of promoting polar arrest, protein-protein interactions may actually perform the opposite function of hindering proper lock formation. Our assay resolves the controversy that still surrounds this protein-DNA complex by providing direct insight into how different DNA processing enzymes in a head-on collision with Tus–*Ter* can exhibit varying blocking efficiencies, in particular by modulating the probability of lock formation through (nonspecific) steric hindrance.

8.2. RESULTS

8.2.1. MIMICKING REPLICATION FORK PROGRESSION USING DNA HAIRPINS

We set up a single-molecule assay using magnetic tweezers and DNA hairpins, which allowed us to controllably invoke the dsDNA unwinding that normally accompanies DNA replication, only now in the absence of the replication proteins. Our experiments initially comprised three DNA hairpin designs with specific sequences inserted at their midpoints: the first hairpin contained a single *TerB* site in the permissive orientation (**Fig. 8.3a**); the second had the *TerB* site inverted, forming the nonpermissive orientation (**Fig. 8.3b**); and lastly, the third hairpin contained a *TerB* site in the nonpermissive orientation, but included a point mutation at the G-C6 site in which the highly conserved C6 base was replaced by a guanine (referred to here as the 'GC flip') (Supplementary Information (SI), **Fig. ??b**). As the mousetrap model suggests a purely mechanical interaction for Tus—*Ter* upon strand separation, it predicts that, in our setup, lock formation should still occur in one direction (nonpermissive) but not the other (permissive). If protein–protein interactions are essential for proper lock formation, the phenomenon should, at most, be infrequent in our assay. We detected lock formation by measuring the difference in extension between a hairpin that is fully opened and one that is blocked halfway.

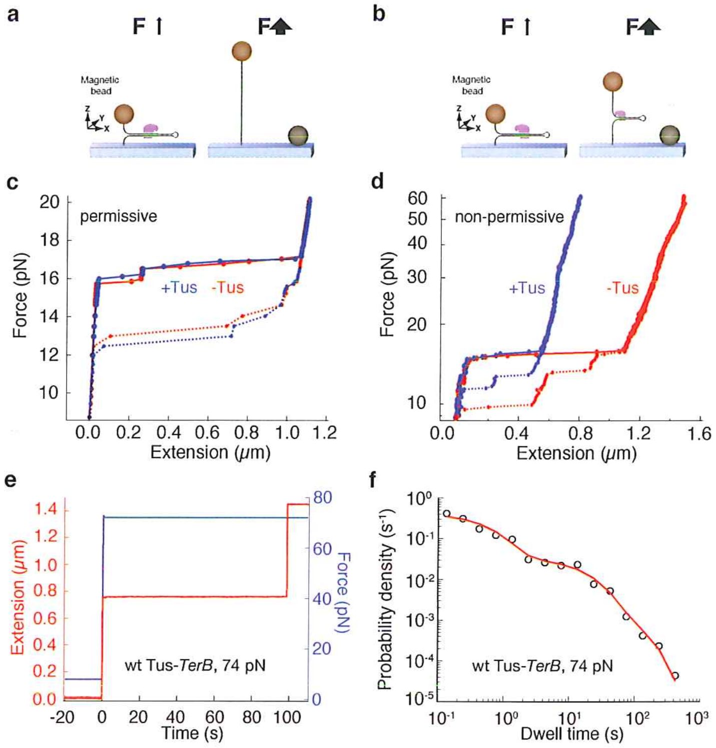


Figure 8.3: **Magnetic Tweezers (MT) assay used to quantify the Tus-*Ter* lock mechanism.** **(a,b)** Schematic overviews of a permissive **(a)** and a nonpermissive **(b)** Tus–*Ter* experiment. In the permissive experiment **(a)**, force-induced DNA strand separation causes weakening of the interaction between the Tus DNA-binding domain and the *Ter* site, with subsequent disruption of the Tus–*Ter* interaction. **(b)** With the nonpermissively oriented *Ter* on a DNA hairpin, strand separation will cause the C6 base to flip into the high-affinity protein-binding pocket of Tus, resulting in a locked Tus–*Ter* complex. **(c)** Permissive MT DNA hairpin experiment yielding almost identical force extension curves for hairpins with (blue) or without (red) Tus, implying full opening of the hairpin. Dotted lines represent reannealing of the hairpin during force decrease. **(d)** Nonpermissive MT DNA hairpin experiment resulting in an extension that is halved in the presence of Tus (blue), suggesting that strand separation is blocked, compared to the same hairpin in the absence of Tus (red). **(e)** Quantification of lock behavior by measurement of lock dwell times under constant force. **(f)** Multiexponential distribution of dwell times (black circles) and fit of the kinetic model (red) (see text for explanation).

8.2.2. TUS–*TER* BLOCKS FORCE-INDUCED UNWINDING

At low forces (<16 pN), base-paired DNA is energetically more favorable than ssDNA, so the hairpin remains closed [25–27]. Upon increasing the force (>16 pN) in the absence of Tus, the hairpin opens, and this could be seen as a rapid increase in extension for both the permissive and the nonpermissive hairpins (**Fig. 8.3c,d**, red traces). Repeating this experiment in the presence of Tus resulted in an almost identical outcome for the permissive *Ter* hairpin (**Fig. 8.3c**, blue trace). Here the Tus–*Ter* interaction left only a transient signal upon hairpin opening (**Fig. 8.7h**). In contrast, the results were very different for the nonpermissive hairpin (**Fig. 8.3d**, blue trace)—here the maximal extension in the presence of Tus was only half of the fully opened hairpin, indicating that strand separation was blocked exactly at the *Ter* site (**Fig. 8.6a**). This behavior was observed for nonpermissive hairpins in 100% of the experiments at 50 mM KCl and a Tus concentration of 2 nM. Increasing the ionic strength to 350 mM resulted in a modest decrease in the occurrence of blocking, but did not affect the lock strength (**Fig. 8.7d-f**). The high efficiency of lock formation still occurred despite the fact that in our experiments the DNA helix was unwound at a rate of ~30 kbp/s (**Fig. 8.6c**), which is at least tenfold faster than any replisome would unwind DNA. Increasing the force showed that the Tus—*Ter* lock could remain in place at forces up to 60 pN, demonstrating the remarkable strength of this locked complex. This experiment thus validated the proposed protein-protein independency [22] for fork arrest and strongly suggested that the Tus-*Ter* locking mechanism alone is readily equipped for the task of blocking an approaching replication fork, other helicases and the transcription machinery alike.

8.2.3. LOCK BREAKAGE SHOWS DIFFERENT DEGREES OF LOCK FORMATION

We measured the distribution of constant-force lock dwell times (**Fig. 8.3e,f**) for all protein and *Ter* variants (**Fig. 8.4a**) taking advantage of the natural force clamp mode and multiplexing capacity of magnetic tweezers. The distributions were highly reproducible and contained force-specific as well as mutant-specific signatures (**Fig. 8.4b,c** and **Fig. 8.7a-c**). A main feature of all lock dwell time distributions was that they were (multi)exponentially distributed, reflecting the stochastic nature of lock rupture (**Fig. 8.3f**). We found that the distributions contained, according to the Bayes-Schwarz information criterion[28], two or three exponentially distributed states depending on the type of Tus–*Ter* interaction investigated. We used maximum-likelihood estimation to fit the data (**Fig. 8.3f**)[29] and obtained confidence intervals through bootstrapping[30] as described previously[31]. As these high-force measurements place a large tension on the DNA tethers, choosing the right force required a trade-off between tether lifetime[32] and the ability to resolve the different lock states (**Fig. 8.6d**). Typically, we chose the force such that the three-exponential data sets exhibited a first short-lived exponential with a lifetime of ~1 s, a second exponential with a lifetime on the order of 10 s and a third, long-lived exponential on the order of 100 s.

In examining Tus locking behavior, it is convenient to consider this system through a lock-and-key analogy, where the C6 base is the key that fits into the Tus lock pocket (**Fig. 8.4a**). In this analogy, the interaction between WT Tus and nonpermissive *Ter* should provide a signature analogous to a perfect match between key and keyhole (**Fig. 8.4a**, WT Tus—emphTer). We found the force-dependent dwell times of the WT Tus–*Ter* lock to be distributed over three states, with the longest-lived exponential distribution having a lifetime of ~720 s at 59 pN (**Fig. 8.4b,c**, purple circles). The shortest-lived exponential state at 59 pN had a lifetime of ~1 s, and the intermediate state had a lifetime of ~30 s. The lifetimes of all three states decreased in a concerted fashion as the force was increased, with the longest-lived distribution having a lifetime of 54 s at 93 pN (**Fig. 8.4c** and Table 8.11). We also observed a force-dependent probability of forming the longest-lived state: although at 93 pN there was a mere 7% chance for a dwell time to belong to the longest-lived state, this probability increased to 73% at 59 pN (**Fig. 8.4d,e**). Conversely, trapping of the system in one of the shorter-lived states became progressively less likely as the force was decreased (**Fig. 8.7i**). The force-dependent probability of all states also indicated that the first two states are likely to represent intermediate conformations that occur at all forces, while the longest-lived state represents the full lock. Thus when the magnetic tweezers exerted their highest forces, they prevented the short-lived conformations from proceeding to the fully locked state, while the longest-lived state predominated at low forces.

Experiments on permissive WT Tus—*Ter* resulted in sharply reduced dwell times that obeyed a single-exponential distribution with a mean of 0.8 s at 19 pN (**Fig. 8.4c** (purple square) and **Fig. 8.7h**); at higher forces dwell times were too short to be detected. In fact, there was no single force at which both nonpermissive and permissive dwell times could be measured; the dwell times of nonpermissive WT Tus—*Ter* became too long at 19 pN for practical measurements (**Fig. 8.6d**). This implies that none of the states we found for the nonpermissive orientation can be attributed solely to binding by Tus. To further investigate the origin of the observed states, we compared the changes in lifetime and probability invoked by mutations in Tus and/or *TerB*.

BINDING-DOMAIN MUTATION DOES NOT HAMPER LOCK FORMATION

The crystal structure of the locked WT Tus—*Ter* shows that DNA sequence recognition and binding can be largely attributed to a Tus DNA-binding domain that primarily consists of two antiparallel β-strands that interact with the major groove of *Ter* DNA (**Fig. 8.2b**). A site-specific mutation in the DNA-binding domain (Q250A, **Fig. 8.8e**) is known to result in a sharp increase in the Tus–*TerB* K_D [23], but whether it affects lock kinetics is unknown. In our experiments, Q250A exhibited dwell time distributions very similar to those of WT Tus at the same forces (**Fig. 8.8c**, cyan). We saw no correlation between the K_D of Tus-ds*TerB* and lock strength. From this we concluded that lock formation is not severely affected by a change in Tus' DNA-binding domain.

8.2.4. C6 BASE IS CRUCIAL BUT NOT RATE-LIMITING FOR LOCK FORMATION

We subsequently set out to examine the effect that changing the key (the C6 base) (**Fig. 8.4a**, switch from green to magenta key) had on WT Tus-*Ter* lifetimes. A single-base-pair inversion of the *TerB* sequence at position 6 profoundly affects the fork-arrest efficiency[33]. In our experiments inversion of G-C6 indeed had a dramatic effect on the lifetimes (**Fig. 8.4c**, purple triangles), as the dominant lifetime was no greater than 1 s at 40 pN. By comparison the dominant lifetime of WT Tus–*Ter* was at least two orders of magnitude higher, as indicated by extrapolation of the lifetimes of the fully locked state observed in the 59–93 pN range (**Fig. 8.4c**, purple circles). Despite the decrease in observed lifetimes, the G6 *Ter* site continued to impose an increased barrier to hairpin opening because the lifetimes remained well above those found for binding of WT Tus only (**Fig. 8.4c**, purple square). For WT Tus with the modified key, we found two states (**Fig. 8.7g**, purple), with the longest-lived lifetime decreasing from 39 to 0.7 s in the 29–40 pN range (**Fig. 8.4c**, purple triangles). We also assessed whether uncoupling lock formation from mechanical probing (by creating a hairpin with an unpaired region of five bases containing C6; see **Fig. 8.6b**) would populate the fully locked state even at high forces, as it is known that this 5-base mismatch dramatically increases the affinity of the Tus–*Ter* complex.[22] The resulting state-probabilities, however, were identical to those of normal WT Tus-*Ter* (**Fig. 8.5a**), indicating that preformation of the lock did not alter the occupancy of the different states.

8.2.5. PROBING MECHANISM VIA MUTATIONS IN OR NEAR THE LOCK DOMAIN

To investigate how the *Ter* key enters the Tus lock, we performed experiments on a series of Tus mutants altered in or near the lock domain. Two amino acid residues, H144 and F140, are situated directly in the lock domain, and the crystal structure[22] suggests that their roles differ in a subtle, though significant, manner. Residue H144 lies deep within the lock pocket and interacts only with the C6 base. The H144A mutation removes the imidazole ring as well as a positive charge, leaving a cavity deep within the pocket (**Fig. 8.8b**, in blue). In our lock-and-key representation, we depicted this by changing the inner shape of the lock (**Fig. 8.4a**, light blue). Amino acid residue F140 lies closer to the outer edge of the lock pocket than does H144. F140 still interacts with C6, and a stacking interaction of the phenyl ring with the adjacent A7 base is also present (**Fig. 8.8c**, in orange). Removal of the phenyl ring in the F140A mutant would thus lead to a gap at the edge of the lock pocket, which we depict as a widening of the keyhole (**Fig. 8.4a**, orange). Residue E49, linked to the putative specific protein-protein interaction between Tus and the *E. coli* DnaB helicase [14, 15] lies just outside the lock domain (**Fig. 8.8d**, in green), although it does make a water-mediated hydrogen-bonding contact with the 5' phosphate of A7 in the locked complex[22]. The shape of the lock pocket remained unaffected by this mutation (**Fig. 8.4a**, green keyhole identical to WT).

8.2.6. F140A AFFECTS SPECIFICITY AND H144A AFFECTS STRENGTH OF LOCK

We found that mutant F140A showed a marked decrease in dwell times at 59 pN (**Fig. 8.4b**, orange). Fitting revealed that the longest-lived exponential now had a lifetime of ~55 s, compared to 720 s for WT Tus at the same force (**Fig. 8.4c**, orange circles). We also observed that this third, longest-lived state had all but disappeared as the probability of entering this state was reduced from 73% for WT Tus to 1.8% for F140A in the same regime (**Fig. 8.4b** (orange; note the absence of counts > 100 s) and **Fig. 8.4e** (orange bar in 'full lock' column)). Thus F140 appeared to give rise to the third, long-lived state observed in the Tus species with an intact lock pocket and, as such, seems to play a role in the probability of forming a fully locked state. Similarly to that of WT Tus, F140A's probability of attaining the full lock state exhibited a clear force-dependence: decreasing the force to 47 pN increased full lock probability to 31% (**Fig. 8.4d**, orange). Combining F140A with the mutated *Ter* site (**Fig. 8.4a**, magenta key with orange lock) further reduced the force-dependent lifetimes, but the resulting force-dependent lifetimes exceeded those of WT Tus with the mutated *Ter* site (**Fig. 8.4c** (orange triangles) and **Fig. 8.7g**, orange circles).

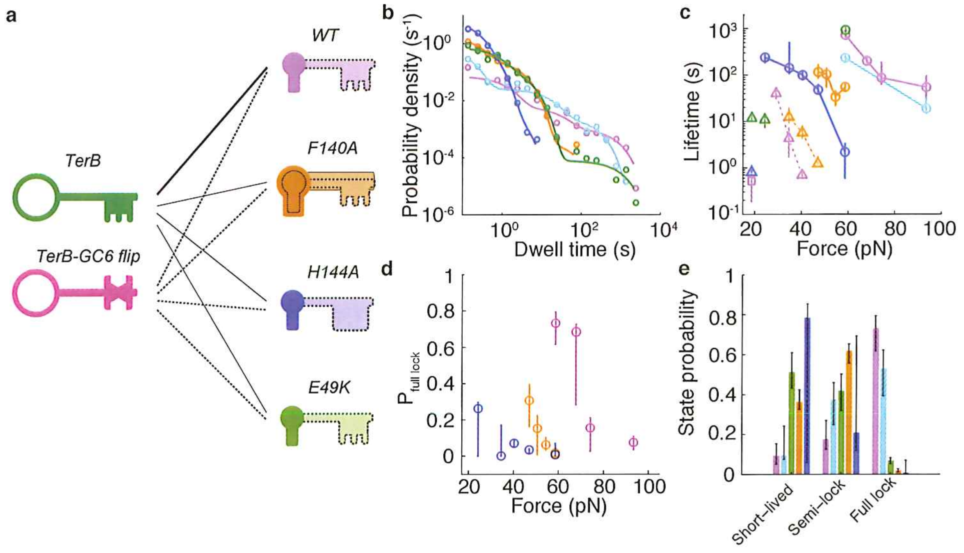


Figure 8.4: **The effect of Tus mutations on lock formation.** **(a)** The Tus–*Ter* lock domains, depicted as keyholes, forming locked complexes with either *Ter*B (green key) or *Ter*B with a mutated GC6 bp (*Ter*B–GC6 flip, magenta key). WT Tus (purple) and *Ter*B are a perfect match, H144A (blue) is a mutation deep within the lock (blue, cavity behind keyhole becomes larger), F140A is a mutation at the edge of the lock pocket (orange, keyhole becomes larger) and e49K is a mutation that lies close to the lock domain (green, keyhole identical to WT). **(b)** The distribution of dwell times of WT Tus (purple, $N = 94$), the lock-domain mutants (H144A, blue, $N = 642$; F140A, orange, $N = 344$; e49K, green, $N = 323$) and binding-domain mutant Q250A (cyan, $N = 90$) when bound to *Ter*B at 59 pN (circles are binned data, solid lines are fits). **(c)** The force-dependent lifetimes of the eight lock domain investigations depicted in a, as well as those of Q250A (cyan) and WT tus in the permissive orientation (purple square). Lifetime of the longest-lived exponential is shown (see **Fig. 8.10** for other lifetimes). Solid lines and circles are trends with *Ter*B, while dashed lines and triangles are trends with *Ter*B–GC6 flip. **(d)** the probability of entering the third, full-lock state (same color scheme as **c**). **(e)** The state-associated probabilities extracted (**SI, eqns. 8.4–8.13** and **Fig. 8.11**) for all Tus species on *Ter*B at 59 pN (purple, cyan, green, orange and blue bars represent WT, Q250A, E49K, F140A and H144A, respectively). in **c–e**, error bars represent the 1-σ confidence interval (CI).

This apparent increase in lock strength in the presence of an altered key indicated that mutation of F140 led to a decreased specificity for allowing only the C6 base into the lock.

Replacing H144 led to a more substantial decrease in Tus-*Ter* lock dwell times than resulted from the F140A mutation (**Fig. 8.4b,c**, blue circles). The dominant lifetime extracted at 59 pN was found to be ~2 s, whereas those of WT Tus and Q250A at the same force were two orders of magnitude higher. The data sets were found to exhibit lifetimes measurable over a wide range of forces (24–59 pN), and all retained three exponential states. At 59 pN, H144A led to a larger drop in the probability of entering the third, longest-lived state than that for F140A, from 73% for WT Tus to 0.7% for H144A (**Fig. 8.4e**, blue). As observed for F140A and WT Tus, there was also a clear force-dependence in the probability to form a fully locked state for H144A: the probability increased to 26% at 24 pN (**Fig. 8.4d**, blue line). While H144A is the single-site mutation with the greatest effect on lock lifetimes, the decrease was not as severe as that for WT Tus with the G6 *Ter* site. When combining the H144A lock mutant with the mutated *Ter* site (**Fig. 8.4a**, magenta key with blue lock), the resulting distribution (0.5 s at 19 pN; **Fig. 8.4c**, blue circle) was not unlike that obtained for permissively oriented WT Tus (0.8 s at 19 pN) (SI **Fig. 8.7h**, blue circles). This indicated that H144A–WT *Ter* lifetimes result from the specific interactions of the C6 base with the remaining amino acid residues in the lock pocket, and that further modification of the key within this altered lock result in the loss of all locking interactions.

8.2.7. E49K DECREASES THE PROBABILITY OF LOCK FORMATION

For E49K we obtained a distribution of dwell times at 59 pN that contained the same three exponentially distributed states as WT Tus-*Ter* (**Fig. 8.4b**, green circles), with a longest-lived state lifetime of 933 s (720 s for WT). However, the probability of this state decreased markedly from 73% for WT to 6% for E49K (**Fig. 8.4e**). The first two exponentials fully overlapped with the two shortest-lived ones for F140A (**Fig. 8.4b**, orange circles). This suggested that the DNA-phosphate interaction with E49 is crucial for guiding the C6 base into the Tus lock pocket to form the fully locked state. Mutating the *TerB* site (**Fig. 8.4a**, magenta key with green lock) caused a loss of almost all dwell times above ~1 s for forces above 26 pN, similar to the barrier imposed by Tus-*Ter* in the permissive orientation (SI **Fig. 8.7g**). However, closer inspection revealed longer-lived events with a low probability of ~1.5% (**Fig. 8.7h** and **Fig. 8.11**). When extrapolated to higher forces, the longer lifetimes (**Fig. 8.4c**, green triangles) resembled the much more probable states found for the interaction between WT Tus and the mutated *Ter* site in the 29–40 pN range (**Fig. 8.4c**, purple triangles, and **Fig. 8.11**). This indicated that while the WT lock domain continued to interact with the incorrect G6 key, mutation of E49 rendered such an interaction unlikely. Our observations clearly link the change invoked by E49K to a change in the probability of forming the third, fully locked state.

On the basis of our observations, we propose a kinetic model for WT lock formation involving three states (**Fig. 8.5b**). In this model, the Tus-*Ter* complex strengthens progressively and irreversibly upon passage from one stable state to the next until the final, fully locked and longest-lived state is reached. Fitting this kinetic model to the data revealed that loss of the long-lived lock state as the force is increased is not due to slower transitions to stronger lock states (**Fig. 8.5c**, k_{12} and k_{23}) but rather to increased rates of disruption (k_{10}, k_{20} and k_{30}) of the lock states, as the force-dependent trends in these rates showed.

8.3. DISCUSSION

Our results have important implications for understanding of how the Tus—*Ter* lock is formed. We can directly discard the notion that Tus—*Ter* requires specific interaction with DnaB to form a stable lock and block replication fork progression for extended times. Our results strongly suggest that strand separation followed by specific interaction of the *Ter* C6 base with the Tus lock domain is the only mechanism needed for polar

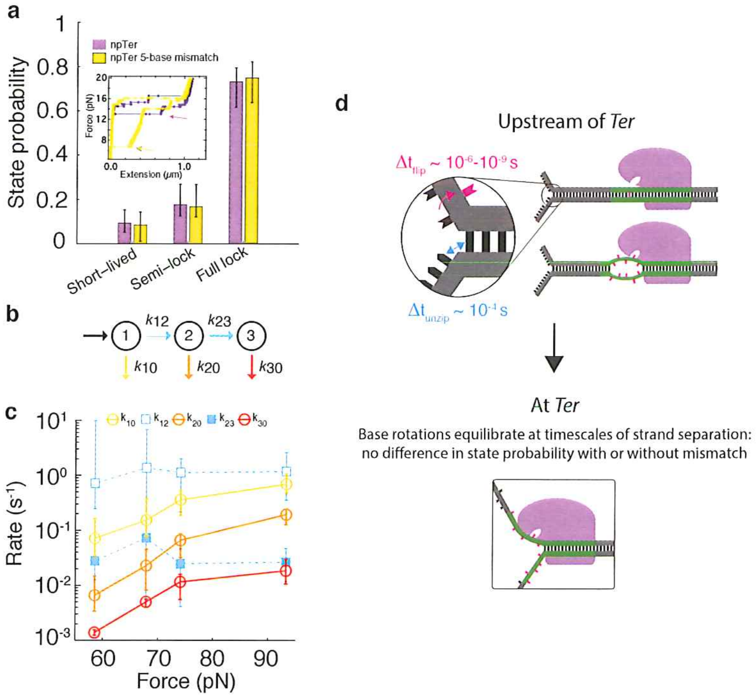


Figure 8.5: **Modeling Tus–*Ter* lock formation and extracting state-associated rates and probabilities. (a)** State probabilities extracted through fitting the dwell time distributions at 59 pN for Wt tus–*Ter* (purple, same data as **Fig. 3e**) and Wt tus on the *Ter* site containing the mismatched bases 3–7 (yellow). Inset, the force-extension curves of both types of hairpins showing that the force at which the mismatch hairpin returns to its fully closed state (yellow arrow, ~7 pN) consistently lies in the entropic regime (i.e., the forces at which SSDNA is no longer fully stretched), whereas the closing force lies markedly higher than the entropic regime for the normal hairpin (purple arrow, ~13 pN). Np, nonpermissive. Error bars indicate 1-σ confidence intervals. **(b)** The 3-state kinetic model used to fit the data sets (**Fig. 8.9**). **(c)** The force-dependent trends in the kinetic rates extracted by fitting the exponential states of Wt tus–*Ter* (**SI, eqns. 8.4–8.13** and **Table 8.11**). We observed an exponential increase in all state exit rates (k_{10}, k_{20}, k_{30}, represented by yellow, orange and red, respectively) with increasing force, while the rates connecting states 1 and 2 (cyan) as well as states 2 and 3 (cyan, fill) remain roughly constant (error bars indicate 1-σ confidence intervals). **(d)** Schematic representation of the estimated rate of base flipping versus the rate of base pair unwinding in our experiments for the normal *Ter* hairpin and the *Ter* site containing a mismatch region. The identical probability distributions for Wt tus–*Ter* and Wt tus on *Ter* with five mismatched bases shown in a are a result of the much faster process of base flipping compared to strand separation.

arrest. Evidence for this can be found in the fact that lock probabilities and lifetimes are affected by mutations in the lock domain or by mutation of the C6 base, but not by a mutation in the DNA-binding domain. We further observed that mutant E49K, which is hypothesized to be deficient in polar replication-fork arrest due to the elimination of specific protein-protein interactions[14, 15], gave rise to lifetimes identical to that of WT Tus, only now with a severely decreased probability of entering the longest-lived state. This tied the observed deficiency of in vivo fork arrest to the decrease in occurrence of the longest-lived state found in our experiments. In other words, the longest-lived state is likely to be the native lock conformation implicated in in vivo fork arrest, and we showed residue E49 to be part of the mechanism that facilitates formation of a tight interaction between C6 and the lock pocket.

Our experiments also demonstrated that the flipping of the C6 base and *Ter*'s subsequent interactions with the Tus lock pocket is not a rate-limiting step in the lock formation process. This is supported by our observation that there was little force-dependence in the inter-state rates (k_{12} and k_{23}) (**Fig. 8.5b,c**); this means that the force-dependence of state probabilities was caused solely by the force-dependent state exit rates (k_{10}, k_{20} and k_{30}). As force influences the speed of strand separation and thus the time available for C6 flipping while not affecting the inter-state transfer rates, the C6 flipping is likely not rate limiting at the comparatively low unwinding velocities (**Fig. 8.6c**) of the *E. coli* replisome.

This notion is further strengthened by our observation that preformation of the lock yielded a distribution of dwell times identical to those of normal Tus—*Ter* (**Fig. 8.5a**). In the preformed lock situation, it could be assumed that the C6 base has reached its equilibrium lock position. Our results imply that our pulling experiment provided sufficient time for this equilibration, although the pulling experiment as a whole is a system in nonequilibrium. This can be directly understood by comparing the typical timescales of DNA unwinding and the timescale of molecular rearrangement of the DNA bases upon disruption of Watson-Crick base pairing (**Fig. 8.5d**). The unwinding rate of DNA by a replisome is of the order of 1 kbp/s, and in our pulling experiments this rate is ~30 kbp/s. Typical molecular single-bond rotations are known to take place on femtosecond to picosecond timescales, with larger-scale motions such as lock formation likely occurring in the nanosecond to microsecond range[34–36]. This implies that there is at least several orders of magnitude of difference between the rate of unwinding and the rate at which flipping of C6 and concomitant molecular rearrangements take place, leaving ample time for C6 to reach its equilibrium-locked conformation.

The interpretation that the longest-lived state is the native, or full lock, state implies the probability of native lock formation is substantially lower than 100% at the highest forces measured (**Fig. 8.4d**, purple). The force-dependent lock probabilities do, however, suggest that the full lock is the dominant state at low forces for WT Tus, and suggest that the mutant with the lowest full-lock probability (H144A) might still have a significant chance of blocking an approaching replisome. Identifying a more direct link between our probabilities and *in vivo* arrest efficiencies would require knowledge of, for example, the amount of work performed by a replisome. It remains to be determined to what extent the two shortest-lived lock states are capable of causing arrest of DNA-processing enzymes, although the reported replisome arrest deficiency of E49K[14, 15] and our observation that E49K affected only the longest-lived state suggest that these intermediate states are not sufficient to block replication fork progression. It is clear, though, that these two 'lesser' lock states still pose a significant barrier to strand separation, much more so than the mere binding of Tus alone.

The difference between the high efficiency of reaching the full-lock state that we observed and the lower efficiencies of replisome arrest observed *in vivo*[13] must have a cause arising from interactions not captured in our experiments. These interactions might be invoked by the presence of an enzyme running into Tus—*Ter*, and steric effects (through functional protein-protein interactions) could then be the cause of the observed decrease in efficiency. Thus, instead of providing the basis of fork arrest, functional interactions could have an antagonistic effect in vivo. Our experiments with mutant E49K suggested a possible mechanism: just as the mutation in the Tus protein modulates the probability of forming the fully locked state without affecting the lifetime of the lock, an enzyme running into Tus-*Ter* could invoke a similar effect through functionally interacting with that same residue. Our experiments with WT and E49K Tus, respectively, thus set the upper (no interaction, high lock probability) and lower (E49 function completely disrupted, low lock probability) boundaries of blocking probabilities. Two different enzymes that run into nonpermissive Tus-*Ter* could then, in turn, have their own characteristic probabilities of being blocked due to the their differential modes of nonspecific interaction with Tus residues (such as E49) upon collision.

8.4. METHODS

8.4.1. DNA HAIRPINS

Plasmids pTER and pTER_{Rev}, containing the *TerB* site in either the nonpermissive or permissive orientation, respectively, and flanked by phage λ sequences, were obtained from Invitrogen. Plasmid $\text{pTER}_{\text{mutant}}$, encoding mutated *TerB* (C6–G6) was generated from pTER by site-directed mutagenesis using primers 1 and 2 (primer sequences are in listed in **Table 8.1**). Hairpins were constructed in a multistep process (**Fig. 8.6**). First, 1-kb fragments containing the *TerB* site were amplified from the three pTER plasmids using primers 3 and 4. These fragments were digested with the nonpalindromic restriction enzyme BsaI (New England BioLabs Inc., Ipswich, MA) and ligated at one end with a 42-bp oligonucleotide to form a U-turn (oligonucleotide 5). To create a 1-kb fragment containing a 5-base mismatch between bases 3–7 in the *Ter* site, two fragments of 500 bp were generated by PCR using pTER as template and primer combinations 3 and 12 and 4 and 13, respectively. These fragments were digested with BsaI and ligated to each end of the annealed primer pair 14 and 15 containing the wobble. Hairpin handles were created by PCR amplification of a 1.2-kb pBluescript SK+ (Stratagene–Agilent Technologies, Santa Clara, CA) fragment using primers 6 and 7 in the presence of either biotin-16-dUTP or digoxigenin-11-dUTP (Roche Diagnostics, Basel, Switzerland). Prior to ligation to spacer oligonucleotides, handles were digested with either *BamHI* or *NotI*. The upper spacer of the hairpin was generated by annealing 5'-phosphorylated oligonucleotides 8 and 9 and ligating this double-stranded DNA fragment to the NotI-digested biotin-labeled handle. The lower spacer was made by annealing 5'-phosphorylated primers 10 and 11 and ligating them to the BamHI-digested digoxigenin-labeled handle. Finally, the overhangs of these handle-spacer constructs were allowed to anneal to form a short (50-bp) stem with a 5'-GCAA overhang that was ligated to the complementary BsaI site of the 1-kb *TerB* fragment. Oligonucleotides were obtained from Biolegio B.V., Nijmegen, the Netherlands

Primer no.	Oligonucleotide sequence
1	5’-CACCACGACTGTGCTATAAAATAACTATGTTGTAACTAAAGTGGTTAATAT-3’
2	5’-ATATTAACCACTTTAGTTACAACATAGTTATTTTATAGCACAGTCGTGGTG-3’
3	5’-CTGCGGTCTCGTTGCTTACCGTCACCAGAAATTACCGTCAC-3’
4	5’-CCATCTTGGTCTCCTAGGTTTTTAGCAGCGAAGCGTTTGATAAG-3’
5	5’-CCTAAGCTCGCCGAGGCGAGCGAAAGCTCGCCTCGGCGAGCT-3’
6	5’-GACCGAGATAGGGTTGAGTG-3’
7	5’-CAGGGTCGGAACAGGAGAGC-3’
8	5’-GGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTT-3’
9	5’-GGCCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCATGCTCTTTACAACCGGTTGACTGCTTCAGGGGTCGATCCCGCTTTGTAC-3’
10	5’-GATCTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGC-3’
11	5’-GCAAGTACAAAGCGGGATCGACCCCTGAAGCAGTCAACCGGTTGTAAAGAGCATCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGA-3’
12	5’-CCATCTTGGTCTCCGACATTATAGCACAGTCGTGGTGAC-3’
13	5’-CTGCGGTCTCGAGGCGGTTAATATTATGGCGCGTTG-3’
14	5’-P-GCCTACTTTAGTTACAACATACTTATT-3’
15	5’-P-TGTCAAACCTCATGTTGTAACTAAAGT-3’

Table 8.1: Primer sequences for PCR amplifications and oligonucleotides that contribute to the hairpin structure

and from Ella Biotech GmbH, Martinsried, Germany.

8.4.2. TUS PROTEINS

N-terminally His6-tagged Tus and mutant derivatives were prepared as described [22, 23]; their concentrations were determined spectrophotometrically ($\epsilon280 = 39.700\mathrm{M}^{-1}\mathrm{cm}^{-1}$).

8.4.3. MAGNETIC TWEEZERS—EXPERIMENTAL CONFIGURATION.

The magnetic tweezers implementation used in this study has been described [31, 32, 37]. In short, light transmitted through the sample was collected by an oil-immersion objective (Olympus UPLSAPO60XO 60×, numerical aperture (NA) = 1.35, Olympus, USA) and projected onto a 12-megapixel CMOS camera (Falcon FA-80-12M1H, Teledyne Dalsa, Canada) with a sampling frequency of 58 Hz at full field of view, or higher when cropped. A 2-inch 200-mm tube lens between objective and camera made the effective magnification 67×. The applied magnetic field was generated by a pair of vertically aligned permanent neodymium-iron-boron magnets (SuperMagnete, Switzerland) separated by a distance of 1.0 or 0.5 mm and suspended on a motorized stage (M-126.PD2, Physik Instrumente, Germany) above the flow cell. Additionally, the magnet pair could be rotated about the illumination axis by an applied DC servo step motor (C-150.PD, Physik Instrumente, Germany).

8.4.4. DATA PROCESSING

Image processing of the collected light was used to track the real-time position of both surface-attached reference beads and superparamagnetic beads coupled to DNA tethers in three dimensions. We implemented custom written software in C++, CUDA and LabView (2011, National Instruments Corporation, USA) that is suited for high-throughput tracking in magnetic tweezers[37]. In short, tracking of the x, y coordinates is performed using center-of-mass computation followed by a further refinement using the quadrant interpolation algorithm. Localization of the bead's z-coordinate is achieved by creating a radial profile using the refined x, y coordinates and comparing this profile to a prerecorded LUT of radial profiles. After subtraction of the reference bead position to correct for instrumental drift, the x, y and z positions of the DNA-tethered beads were determined with a spatial accuracy of < 3 nm. The upward stretching forces on the DNA tethers by the superparamagnetic beads were calibrated from analysis of the extent of its Brownian motion whereby spectral corrections were employed to correct for camera blur and aliasing[38, 39].

8.4.5. SAMPLE PREPARATION AND DATA ACQUISITION

The sample preparation used in this study has been described in detail elsewhere (see Chapters 3 and 4).[32] In short, the DNA hairpins (final concentration ~ 50 pg/μl) were mixed and incubated for 2 min with 20 μl streptavidin-coated paramagnetic polystyrole beads (M270 Dynabeads) at room temperature in Tris buffer (50 mM Tris-HCl pH 7.9, 50 mM KCl, 0.1 mM EDTA, 0.01% Triton X-100). The supernatant was replaced by 50 μl Tris buffer followed by a 15 min incubation of the bead-DNA solution in the flow cell containing an anti-digoxigenin–coated nitrocellulose surface. Nontethered beads were removed by flushing with 1 ml Tris buffer, applying a high (30–40 pN) force while rotating the magnets (10 Hz), and followed by flushing with more buffer until all nontethered beads had been flushed out. All KCl buffers used in this study exclusively contained 50 mM Tris-HCl pH 7.9, 0.1 mM EDTA, 0.01% Triton X-100 unless noted otherwise. Tus proteins were diluted 1000-fold from stock (to ~ 10 nM) unless high salt concentrations required higher concentrations. Data were acquired at 100 Hz with a 10-ms acquisition time. Force-extension curves were obtained through changing the magnet position in an exponential fashion such that the force change was linear. Constant-force dwell time experiments were obtained by lowering the magnets in a linear fashion (10 mm/s) to the desired distance. The dwell time is the time measured between arrival of the magnets at their final position and the further opening of the hairpin from the locked to the fully opened state.

8.4.6. DATA ANALYSIS AND STATISTICAL PROCEDURE

Rupture of the Tus—*Ter* lock results in a sudden opening of the DNA hairpin: rupture points were easily identified as a sharp peak in the derivative of the z-trace. The dwell-time distribution

$$P(t) = \sum_{n=1}^{M} A_n e^{-k_n t} \tag{8.1}$$

with M number of exponentials (as determined by the Bayes–Schwarz information criterion [28]) is fit to the data set containing N experimentally collected dwell times t_i by minimizing the likelihood function[29]

$$L = -\sum_{i=1}^{N} \ln P(t_i) \tag{8.2}$$

with respect to rates and probabilistic weights (**SI, eqns. 8.1** and **8.2**). We calculated the errors in our parameter estimates by bootstrapping the system 1000 times, and reported the one-sigma confidence intervals (1-σ CI) among the bootstrapped data sets (**Figs. 8.10 and 8.11**).

8.5. ACKNOWLEDGEMENTS

We thank J. Cnossen for his countless efforts in adjusting the MT software to our needs, J. Kerssemakers for discussions and S. Hamdan for his contribution to the discussion by generously sharing his own experimental observations. This study was supported by a grant from the Australian Research Council (DP0877658) to N.E.D. and a Vici grant from the Netherlands Organisation for Scientific Research to N.H.D.

8.6. SUPPLEMENTARY INFORMATION

8.6.1. SALT-DEPENDENCE OF TUS-TER LOCK

As the reported dissociation constant (K_d) of the Tus-ds*TerB* complex has been shown to be highly salt-dependent, we investigated whether lock formation also exhibits a strong salt dependence. We observed that the fraction of rupture events recorded with a lifetime below our cutoff time of 10^{-2} s (i.e., the fraction of open hairpins at $t = 0$ s) increased from 0% at 50 mM to 14% at 350 mM KCl, while during these experiments care was taken to keep [Tus] well above (at least an order of magnitude) the reported salt- dependent K_d, thereby ensuring the continuous binding of Tus to *Ter*. Concomitantly, we observed that the lifetimes of the two longest-lived exponentials for wt Tus remain virtually unaffected when increasing the [KCl] from 50 to 350 mM, indicating that the lock strength is hardly affected by salt concentration (**Figs. 8.7d–f**). In contrast, the reported K_d of the Tus–ds*TerB* complex increases from $\sim 10^{-13}$ to $\sim 10^{-8}$ M within the 50 to 350 mM range. We conclude from this that the rate of lock formation is slightly affected by ionic screening, but once the lock is formed its strength remains unaffected. This is in accord with SPR data.

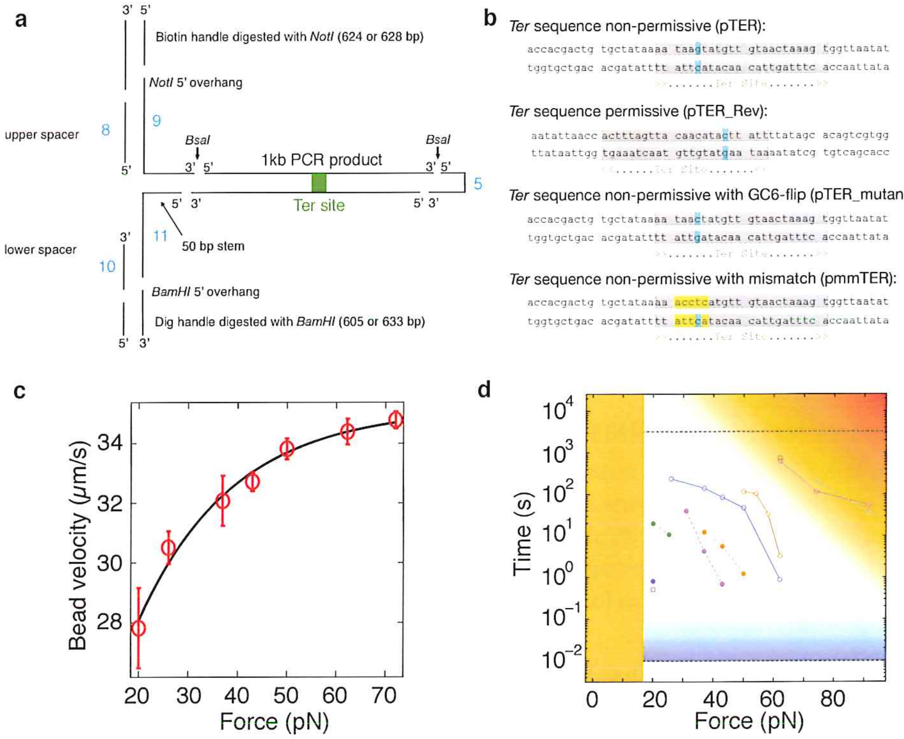


Figure 8.6: **DNA hairpin design and characteristics.** **(a)** The DNA hairpin construct made as described in the Online Methods. **(b)** The used *Ter* sequences. **(c)** Velocity versus force profile of a hairpin opening in the magnetic tweezers. We compute velocities by determining the maximum value of the central derivative of the extension versus time traces, i.e. the instantaneous apparent velocity upon lock rupture. Each data point in the figure is the average of hundreds of rupture events (the data here are from ~ 10^4 rupture events). The data have been fit with a single exponential (black line) to provide a guide to the eye. Note that our computations only provide a lower bound to the velocity, since our 100 Hz sampling frequency is not sufficiently high to capture the opening dynamics over a typical distance of ~ 0.6μm (500 bp opening). Nonetheless, these lower bounds suffice to indicate that the hairpin-opening rate exceeds the DNA unwinding rate of the E. coli replisome by at least 10-fold at 20 pN force. **(d)** Here we visualize the constraints on the experimental time–force window due to biological (orange) or instrumentation (blue) limits. The data, identical to Figure 3c, is added as a frame of reference. Below ~16 pN, base-paired DNA is energetically more favorable, therefore the hairpin remains closed (orange fill). With an acquisition rate of 100 Hz, the cutoff time is in principle 10^{-2} s (black dashed line); however, the error already becomes relatively large for lock lifetimes shorter than 0.1 s (blue gradient). Measurements are further limited by the lifetime of the DNA hairpin since DNA tethering relies on electrostatic interactions. This implies that very long measurement times, high forces or a combination of both (orange gradient) should be avoided. Typically we avoided having to measure lifetimes exceeding an hour (grey dashed line). Here we are able to see that the force–lifetime behavior exhibited by wt Tus already approaches the limits of the assay.

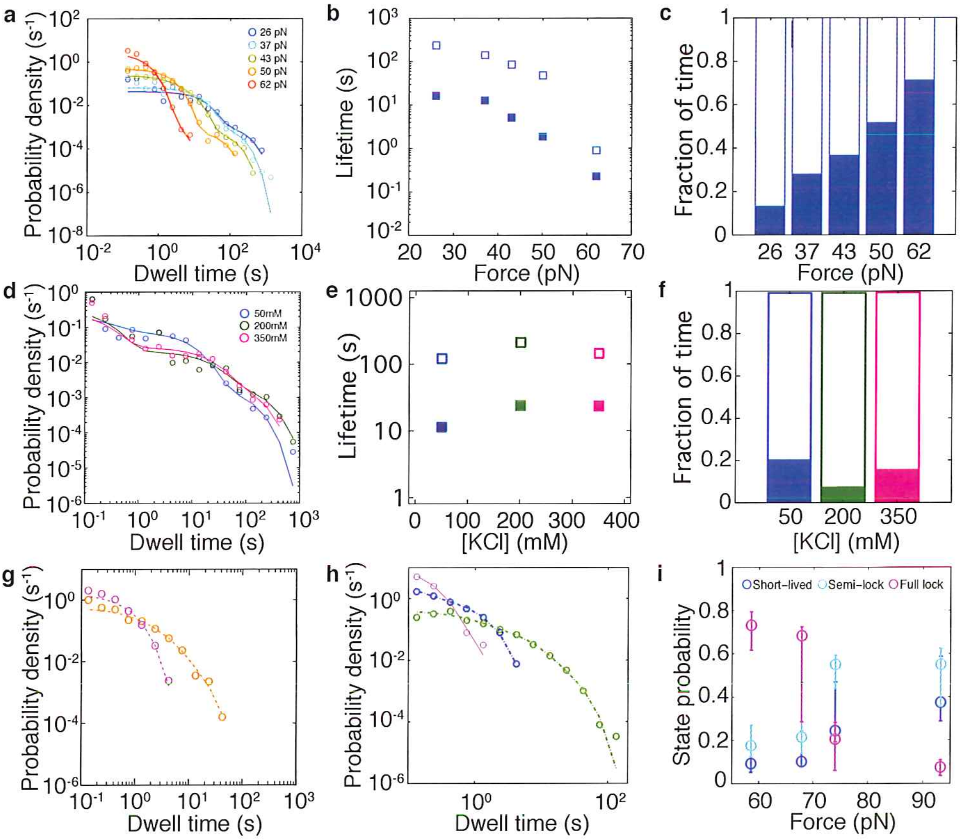


Figure 8.7: **Extended characteristics of Tus–*Ter* force-dependent lifetimes.** Distribution of lock lifetimes with varying force **(a–c)** or KCl concentration **(d–f**, see text below). The force-dependency was acquired with H144A Tus–*TerB* at 50 mM KCl, and the salt-dependency was acquired with wt Tus–*TerB* at 74 pN. **(a, d)** Probability densities (circles) including fits (solid lines). **(b, e)** The two longest-lived lifetimes extracted through fitting all datasets (filled squares are the second intermediate state, open squares represent the full lock state). **(c, f)** The normalized contribution of each of the fit states to the total experimental time (filled bar is the second state, open bar the full lock state). **(g, h)** Lifetime distributions obtained for the *TerB* GC flipped mutant and permissive orientation. **(g)** Fit probability density distributions of wt Tus and F140A with the GC-flipped *TerB* sequence at 40 pN (purple and orange, respectively). **(h)** Fit probability density distributions of bound-only wt Tus (i.e. in the permissive orientation, purple circles and solid line), H144A combined with the GC flipped *Ter* (blue), and E49K (green) with the GC flipped *Ter* at 19 pN. The lines represent fits to the data. **(i)** Trends in probabilities of all 3 lock states obtained through fitting our 3-state model to the wt Tus–*Ter* dataset.

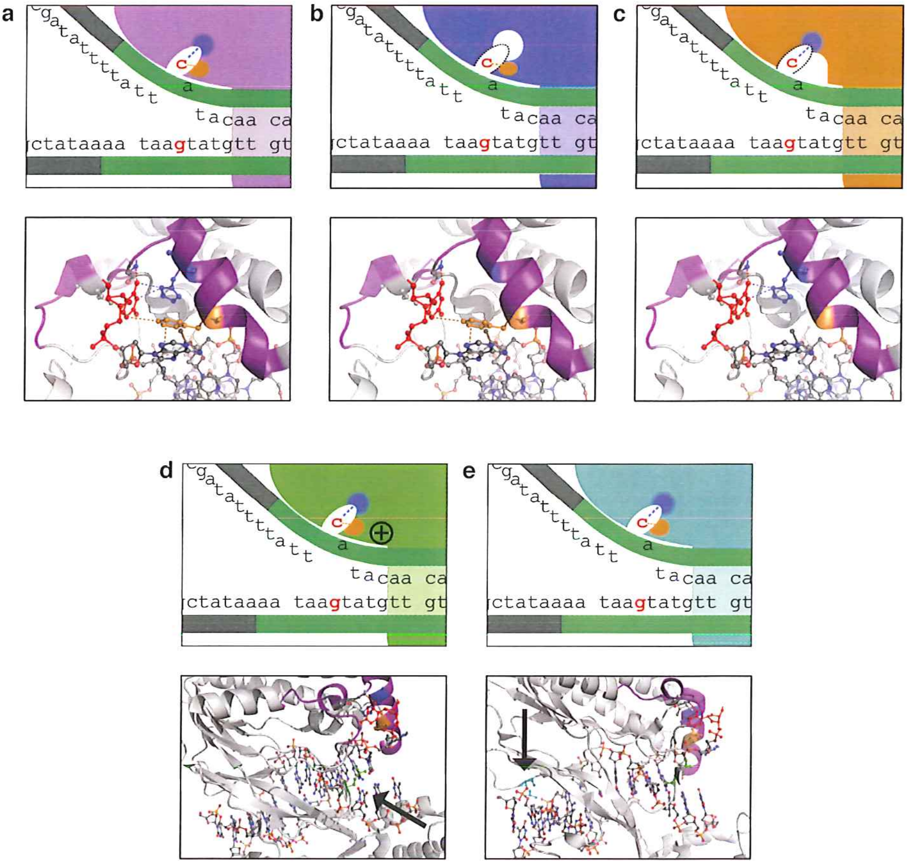


Figure 8.8: **Schematic representations of Tus mutations and the corresponding domains in the crystal structure.** Shown is the schematic representation of the lock domain of wt Tus (**a**, purple), inner lock domain mutant H144A (**b**, blue), edge lock domain mutant F140A (**c**, orange), outside lock domain mutant E49K (**d**, green) and binding domain mutant Q250A (**e**, cyan), as well as the relevant areas of the crystal structure directly below. Shown is how the C6 base (red) interacts with various amino acids of the lock domain (purple cartoon representation). Since amino acids E49 and Q250 are not part of the lock pocket both amino acids are indicated with an arrow.

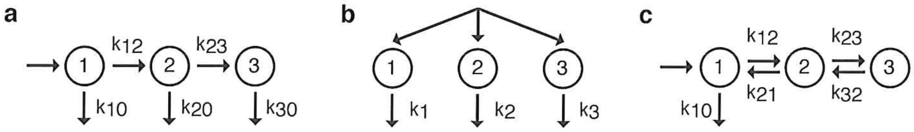


Figure 8.9: **Kinetic models that could fit the lifetime distributions.** Although there are many different three-state models that can fit our data, we have good reason to assume that our sequential model **(a)** is the simplest that can explain it. We consider it more likely that these substructures are on-pathway intermediate states towards a fully locked state instead of three completely independent structures induced by our pulling experiments, as explained in the Discussion section. By extracting the rates we find that only the exit rates show a dependence on force (**Fig. 8.5b,c**). Interestingly, the inter-state rates (k_{12} and k_{23}) are shielded from the force we subject the Tus–*Ter* system to in our assay. The fact that we observe an exponential dependency on force of only the off-rates implies that the force dependency of the probabilities (**Eqns. 8.14–8.17**) and lifetimes we observe only depend on the state exit rates. **Table 8.11** contains all the rates of the datasets presented in this work. Depending on the outcome of the Bayes-Schwartz information criterion, either a two or three exponential fit was used. **(b)** If sequential, progressive strengthening of states were not the case, a model with 3 independent states can also fit the data, yielding probabilities (the A's in Table 8.10) very similar to those obtained by fitting our sequential model (P's in Table 8.11). However, now all the parameters of this model display force-dependent trends, as we show that both the state lifetimes as well as the rates underlying the probabilities are force- dependent (**Fig. 8.7a–c** and **Fig. 8.4d**, respectively). **(c)** While there are many more three-state models that can fit our data, all add an additional complexity that cannot be verified by our experiments. This also holds true in the case of the reverse-exit model shown here. But since we are applying large forces to the system, we might modify the exit pathway out of the Tus–*Ter* lock state. While in vivo there might be a reverse order of exiting, our high pulling forces likely deform the energy landscape in such a way that other exit pathways also become available. Intuitively this makes sense since reverse exit implies a return of the *Ter* bases at the fork to their base-paired conformation – the forces applied in our experiments will always prevent this from happening.

8.6.2. RATES

The exponential fit has the general form of:

$$P(t) = \sum_{i=1}^{N} A_i e^{-k_i t} \tag{8.3}$$

where N is the number of exponentials determined by the BIC. In our kinetic model (**Fig. 8.9a**) the general rates (k_i) and probabilities (A_i) are expressed in terms of the five state associated rates, with

$$k_1 = k_{10} + k_{12} \tag{8.4}$$

$$k_2 = k_{20} + k_{23} \tag{8.5}$$

$$k_3 = k_{30} \tag{8.6}$$

$$A_1 = \frac{k_{10}^2 + k_{10}k_{12} - k_{10}k_{20} - k_{12}k_{20} - k_{10}k_{23} - k_{10}k_{30} + k_{20}k_{30} + k_{23}k_{30}}{(k_{10} + k_{12} - k_{20} - k_{23})(k_{10} + k_{12} - k_{30})} \tag{8.7}$$

$$A_2 = \frac{k_{12}k_{20} - k_{12}k_{30}}{(k_{10} + k_{12} - k_{20} - k_{23})(k_{20} + k_{23} - k_{30})} \tag{8.8}$$

$$A_3 = \frac{k_{12}k_{23}}{(k_{10} + k_{12} - k_{30})(-k_{20} - k_{23} + k_{30})} \tag{8.9}$$

for the 3-exponential model, and

$$k_1 = k_{10} + k_{12} \tag{8.10}$$

$$k_2 = k_{20} \tag{8.11}$$

$$A_1 = \frac{k_{10} - k_{20}}{k_{10} + k_{12} - k_{20}} \tag{8.12}$$

$$A_2 = \frac{k_{12}}{k_{10} + k_{12} - k_{20}} \tag{8.13}$$

for the 2-exponential model.

$$P(1) = \frac{k_{10}}{k_{10} + k_{12}} \tag{8.14}$$

$$P(2) = \frac{k_{12}}{k_{10} + k_{12}} \qquad \text{(2-exp fit)} \tag{8.15}$$

$$P(2) = \frac{k_{12}}{k_{10} + k_{12}} \cdot \frac{k_{20}}{k_{20} + k_{23}} \qquad \text{(3-exp fit)} \tag{8.16}$$

$$P(3) = \frac{k_{12}}{k_{10} + k_{12}} \cdot \frac{k_{23}}{k_{20} + k_{23}} \tag{8.17}$$

Dataset	k1 (1/s)	k2 (1/s)	k3 (1/s)	A1	A2	A3	τ1 (s)	τ2 (s)	τ3 (s)
E49K 50mM 59 pN	1.3974 1.0512 1.8991	0.2432 0.2062 0.2793	0.0011 0.0009 0.0013	0.4254 0.3463 0.5354	0.5055 0.3969 0.5865	0.0691 0.0521 0.0831	0.72 0.53 0.95	4.11 3.58 4.85	932.63 772.62 1097.00
E49KGCf 50mM 19 pN	1.3655 0.7784 2.0261	0.2237 0.1411 0.3062	0.0646 0.0430 0.0774	0.1912 0.1253 0.3331	0.5535 0.4822 0.6200	0.2554 0.0898 0.3741	0.73 0.49 1.28	4.47 3.27 7.09	15.47 12.92 23.24
E49KGCf 50mM 24 pN	1.7844 1.7052 2.0610	0.4405 0.3415 1.5680	0.0519 0.0368 0.0919	0.9070 0.0007 0.9301	0.0748 0.0543 0.9410	0.0182 0.0103 0.0348	0.56 0.49 0.59	2.27 0.64 2.93	19.26 10.88 27.16
F140A 50mM 47 pN	7.3335 5.9305 12.7417	0.0731 0.0404 0.3144	0.0086 0.0060 0.0116	0.3966 0.3340 0.5092	0.2549 0.1442 0.3839	0.3486 0.1902 0.4282	0.14 0.08 0.17	13.68 3.18 24.75	116.29 85.93 167.64
F140A 50mM 50 pN	1.6428 1.2645 73.4507	0.0536 0.0457 0.1169	0.0097 0.0072 0.0183	0.2275 0.1837 0.9808	0.5843 0.0085 0.6500	0.1882 0.0096 0.2821	0.61 0.01 0.79	18.67 8.55 21.89	103.08 54.65 139.58
F140A 50mM 54 pN	19.6676 9.8855 31.1323	0.1630 0.1476 0.1933	0.0298 0.0214 0.0463	0.3799 0.2361 0.6245	0.5430 0.3010 0.6643	0.0771 0.0391 0.1347	0.05 0.03 0.10	6.13 5.17 6.77	33.61 21.61 46.62
F140A 50mM 59 pN	4.7211 3.4487 6.6327	0.3507 0.3168 0.3873	0.0181 0.0153 0.0216	0.3139 0.2766 0.3735	0.6671 0.6058 0.7035	0.0190 0.0107 0.0287	0.21 0.15 0.29	2.85 2.58 3.16	55.27 46.34 65.34
F140AGCf 50mM 35 pN	0.5269 0.3206 0.9963	0.0799 0.0509 0.1070		0.5382 0.3576 0.7469	0.4618 0.2526 0.6421		1.90 1.00 3.12	12.52 9.35 19.63	
F140AGCf 50mM 40 pN	1.0714 0.7378 2.8859	0.1757 0.1340 0.2331		0.5317 0.3418 0.6837	0.4683 0.3163 0.6575		0.93 0.35 1.36	5.69 4.29 7.46	
F140AGCf 50mM 47 pN	6.1670 5.1176 7.3741	0.8115 0.7731 0.8504		0.2588 0.2269 0.2986	0.7412 0.7011 0.7725		0.16 0.14 0.20	1.23 1.18 1.29	
H144A 50mM 24 pN	14.9928 0.4369 100.0000	0.0579 0.0428 0.0691	0.0042 0.0034 0.0054	0.1094 0.0725 0.9981	0.6081 0.0013 0.6432	0.2825 0.0005 0.3258	0.07 0.01 2.29	17.29 14.46 23.36	238.43 184.54 290.91
H144A 50mM 35 pN	100.0000 0.0897 100.0000	0.0779 0.0137 0.0802	0.0071 0.0020 0.0092	0.9952 0.0778 0.9927	0.0039 0.0057 0.7413	0.0009 0.0016 0.1938	0.01 0.01 11.15	12.84 12.47 72.77	141.63 108.65 488.25
H144A 50mM 40 pN	0.9332 0.6735 14.1291	0.1513 0.1363 0.1832	0.0101 0.0087 0.0126	0.2001 0.1235 0.3013	0.7238 0.6263 0.7954	0.0761 0.0572 0.0915	1.07 0.07 1.48	6.61 5.46 7.34	98.74 79.21 114.57
H144A 50mM 47 pN	99.0198 0.5996 100.0000	0.5380 0.3750 0.5335	0.0211 0.0164 0.0255	1.22E-04 0.0009 0.9416	0.9645 0.0481 0.9587	0.0354 0.0195 0.0418	0.01 0.01 1.67	1.86 1.87 2.67	47.42 39.23 60.86
H144A 50mM 59 pN	6.0193 5.3403 99.7892	2.3191 1.9081 4.5799	0.4642 0.3036 1.6819	0.6550 0.0009 0.7652	0.3352 0.2067 0.7556	0.0098 0.0037 0.1306	0.17 0.01 0.19	0.43 0.22 0.52	2.15 0.59 3.29
H144AGCf 50mM 19 pN	10.8718 6.0554 31.3382	1.2214 1.1389 1.2910		0.2124 0.1901 0.5849	0.7876 0.4051 0.8098		0.09 0.03 0.17	0.82 0.77 0.88	
Q250A 200mM 59pN	50.9000 4.5013 67.2385	0.0925 0.0780 0.1032	0.0046 0.0038 0.0057	0.8098 0.4415 0.9364	0.1540 0.0266 0.4545	0.0362 0.0042 0.1344	0.02 0.01 0.22	10.81 9.69 12.82	219.54 176.19 262.00
Q250A 50mM 59 pN	7.6697 0.2996 10.6434	0.0657 0.0302 0.1375	0.0043 0.0035 0.0051	0.0928 0.0720 0.2221	0.3399 0.2205 0.4223	0.5673 0.4427 0.6560	0.13 0.09 3.34	15.23 7.27 33.13	232.29 194.67 286.12
Q250A 50mM 93 pN	100.0000 0.9019 100.0000	0.7251 0.1305 0.7635	0.0530 0.0196 0.0675	0.9980 0.3342 0.9985	0.0016 0.0012 0.5328	0.0004 0.0003 0.1073	0.01 0.01 1.11	1.38 1.31 7.67	18.85 14.82 50.97
WT 200mM 74 pN	13.7971 0.7918 25.0195	0.0486 0.0252 0.2076	0.0048 0.0036 0.0066	0.2245 0.1212 0.4298	0.3365 0.1575 0.4679	0.4390 0.2483 0.5759	0.07 0.04 1.26	20.56 4.82 39.69	206.86 151.47 279.95
WT 200mM 93 pN	1.0919 0.9192 1.7879	0.1625 0.1472 0.6839	0.0311 0.0224 0.0957	0.4163 0.2012 0.4782	0.5291 0.4341 0.6032	0.0547 0.0311 0.3634	0.92 0.56 1.09	6.16 1.46 6.79	32.19 10.45 44.71
WT 350mM 74 pN	12.6376 8.8062 17.8469	0.0469 0.0395 0.0576	0.0071 0.0063 0.0082	0.1848 0.1076 0.3063	0.4464 0.3402 0.5134	0.3687 0.2751 0.4523	0.08 0.06 0.11	21.31 17.35 25.32	140.11 122.26 157.79
WT-mmTer 50mM 59 pN	0.3550 0.1608 3.3106	0.0247 0.0098 0.0909	0.0018 0.0017 0.0021	0.0998 0.0250 0.1608	0.2101 0.1529 0.3477	0.6900 0.5258 0.7925	2.82 0.30 6.22	40.51 11.00 101.61	541.90 481.12 600.87
WT 50mM 59 pN	0.7843 0.2804 16.0718	0.0343 0.0175 0.1149	0.0014 0.0012 0.0016	0.0843 0.0454 0.1429	0.1514 0.1021 0.2301	0.7643 0.6691 0.8218	1.27 0.06 3.57	29.16 8.70 57.16	720.22 636.56 800.91
WT 50mM 68 pN	1.5233 0.4821 7.3087	0.0956 0.0143 0.2429	0.0050 0.0040 0.0056	0.0874 0.0649 0.2198	0.1887 0.1221 0.3548	0.7239 0.3953 0.7643	0.66 0.14 2.07	10.46 4.12 70.08	199.68 178.31 250.41
WT 50mM 74 pN	23.7187 0.2335 44.2823	0.1055 0.0495 0.1323	0.0087 0.0051 0.0129	0.3150 0.1597 0.7889	0.5152 0.1345 0.6289	0.1697 0.0323 0.2382	0.04 0.02 4.28	9.48 7.56 20.19	114.44 77.53 194.85
WT 50mM 93 pN	1.8677 0.8525 3.6589	0.2205 0.1434 0.2883	0.0185 0.0107 0.0300	0.3008 0.2281 0.5206	0.6167 0.4123 0.6732	0.0825 0.0396 0.1219	0.54 0.27 1.17	4.53 3.47 6.98	54.00 33.30 93.39
WTGCf 50mM 29 pN	0.3818 0.3598 0.4072	0.0250 0.0198 0.0332		0.9539 0.9368 0.9689	0.0461 0.0310 0.0631		2.62 2.46 2.78	40.00 30.15 50.53	
WTGCf 50mM 35 pN	1.0263 0.9606 1.4694	0.2331 0.1381 0.5954		0.9370 0.5704 0.9713	0.0630 0.0286 0.4233		0.97 0.68 1.04	4.29 1.68 7.24	
WTGCf 50mM 40 pN	3.6937 3.2017 4.4938	1.4405 1.2980 1.5661		0.4913 0.3824 0.6074	0.5087 0.3924 0.6168		0.27 0.22 0.31	0.69 0.64 0.77	
WTperm 50mM 19pN	7.1863 6.8488 100.0000	1.9324 1.5950 5.4160		0.9528 0.9176 0.9959	0.0472 0.0039 0.0823		0.14 0.01 0.15	0.52 0.18 0.63	

Figure 8.10: **The fit parameters to generate the fits ($k1$–3 and $A1$-3) as well as the lifetimes associated with each state ($\tau 1$-3; the inverse of the respective ks).** Lower and upper 1-σ CIs are shown left and right below each value respectively.

Dataset	k10 (1/s)	k12 (1/s)	k20 (1/s)	k23 (1/s)	k30 (1/s)	P(1)	P(2)	P(3)
E49K 50mM 59 pN	0.7175 0.6178 0.8735	0.6799 0.4164 1.0609	0.2088 0.1683 0.2431	0.0344 0.0260 0.0431	0.0011 0.0009 0.0013	0.5134 0.4313 0.6088	0.4178 0.3189 0.5016	0.0687 0.0518 0.0825
E49KGCf 50mM 19 pN	0.4013 0.3370 0.4534	0.9642 0.4289 1.5777	0.1689 0.1259 0.2018	0.0548 0.0145 0.1071	0.0646 0.0430 0.0774	0.2939 0.2210 0.4419	0.5331 0.4632 0.5924	0.1730 0.0594 0.2666
E49KGCf 50mM 24 pN	1.6523 1.5699 1.7387	0.1321 0.0962 0.2959	0.3476 0.2455 0.9245	0.0929 0.0446 0.1390	0.0519 0.0368 0.0919	0.9260 0.8486 0.9445	0.0584 0.0425 0.1294	0.0156 0.0087 0.0311
F140A 50mM 47 pN	2.9297 2.0672 5.9511	4.4037 3.5706 6.4214	0.0357 0.0262 0.0896	0.0374 0.0125 0.2314	0.0086 0.0060 0.0116	0.3995 0.3375 0.5115	0.2933 0.1539 0.4174	0.3072 0.1632 0.3969
F140A 50mM 50 pN	0.4068 0.3102 70.4627	1.2360 0.5425 2.0994	0.0427 0.0369 0.0697	0.0109 0.0063 0.0462	0.0097 0.0072 0.0183	0.2476 0.2049 0.9809	0.5992 0.0083 0.6533	0.1532 0.0085 0.2249
F140A 50mM 54 pN	7.5623 2.3778 19.4368	12.1052 6.9996 12.8166	0.1464 0.1351 0.1633	0.0167 0.0104 0.0330	0.0298 0.0214 0.0463	0.3845 0.2472 0.6264	0.5526 0.3180 0.6683	0.0629 0.0329 0.1050
F140A 50mM 59 pN	1.7161 1.3277 2.3275	3.0050 2.0426 4.3437	0.3408 0.3072 0.3759	0.0099 0.0057 0.0156	0.0181 0.0153 0.0216	0.3635 0.3258 0.4231	0.6185 0.5566 0.6532	0.0180 0.0102 0.0271
F140AGCf 50mM 35 pN	0.3205 0.2516 0.4502	0.2064 0.0661 0.5809	0.0799 0.0509 0.1070			0.6082 0.4230 0.7902	0.3918 0.2094 0.5746	
F140AGCf 50mM 40 pN	0.6519 0.5387 1.1811	0.4195 0.1928 1.7212	0.1757 0.1340 0.2331			0.6085 0.3942 0.7371	0.3915 0.2628 0.6057	
F140AGCf 50mM 47 pN	2.1976 1.8948 2.5662	3.9695 3.1183 4.8638	0.8115 0.7731 0.8504			0.3563 0.3258 0.3966	0.6437 0.6031 0.6741	
H144A 50mM 24 pN	1.6761 0.0597 99.8145	13.3167 0.1211 11.2308	0.0408 0.0284 0.0472	0.0171 0.0101 0.0244	0.0042 0.0034 0.0054	0.1118 0.0948 0.9981	0.6262 0.0014 0.6512	0.2620 0.0005 0.2954
H144A 50mM 35 pN	99.5198 0.0692 98.7552	0.4802 0.0206 3.7824	0.0645 0.0123 0.0654	0.0133 0.0012 0.0164	0.0071 0.0020 0.0092	0.9952 0.1038 0.9927	0.0040 0.0060 0.7292	0.0008 0.0014 0.1699
H144A 50mM 40 pN	0.2970 0.2693 2.1655	0.6362 0.3850 10.4933	0.1357 0.1214 0.1672	0.0156 0.0120 0.0193	0.0101 0.0087 0.0126	0.3183 0.1498 0.4269	0.6115 0.5092 0.7732	0.0702 0.0527 0.0848
H144A 50mM 47 pN	0.5317 0.5321 0.6938	98.4881 0.0348 99.1527	0.5196 0.1296 0.5133	0.0184 0.0133 0.0338	0.0211 0.0164 0.0255	0.0054 0.0066 0.9521	0.9606 0.0454 0.9545	0.0340 0.0144 0.0392
H144A 50mM 59 pN	4.7245 4.2569 5.4003	1.2947 0.7782 93.3624	2.2411 1.7924 3.7682	0.0780 0.0272 0.5556	0.4642 0.3036 1.6819	0.7849 0.0566 0.8541	0.2079 0.1179 0.6938	0.0072 0.0026 0.0699
H144AGCf 50mM 19 pN	3.2712 2.2090 16.4183	7.6006 3.0051 11.3265	1.2214 1.1389 1.2910			0.3009 0.2984 0.6109	0.6991 0.3880 0.7015	
Q250A 200mM 59pN	41.2326 0.1296 62.9968	9.6673 0.2685 15.1011	0.0757 0.0607 0.0850	0.0168 0.0105 0.0227	0.0046 0.0038 0.0057	0.8101 0.4522 0.9365	0.1555 0.0302 0.4397	0.0345 0.0015 0.1114
Q250A 50mM 59 pN	0.7363 0.0685 1.3946	6.9334 0.2295 9.2923	0.0272 0.0120 0.0419	0.0385 0.0158 0.0960	0.0043 0.0035 0.0051	0.0960 0.0754 0.2423	0.3742 0.2475 0.4587	0.5298 0.3921 0.6220
Q250A 50mM 93 pN	99.8000 0.6705 99.8517	0.2000 0.1319 1.8363	0.5982 0.1116 0.6188	0.1269 0.0141 0.1521	0.0530 0.0196 0.0675	0.9980 0.4295 0.9985	0.0016 0.0012 0.4628	0.0003 0.0002 0.0938
WT 200mM 74 pN	3.1162 0.1255 9.2143	10.6808 0.6257 14.0952	0.0238 0.0134 0.0460	0.0248 0.0091 0.1560	0.0048 0.0036 0.0066	0.2259 0.1321 0.4473	0.3789 0.1766 0.5026	0.3952 0.2161 0.5324
WT 200mM 93 pN	0.5422 0.4687 0.6504	0.5497 0.4218 1.0533	0.1486 0.1302 0.1914	0.0139 0.0081 0.2583	0.0311 0.0224 0.0957	0.4965 0.3657 0.5605	0.4605 0.3760 0.5497	0.0429 0.0252 0.2822
WT 350mM 74 pN	2.3591 0.9963 4.6580	10.2785 7.3901 13.0933	0.0289 0.0243 0.0342	0.0180 0.0128 0.0259	0.0071 0.0063 0.0082	0.1867 0.1142 0.3076	0.5009 0.3946 0.5647	0.3125 0.2307 0.3847
WT-mmTer 50mM 59 pN	0.0354 0.0220 0.0819	0.3196 0.1340 3.1424	0.0058 0.0031 0.0149	0.0189 0.0061 0.0759	0.0018 0.0017 0.0021	0.0844 0.0113 0.1427	0.1659 0.1212 0.2675	0.7497 0.6378 0.8222
WT 50mM 59 pN	0.0724 0.0348 0.1660	0.7120 0.2442 14.2307	0.0066 0.0034 0.0147	0.0277 0.0127 0.0925	0.0014 0.0012 0.0016	0.0923 0.0514 0.1529	0.1756 0.1250 0.2689	0.7321 0.6171 0.7943
WT 50mM 68 pN	0.1548 0.0787 0.3782	1.3685 0.3884 6.7703	0.0228 0.0082 0.0454	0.0728 0.0059 0.1986	0.0050 0.0040 0.0056	0.1016 0.0787 0.2401	0.2147 0.1402 0.4683	0.6837 0.2857 0.7249
WT 50mM 74 pN	7.5284 0.1225 31.3348	16.1903 0.1107 15.5270	0.0815 0.0320 0.0958	0.0241 0.0083 0.0388	0.0087 0.0051 0.0129	0.3174 0.1700 0.7980	0.5270 0.1516 0.6338	0.1556 0.0264 0.2113
WT 50mM 93 pN	0.6994 0.4852 1.0736	1.1683 0.3525 2.5548	0.1942 0.1275 0.2459	0.0264 0.0117 0.0462	0.0185 0.0107 0.0300	0.3745 0.2877 0.5864	0.5508 0.3557 0.6237	0.0748 0.0356 0.1077
WTGCf 50mM 29 pN	0.3654 0.3439 0.3888	0.0164 0.0109 0.0228	0.0250 0.0198 0.0332			0.9569 0.9421 0.9711	0.0431 0.0289 0.0579	
WTGCf 50mM 35 pN	0.9763 0.9197 1.1469	0.0500 0.0244 0.3774	0.2331 0.1381 0.5954			0.9513 0.7458 0.9754	0.0487 0.0246 0.2529	
WTGCf 50mM 40 pN	2.5475 2.3883 2.7553	1.1462 0.7434 1.7795	1.4405 1.2980 1.5661			0.6897 0.6008 0.7686	0.3103 0.2313 0.3976	
WTperm 50mM 19pN	6.9384 6.5977 99.5651	0.2479 0.1189 0.4407	1.9324 1.5950 5.4160			0.9655 0.9421 0.9968	0.0345 0.0032 0.0578	

Figure 8.11: **Overview of extracted kinetic rates and probabilities.** The probabilities are calculated from the extracted rates using eqns. 8.14–8.17. Lower and upper 1-σ CIs are shown left and right below each value respectively.

REFERENCES

[1] B. A. Berghuis, D. Dulin, Z.-Q. Xu, T. van Laar, B. Cross, R. Janissen, S. Jergic, N. E. Dixon, M. Depken, and N. H. Dekker, *Strand separation establishes a sustained lock at the tus-ter replication fork barrier,* Nat Chem Biol **11**, 579 (2015).

[2] T. M. Hill, J. M. Henson, and P. L. Kuempel, *The terminus region of the Escherichia coli chromosome contains two separate loci that exhibit polar inhibition of replication,* PNAS **84**, 1754 (1987).

[3] M. Hidaka, T. Kobayashi, S. Takenaka, H. Takeya, and T. Horiuchi, *Purification of a DNA replication terminus (ter) site-binding protein in Escherichia coli and identification of the structural gene,* **264**, 21031 (1989).

[4] T. M. Hill, *Arrest of Bacterial DNA Replication,* Annual review of microbiology **46**, 603 (1992), 10.1146/annurev.mi.46.100192.003131.

[5] K. Kamada, T. Horiuchi, K. Ohsumi, and N. Shimamoto, *Structure of a replication-terminator protein complexed with DNA,* Nature **383**, 598 (1996).

[6] G. S. Khatri, T. MacAllister, P. R. Sista, and D. Bastia, *The replication terminator protein of E. coli is a DNA sequence-specific contra-helicase,* Cell **59**, 667 (1989).

[7] T. M. Hill and K. J. Marians, *Escherichia coli Tus protein acts to arrest the progression of DNA replication forks in vitro,* PNAS **87**, 2481 (1990).

[8] T. M. Hill, *Escherichia coli and Salmonella: Cellular and Molecular Biology* (1996) pp. 1602–1614.

[9] F. F. Coskun-Ari, A. Skokotas, G. R. Moe, and T. M. Hill, *Biophysical characteristics of Tus, the replication arrest protein of Escherichia coli,* **269**, 4027 (1994).

[10] C. Neylon, A. V. Kralicek, T. M. Hill, and N. E. Dixon, *Replication Termination in Escherichia coli: Structure and Antihelicase Activity of the Tus-Ter Complex,* Microbiology and Molecular Biology Reviews **69**, 501 (2005).

[11] J. Y. Lee, I. J. Finkelstein, L. K. Arciszewska, D. J. Sherratt, and E. C. Greene, *Single-Molecule Imaging of FtsK Translocation Reveals Mechanistic Features of Protein-Protein Collisions on DNA,* Molecular Cell **54**, 832 (2014).

[12] S. Kaul, B. K. Mohanty, T. Sahoo, I. Patel, S. A. Khan, and D. Bastia, *The replication terminator protein of the gram-positive bacterium Bacillus subtilis functions as a polar contrahelicase in gram-negative Escherichia coli.* PNAS **91**, 11143 (1994), 10.1073/pnas.91.23.11143.

[13] P. A. Andersen, A. A. Griffiths, I. G. Duggin, and R. G. Wake, *Functional specificity of the replication fork-arrest complexes of Bacillus subtilis and Escherichia coli: significant specificity for Tus–Ter functioning in E. coli,* Molecular Microbiology **36**, 1327 (2000).

[14] T. Sahoo, B. K. Mohanty, M. Lobert, A. C. Manna, and D. Bastia, *The Contrahelicase Activities of the Replication Terminator Proteins of Escherichia coli and Bacillus subtilis Are Helicase-specific and Impede both Helicase Translocation and Authentic DNA Unwinding,* Journal of Biological Chemistry **270**, 29138 (1995).

[15] S. Mulugu, A. Potnis, J. Taylor, K. Alexander, and D. Bastia, *Mechanism of termination of dna replication of* Escherichia coli *involves helicase-contrahelicase interaction,* Proceedings of the National Academy of Sciences of the United States of America **98**, 9569 (2001).

[16] B. K. Mohanty, T. Sahoo, and D. Bastia, *The relationship between sequence-specific termination of DNA replication and transcription,* Embo Journal **15**, 2530 (1996).

[17] B. K. Mohanty, T. Sahoo, and D. Bastia, *Mechanistic Studies on the Impact of Transcription on Sequence-specific Termination of DNA Replication and Vice Versa,* Journal of Biological Chemistry **273**, 3051 (1998).

[18] E. H. Lee, A. Kornberg, M. Hidaka, T. Kobayashi, and T. Horiuchi, *Escherichia coli replication termination protein impedes the action of helicases,* PNAS **86**, 9104 (1989).

[19] E. H. Lee and A. Kornberg, *Features of replication fork blockage by the Escherichia coli terminus-binding protein,* **267**, 8778 (1992).

[20] C. L. Bedrosian and D. Bastia, *Escherichia coli replication terminator protein impedes simian virus 40 (SV40) DNA replication fork movement and SV40 large tumor antigen helicase activity in vitro at a prokaryotic terminus sequence.* PNAS **88**, 2618 (1991).

[21] M. Hidaka, T. Kobayashi, Y. Ishimi, M. Seki, T. Enomoto, M. Abdel-Monem, and T. Horiuchi, *Termination complex in Escherichia coli inhibits SV40 DNA replication in vitro by impeding the action of T antigen helicase,* **267**, 5361 (1992).

[22] M. D. Mulcair, P. M. Schaeffer, A. J. Oakley, H. F. Cross, C. Neylon, T. M. Hill, and N. E. Dixon, *A molecular mousetrap determines polarity of termination of dna replication in e. coli,* Cell **125**, 1309 (2006).

[23] C. Neylon, S. E. Brown, A. V. Kralicek, C. S. Miles, C. A. Love, and N. E. Dixon, *Interaction of the Escherichia coliReplication Terminator Protein (Tus) with DNA: A Model Derived from DNA-Binding Studies of Mutant Proteins by Surface Plasmon Resonance †,* Biochemistry **39**, 11989 (2000).

[24] D. Bastia, S. Zzaman, G. Krings, M. Saxena, X. Peng, and M. M. Greenberg, *Replication termination mechanism as revealed by Tus-mediated polar arrest of a sliding helicase,* Proceedings of the National Academy of Sciences **105**, 12831 (2008).

[25] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco Jr, and C. Bustamente, *Reversible Unfolding of Single RNA Molecules by Mechanical Force,* Science **292**, 733 (2001).

[26] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, *Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins,* PNAS **103**, 6190 (2006).

[27] T. Lionnet, M. M. Spiering, S. J. Benkovic, D. Bensimon, and V. Croquette, *Real-time observation of bacteriophage T4 gp41 helicase reveals an unwinding mechanism,* Proceedings of the National Academy of Sciences of the United States of America **104**, 19790 (2007).

[28] G. Schwarz, *Estimating the Dimension of a Model,* The Annals of Statistics **6**, 461 (1978), 10.1214/aos/1176344136.

[29] S. Cowan, *Statistical Data Analysis* (Oxford University Press, 1998).

[30] W. Press, B. Flannery, S. Teukolsky, and W. T. Vetterling, *NUMERICAL RECIPES IN C: THE ART OF SCIENTIFIC COMPUTING* (Cambridge University Press, 1992).

[31] D. Dulin, I. D. Vilfan, B. A. Berghuis, S. Hage, D. H. Bamford, M. M. Poranen, M. Depken, and N. H. Dekker, *Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase,* Cell Reports **10**, 983 (2015).

[32] R. Janissen, B. A. Berghuis, D. Dulin, M. Wink, T. van Laar, and N. H. Dekker, *Invincible DNA tethers: covalent DNA anchoring for enhanced temporal and force stability in magnetic tweezers experiments,* Nucleic Acids Research **42**, e137 (2014).

[33] F. F. Coskun-Ari and T. M. Hill, *Sequence-specific Interactions in the Tus-Ter Complex and the Effect of Base Pair Substitutions on Arrest of DNA Replication in Escherichia coli,* Journal of Biological Chemistry **272**, 26448 (1997).

[34] M. Dantus, R. M. Bowman, and A. H. Zewail, *Femtosecond laser observations of molecular vibration and rotation,* Nature **343**, 737 (1990).

[35] Y. Arasaki, K. Takatsuka, K. Wang, and V. McKoy, *Energy- and angle-resolved pump–probe femtosecond photoelectron spectroscopy: Molecular rotation,* The Journal of Chemical Physics **114**, 7941 (2001).

[36] C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernández, and B. Kohler, *DNA Excited-State Dynamics: From Single Bases to the Double Helix,* Annual Review of Physical Chemistry **60**, 217 (2009).

[37] J. P. Cnossen, D. Dulin, and N. H. Dekker, *An optimized software framework for real-time, high-throughput tracking of spherical beads,* Review of Scientific Instruments **85**, 103712 (2014).

[38] A. J. W. te Velthuis, J. W. J. Kerssemakers, J. Lipfert, and N. H. Dekker, *Quantitative Guidelines for Force Calibration through Spectral Analysis of Magnetic Tweezers Data,* Biophysical Journal **99**, 1292 (2010).

[39] Z. Yu, D. Dulin, J. Cnossen, M. Köber, M. M. van Oene, O. Ordu, B. A. Berghuis, T. Hensgens, J. Lipfert, and N. H. Dekker, *A force calibration standard for magnetic tweezers,* Review of Scientific Instruments **85**, 123114 (2014).

9

WHAT IS ALL THIS FUSS ABOUT TUS? THAT AND MORE ON NON-GAUSSIAN REPLISOME VELOCITY DISTRIBUTIONS

> *Whatever you will do in life will be insignificant,*
> *but it is very important that you do it.*
>
> Mahatma Ghandi

The convergence of DNA replication machinery at specific termination sites is a tightly coordinated process in bacteria. In *E. coli*, a 'replication fork trap' — a chromosomal region replication forks are allowed to enter but not leave — is set by the protein-DNA roadblock Tus-*Ter*. The exact sequence of events through which Tus–*Ter* blocks advancing replisomes in one direction, but not the other, has been the subject of study for many years – and surprisingly difficult to elucidate. Specific protein-protein interaction between Tus and the approaching replisome was denied as being the most likely candidate as soon as biochemical studies supported by the crystal structure of the locked complex suggested structural rearrangements between Tus and *Ter* could also provide a mechanism. Nevertheless, the subject remained a matter of controversy and recently gained renewed attention as three single-molecule-based approaches scrutinized the elusive Tus–*Ter* mechanism once again - leading to refinement of existing models, but also some new controversy. Here we discuss and compare the findings, merits and weaknesses of each of the single-molecule studies in an effort to find their common ground, pinpoint the crucial differences still existing, and push the understanding of this seemingly simple DNA-protein interaction further.

9.1. INTRODUCTION

9.1.1. TUS-TER

DNA replication in *Escherichia coli* initiates bidirectionally at *oriC*, creating two replication forks that proceed around the circular 4.6 Mbp chromosome in opposite directions. The forks progress at an average speed of 1 kbp/s until they meet again at the terminus region. As the replication forks approach the terminus, each encounters five 23 bp *Ter* DNA sites (denoted *TerA-J*) bound in a specific orientation by a 36 kDa DNA binding protein called Tus[1–4] and proceeds unhindered, implying a dislodging of Tus from *Ter*. However, when a replication fork continues beyond the terminus, Tus–*Ter* is approached from the opposite direction (**Fig. 9.1a**), triggering Tus–*Ter* to form a tightly locked complex, thereby bringing the replication fork to a halt.[1, 5–7] Each *Ter* site is non-palindromic, does not contain any direct repeats and has a strictly conserved GC(6) base pair followed by a highly conserved 13 base-pair core region. Tus is a monomeric protein that forms a simple 1:1 complex with *Ter* (**Fig. 9.1a**).[8] The structure of the Tus–*TerA* complex shows that many of the conserved residues among the *Ter* sites make base-specific contacts with the protein.[4, 9] The Tus–*TerB* complex has a reported dissociation constant (K_D) of 44 pM in 50 mM NaCl.[10] This renders it the most stable complex known between a monomeric sequence-specific DNA-binding protein and a duplex DNA recognition sequence.

Tus-*Ter* is thought to have evolved specifically to block the *E. coli* replisome, a multi-protein subunit complex that catalyzes DNA replication (**Fig. 9.1e**). Many aspects of replisomal organization are conserved throughout all life forms, from viruses to higher order eukaryotes.[13, 14] At the front of the replisome is the hexameric ring-shaped DNA helicase, which encircles the lagging strand and moves in the 5'-3' direction while unwinding DNA, thereby excluding and displacing the 3'-5' leading strand template. DNA synthesis of the two daughter strands is performed by DNA polymerases, enzymes that extend a primer in the 5'-3' direction only, thus move along a template strand in the 3'-5' direction. The antiparallel nature of the double-stranded (ds) DNA helix necessitates simultaneous coordinated synthesis of two DNA strands with opposite polarity, leading to an asymmetric DNA daughter strand synthesis mechanism: a continuous leading strand, versus a discontinuous lagging strand DNA polymerization activity.[15] *In vitro* assays such as those discussed here (**Boxes 1 and 3**) can force the assembly of a leading-strand-only replisome by leaving out enzymes such as the primase. The leading strand DNA polymerase is thought to directly follow the helicase on the leading strand template that the helicase excludes.

The polarity of the DNA in the context of Tus-*Ter* is as pivotal to the full understanding of this system as it is to the DNA replication process itself. Most notably, the C(6) base of *Ter* is located on the leading strand template – the strand excluded by the helicase – when Tus-*Ter* is approached at the nonpermissive face, and on the lagging strand when approached at the permissive face of the protein-DNA complex.

Despite extensive studies over multiple decades, it has been surprisingly difficult to elucidate the role of Tus-*Ter* as well as the exact sequence of events that leads to asymmet-

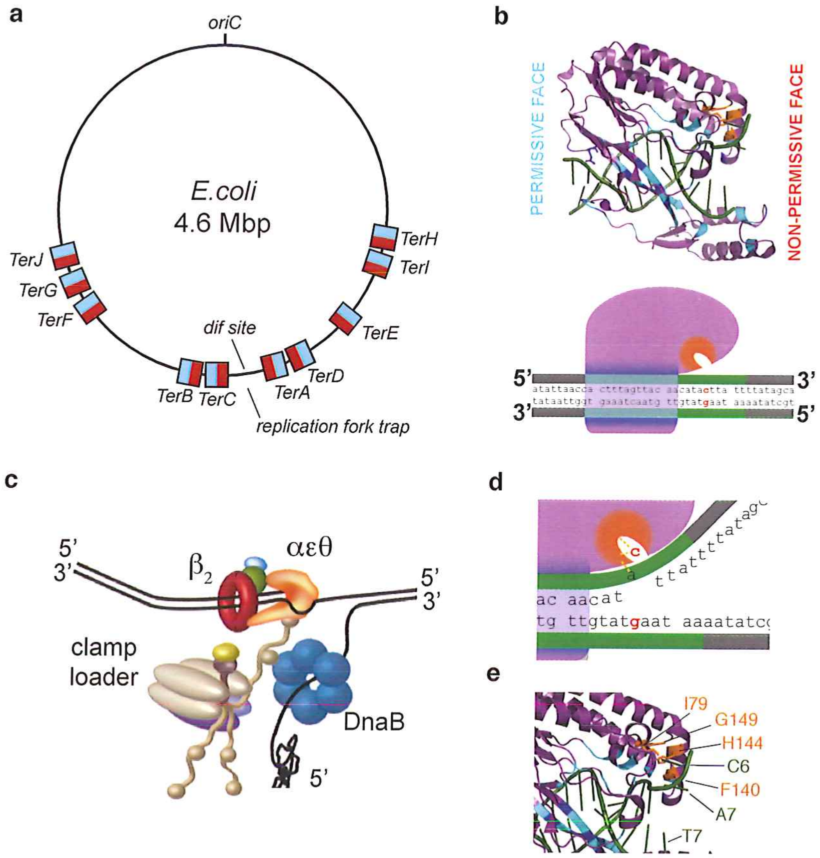


Figure 9.1: **The Tus–*Ter* complex structure and domains, and the *E. coli* replisome.** **(a)** Location and orientation (turquoise for permissive face, red for nonpermissive face) of the 23-bp *Ter* sites in the *E. coli* chromosome. **(b)** The crystal structure (top) of the locked Tus–*Ter* complex (PDB ID: 2EWJ) with a schematic representation (bottom) showing the protein has a DNA-binding domain located mainly in the two antiparallel β-strands interacting with the major groove of *Ter* DNA (dark blue for base-specific interactions, light blue for nonspecific interactions). **c)** Schematic representation of the *E. coli* replisome. The replisome is spearheaded by the DnaB helicase, which unwinds the parental DNA. In its wake, the helicase is followed directly by the Pol III DNA polymerase (α, ϵ, θ heterotrimer) in a continuous fashion on the leading strand template (3'-5'), and in a segmented fashion on the lagging strand template. Both helicase and polymerases are tethered to the clamp loader complex. Adopted from [11]. **(d,e)** schematic representation (**d**) and crystal structure (**e**) showing the lock domain and amino acid residues (in orange) in Tus that interact specifically with C(6) upon strand separation. In **d**, *Ter* nucleotides in green. Panels a, b, d, and e are adopted from [12].

ric replication fork arrest at Tus-*Ter*. Besides this, it remains a mystery why Tus-*Ter* has evolved to be only ~ 50% efficient *in vivo*, and why the arrays of *Ter* sites are spread out over hundreds of kilobases (**Fig. 9.1a**), all together covering roughly half the *E. coli* chromosome. Three mechanistic models that have been proposed and heavily debated over the course of almost three decades, and while one model has been effectively ruled out recently (**Boxes 1 and 2**), it is still unclear whether one of the others is entirely correct.

Most likely, as we will argue, it seems both models have their merits, yet both are incomplete as more detailed mechanistic workings await discovery.

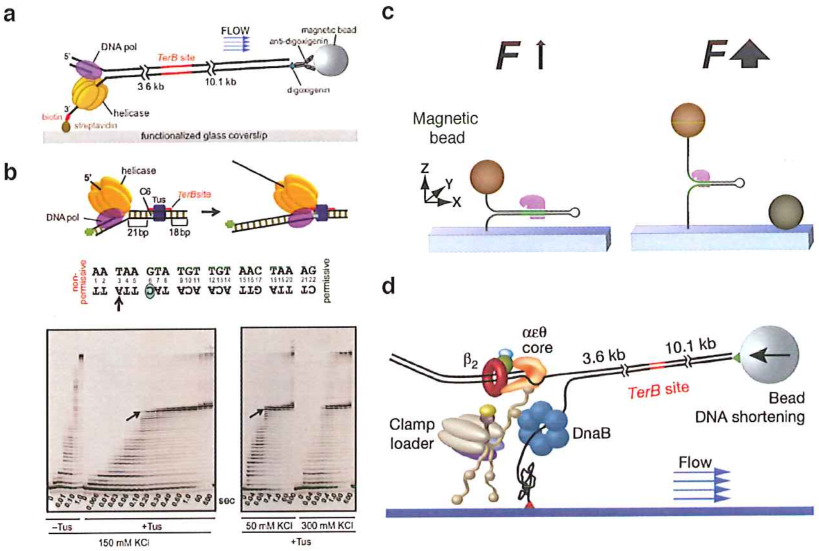


Figure 9.2: **The three single-molecule assays discussed. a,b)** The flow-stretching assay and the quenched-flow assay of Pandey *et al.*, respectively (See **Box 1** for more information).[16] **c)** The magnetic tweezers DNA hairpin assay of Berghuis *et al.* (See **Box 2** or Chapter 8 for more information). **d)** The flow stretching assay of Elshenawy *et al.* (See **Box 3** for more information).[11]

Here we specifically discuss the most recent efforts that have been made to unravel the workings of Tus-*Ter* polar replication fork arrest. A first study, performed by Pandey *et al.*, examines the polar interaction of Tus-*Ter* when the evolutionarily unrelated T7 replisome, the isolated T7 replicative helicase or the T7 DNA polymerase collide with this DNA roadblock.[16] Two approaches are used here: a single-molecule flow stretching assay, as well as a quench flow assay with single base resolution (**Box 1, Fig. 9.2a,b**). In a second study, Berghuis *et al.* investigate the sequence of events that lead to polar blocking of strand separation by probing the Tus-*Ter* lock in an isolated fashion in a single-molecule force spectroscopy assay (**Box 2, Fig. 9.2c**)[1].[12] And finally, in a third study Elshenawy *et al.* reconstitute a leading strand *E. coli* replisome and observe single collision events with Tus-*Ter* in a single-molecule flow-stretching assay (**Box 3, Fig. 9.2d**). All studies probe either a wide variety of Tus mutants, altered *Ter* sites, or both, as to discover the missing pieces of the Tus-*Ter* puzzle.

---

[1] And chapter 8 of this thesis.

Box1

Pandey *et al.* use a single-molecule flow-stretching assay as well as a quenched flow assay to examine the efficiency and details of Tus-*Ter* blocking the T7 bacteriophage helicase and DNA polymerase progression. Both the coordinated activity of the helicase-polymerase as well as the isolated enzymatic activities are examined. In the flow-stretching assay three types of experiments are performed: either helicase-polymerase induced bursts of DNA replication, T7 gp5 DNA polymerase strand-displacement activity, or T7 gp4 helicase unwinding activity. All three experimental assays give rise to single-molecule trajectories of number of synthesized or progressed base pairs over time. The quench flow assay provides additional information on the position of the DNA polymerase active site with single-nucleotide resolution, thereby providing details on the exact location of polymerase or replisome stall sites.

Nonpermissive Tus-*Ter* was shown to block the progression of the T7 helicase-polymerase as well as the T7 helicase. The DNA polymerase showed significant pausing at the nonpermissive face; however, arrest of DNA polymerase strand-displacement activity occurred at the permissive face of Tus-*Ter*. This reverse blocking effect seen for the DNA polymerase points towards a sensitivity of the Tus-*Ter* mechanism towards the translocation polarity of the enzyme in question. The quenched flow data show that the last nucleotide incorporated by the DNA polymerase of the replisome is mainly the A(4) base, with smaller fractions being T(3) or A(5), but never G(6). This last result is indicative of the specific interactions C(6) forms upon strand separation, leading to replication fork arrest. These results are further supported by *Ter* C(6) mutations that show to be deficient in polar arrest. The authors also find that a slowing of progression occurs upon interaction of the helicase-polymerase at nonpermissive Tus-*Ter*: first only T(3) and A(4) are observed, after a couple of minutes A(5) incorporation is also observed in the quench flow assay.

In summary, Pandey *et al.* show that Tus-*Ter* effectively blocks the progression of the helicase-polymerase of bacteriophage T7, showing that the Tus-*Ter* blocking mechanism occurs independently of specific interaction with the (*E. coli* DnaB) helicase. The C(6) base is crucial to full arrest of the T7 replication machinery, yet the blocking of polymerase progression at the permissive face of Tus-*Ter* indicates that this DNA-protein system has also evolved to block entities with a specific translocation polarity, as the C6 lock does not form in this case. The variation in the polymerase stall sites is suggested to be a result of an underlying multi-step process towards C(6) lock formation.

Box2

Berghuis *et al.* examine the interaction strength of Tus-*Ter* upon strand separation by applying a force on single DNA hairpins containing a *Ter* site in either the permissive or nonpermissive orientation. This implies not that DNA replication enzymes unwind duplex *Ter* DNA, but that force-induced unzipping mimics the strand separation that normally accompanies DNA replication activity. Upon performing unzipping in the presence of Tus, the authors observe nearly unimpeded strand separation for *Ter*-bound Tus in the permissive orientation, yet a blocked strand separation for *Ter*-bound Tus in the nonpermissive orientation, directly showing that replication enzymes are not required for the Tus-*Ter* lock to form. In fact, the lock forms much more efficiently than *in vivo* observations report on (100% blocking of hairpin strand separation vs. ~50% reported *in vivo* replication arrest efficiency), even though the rate of strand separation is ~20-fold higher than during replisome-mediated strand separation.

The strength of the Tus-*Ter* interaction is quantified by measuring the dwell time of the blocking of strand separation at a given force (typical $N \sim 100$). The distribution of these dwell times typically consists of three single exponential distributions, implying that three (lock) states, governed by single respective rates, underlie the Tus-*Ter* nonpermissive interaction. The state lifetimes were shown to be the longest for the WT Tus-*Ter* interaction, while they decreased as mutations in the lock domain of Tus were introduced. Mutations in the DNA binding domain of Tus did not affect the state lifetimes, mutation of the *Ter* C(6) base disrupted lock formation more than any single mutation of Tus could induce, and specific mutation of residue E49(K) led to a decrease in probability of the longest-lived lock state.

Together, these results indicate that Tus-*Ter* lock formation at the nonpermissive face is a 3-step process, with the longest-lived state reflecting full lock formation: i.e. the final state where C(6) interacts tightly with the Tus lock residues. As E49K influences only the probability of full lock formation, this residue is proposed to play a role in guiding C(6) to its final (full lock) position. The authors propose that the E49K experiments also suggest how enzymes could decrease the probability of full lock formation by sterically hindering the lock formation process through non-specifically interacting with residues such as E49 upon collision.

Box3

Elshenawy *et al.* investigate how efficient Tus-*Ter* is at blocking the progression of DNA replication by the *Escherichia coli* replisome through monitoring single replication events of a DNA sequence containing a *Ter* site in a single-molecule flow-stretching assay. Data emerging from this assay consist of the number of synthesized base pairs over time. The *Ter* site is 3.5 kb downstream of the initiation site. Nonpermissively oriented *Ter* DNA constructs show a significant increase in the occurrence of termination of DNA synthesis at this 3.5 kb site in the presence of Tus (~ 50%) compared to permissive Tus-*Ter* (~11%), or termination at this site without Tus (~5%) (**Fig. 9.4a-c**).

The single-molecule replication events are divided into groups based on whether DNA replication terminates (T), temporarily pauses (P), or proceeds unimpeded (B(ypass)) at Tus-*Ter*. By correlating these occurrences to the estimated replication fork velocities ($v \sim 500-1500$ bp/s), the authors find a positive correlation between the fork velocity and the probability of an unimpeded bypass of Tus-*Ter* (**Fig. 9.5a**). Altering the *Ter* site at or around C(6) changes the $T:P:B$ ratio, as do specific mutations at the nonpermissive face of Tus. Pre-formation of the Tus-*Ter* lock (by introducing a mismatch around C(6)) leads to a significant increase of termination events (~ 90%), replacing C(6) with G(6) leads to the majority of WT T events becoming P while the fraction of $P+T$ remains constant, and mutation of R198(A) leads to the largest decrease in T (and $T+P$) measured, for instance. However, lock pre-formation restores the detrimental effect of the R198A mutation completely.

Taken together, these single-molecule experiments suggest that the efficiency of Tus-*Ter* lock formation depends on the velocity of the replication fork that collides with Tus-*Ter*, and that C(6) interaction is a crucial part of this lock formation. The observations also suggest that Tus-*Ter* interaction upon strand separation at the nonpermissive face is a multi-step process, and the interactions of R198 with the strand complementary to C(6) are likely to form one of these steps. The authors propose that R198 interaction with this strand needs to be disrupted by the replisome, and that this process is essential to buy time for the inefficient C(6) rearrangements necessary for full lock formation.

9.1.2. EXISTING TUS-TER MODELS

Three mechanistic models have been proposed to explain polar arrest of a replisome by Tus-*Ter*. In the helicase interaction model (**Fig. 9.3i**), the DnaB helicase interacts specifically and physically with the nonpermissive face of Tus-*Ter*. The amino acid residue E49 of Tus is proposed to be the main point of contact with DnaB.[17] The dynamic clamp model (**Fig. 9.3ii**) proposes an intrinsic difference in the Tus-DNA interactions at the two ends of *Ter*, leading to a different blocking efficiency. Finally, the mousetrap model (**Fig. 9.3iii**) proposes a specialized mechanism that involves the re-orientation of the highly conserved C(6) base into a binding pocket of Tus upon strand separation.[18] These models are not necessarily mutually exclusive, as dynamic clamping could e.g. co-exist with a helicase interaction or a mousetrap mechanism.

9.2. DISCUSSION

9.2.1. HELICASE INTERACTION EFFECTIVELY RULED OUT

Of the three Tus-*Ter* models, the helicase interaction model has been effectively shown not to be the dominant mechanism for polar Tus-*Ter* polar arrest by the studies of Pandey *et al.* and Berghuis *et al.* (**Boxes 1 and 2**, resp.). The former show that the T7 replication machinery is blocked at nonpermissive Tus-*Ter*, the latter show that Tus-*Ter* effectively blocks DNA strand separation at the nonpermissive face in the absence of any replicative enzymes whatsoever. In addition, Pandey *et al.* report to have sought to detect physical interaction between Tus and DnaB in a variety of formats using surface plasmon resonance, yet to date no manifestation thereof has been found.

Besides the argument that the T7 system could not have co-evolved with Tus-*Ter* to form a helicase specific interaction, the quenched flow T7 study further strengthens the argument against the helicase interaction model by observing different blocking efficiencies for the separate isolated components of the T7 replisome. It is shown that while the 5'-3' translocating helicase is effectively blocked at the nonpermissive face when operating in an isolated fashion, the 3'-5' moving polymerase is rather surprisingly blocked at the permissive face, while allowed an unimpeded pass at the non-permissive face. The blocking of polymerase activity at the permissive face is explained by specific interactions between Tus and the leading strand template (i.e. the 3'-5' directed strand, **Fig. 9.3i**). This observation implies there are inherently different types of DNA-Tus interactions at the respective faces of Tus-*Ter*, thus argue in favor of dynamic clamping. More specifically, these observations suggest Tus-*Ter* has evolved to block a strand-separation process that allows the 3'-5' oriented strand more degrees of freedom than the complementary strand. In other words, Tus-*Ter* seems to have evolved to block the process of strand separation at the nonpermissive face induced by an enzyme moving along the 5'-3' lagging strand template. The canonical process through which this occurs is of course *E. coli* DNA replication: a replisome spearheaded by the DnaB helicase.

The only study that can still be interpreted as implying a specific helicase-Tus interaction is one where a double-stranded (ds) DNA translocating DnaB helicase was blocked at the nonpermissive face of Tus-*Ter*.[17] In an attempt to rule out any involvement of C(6), the

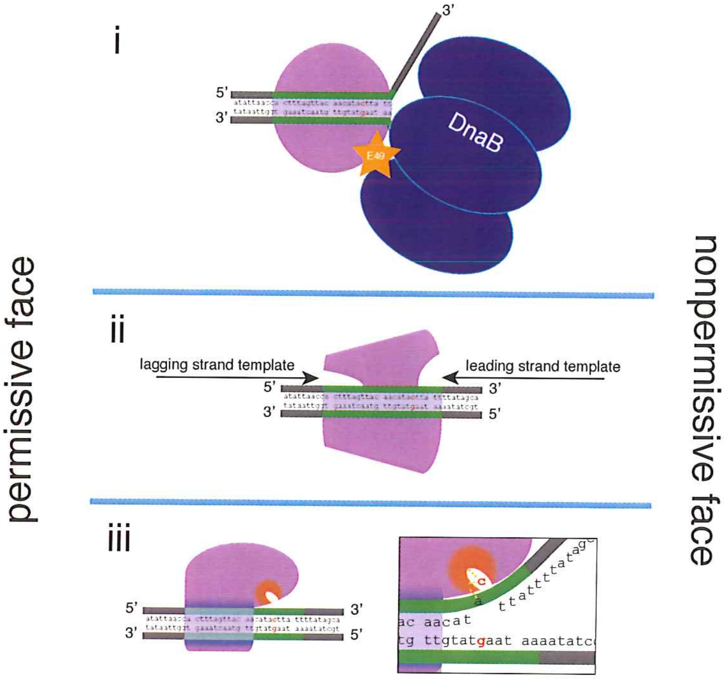


Figure 9.3: **The existing Tus-*Ter* models.** **i)** The helicase interaction model. Direct interaction between the DnaB helicase of *E. coli* and Tus at the nonpermissive face of Tus-*Ter* is what triggers replication fork arrest. Tus residue E49 has been proposed to be the key point of interaction. **ii)** In the dynamic clamp model, it is the inherently different interaction strengths at the permissive and the nonpermissive faces of Tus-*Ter*, as well as differences in interaction depending on strand polarity that determine the barrier imposed on enzymes colliding with Tus-*Ter*. Here this is portrayed as the Tus 'thickness': at the nonpermissive face Tus is thicker, implying a stronger clamping. In addition, the polarity effect is portrayed as one strand (containing C(6)) leaving room for an enzyme to wedge itself between Tus and *Ter*, thereby making dislodging Tus from *Ter* more probable for enzymes translocating along this strand (this has nothing to do with the actual protein structure). **iii)** The mouse trap model suggests that strand separation invoked by the approaching unwinding enzyme is what triggers tight interaction between the *Ter* C(6) base and specific amino acid residues in a Tus lock domain. The C(6) lock formation is a final of several conformational changes, of which rearrangement of R198 is one.

authors covalently linked bases 4 and 5 of *Ter*, preventing any strand separation from occurring before C(6). An interesting additional experiment that could have been performed is one with a covalently linked GC(6) base pair. With GC6 covalently linked, any C(6) rearrangement for polar arrest can now certainly no longer take place. In addition, in the light of new evidence presented by Elshenawy *et al.*, it might well be that covalently linking the base pairs preceding GC(6) is what actually promoted polar arrest, as a rate-limiting process involving the fifth base pair of Tus-*Ter* might exist, as we discuss further below.

9.2.2. DYNAMIC CLAMPING AND C(6) INTERACTIONS

The mousetrap model and the dynamic clamp model both do not require specific helicase-Tus interactions. The strongest arguments in favor of dynamic clamping are presented by Pandey *et al.*, as the isolated DNA polymerase experiments highlight the Tus-*Ter* interactions with the leading strand are inherently different from those with the lagging strand, as the unlocked crystal structure had also revealed.[4] The fact that the polarity of the strand being operated on matters so much is a clear indication that Tus-*Ter* has evolved such that a 5'-3' translocating entity is effectively blocked in one direction, but not the other. However, this dynamic clamping does not propose the existence of structural rearrangements taking place upon strand separation.

The force-induced DNA hairpin unzipping experiments, which directly show that Tus-*Ter* blocks strand separation at the nonpermissive but not the permissive face, are hard to explain through dynamic clamping alone. Not only would the enormous difference in observed blocking efficiency in one orientation relative to the other be hard to explain through dynamic clamping, Berghuis *et al.* directly show how Tus amino acid residues not showing any interaction with *Ter* DNA in the unlocked crystal structure[4] have a dramatic effect on the polar blocking of strand separation when altered. In fact, all three studies in focus here present overwhelming evidence for a specific interaction of C(6) with its lock pocket. For instance the quench flow assay, which allows for monitoring the last base incorporated before termination or pausing at Tus-*Ter* by the leading strand polymerase at single-base resolution, showed the formation of a variety of products (from T(3) through A(5), with the majority being A(4)), while incorporation of the G(6) was never observed. Both Pandey *et al.* and Elshenawy *et al.* show that inverting the GC(6) base pair has dramatic consequences for the incidences of (permanent) arrest in the single-molecule flow stretching assays, and improvements in arrest efficiency are observed when the C(6) lock is allowed to form before any interactions with the replication fork occur (**Fig. 9.4a**).

9.2.3. MORE THAN A SINGLE BARRIER, MORE THAN A SINGLE SWITCH

What all three studies also show is that when the C(6) lock is impaired -- by altering the C(6) base or impairing lock formation through lock pocket mutation H144A -- is that interactions weaken or a halt turns from being permanent to transient (**Fig. 9.4a-c**). However, it remains clear that a significant barrier remains, much more so than the barrier imposed by permissively oriented Tus-*Ter* for instance (**Fig. 9.4e**). In the hairpin pulling experiments, the barrier reduces to permissive Tus-*Ter* proportions when H144A is combined with a CG(6) inversion: a strong indication that C(6) interaction plays a dominant role in lock formation. When the CG(6) locking interaction is allowed to form in the absence of any actively DNA-unwinding enzyme at the fork (**Fig. 9.4b**, 'CG(6) 5b mm'), the occurrence of bypass events is indistinguishable from the occurrence of bypass at a preformed wt lock (**Fig. 9.4b**, 'WT', cyan). This has the important suggestion that during the hairpin pulling experiments C(6) locking and other interactions are always allowed to take place, as there are no DNA-unwinding enzymes present in these experiments. This was confirmed by the identical results for experiments with a pre-formed lock and with a WT *Ter* site (**Fig. 9.4b**, in purple: 'WT' and 'WT 5b mm'). This is contrary to the experiments carried out with enzyme-induced fork propagation, where alterations of the *Ter* site allow the lock to pre-form, but will thus always need to be performed with a non-native *Ter* site as there needs to be a region of unpaired bases around C(6). The hairpin pulling experiments consistently find the existence of three states that occur during the non-permissive blocking of strand separation, while any of the impaired lock experiments (i.e. experiments using H144A and CG(6)) rendered only two states.

9.2.4. OTHER POSSIBLE KEY ACTORS IN POLAR ARREST

Besides the amino acid residues that directly interact with C(6) in the lock pocket, there are several other key players that are known to, or at least suspected of playing a pivotal role in forming the Tus-*Ter* lock. Two residues that have been suggested are E49 and R198. From the crystal structure it is apparent that structural rearrangements take place in the R198 region that interacts with the lagging strand template upon lock formation. Specifically, R198 interacts with the A(5) and G(6) bases of this strand, with Elshenawy *et al.* suggesting additional remote interactions with T(5) on the leading strand. In the locked complex, these R198-A(5)-G(6) interactions have made place for interactions of R198 with the phosphate backbone of the lagging strand, between bases 6 and 7. Observing severely impaired replisome blocking behavior for Tus mutant R198A, Elshenawy *et al.* suggest that R198 transiently holds the two strands together by interacting with base pair 5 of *Ter*, making it a rate-limiting step in the process towards C(6) lock formation. The flow stretching experiments show that like for the impaired CG(6) lock, the R198A blocking behavior is restored to WT-like proportions when allowed to form in the absence of replication machinery at the fork (**Fig. 9.4b**, bypass of 'R198A 5b mm' equals 'WT 5b mm' (cyan)). When R198A was subjected to hairpin pulling experiments, we observed full C(6) lock formation (see **Box 2** for explanation on 'full lock formation') in instances similar to those observed for wt Tus-*Ter*, but now showing a marked difference in the lifetime or probability of the first and second lock states, respectively (**Fig. 9.4b,e** new data), implying that either one of these states reflects the magnitude of R198-*Ter* interactions. Taking both observations into account, it indeed appears that rearrangements involving R198 are a step towards C(6) lock formation.

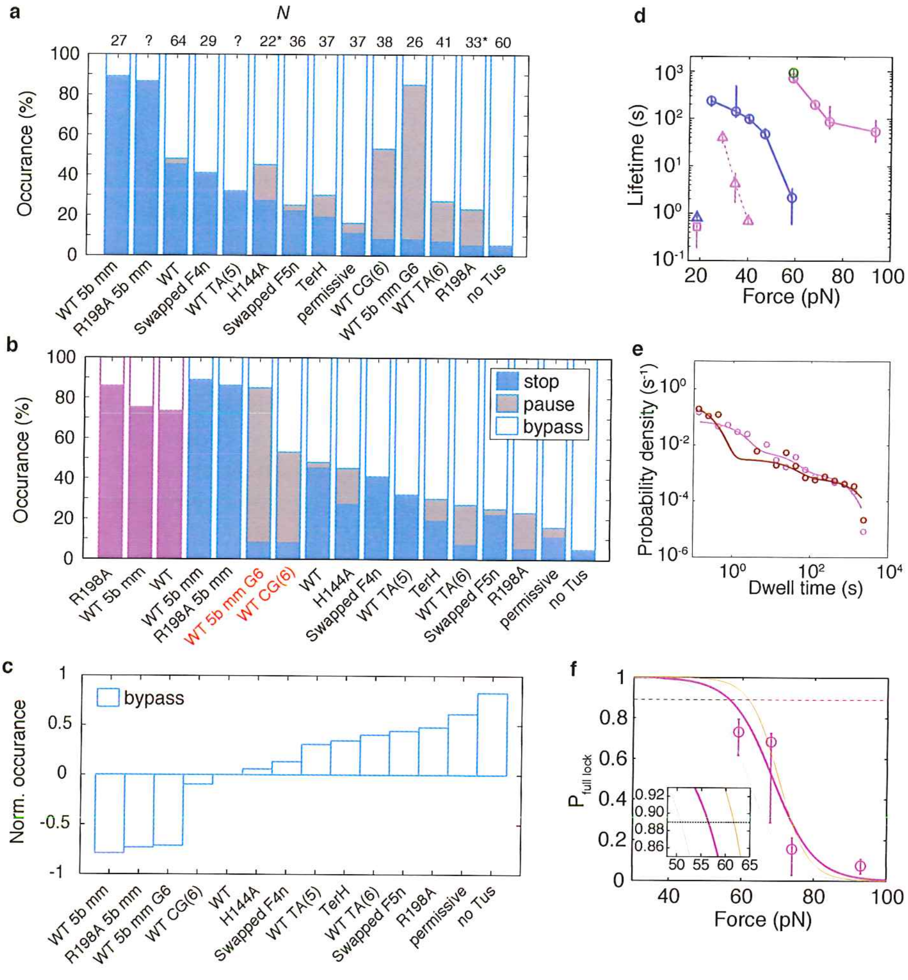
The E49 residue of Tus has been reported to be deficient in polar arrest of replication forks *in vivo*. Specific interaction between DnaB and this Tus residue was hypothesized, but effectively ruled out by the isolated hairpin pulling experiments. Interestingly, a more subtle and refined role was hypothesized based on the observations made using single-molecule tweezers experiments ([12], **Box 2**). These experiments revealed that full WT-like lock formation still occurred, yet with a dramatically decreased probability (P_{fulllock}(WT)= 79%; P_{fulllock}(E49K)= 7% at 59 pN). This directly shows that residues outside the Tus lock pocket are essential in guiding C(6) towards its final, fully locked position. In addition, the decrease in probability observed *in vivo* also suggests that the the presence of replicative enzymes at the nonpermissive face of Tus-*Ter* could influence the reaction equilibrium of lock formation. In other words, a certain enzyme at the fork could bring with it its own particular fork stochiometry, that could concomitantly lead to a specific blocking probability.

In the light of these observations, it is not entirely surprising that Elshenawy *et al.* find completely impaired blocking behavior in their flow-stretching replisome experiments (data not shown, personal communication). This situation couples a E49K-induced decrease (by an order of magnitude) in likelihood of lock formation directly observed in the hairpin experiments, to an additional decrease in blocking probability caused by the presence of an enzyme moiety at the fork. In other words, *E. coli* replisome blocking by E49K-Tus will become even less likely than the E49K blocking of strand separation observed in the hairpin pulling experiments. A prediction would be that pre-formation of the lock (e.g. 'E49K 5b mm') in context of the replisome colliding with it, would restore the probability of observing a blocking event to a similar order of magnitude observed for E49K in the hairpin experiments.

9.2.5. PROGRESSIVE STRENGTHENING OF THE MOUSETRAP: EXTENSION OF THE MOUSETRAP MODEL

In light of the studies discussed here, it is now possible to dovetail parts of the models proposed. A common feature of the three single-molecule studies is that Tus-*Ter* blocking is a multi-step process. Both Berghuis *et al.* and Elshenawy *et al.* propose an extended version of the mousetrap model, where C(6) interaction with the Tus lock pocket is the third and final step towards a locked state. Both works hypothesize that the series of events leading to this C(6) lock state are most likely to be sequential, though they do not necessarily have to be. As a result of the intrinsicly different nature of both assays, both studies propose slightly different versions of a locking mechanism, making assumptions their own data supports. Berghuis *et al.* show that the C(6) state is the strongest interaction and propose a three state kinetic model where interaction progressively strengthens going from one state to the next. Elshenawy *et al.* emphasize and propose that R198 rearrangement is rate limiting, so as to allow time for C(6) to find its pocket, as they observe a clear correlation between replisome velocity and probability of fork stalling at nonpermissive Tus-*Ter*.

In addition, it might for instance now be possible to quantify the stall force of an *E. coli* leading strand replisome for the first time. The probability for WT Tus-*Ter* full C(6) locking behavior is obtained by fitting the kinetic three-state model to the hairpin pulling datasets, which turned out to be 79% at 59 pN, going down to 8% at 93 pN. As experiments have shown that the situation obtained in the hairpin pulling experiments is the same as (and should be compared directly with) the probability of blocking observed when the Tus-*Ter* lock is allowed to pre-form, the 89% blocking behavior of the *E. coli* replisome with a pre-formed lock (**Fig. 9.4a,b**) suggests that the stall force is lower than 59 pN (which shows 79% lock formation). Assuming that full lock formation is the result of a competition of rates between a fully locked and an unlocked state, we can fit the expression $P_{\text{lock}} = 1/(1+\exp[-(F_{\text{eq}} - F)\Delta x/k_B T])$ to the trend in the force-dependence of the full lock state (**Fig. 9.4f**). Here F_{eq} is the force at which $P_{\text{lock}} = 0.5$, F is the applied force, Δx the distance between the locked and the unlocked state, k_B is the Boltzmann constant and T the absolute temperature (see **Supplementary Discussion** for details). The force at which $P_{\text{lock}} = 89\%$ shows that the upper boundary of replisome stall force is ~56 pN (**Fig. 9.4f**). The effective distance to the transition state is ~0.7 nm.



* derived from an indirect mention in the manuscript

Figure 9.4: **Merging the *in vitro* flow stretching (a-c) and DNA hairpin (b,d-g) experimental data. a)** The experimental assays ranked from highest to lowest permanent arrest occurrence ('5b mm' = 5 base mismatch bubble to pre-form lock, 'F[*n*]n' = first [*x*] nucleotides of *Ter*). **b)** The same results ranked from lowest to highest bypass frequency, only now with the addition of the full lock probabilities extracted from the hairpin pulling experiments as well (purple). **c)** The change in replisome bypass frequency relative to WT Tus-*Ter*. **d)** Force-dependent lifetimes derived from the hairpin experiments show that wt Tus-*Ter* (purple, circles) has the longest lifetime at the highest force, and that E49K (green) has an identical lifetime as WT. Disrupting the lock domain (H144A, blue circles) or inversion of GC(6) (CG(6)), purple triangles and dashed lines) severely shortens the lifetimes at a given force. Combining H144A with CG(6) (blue triangle) leads to lifetimes similar to that of permissive Tus-*Ter* (purple square). **e)** The distribution of lifetimes obtained for R198A (brown) show that the fully locked state lifetime is similar to that of WT Tus-*Ter* (purple), yet the first state now has a much shorter lifetime ($\tau_{\text{wt}} = 1.3$ s, $\tau_{\text{R198A}} = 0.2$ s). **f)** The full lock probability of WT Tus-*Ter* (purple) versus the applied force. As the lock equilibrates when triggered in the absence of enzymes at the fork, we can use the state Boltzmann weights to extrapolate the force-dependent locking probability ($P_{\text{lock}} = 1/(1 + \exp[-(F_{\text{eq}} - F)\Delta x/k_B T])$, with F_{eq} the force at which $P_{\text{lock}} = 0.5$ and Δx the effective distance to the transition state) to get an idea of the force at which P_{lock} (purple) equals the probability of permanently blocking a replisome with a pre-formed lock (**a**, 'WT 5b mm', black dashed line, see inset). Cyan and orange lines are fits to the lower and upper 1-sigma confidence intervals, respectively. The force corresponding to the upper boundary of the force exerted by a replisome ~56 pN, Δx was found to be 0.70 nm (0.70 and 1.00 nm for the lower and upper boundary, resp.). The free energy difference at zero force ΔG_0 was found to be -29 kJ/mol (CIs were -28 and -42 kJ/mol, resp.).

9.2.6. OVER VARIOUS ASSAYS, THE VELOCITY CORRELATION IS ABSENT

The question whether there is also a causal relationship between the difference in observed apparent velocities and (replication) fork stalling is pivotal. The correlation observed in the single-molecule E. coli replication fork experiments is intriguing and convincing, yet drawing any analogy with macroscopic phenomena such as momentum is incorrect and potentially misleading in an over-damped environment. A causal relationship between fork velocity and blocking would only occur if the rates of both events directly compete with each other – a causal relationship implies the unwinding and C(6) rearrangement rates are of the same order of magnitude. This means that the fork velocity dependence should be observed in every type of single-molecule Tus-*Ter* assay, irrespective of the enzyme present at the DNA fork, if any at all. Such a correlation has thus far not been reported, as the two other single-molecule assays performed have not observed a fork velocity dependence. In the study performed by Pandey *et al.*, where the on average order of magnitude slower than E. coli replisome progressing T7 replication fork (~100 bp/s) collides with Tus-*TerB*, a velocity dependence is not observed or reported on. A significant fraction (89%) of the T7 leading strand synthesis is blocked, and a velocity dependence of fork arrest would suggest that the remaining 11% of forks that proceed through unhindered would on average progress at higher velocities than the average blocked fraction. It would be possible to start looking for the existence of a correlation between fork velocity and blocking probability in these single-molecule T7 datasets as well.

The experiments performed by Berghuis *et al.* are on another extreme when it comes to fork velocities. These magnetic tweezers induced hairpin openings have (very narrowly distributed) fork progression rates of ~20 kb per second, and 100% of events were blocked at least temporarily during these experiments. A fraction of these events might be accounted for by the difference in acquisition rate, which is 100 Hz in these hairpin experiments, compared to 2 Hz with the flow stretching experiments. This higher acquisition rate allows for a much more detailed observation of the most transient blocking events and might account for a bias in the instances counted as being blocked. In other words: the shortest events counted as blocking by Berghuis *et al.* will be seen as an unimpeded pass by the other studies due to the difference in temporal resolution. This does imply that Pandey *et al.* and Elshenawy *et al.* cannot observe the short-lived state observed in the experiments of Berghuis *et al.* But even if the shortest-lived events would be counted as an unimpeded pass in the latter study, the majority of WT Tus-*Ter* events are still blocked in the hairpin experiments – even though the fork opening occurs at velocities an order of magnitude faster than the fastest-moving leading strand replication fork observed by Elshenawy *et al.* This strongly suggests that the rate of fork progression does not compete with the rates at which the molecular rearrangements take place.

Another strong indication that the rates that govern Tus-*Ter* lock formation occur at totally different timescales than base pair unwinding is the observation that in the hairpin experiments, lock pre-formation produced a distribution of lock lifetimes identical to WT Tus-*TerB*. This indicates that given the opportunity to form in an isolated fashion, the Tus-*Ter* lock behaves identically to a lock that has to form within the timeframe allowed for an entity moving at 20 kbp/s. This strongly suggests that the observed correlation between replisome velocity and bypass probability observed by Elshenawy *et al.* does not imply a causation.

9.2.7. SLUGGISH MOLECULE MOTION OR PROTEIN-INDUCED CHANGE OF CHEMICAL EQUILIBRIUM?

So far two out of three single-molecule studies have not observed a correlation between the apparent replication fork progression velocity and the probability of being blocked at Tus-*Ter*. What has been observed is that each different enzyme at the fork (Pandey *et al.*), or the absence of any enzyme at the fork whatsoever (Berghuis *et al.*), comes with its own characteristic blocking probability. Pandey *et al.* show that the probability of being blocked heavily depends on the strand excluded by the enzyme at the fork. Berghuis *et al.* also observe efficient and reproducible blocking of strand separation, even when the time window allowed for lock formation is cut down to one tenth of the time allowed for lock formation by the fastest moving *E. coli* replisome. Besides this, by showing that a chemical change in the protein structure (using Tus mutant E49K), the hairpin experiments show that the probability of full locking (C(6) entering lock pocket) can be altered without affecting the actual strength of the lock interaction. This allows for a different interpretation of the variation in blocking efficiency of Tus-*Ter*: the physical presence of a specific enzyme at Tus-*Ter* creates a unique blocking efficiency signature due to the unique chemical environment brought forward by the steric interactions of the colliding polypeptide structures.

9.2.8. NON-GAUSSIAN DISTRIBUTION OF REPLISOME VELOCITIES

Elshenawy et al. tested their Tus-*Ter* roadblock with the native replisome components only, yet observe two distinct Gaussian distributed subpopulations in the replisome velocity histograms which are either blocked or not (**Fig. 9.5a**). As the means of the two populations differ significantly, this implies that the population as a whole cannot have a Gaussian distribution of velocities – an observation that strongly suggests that the process giving rise to the observed Gaussian spread of one subpopulation is kinetically different from the other. The existence of different populations amongst *E. coli* replication forks has recently also been very convincingly demonstrated *in vivo* through a DNA combing approach (**Fig. 9.5c,d**).[19] The low total number of events (as well as the low temporal resolution, see **Supplementary discussion** 'Future directions: a need for higher resolution') in single-molecule replisome experiments of Elshenawy *et al.* makes determination of the exact number of subpopulations uncertain, yet the *in vivo* data of Pham *et al.* suggest there are predominantly three conformations in which the replisome can be found (**Fig. 9.5c,d**).

As the replisome is known to be a highly dynamic complex, it is very well possible that there exist a limited variety of replisome conformations, each with its own characteristic velocity. In addition, the fork velocity does not seem to be what causes a characteristic blocking probability, and different enzymes give rise to the differences in blocking efficiency at Tus-*Ter*. Taken together, a possible explanation for the populations that are either blocked or not in the work of Elshenawy *et al.* could be that it is not the replisome velocity, but the replisome composition that gives rise to a specific probability of being blocked. In context of replisome structure, it could be variations in the exact position of the polymerase with respect to the helicase that leads to not only a characteristically different velocity of the complex, but also a difference in the stoichiometry of the moiety at the fork. In other words, the fact that slower moving *E. coli* replication forks are blocked more often may just be coincidental: it might as well have been the other way around. It is because one population is conformationally different from the other that the interactions upon collision give rise to a different blocking probability. If there would be a single kinetic conformation governing replication fork progression, a single Gaussian distribution – or at least a distribution with a single peak – would be expected to best describe the single-molecule datasets. The fraction of blocked replication forks would be some fraction of that single-peak (Gaussian) distribution (**Fig. 9.4b**).

9.2.9. TER REDUNDANCY PARADOX

The reason for the existence of two groups of *Ter* sites oriented in an opposite orientation on both sides of the *dif* site on the circular *E. coli* chromosome has long been a curiosity. The binding affinity Tus has to these sites increases towards the *dif* site on both sides, as the number of mutations with respect to the highest affinity *TerB* site increases going outward from the *dif* site.[21] At first glance a satisfactory explanation might seem that, given the low efficiency of polar arrest of Tus-*Ter* (~50% in vivo), the other *Ter* sites must act as a backup mechanism when polar arrest at the preceding Tus-*Ter* site has failed. What weakens this argument however is that the *Ter* sites, instead of being localized in a relatively compact array, are spread out over roughly half the *E. coli* chromosome. This implies a distance of more than 1 Mb between the innermost and outermost *Ter* sites on both sides of the *dif* site (**Fig. 9.1a**).

Another possible explanation could be that the only *Ter* sites that still really function in the light of the replication fork trap are the ones directly flanking the dif site (*TerA* and *TerC*), and that these *Ter* sites are efficient enough to perform their task at creating a replication fork trap. Strong evidence for this comes from a recent study seeking to answer to how the *E. coli* adaptive immune system CRISPR-Cas distinguishes self DNA from non-self DNA to insert as spacers in the CRISPR arrays.[20] It was found that spacer acquisition is dependent on the presence of DNA replication. By performing chromosome wide sequencing of spacers, the researchers show that spacer acquisition from the self chromosome, though rare, occurs at replication fork stall sites. Remarkably, only two of these 'proto-spacer hotspots' showed up on the chromosome: directly upstream of *TerA* and *TerC*, the *Ter* sites directly flanking the *dif* site (**Fig. 9.5e**). Engineering a nonpermissively oriented *TerB* site on the chromosome led to the emergence of a new, artificial proto-spacer hotspot (**Fig. 9.5f**).

It thus might well be that the more outward situated *Ter* sites have become obsolete over evolutionary timescales as, hypothetically, addition of information to the *E. coli* chromosome mainly occurs around the terminus region. This would cause an outward growth of the *dif* site flanking *Ter* sites, every now and then creating the necessity to insert a new *Ter* site closer to the *dif* site. Evidence for this might be found in the fact that there are other *Ter*-like sites (with a still conserved C(6)(!)) such as *TerL* located on the upper half of the chromosome.[21] Nevertheless, this suggested explanation of the outer *Ter* sites being remnants of once important blocking sites in an evolutionary earlier era is purely hypothetical and would have to be further investigated.

9.3. CONCLUSION

The works presented here define a new and clear step towards unraveling the mechanism behind Tus-*Ter* asymmetric blocking of replication forks. They also reveal the power of single-molecule studies, for the kinetic details of the dynamic, strand separation dependent biochemical reaction at the non-permissive face of Tus-*Ter* would not have revealed itself using classical ensemble-averaged techniques. While an inherent different nature of protein- DNA interactions at opposing faces of Tus-*Ter* (i.e. dynamic clamping) is present, this mainly points towards evolution of Tus-*Ter* to specifically block entities moving along the 5'-3' directed strand. However, it is evident that specific blocking of Tus-*Ter* at the nonpermissive face cannot occur without structural rearrangements taking place – rearrangements culminating in insertion of C(6) into the Tus lock pocket. What was first coined the mousetrap model still remains the most convincing model for Tus-*Ter* to date, yet it turned out to be incomplete, hence the emergence of the extended (multi-step) mousetrap model. We now know that the mousetrap, once triggered, is not a single hinge that flips and blocks fork progression once and for all, but a multi-state process with delicate reaction equilibria, equilibria easily perturbed by enzyme-induced steric interactions. Rearrangement of DNA-protein interactions at the lagging strand involving Arg198 is one of these steps, and there is likely an additional rearrangement that awaits discovery, as Berghuis *et al.* find three rate-limiting processes that give rise to Tus-*Ter* lock formation. Single-molecule assays will quite likely be the best way to go forward here.

9.3.1. FUTURE DIRECTIONS: QUESTIONS IN NEED OF AN ANSWER

Linking the observations of hairpin pulling experiments and replisome flow stretching experiments can lead to a mutual improvement of experimental interpretation. In this review, we have for instance demonstrated that the conclusions involving R198A of Elshenawy *et al.* facilitate the interpretation of the change in the distribution of dwell times R198A hairpin pulling experiments invoke. The other way around, the hairpin pulling experiments offer a unique insight into the dynamics of isolated lock formation, as preformation of a lock always necessitates an alteration of *Ter*. Using the trend in WT 'full locking' probability and comparing it with the stall probability of a replisome upon collision with a pre-formed Tus-*Ter* lock, we can now (roughly) quantify the stall-force of a replisome for the first time. Most importantly, as the pulling experiments show that the rate of fork progression is likely not in direct competition with the rates involved in lock formation, we can now view the mousetrap efficiency in terms of enzyme stoichiometry at the replication fork. A next step towards more intricate linking of the single-molecule experiments would e.g. be to follow *E. coli* replication fork progression as it collides with a pre-formed E49K Tus-*Ter* site, or examining the barrier G(6) imposes on the lagging strand through subjecting *Ter* GG(6) to the hairpin assay.

Further investigation of the variation in replisome conformation would require different kinds of assays than those discussed here. Single-molecule FRET might offer additional insights, as there might be a correlation between replisome stall probability and Tus–Pol III distance. Finally, as Tus-*Ter in vivo* stalling efficiency has been reported to be ~50% on multiple occasions, an answer must be sought to why *TerA* and *TerC* emerge as the only replisome stall sites *in vivo*, and why a system would evolve to work only roughly half of the time. The answer might be that 50% is just sufficient (and inherent to chemical equilibria) given the low probability of one of two replication forks being so seriously stalled midway in the first place. Another possibility might be that for some reason the *in vivo* stall probability is actually higher than predicted. This would be a question *in vivo* single molecule fluorescence experiments could provide an answer to.

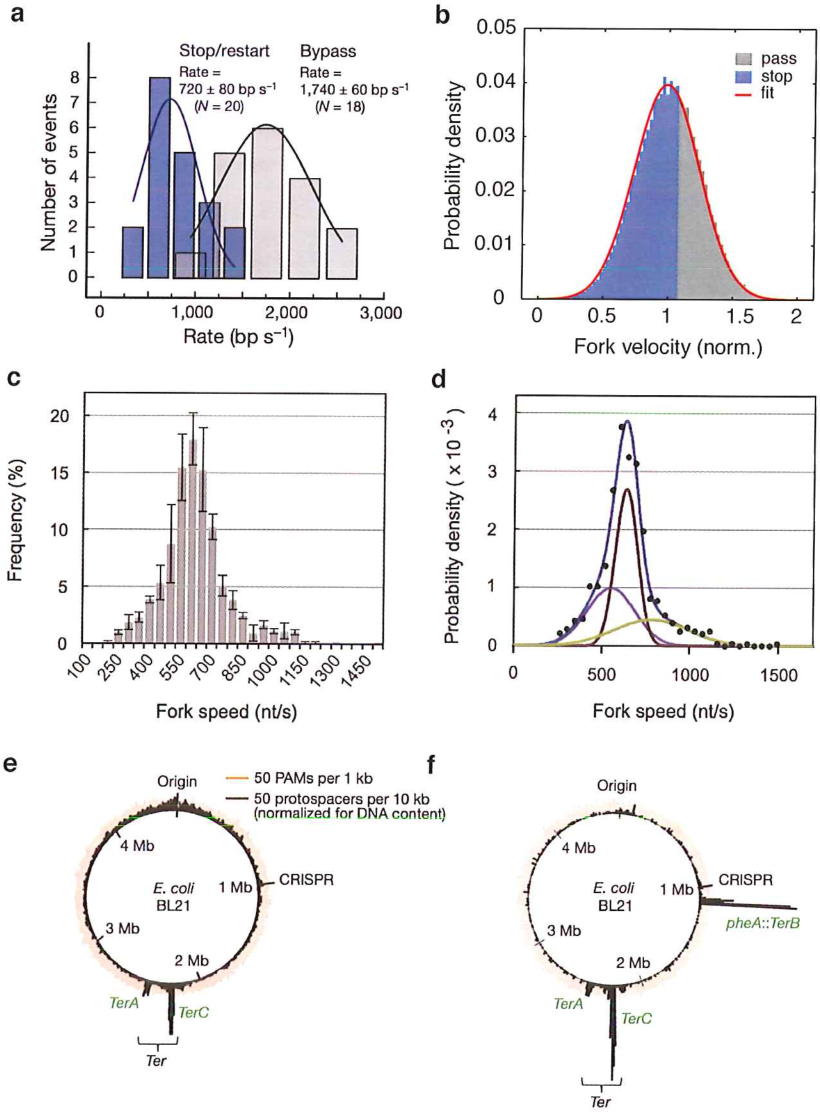


Figure 9.5: **Observed *in vitro* and *in vivo* replisome velocity distributions reveal the existence of multiple populations.** **a)** The distribution of individual replisome velocities as measured by Elshenawy *et al. in vitro*. (Taken from Elshenawy *et al.* [11]) **b)** The expected (though idealized) theoretical distribution of a single identical replisome population, where the fork progression rate directly competes with molecular rearrangement rates of Tus-*Ter*. **c,d)** Distribution of replisome velocities originating from 667 individual replication events *in vivo* by Pham *et al.* (Adapted from Pham *et al.*) **e)** Origins of 'self-DNA insertion sites' (i.e. protospacer hotspots) revealed through chromosome-scale sequencing of spacers. Spacer acquisition coincides with dsDNA breaks that occur at replication fork stall sites. From this it becomes apparent that only the *Ter* sites directly flanking the *dif* site have a large incidence of stalled replication forks. It also shows that replication fork stalling occurs more frequently at *TerC*, for the counter clockwise moving replication fork has the shortest distance to travel. **f)** Insertion of a novel (nonpermissively oriented) *Ter* site caused another proto-spacer hotspot to emerge (Adapted from Levy *et al.* [20]).

9.4. SUPPLEMENTARY DISCUSSION

9.4.1. FITTING THE FORCE DEPENDENT LOCKING PROBABILITY

The fit to the force dependence of the Tus full lock probability as shown in figure 9.4e we assume that the probability of full locking is a result of a direct competition between two rates: the rate of entering the locked state k_l the rate of not going into the locked state $k_\mathrm{u} = 1 - k_\mathrm{l}$, which makes the probability of locking:

$$P_\mathrm{l} = \frac{k_\mathrm{l}}{k_\mathrm{l} + k_\mathrm{u}} \tag{9.1}$$

From Kramer's theory of reaction kinetics we have an expression for the rates k_l and k_l:

$$k_\mathrm{l} = k_\mathrm{l}^0 \mathrm{e}^{(E_\mathrm{l}^\ddagger - \delta_\mathrm{l} F)\cdot\beta} \tag{9.2}$$

$$k_\mathrm{u} = k_\mathrm{u}^0 \mathrm{e}^{(E_\mathrm{u}^\ddagger - \delta_\mathrm{u} F)\cdot\beta} \tag{9.3}$$

where the k^0's are the pre-exponential factors at zero force, $E^\ddagger$ the energy difference between the respective states and the transition state, δ the distance to the transition state, F the applied force and $\beta = [k_B T]^{-1}$. Defining a reaction rate at zero force $K^0 = k^0 \mathrm{e}^{E^\ddagger \cdot \beta}$, and combining eqns. 9.1 and 9.2, we obtain:

$$\begin{aligned} P_\mathrm{l} &= \frac{K_\mathrm{l}^0 \mathrm{e}^{-\delta_\mathrm{l}\cdot F\cdot\beta}}{K_\mathrm{l}^0 \mathrm{e}^{-\delta_\mathrm{l} F\beta} + K_\mathrm{u}^0 \mathrm{e}^{-\delta_\mathrm{u} F\cdot\beta}} && (9.4) \\ &= \frac{1}{1 + \frac{K_\mathrm{u}^0}{K_\mathrm{l}^0} \mathrm{e}^{-F(\delta_\mathrm{u} - \delta_\mathrm{l})\cdot\beta}} && (9.5) \\ &= \frac{1}{1 + \mathrm{e}^{(\ln \frac{K_\mathrm{u}^0}{K_\mathrm{l}^0} \cdot \frac{1}{(\delta_\mathrm{u} - \delta_\mathrm{l})} - F)(\delta_\mathrm{u} - \delta_\mathrm{l})\cdot\beta}} && (9.6) \\ &= \frac{1}{1 + \mathrm{e}^{(F_\mathrm{eq} - F)\Delta x\cdot\beta}} && (9.7) \end{aligned}$$

where $F_\mathrm{eq} = \ln \frac{K_\mathrm{u}^0}{K_\mathrm{l}^0} / (\delta_\mathrm{u} - \delta_\mathrm{l})$ and $\Delta x = \delta_\mathrm{u} - \delta_\mathrm{l}$. Berghuis *et al.* extract the full lock probability by fitting a 3-state kinetic model to the data, making the full lock probability a result of two preceding exponential processes. This would add to another term in P_l, but since the trend is course-grained, we will not be able to fit out the additional parameters. This is why we denote the distance to the transition state Δx as the *effective* distance to the transition state.

9.4.2. FUTURE DIRECTIONS – A NEED FOR HIGHER RESOLUTION

The single-molecule flow-stretching assays offer a unique insight into the direct, label-free real-time dynamics of DNA replication, as Pandey *et al.* and Elshenawy *et al.* beautifully demonstrate. They would however, benefit from higher temporal resolution. Of course most single-molecule force spectroscopy assays to date would benefit from increased temporal and spatial resolution, but the temporal resolution here (2 Hz) is most in need of improvement – and it is by far the improvement that is technically most feasible to achieve (i.e. better cameras). In the specific case of Elshenawy *et al.*, the combination of an acquisition rate of 2 Hz, the Tus-*Ter* barrier being 3.5 kb downstream of the initiation site, and a replisome velocity of ~ 1 kb a second, this would imply on average 7 data points to infer a replisome velocity. Replisomes moving at identical velocities of exactly 1 kb a second will have between 6 and 8 data points between initiation and Tus-*Ter* due to the random phase difference between acquisition rate and moment of initiation. The apparent velocity would then be interpreted to be either be 875, 1000 or 1166 bp/s just due to this one uncertainty in the measurement. The obvious limits to the spatial resolution, as well as stochastic pauses in the bursts of replication respectively will increase (or at least influence) the uncertainty and inaccuracy in the apparent velocity further. In short, the limited temporal resolution broadens the current replisome velocity distribution such that valuable information – like the existence of replisome subpopulations – is obscured.

REFERENCES

[1] T. M. Hill, J. M. Henson, and P. L. Kuempel, *The terminus region of the Escherichia coli chromosome contains two separate loci that exhibit polar inhibition of replication,* PNAS **84**, 1754 (1987).

[2] M. Hidaka, T. Kobayashi, S. Takenaka, H. Takeya, and T. Horiuchi, *Purification of a DNA replication terminus (ter) site-binding protein in Escherichia coli and identification of the structural gene,* **264**, 21031 (1989).

[3] T. M. Hill, *Arrest of Bacterial DNA Replication,* Annual review of microbiology **46**, 603 (1992), 10.1146/annurev.mi.46.100192.003131.

[4] K. Kamada, T. Horiuchi, K. Ohsumi, and N. Shimamoto, *Structure of a replication-terminator protein complexed with DNA,* Nature **383**, 598 (1996).

[5] G. S. Khatri, T. MacAllister, P. R. Sista, and D. Bastia, *The replication terminator protein of E. coli is a DNA sequence-specific contra-helicase,* Cell **59**, 667 (1989).

[6] T. M. Hill and K. J. Marians, *Escherichia coli Tus protein acts to arrest the progression of DNA replication forks in vitro,* PNAS **87**, 2481 (1990).

[7] T. M. Hill, *Escherichia coli and Salmonella: Cellular and Molecular Biology* (1996) pp. 1602–1614.

[8] F. F. Coskun-Ari, A. Skokotas, G. R. Moe, and T. M. Hill, *Biophysical characteristics of Tus, the replication arrest protein of Escherichia coli,* **269**, 4027 (1994).

[9] C. Neylon, A. V. Kralicek, T. M. Hill, and N. E. Dixon, *Replication Termination in Escherichia coli: Structure and Antihelicase Activity of the Tus-Ter Complex,* Microbiology and Molecular Biology Reviews **69**, 501 (2005).

[10] J. Y. Lee, I. J. Finkelstein, L. K. Arciszewska, D. J. Sherratt, and E. C. Greene, *Single-Molecule Imaging of FtsK Translocation Reveals Mechanistic Features of Protein-Protein Collisions on DNA,* Molecular Cell **54**, 832 (2014).

[11] M. M. Elshenawy, S. Jergic, Z.-Q. Xu, M. A. Sobhy, M. Takahashi, A. J. Oakley, N. E. Dixon, and S. M. Hamdan, *Replisome speed determines the efficiency of the tus-ter replication termination barrier,* Nature (2015).

[12] B. A. Berghuis, D. Dulin, Z.-Q. Xu, T. van Laar, B. Cross, R. Janissen, S. Jergic, N. E. Dixon, M. Depken, and N. H. Dekker, *Strand separation establishes a sustained lock at the tus-ter replication fork barrier,* Nat Chem Biol **11**, 579 (2015).

[13] S. J. Benkovic, A. M. Valentine, and F. Salinas, *Replisome-mediated dna replication,* Annual review of biochemistry **70**, 181 (2001).

[14] A. Johnson and M. O'Donnell, *Cellular dna replicases: components and dynamics at the replication fork,* Annu. Rev. Biochem. **74**, 283 (2005).

[15] A. M. van Oijen and J. J. Loparo, *Single-molecule studies of the replisome,* Annual review of biophysics **39**, 429 (2010).

[16] M. Pandey, M. M. Elshenawy, S. Jergic, M. Takahashi, N. E. Dixon, S. M. Hamdan, and S. S. Patel, *Two mechanisms coordinate replication termination by the escherichia coli tus–ter complex,* Nucleic acids research , gkv527 (2015).

[17] D. Bastia, S. Zzaman, G. Krings, M. Saxena, X. Peng, and M. M. Greenberg, *Replication termination mechanism as revealed by Tus-mediated polar arrest of a sliding helicase,* Proceedings of the National Academy of Sciences **105**, 12831 (2008).

[18] M. D. Mulcair, P. M. Schaeffer, A. J. Oakley, H. F. Cross, C. Neylon, T. M. Hill, and N. E. Dixon, *A molecular mousetrap determines polarity of termination of dna replication in e. coli,* Cell **125**, 1309 (2006).

[19] T. M. Pham, K. W. Tan, Y. Sakumura, K. Okumura, H. Maki, and M. T. Akiyama, *A single-molecule approach to dna replication in escherichia coli cells demonstrated that dna polymerase iii is a major determinant of fork speed,* Molecular microbiology **90**, 584 (2013).

[20] A. Levy, M. G. Goren, I. Yosef, O. Auster, M. Manor, G. Amitai, R. Edgar, U. Qimron, and R. Sorek, *Crispr adaptation biases explain preference for acquisition of foreign dna,* Nature **520**, 505 (2015).

[21] M. J. Moreau and P. M. Schaeffer, *Differential tus–ter binding and lock formation: implications for dna replication termination in escherichia coli,* Molecular BioSystems **8**, 2783 (2012).

10

CONCLUDING REMARKS AND PERSPECTIVES

As I sit here writing what will be the concluding chapter of my dissertation, I cannot help thinking: "There's so much left to write, so much left to do, so much left to figure out!" Apart from the biological questions we attempted to answer – with different degrees of success of course – during the course of my thesis work, all of us were continuously in the process of improving the magnetic tweezers instrumentation, the methods, the software, the data quality and quantity, the data processing and the data analysis methods. Many of these matters are elaborately discussed elsewhere either in this thesis (methods: Chapters 3 and 4; instrumentation: Chapter 5) or in other publications (software: [1]; instrumentation: [2–5]). All these matters are a continuous effort to improve the quality of the datasets, while ensuring that extensive datasets can be obtained within manageable timeframes. Yet then comes data processing, analysis of the processed data, and interpretation. So where do we stand when it comes to data quality, or more specifically, the quality of the analyzed data?

As discussed in the introduction, tracking polymerase activity along a template no longer a huge achievement these days, but actually making sense of the traces emerging from these machines can still be considered as a big achievement, even in 2016. However sophisticated your single-molecule setup, the data will contain noise, so processing has to deal with this one way or another. From the engineering perspective one could argue that we just need to keep improving our setups, gaining better resolution until one day we will get where we want to be. For example, ideally one would want to track hundreds of tethers simultaneously at acquisition rates in the order of 50 kHz (using beads that do not limit your acquisition bandwidth) and a magnification that yields a resolution of 1 nm per pixel[1]. Apart from the problem that would arise from having to process terabytes

---

[1] Current state of the art MT multiplexing setups can track hundreds of tethers at 58 Hz, at typical magnifications that result in ~100 nm/pixel (the physical pixel size also contributes to the resolution, but a full discussion of what contributes to improving resolution is besides the point here)

of data, technology will not allow these kind of acquisition frenzies for at least a while. Besides this, more is not always better, and being trained as a scientist over the past years, I plead for a scientific approach. Thorough understanding of what the possibilities and limits of our current datasets are will – in my opinion – make a larger contribution to scientific advance than a brute force 'measuring everything more and faster'.[2]

10.1. CONSIDERATIONS FOR DESIGNING A MT EXPERIMENT

Over the past years of working with MT, having performed various multiplexed MT experiments, and having observed others embarking on complicated missions towards reading out enzymatic activity along nucleic acid templates, I have often found myself at a search for a systematic approach as to what the optimal experimental design strategy would be for the particular biological system of interest. Typical questions one seeks an answer to are: What are the timescales of my system of interest and how do they depend on force? How fast should I acquire my data at the very least to be able to make statements on these timescales? How much should I subsequently filter my data, if at all? What magnification should I use? What is the optimal construct length and bead size? These are just a few of many questions that typically arise, each usually requiring finding an answer to a subset of questions. There are many parameters, and not all parameters are equally important. Not all parameters deserve tweaking, even if tweaking leads to an improved spatiotemporal resolution. Parameter space, in other words, is quite extensive and developing an intuition for it takes time.

To this end, I propose a parameter space characterization method that could help to give a rough estimation of the performance of your (ideally, future) MT experiment, with respect to the biological system of interest. The first basic assumption here is that the system has a basic unit of discrete stepping size – for a polymerase this would be the length of a single base pair, for the Tus-*Ter* on hairpins this is the further opening of 500 bp dsDNA. The second assumption is that there are a set of processes that occur at certain characteristic timescales (e.g. a stepping rate). The system of interest thus occupies a specific region in space and time (z, t-space), and magnetic tweezers based force spectroscopy can be simply regarded as a tool that makes a specific region of z, t-space accessible. In other words, we have to map the performance of a specific bead-tether-force-microscope-camera system onto z, t-space to gain insight into which region of the biological system can be made accessible, and which will never be no matter how much certain parameters are tweaked. Mapping of MT-system performance has been readily performed. [3] We can now attempt to compare the length and timescales accessible through MT with the typical length and timescales of the biological system of interest.

10

10.1.1. SIMULATING MAGNETIC TWEEZERS TO DETERMINE THE APPROPRIATE WINDOW SIZE

Making use of the MT-simulation described in detail by Burnham *et al.* [6], we can get a good idea of what a typical trace would be for a particular system. Using a simulation

---

[2] This is what makes science science, and not engineering.

timestep of 5 μs, a Faxén-corrected friction coefficient of a bead with radius R, Worm-like chain parameters for dsDNA or ssDNA of a given length to determine force-extension characteristics and trap stiffness, the thermal force acting on the system at 296 K and a simulated camera frequency including optical noise (taken to be 1 nm), we simulate traces for 5×10^6 time steps (i.e. 25 s). Subsequently we compute the Allan deviation (which more or less equals the standard deviation) of a trace over time as the averaging window increases to get an idea to how (boxcar) filtering improves the resolution. Finally, to get an idea of the required window size to successfully monitor a static process over time, we impose the criterion that all data within 3 standard deviations of the mean needs to fall within the dwell time window size. This implies the minimal window size is 6 times the Allan deviation at a given time (6-σ window).[3] The results of this exercise can be observed in **Fig. 10.1**.

For the P2 measurements, the acquisition rate was 25 Hz on RNA constructs of 2.8 kbp in length and a bead diameter of 2.8 μm (**Fig. 10.1a**), WLC parameters of DNA taken here). Each experiment starts off with dsRNA (solid lines) at a minimum force of 16 pN (orange) and a maximum force of 36 pN (blue). The stiffness of the system at both forces is comparable for dsDNA, leading to similar 6-σ window sizes. However, the contribution of the much more flexible ssRNA (Lp of ssDNA taken here 0.5 nm) increases as more of the template is transcribed, and the lower stiffness in z results in much larger fluctuations at 16 pN than at 36 pN. This leads to a proportionally larger increase of the 6-σ window size at 16 pN compared to 36 pN for 2800 bases of ssRNA at the end of the experiment (dashed lines). As the standard deviation is proportional to $1/\sqrt{N}$, the slope of the 6-σ window is a power law with a -0.5 slope. If a base pair is taken to be 0.34 nm we can get an idea of the required window size in base pairs (cyan, window size of 1-10 bp shown). Taking these observations into account **Fig. 10.1a** suggests that, to accurately monitor a static polymerase in these experimental conditions over this full force range at all times during an experiment, we would require filtering the data to 1 Hz and use a dwell time window of 10 bp (exactly the used window size in **Chapters 6 and 7**). If only the data at 36 pN would be used, one could suffice with either 1Hz filtering and a dwell time window of 6 bp, or ~2 Hz filtering and a window of 10 bp, for instance.

Similar predictions (or, in hindsight, justifications) can be made for other types of experiments (**Fig. 10.1b**). Canonical camera acquisition rates now lie around 58 Hz at full field of view, such that the first data point of the 6-σ window slope lies at shorter timescales than obtained for the P2 datasets. For an experiment starting with a 7 kb dsDNA tether at 30 pN (blue, solid) it might for instance make sense to include only the first couple of transcribed kbs in the dwell time analysis, as 7 kb ssDNA at this force (blue, dashed line) will probably require filtering extensively (to 0.2 Hz) in order to still be able to use a 10 nt dwell time window. Or one might want to reconsider bead size. A typical experiment where this would be a serious consideration is in cases where low forces are used. The 6-σ window size for a 2.8 μm diameter bead, 2.8 kbp dsDNA tether system at 5 pN (green,

---

[3] I justify this as follows: a 6-σ window implies that 99.7% of all data would fall within this window, or conversely 1 point in ~370 would not. For 1 Hz filtered data – a typical value for our datasets – this would imply an incorrect window crossing of once every 6 minutes, which seems a reasonable value for 30' measurements.

solid), filtered at 1 Hz would be ~8 bp, while the same measurement using a 1 μm diameter bead (green, dashed) would allow for a 4 bp dwell time window (or much less filtering).

However, this is only part of the story. The Allan deviation (and corresponding 6-σ window) yield what could be considered as the maximum resolution given a certain amount of filtering, but this resolution does not improve *ad infinitum*. A major determinant of the maximum obatinable resolution in z is the magnification, which determines the tracking noise and sets a floor to the obtainable resolution (**Fig. 10.1d**). At 50×, the maximum resolution was found to be ~3.3 nm by van Loenhout *et al.* [5], which is roughly a window size of 10 bp. The simulations described here also considers drift-free situations only. While drift is cut down to a minimum by using a reference bead, there is likely to be drift in experiments that last in the order of 30 mins.

10.1.2. MAPPING THE BIOLOGICAL SYSTEM TO z, t-SPACE

As we know the typical timescales and step sizes of the processes observed for P2 and Tus-*Ter*, we can map these onto z, t-space as explored above (**Fig. 10.1d**). We then see that single-base pair resolution for P2-like experiments is unattainable mainly due to the magnification used, but even if magnification would not set this resolution floor, single base pair resolution would be obtainable for steps that take in the order of 100 s (the blue area indicates the inaccessible length scales at a given time, the slope is obtained through a fit to the 6-σ window for the 36 pN 2.8 kb SSDNA P2 experiment shown in **Fig. 10.1a**). In fact, the 10 bp window roughly corresponds to the maximum obtainable resolution with the magnification used for these experiments. The amount of filtering for these particular experiments (0.5 Hz, blue dashed vertical line) could have been lower, as the floor seems to be reached with, 1 Hz filtering as well. The mapping also shows that the timescale of the nucleotide addition rate (k_{na}) is not readily accessible during these experiments and thus will be noisy, implying that statements on the nucleotide addition rate should be considered with caution. The P2 experiments are also typically designed such that the applied force assists polymerase movement, so an increase in force will lead to shortening of all timescales. To make k_{na} more accessible one could lower the force to lengthen the timescales (though this leads to an increase in the 6-σ window size) or – more effectively perhaps – make an opposing force experimental design. The most easily implementable adaptation would probably be to ramp up the acquisition frequency, though this would require cropping – thus lowering of the multiplexing capacity.

The enormous 'step size' used in the Tus-*Ter* hairpin experiments (an added length of 1000 SSDNA bases upon lock rupture) is easily detectable without filtering, such that the used acquisition frequency of 100 Hz sets the lower boundary to the measurable timescales (purple dashed vertical line). Whether at 18 pN (6-σ window purple area) or 95 pN (purple solid line in blue area), the region occupied by Tus-*Ter* in z, t-space is readily probable with the MT settings used.[4] In fact, for this assay we could in hindsight

---

[4] With the exception of some short-lived states found for mutants, see supplemental information in Chapter 8.

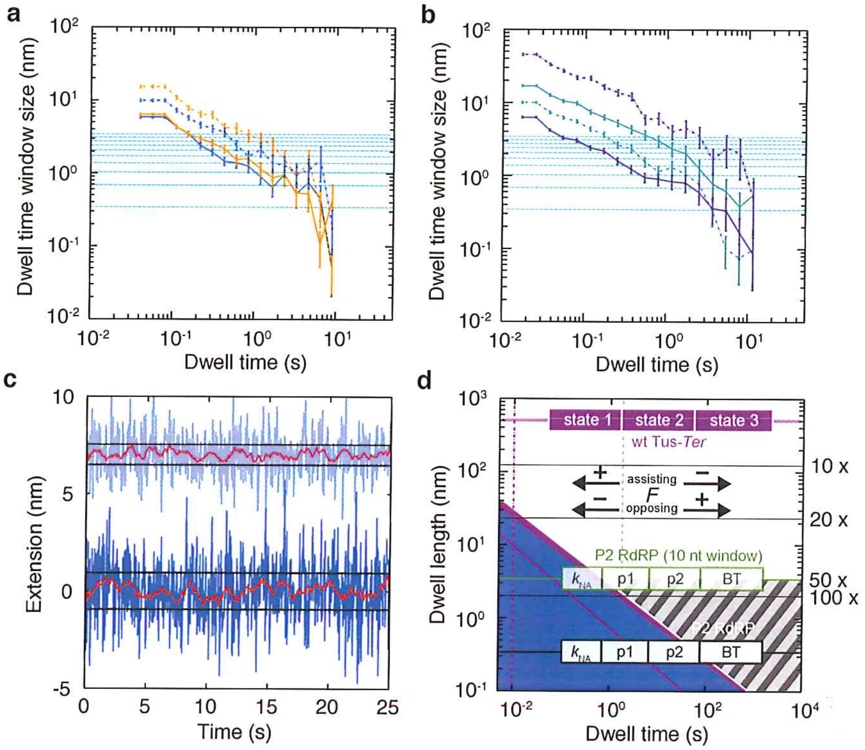


Figure 10.1: **Dwell time analysis on noisy signals arising from stochastic processes.** (**a**) The suggested minimal dwell time window size is emperically found to be roughly 6× the Allan deviation of simulated magnetic tweezers data for P2-like situations (i.e. a 6-σ window). P2 experiments start with 2800 bp dsDNA tethers (solid lines) at a lower limit of 16 pN (orange) or upper limit of 36 pN (blue), and end with a maximum of 2800 bases of ssDNA (dashed lines). The simulated camera acquisition rate was 25 Hz, while the timestep for the simulation was 5 μs, for 1.4 μm bead radius, a Faxén-corrected friction coefficient and the WLC parameters of ds and ssDNA, respectively. The required window size roughly decreases with $1/\sqrt{t}$ as expected for drift-free situations (slope ≈ -0.5). Horizontal lines (cyan) show multiples of 0.34 nm (from 1 to 10) as a rough indication of the window size in base pairs. (**b**) Dwell time window size for a P2-like experiment starting with a 7 kb dsDNA construct (blue, solid) at 30 pN, as it gets converted into ssDNA (blue, dashed). At low forces of e.g. 5 pN (green) the noise is relatively high for a 2.8 kbp dsDNA construct with 1.4 μm radius beads (solid line), so switching to 0.5 μm radius beads can improve the resolution substantially (dashed line). Horizontal lines (cyan) show the window size from 1-10 bp. The higher acquisition rate (58 Hz) shifts the starting point of the -0.5 power law to shorter timescales than in the measurements in (**a**). (**c**) The simulated data of 2.8 kbp ssDNA (blue) and dsDNA (blue-grey, 7 nm offset) at 36 pN as described in (**a**). The 25 Hz data is smoothed to 1 Hz (red), the 6-σ window at 1 Hz is shown by the black horizontal lines for both nucleic acid species. (**d**) The typical time scales and step sizes of the P2 and Tus experiments placed in perspective of the accessible timescales with MT experiments. Shown are the timescales inaccessible to P2 experiments (blue area, derived from a fit to the 36 pN 2.8 kbase ssDNA trace shown in (**a**) and Tus experiments (purple area and line, also derived form fits to 18 pN and 95 pN simulated 6-σ windows). The vertical position shows the basic stepping unit of the system (0.5 μm for Tus, 1 bp for P2), the effective experimental bandwidth is indicated by the vertical dashed lines (no filtering for Tus, 0.5 Hz filtering for P2). The tracking noise (related to the magnification, camera pixel size, etc.) determines the experimental resolution floor (grey shaded area, see text for further details).

argue that a 10× magnification) would have been more apt, as this would vastly increase the multiplexing capacity of the MT setup, computaional power allowing.

10.2. MAPPING AS AN ASSESSMENT AND TIME SAVER, AN EXAMPLE

As shown and discussed above, performing a MT simulation can be useful as a guide to come up with an appropriate bead-tether-force-filtering design strategy for a future MT experiment. This can save time, but mapping also allows for time saving by visualizing what the effect of tweaking of parameters of e.g. data processing would be. For example, one could consider spending a considerable amount of time on improving the accuracy of the nm to bp conversion by correcting for the bead attachment point as discussed by Klaue and Seidel [7]. Klaue and Seidel realized that an off-pole attachment of a DNA tether to a magnetic bead gives rise to considerable increase in δz at applied forces above 2 pN due to a presence of as well as a leveling off (saturation) of the torsional stiffness of the bead (leveling off occurs at ~5 pN). The authors note that at high forces, shear forces of up to 10 pN can easily be caused at the bead surface, which of course can have major implications for situations where object of interest is rigidly attached to the bead. Another implication is that the measured bead height will decrease as the attachment point moves to a more off-pole position. If one considers the measured bead height to encompass the full contour length of the DNA molecule, the relative extension at a given force is shown to vary ~ ±10%, which is considerable.

In a P2-like situation, where the bead is attached to the DNA/RNA and polymerase activity is read out indirectly, the activity-induced tether extension changes will be a relative measure: the difference between ds and ssDNA. The exact extension is thus less relevant[5]: the only possible worry could be the increased noise in z for the off-pole tethered beads. As all polymerase traces of a particular dataset have to be processed in an identical manner, the minimal dwell time window size depends on the noisiest trace of the dataset. The larger the dwell time window, the larger the chance of multiple pauses occurring within the dwell time window and the larger the probability that any given dwell time will contain a pause (which will lead to a merging of short, otherwise measurable pauses with the 'Gamma peak'). Therefore increased 'noisiness' as a result of off-pole attachment is likely the only reason for worry in the case of indirect polymerase readout for current MT assays.[6] As the noise level can not be corrected for, the attachment point can be a quantitative rule of thumb for weeding out particularly noisy traces.

10.2.1. THE CASE OF DIRECT POLYMERASE READOUT

The same arguments apply for experiments where the polymerase is attached directly to the bead, except for the fact that tether length will not only vary depending on the attachment point, but depending on where the polymerase initiates on the construct. The bead height at a given force can be directly translated into number of base pairs

---

[5] To the extent that that the construct appears to be the correct one.

[6] Though from the P2 measurements in Chapter6, we have no reason to assume this is a huge effect.

using a force-extension curve for that particular type of nucleic acid template. If one assumes that the measured bead height reflects a contour length as inferred from the nucleic acid force extension characteristics, the ~20% variation in bead height will then directly translate into a 20% variation in absolute extension, as Klaue and Seidel point out. The more off-pole the attachment point is, the shorter the tether length will be at a given force, the more apparent number of base pairs per nanometer, thus the larger the number of base pairs that fit in a given dwell time window. Or, if the dwell time window is set to be a given number of bp, the faster the apparent velocity of the polymerase will be the more off-pole the attachment point becomes. In addition, if off-pole attachment leads to a larger deviation in z this would require larger dwell time windows, thus a (small) increase in the probability that two pauses will fall in the same dwell time window.

So what is important here? The change in apparent velocity? Since the nucleotide addition rate generally lies at timescales that are not fully accessible with the current setups – or at least should be approached with caution – the increased variation in apparent velocity might be of minor importance. The variation in nm to bp conversion will not influence the (measurable) pause lifetimes as pausing implies 0 velocity. In this way, one could argue that the effect of bead attachment point on the dwell time distribution will be very minor, leaving the most important feature – the pause spectrum – as it is. The correction of Klaue and Seidel could yield a more accurate determination of the exact position of the polymerase along the template if the exact sequence would matter. In that case, dwell time analysis could perhaps be performed taking into account the sequence. However, a clear marker along the nucleic acid template (e.g. a GC-rich barrier, an initiation or termination site) is likely to be more apt to determine the exact location of a polymerase along a template.[8]

In summary, using the bead attachment point as a *correction* method in cases of direct polymerase attachment to the bead however, might actually be inappropriate. Not only do the noise motivated reasons discussed earlier apply here as well, the shear forces as a result of the relatively weak torsional bead stiffness are now directly applied to the polymerase. In this type of experimental assay – with shear forces up to several pN and the polymerase pause dynamics being sensitive to forces in the pN range – the bead attachment point determination should perhaps – like in the P2-like experiments – be a *selection* method rather than a *correction* method. Considering the accessible biological time and length scales (as e.g. is shown in **Fig. 10.1d**) before embarking on time-consuming corrections that give rise to an improved precision without actually affecting the accuracy of the final dataset (i.e. the dwell time distribution) in a significant or positive manner. In other words, correcting – and therefore lowering the barrier for what is considered an acceptable experiment – might actually decrease the overall accuracy of the final dataset.

10.3. THE APTITUDE OF DWELL TIME ANALYSIS

Is dwell time analysis the most appropriate method for analyzing extensive polymerase datasets? If we compare dwell time analysis to the other type of polymerase trace analysis method that exists – velocity analysis through fitting trace segments with a line – dwell time analysis has considerable advantages. To start off, both analysis types will suffer from noise in a similar fashion, as timescales that are unresolvable in one method will not become resolvable in the other. Velocity profile analysis intuitively makes more sense, as our macroscopic intuition tells us that movement is special, and immobility is trivial.[7] Velocity analysis throws all pauses onto one heap (i.e. a pause is a datapoint at 0 velocity) and focusses on characterizing the stepping velocity. However, we have shown here that the stepping velocity should be approached with caution and – of all dynamics – is the least reliable parameter we obtain from these experiments! In the nanoscopic world of polymerase movement along a nucleic acid template – a world where everything is continuously in motion – I think our intuition should be reversed: movement is trivial and pausing is special. Dwell time analysis applies this reverse intuition by visualizing the spectrum of pause times, thereby revealing a much more rich an intricately detailed spectrum of the polymerase behavior.

10.4. THE EFFECT OF NOISE ON DWELL TIME DISTRIBUTIONS

The effect of noise and filtering on polymerase dwell time distributions is discussed at various points in this thesis, mainly in Chapters 2, and 6. Nevertheless, the effect of noise on dwell time distributions is a complex issue, has not been formalized, and is a matter of ongoing research. For relatively slow polymerases such as P2 (~20 bp/s), we have argued that the 'Gamma' peak – the peak in the distribution associated with the polymerase nucleotide addition rate – might be a good indicator of the actual polymerase nucleotide addition rate. Even so, there is no 100% guarantee that the found rate is completely pause-free, as **Fig. 10.1d** shows that the pause-free timescales are at the limit of what was experimentally accessible. Other polymerases – like the replicative polymerase III of *E. coli* – can have much higher nucleotide addition rates (~150-700 bp/s). In Chapter 1 we modeled polymerases with a stepping rate of 150 bp/s. To illustrate the range of possible outcomes of dwell time analysis on traces with noise, **Fig. 10.2** compares the probability distribution functions arising from dwell time analysis of two of the polymerases modeled in Chapter 1 (shown in **Chapter 1, Fig. 1.13**, polymerases A and E). What becomes clear is that though the pause spectrum is still accessible, the Gamma peak only coincidentally reflects the actual nucleotide addition rate.

10.5. THE LIMITED APPLICABILITY OF FITTING KINETIC MODELS TO DTDS

The introduction chapter introduces[8] dwell time distributions and shows how they could be fit by kinetic models, but whether a model is the correct one can usually not infer from the dwell time distribution alone. We need structural information, information from bulk biochemical studies to support our assumptions or hypotheses. Even when a relatively simple dwell time distribution arises from a well-studied, generally-well un-

---

[7] Like for instance in the case of cars, we are generally interested in the acceleration, maximum speed, mileage – all things to do with movement – and not in the frequency and length distribution of parking times.

[8] The ensuing chapter 2 elaborates on this, chapters 6, and 7, and 8 show the applications of DTDS

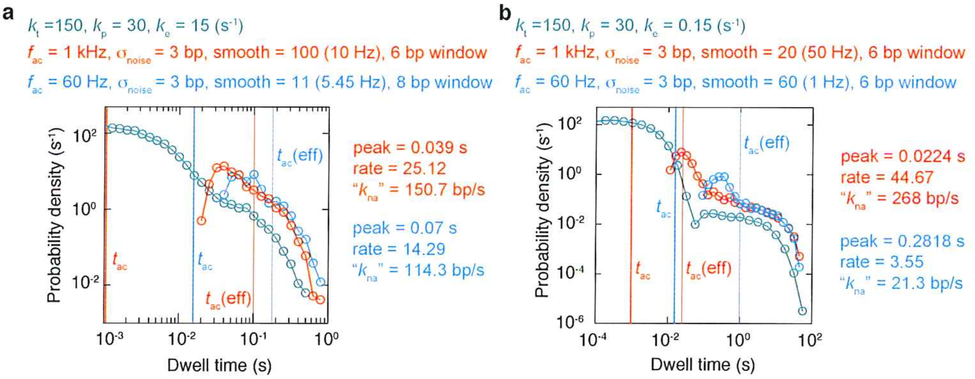


Figure 10.2: **The effect of noise on polymerase dwell time distributions.** **(a)** Dwell time distributions arising from polymerase A (**Fig. 1.13**) in the noise-free situation (green), and the same trace – yet then with added noise (cyan and orange). Gaussian noise with a standard deviation (σ) of 3 bp was added to the discrete simulated trace at an emulated camera acquisition rate (t_{ac}) of either 60 Hz (cyan) or 1kHz (orange). The resulting traces were treated and analyzed in the same manner the experimental data would have been, using a smoothing window ('smooth') and base pair window ('bp window') that was considered to be most appropriate. The effective bandwidth after filtering (t_{ac}(eff)) is also shown. The nucleotide addition rate k_{na} was calculated by dividing the dwell time distribution peak position by the base pair window size. **(b)** The same procedure, only now for polymerase E of **Fig. 1.13**.

derstood system, the model aptitude cannot be expressed with absolute certainty. For most systems detailed information to build kinetic models lacks. Dwell time distributions in addition are usually[9] rather complicated – clearly not always a nice distribution with a small number of well-defined exponentials as e.g. shown in **Fig. 10.2.**[10] In other words: fitting a kinetic model can become a pointless exercise sooner rather than later. So what to do? Just give up on the data? Ignore the importance of statistics for the time being, switch to high resolution MT, and focus on trying obtain base pair resolution (or the small intricate details of enzymatic activity, e.g. scrunching etc.)?

Dwell time distributions could on the other hand also be more than a distribution to fit kinetic models to. Though being able to fit a kinetic model to the data extracted from the system of interest would be great, it will in many cases not be trivial to state with high certainty that a specific model is correct, no matter how high the MT setup resolution is. Yet there can be other ways to gain deeper understanding of the system of interest, even if the MT assay does not reveal the underlying kinetic steps directly. For instance, as we can gather extensive data sets over a broad range of forces, NTP concentrations, drug concentrations, ionic strengths, and e.g. pH gradients, the effect of a changing parameter on the DTD can be quantified. For instance by looking at dwell times over a certain length, the fraction of dwell times within a certain time range, relative weights of various fractions etc., like we embark on in **Chapter 7** for instance. In addition, as more enzymes

---

[9] at least in the case of DTDs arising form polymerase traces

[10] Though sometimes they are simple and well-defined of course, as can be seen for various DTDs in Chapter 8, but then again, this is not tracking of enzymatic activity.

are being quantified in this manner, characteristic differences between e.g. one class of polymerases and the next might emerge. For instance: as viral polymerases generally have a higher error rate than eukaryotic replicative polymerases, this might imply the emergence of a characteristic difference in the pause spectrum of the respective DTDs. DTDs could hence potentially classify different types of polymerases.

10.6. THE FUTURE OF MAGNETIC TWEEZERS, A SPECULATION

The magnetic tweezers technique has established itself in the field of single-molecule biophysics now for well over two decades, and has proven to be a robust, simple and stable alternative to other single molecule force spectroscopy techniques such as atomic force spectroscopy or optical tweezers.[9, 10] Though the technique has been commercialized, it still requires specialist knowledge and tinkering to be of added value to a biological lab. What will be the future of the magnetic tweezers technique? Will it be a niche for specialists or will it become a broad tool for biological research? There are various strategies that can be adopted when it comes to MT, of which multiplexing is one.

Various recent experiments have shown that the other use of MT – tracking particles with high spatiotemporal resolution – can be a great alternative to optical tweezing.[3] The natural force clamping mode of MT allows for easy visualization of real time protein-DNA binding kinetics, or an investigation of fluctuating enzyme conformations on enzymatic activity.[11] High-resolution MT easily combines with fluorescence and will find various niche systems where it can come in as a useful tool to study the dynamic structure-function relationship of various biological processes.

10.6.1. HIGH LEVEL MT SOFTWARE

High-resolution MT assays will typically continue to require highly specialized technical skills and knowledge and therefore will likely remain a tool mainly used by scientists with a background in physics or engineering. Multiplexed MT setups on the other hand have more potential to become a tool used by a broader spectrum of scientists due to the relatively simple nature of the setups. For this transition to occur, significant advances in the software still need to be made. Most MT software developments to date have concentrated on improving the tracking accuracy [4, 5] or the real time multiplexing capacity.[1] At a higher level, there are many more advances in MT software that await development. For instance, a multiplexed experiment now typically involves the manual selection of a region of interest centered around a bead, assessment of the look-up table quality, testing and selecting for e.g. tether length, amplitude of the fluctuations in x, y, z, and force-extension characteristics. While this is manageable for tens of beads, this process becomes cumbersome for current day multiplexed experiments which can contain 600 tethers to start off with. Automating this step of the experimental procedure in a reliable way is a next step towards improved efficiency and user-friendliness of MT setups.

10.6.2. IMPROVING SPATIOTEMPORAL RESOLUTION

An ongoing development during the course of this thesis work was one that coincides with camera developments. Effort was made to expand the field of view without significant loss of resolution (**Fig. 10.3**).[11] while also making sure the concomittant increase in tethers could be dealt with computationally.[1] Near future camera upgrades on current setups could for instance consider increasing the magnification further while keeping the field of view size as it is ($\sim 300 \times 400 \mu$m), as the current magnifications (50-60×) limit the tracking resolution to a couple of nm, as discussed earlier. Probably a magnification of 200× would set the z-resolution floor to sub-nanometer length scales. However, as the current camera chip sizes cover nearly the complete back aperture of the microscope objective, increasing the magnification will come with a loss of field of view size sooner rather than later. To expand the field of view further considerably, the only option is going to lower magnification strengths. Any improvement of tracking resolution that coincides with an expansion of the field of view will thus have to come from higher resolution cameras: a larger number of pixels per covered field of view area. This implies a smaller physical pixel size on the CMOS camera chip and a smaller area to detect incident photons. So for any doubling of the resolution as a result of pixel size, the number of incident photons will have to quadrouple.[12]

One can also consider lowering the magnification strength to increase the allowable tracking frequency for an unchanging field of view size. As the mapping of P2 timescales in accessible z, t-space has shown, the polymerase experiments would benefit from an increase in the acquisition frequency. Cameras with bigger chip and smaller pixel sizes than the current cameras used[13] might make it possible to track in a cropped camera mode (i.e. using only a small region of the camera chip) at e.g. 300 Hz covering a field of view the size of the current full-frame field of view. Expanding the potential field of view size to enhance the spatiotemporal resolution of multiplexed MT setups is necessary, but probably a couple 100 Hz and a floor <1 nm will be an improvement to sufficiently characterize many polymerases in a P2-like manner.

10.6.3. WHEN IS ENOUGH ENOUGH?

Concerning the field of view size, one could consider further expanding it further considerably. GPU processing speeds (and tether selection software) allowing, tracking of thousands of beads simultaneously would then become possible. However, when a field of view starts to span millimeters, the homogeneity of the magnetic field across the field of view is no longer a given as a result of physical limits to the magnet size. Bigger magnets could be an answer to this problem (and a coinciding focus on expansion of the field of view in the longitudinal direction direction of the magnet gap), or maybe other techniques such as acoustic force spectroscopy (AFS) are more suitable for millimeter-sized fields of view. One could consider combining benefits of both techniques: using mag-

---

[11] Chapter 6 is the first demonstration of these efforts

[12] Assuming the pixel light-sensitivity does not change.

[13] Our current multiplexed MT setups have a 12 megapixel camera: 4096 × 3072 pixels with a size of 6μm. Using all pixels allows a maximum acquisition rate of 58 Hz, a region of e.g. 512 × 512 pixels would allow for a maximum acquisition rate of 571 Hz.

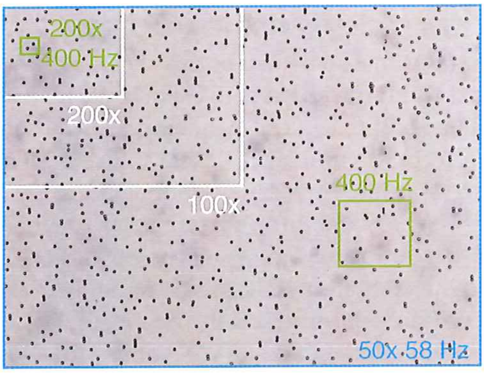


Figure 10.3: **The field of view size as a result of magnification change, acquisition rate change, or both.**

netic beads to be able to perform surface cleaning with an applied magnetic field, then switch to AFS mode for homogeneous fields without e.g. the torsional stiffness present in MT experiments.

Before embarking on these kind of expansions, it is again likely good to consider the biological system of interest. What phenomenon are you trying to observe? Is increased multiplexing needed to collect an entire dataset in a single experiment (i.e. multiplexing to increase the efficiency), or is the multiplexing needed to capture rare events? When is an event rare yet biologically interesting, and when does an event become obsolete? Is an event rare because it is biologically rare (e.g. polymerase error incorporation) or because the assay has not been optimized properly (and thus biologically irrelevant perhaps)? Finding an answer to these types of questions will remain an important step towards a good experimental design strategy.

10.6.4. HYBRID MT SETUPS

A final consideration for MT setups would be to have the best of both worlds – i.e. multiplexing and high frequency tracking – within reach in a single experiment (**Fig. 10.4**). For instance, most current polymerase tracking experiments rely on multiplexing to account for the typically low initiation rates (<5%) of the molecular motors. This means tracking at full frame field of view, where 58 Hz is the maximum tracking rate for the current cameras used. As soon as a tether shows activity however, ideally one would want to focus on this single tether and ramp up the acquisition frequency. As soon as the tether stops showing activity, it might make sense to switch back to full frame mode again, with again the possibility of tracking a different bead at high acquisition rates.[14]

These kind of hybrid, dynamically switching setups are – though a considerable technical challenge – within reach with current technologies. A couple of examples of the developments that would have to be made are the following. The CMOS cameras used

---

[14] Or the same bead again if the halt in activity turned out to be a pause!

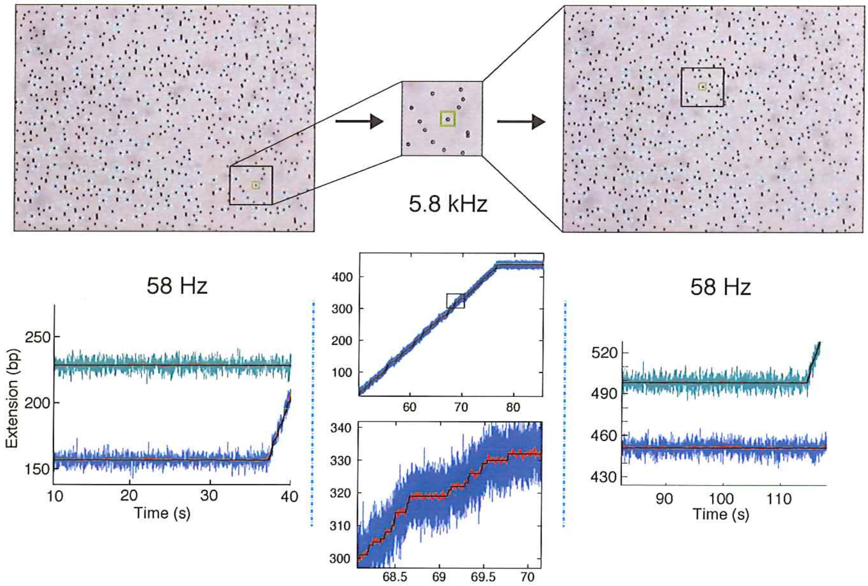


Figure 10.4: **A measurement with dynamically acquiring multiplexed MT.** **(left)** At full frame field of view, hundreds of tethers (200 ROIs shown) are tracked at 58 Hz at the start of an experiment, until activity is detected on one of the tethers (blue trace, 1 Hz filtering (orange) and simulated position (black) also shown). **(middle)** This prompts the software to crop the field of view, adjust the stage position, as well as adjust the light intensity, acquisition time and frame rate to acquire at high frequency (middle, 5.8 kHz) until activity has come to a halt (100 Hz filtered (orange), exact position (black)). **(right)** Then the software switches back to full frame, low acquisition rate mode until activity is once more detected.

crop the field of view towards the middle of the chip, so unless the active tether is located in the middle of the field of view, selection of an active tether would have to prompt a motorized microscope stage to place the tether of interest in the middle of the field of view. Each tether would also be required to have a reference bead nearby to fit in the cropped field of view. At full frame, not all these reference beads would necessarily have to be tracked, but the look-up tables would have to be made and the beads tested. The tracking software would also ideally have to be able to signal the user which tethers are likely showing activity, otherwise one would have to cycle through the entire tether list continuously at the risk of missing valuable events. The light source would probably also have to be modulated such that shorter acquisition times are compensated for the decrease in available photons.

In summary, I think there are still many possibilities for further development of the MT technique into either a user-friendly tool for broad use in biological laboratories, or for making highly advanced and powerful force spectroscopy tools in more specialized biophysics labs.[15] The broadness of MT use as a tool will largely depend on advances in

---

[15] Without directly thinking of combining different techniques, like MT and fluorescence, MT and zero mode wave guides, etc.

higher level software as described above. In my opinion, the most exciting technical advances would lie in advancing the dynamic switching capability of MT. However, also strategic decisions concerning which developments to push forward should probably involve asking questions such as: what extra information will we gain by doing this? Can we not already obtain sufficient information from the slightly higher acquisition rates or slightly better resolution that the next generation of cameras will bring us? It is once again important to consider the pros and cons of hypothesis-driven research vs. technique driven research – and whether the focus should be heavy on engineering or answering a particular biological question – to assess what is going to help you get where you want to be.

REFERENCES

[1] J. P. Cnossen, D. Dulin, and N. H. Dekker, *An optimized software framework for real-time, high-throughput tracking of spherical beads,* Review of Scientific Instruments **85**, 103712 (2014).

[2] D. Dulin, J. Lipfert, M. C. Moolman, and N. H. Dekker, *Studying genomic processes at the single-molecule level: introducing the tools and applications,* Nat Rev Genet **14**, 9 (2013).

[3] D. Dulin, T. J. Cui, J. Cnossen, M. W. Docter, J. Lipfert, and N. H. Dekker, *High spatiotemporal-resolution magnetic tweezers: Calibration and applications for dna dynamics,* Biophysical journal **109**, 2113 (2015).

[4] A. J. W. te Velthuis, J. W. J. Kerssemakers, J. Lipfert, and N. H. Dekker, *Quantitative Guidelines for Force Calibration through Spectral Analysis of Magnetic Tweezers Data,* Biophysical Journal **99**, 1292 (2010).

[5] M. T. vanÂ Loenhout, J. W. Kerssemakers, I. D. Vlaminck, and C. Dekker, *Non-bias-limited tracking of spherical particles, enabling nanometer resolution at low magnification,* Biophysical Journal **102**, 2362 (2012).

[6] D. R. Burnham, I. De Vlaminck, T. Henighan, and C. Dekker, *Skewed brownian fluctuations in single-molecule magnetic tweezers,* PloS one **9**, e108271 (2014).

[7] D. Klaue and R. Seidel, *Torsional Stiffness of Single Superparamagnetic Microspheres in an External Magnetic Field,* Physical Review Letters **102**, 028302 (2009).

[8] K. M. Herbert, A. L. Porta, B. J. Wong, R. A. Mooney, K. C. Neuman, R. Landick, and S. M. Block, *Sequence-resolved detection of pausing by single {RNA} polymerase molecules,* Cell **125**, 1083 (2006).

[9] C. Bustamante, Z. Bryant, and S. B. Smith, *Ten years of tension: single-molecule DNA mechanics,* Nature **421**, 423 (2003).

[10] C. Gosse and V. Croquette, *Magnetic Tweezers: Micromanipulation and Force Measurement at the Molecular Level,* Biophysical Journal **82**, 3314 (2002).

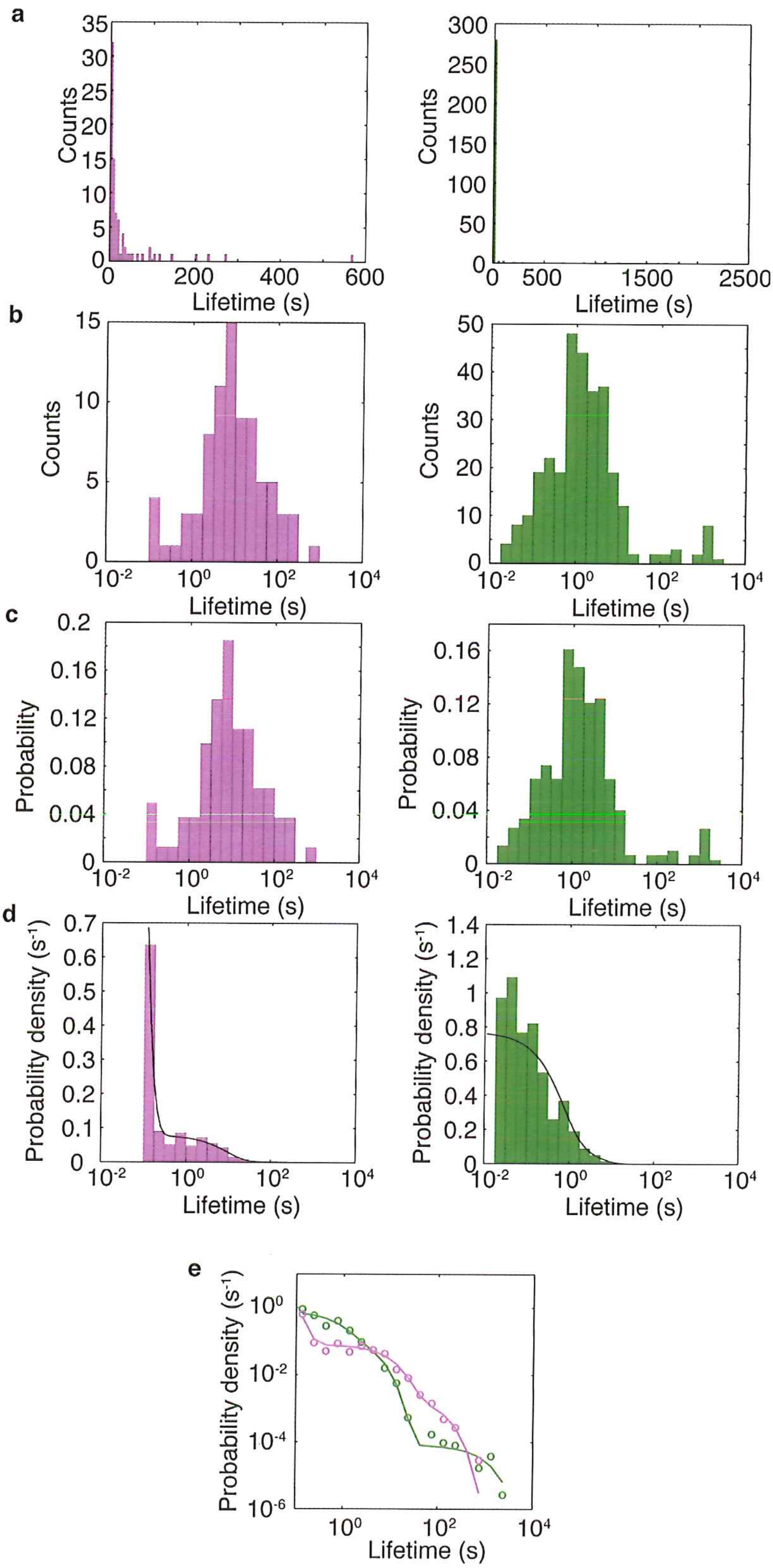
[11] Q. Guo, Y. He, and H. P. Lu, *Interrogating the activities of conformational deformed enzyme by single-molecule fluorescence-magnetic tweezers microscopy,* Proceedings of the National Academy of Sciences **112**, 13904 (2015).

A

APPENDIX A: DWELL TIME DISTRIBUTIONS

A distribution of dwell times arising from stochastic events are best represented as a probability density distribution. This implies a logarithmic binning of the data and subsequent normalization to the bin width and bin height. The probability density has units of s^{-1}. Shown here is an example of two Tus-*Ter* datasets.

The force-dependent lifetimes of WT (purple, left) or E49K (green, right) at 59 pN, 50 mM KCl. a) The occurrence of lifetimes decays exponentially, rendering a normal histogram count not very informative. **b)** With a spread of lifetimes over several orders of magnitude, logarithmic binning is more appropriate. **c)** Normalization by the total number of events yields the probability. However, the bins now are not equal in size, making statements about probability uninformative as well (larger bins have a higher probability of including an event than smaller bins). **d)** By normalizing the probabilities by their respective bin widths the probability density is obtained (black line is a fit). **e)** A log-log represtentation is the most informative manner of visualizating the data.


a
Counts
Lifetime (s)
0
200
400
600
0
500
1500
2500
0
5
10
15
20
25
30
35
0
50
100
150
200
250
300
b
Counts
Lifetime (s)
10^{-2}
10^{0}
10^{2}
10^{4}
0
5
10
15
0
10
20
30
40
50
c
Probability
Lifetime (s)
0
0.04
0.08
0.12
0.16
0.2
d
Probability density (s^{-1})
Lifetime (s)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
0.2
0.4
0.6
0.8
1
1.2
1.4
e
Probability density (s^{-1})
Lifetime (s)
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{0}
10^{2}
10^{4}

SUMMARY

THE ability to observe and manipulate biological processes at the level of single molecules is a research approach that has established itself as a valuable addition to the traditional bulk biochemical assays over recent decades. Single-molecule force spectroscopy techniques such as atomic force microscopy (AFM), optical tweezers (OT) and magnetic tweezers (MT) can infer the forces involved in molecular reactions and interactions, or use force as a reaction coordinate to drive biochemical reactions in a specific direction. In the brownian motion dominated world of enzymatic or (nucleic acid) protein interactions, the outcome of a single molecular reaction is stochastic in nature, which implies that single-molecule observations need to be repeated extensively – typically at least several hundreds of times – to obtain a statistically compelling view of the scope of possible reaction outcomes. Many of the processes that are typically of interest – e.g. polymerase pause dynamics or protein-DNA dissociation rates – involve several kinetic steps that are stochastic in nature, thus typically cover a wide range of timescales to complete: a single event can take anywhere between a microsecond and several thousands of seconds to take place. Sequential repetition of single-molecule measurements can thus become time consuming and cumbersome: this has been the driving force behind parallelization of single-molecule measurements.

A single-molecule technique specifically well suited to pursue the development of parallel single-molecule measurements – i.e. multiplexing – is magnetic tweezers. In magnetic tweezers measurements, micron-sized magnetic particles (beads) are tethered to the surface of a fluid chamber (flow cell) through a single polymer molecule, typically a nucleic acid polymer such as DNA or RNA. The surface area of the flow cell is typically in the order of a few cm^2, implying that – if optimized sufficiently – the fluid chamber will contain several hundred thousand individual tethered beads. The micron-sized beads are, unlike the DNA or RNA molecule attaching it to the surface, visible through conventional light-microscopy by using a (non-coherent) light source and a microscope objective. The beads are continuously in motion as a result of the brownian motion of the surrounding molecules, yet their motion is limited by the molecule tethering it to the surface. The image of the flow cell surface leaving the objective back focal plain is projected onto a camera chip, such that the field of view image can be recorded. The field of view size depends mainly on the magnification used and the fraction of the back focal plain covered by the camera chip. Current state of the art setups typically cover an area of $\sim 300 \times 400 \mu$m at 50× magnification and can contain up to ~ 800 tethers.

The actual magnetic tweezing comes in by mounting a pair of permanent magnets directly above the flow cell. Lowering the magnets towards the flow cell surface increases the local magnetic field strength and thus the upward pulling force on the magnetic beads, a force that is directly applied on the single molecules tethering the beads to the surface. The tethered beads can be regarded as an inverted pendulum, where the lateral bead movement is proportional to the length of the tethering molecule, and inversely proportional to the applied force. This lateral bead movement – or more precisely the spectral bandwidth of the lateral movement – is used as a measure to infer the applied force on a single molecule. Contrary to optical tweezers – where a single particle is trapped in a focussed laser beam similar to the way a macroscopically scaled tweezers can 'trap' or pinch a macroscopic object – the term magnetic tweezers is thus somewhat misleading in the sense that this technique treats all the tethered molecules within the field of view in an identical manner by 'clamping' them with a magnetic force; there is no possibility to precisely control the position of the beads, nor can the beads be treated on an individual basis. This does imply however, that magnetic tweezers – contrary to optical tweezers – has a natural multiplexing capacity. In other words, magnetic tweezers is ideally suited to probe a broad spectrum of outcomes of a single molecular reaction in a single experiment. The multiplexing extent of magnetic tweezers purely depends on the amount of trackable molecules per unit area, as well as the total area that is covered by the camera.

This thesis encompasses the efforts to improve the multiplexing capacity of magnetic tweezers for measurements that require the application of relatively high tensile forces (e.g. 15-95 pN) on DNA or RNA. All the experimental work covered in this thesis does not focus on inferring the physical properties of the tethered DNA or RNA molecule in particular, but rather uses the properties of the different forms the nucleic acid polymers can adopt (e.g. conversion of double-stranded DNA into single-stranded DNA) to infer the dynamics and kinetics of proteins or enzymes interacting with the nucleic acid molecules. As such, we indirectly infer the kinetic properties of a protein or enzyme by monitoring how it modifies DNA or RNA. The transitions DNA or RNA undergo are, in turn, indirectly inferred through the bead movement.

After introducing this particular field of single-molecule biophysics in Chapter 1 and discussing the multiplexing developments, applications, and analysis methods in greater depth in Chapter 2, Chapters 3 and 4 have a particular focus on improving the quantity and quality of the tethered beads per experiment. More specifically, Chapter 3 explores the possible advantages of covalent linkages between surface, DNA, and bead. Covalent linkage has the potential to improve tether lifetime at high forces, as covalent bonds can withstand nanoNewton forces while the strongest traditional electrostatic anti-body interactions used for tethering can only withstand up to ~ 150 pN for a limited amount of time. Chapter 4 covers a more pragmatic approach to improving the lifetime of tethers at high forces, while keeping the tether density to a maximum. We discuss the method application to two types of experiments, of which one is the characterization of the Tus-*Ter* system described in detail in Chapter 8, the other experiment shows how we measure and analyze the DNA-unwinding activity of the Large T antigen (LTag) helicase of simian virus 40 (SV40). Though the multiplexing capacity increase that coincides with the projects featured in this thesis goes hand in hand with major technical developments such as camera and tracking software improvements, these are not features covered in this thesis. Needless to mention, the experiments in this thesis would not have been possible without these instrumentation advances.

In Chapter 5 we use the equilibrium folding and unfolding dynamics of a short DNA hairpin to demonstrate the workings of a home built, Arduino-controlled magnetic tweezers temperature control system. The ability to control and stably maintain a specific temperature does not only reduce experimental drift, but also allows us to measure enzymatic activity at temperatures that are closer to physiological temperatures (e.g. 37°C) than room temperature measurements are. We demonstrate that the flow cell temperature stability improves significantly when the temperature control system is enabled and, following calibration, the free-energy difference between the folded and unfolded state of the hairpin decreases linearly with linearly increasing temperature.

The ensuing chapters (6-9) cover the main experimental body of this thesis. In chapter 6 we measure the transcriptional activity and pause dynamics of the viral RNA dependent RNA polymerase P2 of bacteriophage $\phi 6$. Transcriptional activity is tracked by monitoring the length increase that coincides with conversion of dsRNA into ssRNA by the polymerase in the 20-35 pN force range. Using extensively multiplexed magnetic tweezers, we are able to collect a large amount of these single polymerase traces at a range of forces and nucleotide (analog) concentrations. We observe the pausing behavior of P2 recording the time the polymerase takes to move through a window of 10 nucleotides. We observe two types of pauses with a distinct difference in lifetime and probability, as well as backtracking behavior. We propose a kinetic model for this enzyme by combining our observations with structural and biochemical data. As polymerase pausing is relatively rare, full characterization of the pause states would not have been possible without multiplexed MT measurements.

Chapter 7 zooms in on the rare phenomenon of backtracking in more detail, as the kinetic model proposed in chapter 6 does not include the P2 backtracking behavior. While backtracking has been characterized for multi-subunit RNA polymerases, this work describes the first coverage of a single-subunit RdRP. We hypothesize how backtracking could facilitate the emergence of another initiation site for RdRP transcription.

Chapter 8 describes the efforts to investigate a hypothesis dating from 2006, suggesting that Tus-*Ter*, a protein-DNA complex found to play a role in the termination of DNA replication in *E. coli*, blocks replisome progression by forming a tight interaction between a *Ter* DNA base and specific amino acid residues of Tus that is induced by strand separation invoked by the approaching replication machinery. As this hypothesis was surrounded by controversy (others favored a replisome-Tus interaction mechanism) and MT is a technique well suited to invoke the force-induced strand separation of DNA hairpins that mimicks replisome progression, we set out to test the strand-separation hypothesis. Soon enough we discovered that the Tus-*Ter* lock forms readily in hairpin unzipping experiments, such that we set out to characterize Tus-*Ter* lock formation. This chapter presents a detailed analysis of the Tus-*Ter* lock by measuring the lock lifetimes under an applied constant force for a range of experimental conditions, Tus mutants, and *Ter* variants. We validate the 2006 hypothesis, thereby providing a clear-cut answer to an existing controversy. However, new controversies emerge as two other single-molecule studies shed light on different aspects of the Tus-*Ter* system.

This is why Chapter 9 sets out to re-evaluate the status of Tus-*Ter* in light of the findings presented in Chapter 8 and two other single-molecule studies investigating Tus-*Ter*. The three studies turn out to be complementary in many aspects, allowing us to gain additional insights by combining the information available. The chapter touches upon the remaining outstanding questions surrounding the Tus-*Ter* system and its role in *E. coli*, as well as the technical challenges lying ahead.

In the concluding Chapter 10, I express my personal view on the challenges that remain in the field of multiplexed MT measurements. I discuss the considerations to make before embarking on measuring a certain biological process at the single-molecule level. I then proceed to discuss the pro's and cons of dwell time analysis compared to velocity analysis and the effect of experimental noise. Though the effect of noise on data analysis is a matter of ongoing research and still needs to be formalized, I elaborate briefly on which aspects of e.g. polymerase readout should be treated with caution. Finally, I touch upon what I think are some of the most exciting future challenges and possibilities of multiplexed single-molecule magnetic tweezers.

SAMENVATTING

IN traditioneel biologisch onderzoek worden de eigenschappen van een systeem – bijvoorbeeld een molecuul of een complex van verschillende moleculen – onderzocht door het gezamenlijke gedrag van minimaal miljarden identieke kopieën van dat systeem uit te lezen. Sinds enkele decennia hebben met name experimentele fysici de ontwikkeling van enkel-molecuultechnieken voortgestuwd. Enkel-molecuultechnieken zijn – zoals de term suggereert – technieken waarbij het gedrag van een molecuul wordt onderzocht door één enkel exemplaar van het systeem te bekijken. Dit geeft een drastisch andere kijk op biologische systemen door het directe inzicht in de dynamiek en stochastiek – een mate van willekeur die ontstaat door de Brownse beweging van moleculen waaraan alle deeltjes op deze schaal onderhevig zijn – van het biologische proces die je op deze manier verkrijgt. Met andere woorden, traditioneel biologisch onderzoek toont de gemiddelde uitkomst van een biochemische reactie, enkel-molecuul biofysica laat de rijke schakering zien waaruit dat gemiddelde doorgaans is opgebouwd.

Enkel-molecuul krachtspectroscopietechnieken – zoals atomaire krachtspectroscopie, de optische- of de magnetische pincet – zijn technieken waarbij de kracht die moleculen op elkaar uitoefenen wordt uitgelezen. Daarnaast kunnen deze technieken omgekeerd ook kracht uitoefenen op moleculen om de uitkomst van een biochemische reactie te een bepaalde kant op te sturen, net zoals bijvoorbeeld temperatuur daar ook voor gebruikt kan worden. Een bijkomende uitdaging van het meten aan een enkel molecuul is dat de meting vaak herhaald moet worden (typisch minimaal enkele honderden malen) om een compleet beeld te krijgen van het spectrum aan mogelijke uitkomsten. Herhaling van een experiment kan sequentieel gedaan worden, maar dit is vaak een tijdrovende bezigheid gezien één meting niet zelden enkele duizenden seconden kan duren. Parallellisatie van enkel-molecuulmetingen is daardoor een efficiënte strategie, met als bijkomend voordeel dat je meestal beter kunt garanderen dat alle enkele moleculen onder *exact* dezelfde omstandigheden gemeten zijn. Van de drie genoemde krachtspectroscopietechnieken is de magnetische pincet van nature het meest geschikt voor parallellisatie.

De magnetische pincetopstelling is een geïnverteerde lichtmicroscoop-opstelling waarbij een permanent magneetpaar een magnetisch veld in een door het microscoopobjectief gevisualiseerde vloeistofcel aanlegt. In de vloeistofcel zelf worden magnetische bollen met een typische diameter van één tot enkele micrometers aan het oppervlak geboeid door een enkel polymeer – in de praktijk is dit polymeer meestal DNA of RNA. Onder geoptimaliseerde omstandigheden bevat de hele vloeistofcel enkele honderdduizenden aan het oppervlak geboeide bollen. De bollen zijn klein genoeg om onderhevig te zijn aan de Brownse beweging van de omringende vloeistofmoleculen, maar – in tegenstelling tot DNA of RNA – groot genoeg om met een lichtmicroscoop waargenomen te kunnen worden. De mate van beweging wordt voorts beperkt door het aan de bol gebonden DNA of RNA molecuul, hierdoor is de mate van bolbeweging een afgeleide van de DNA/RNA beweging. Middels een camera en geavanceerde volgalgoritmes kan de bolpositie in 3 dimensies gevolgd worden en kunnen zo de eigenschappen van het enkele molecuul uitgelezen worden.

Het aantal enkel-molecuulboeien dat tegelijk gevisualiseerd kan worden is afhankelijk van de gebruikte vergroting van het objectief alsmede de mate waarin een camera het beeld dat het objectief verlaat kan opvangen. In huidige opstellingen komt dat in de meest gunstige omstandigheden neer op ongeveer 800 boeien. De in de vloeistofcel heersende magnetische veldsterkte resulteert in een opwaartse trekkracht die op de bollen uitgeoefend wordt – een kracht die dus ook op de moleculen die de bollen aan het oppervlak geboeid houden uitgeoefend wordt. De krachtgrootte neemt exponentieel toe naarmate de afstand tussen het magneetpaar en de magnetische bollen in de vloeistofcel verkleint. In feite moet de term 'pincet' dus niet te letterlijk worden genomen, gezien het magneetveld alle enkel-molecuulboeien dus op eenzelfde manier aan een kracht onderwerpt. Maar juist hierin schuilt de natuurlijke parallellisatiepotentie van de magnetische pincettechniek.

Dit proefschrift bundelt de in de afgelopen jaren geleverde inspanningen om de kwaliteit en kwantiteit van geparallelliseerde enkel-molecuul magnetische pincetmetingen op te schroeven. De nadruk ligt in dit werk (hoofdstukken 3-5) voornamelijk op het optimaliseren en verbeteren van de efficiëntie en de sterkte waarmee DNA (of RNA) aan het vloeistofceloppervlak en de magnetische bol gebonden wordt. Het verbeteren van de bindingssterkte was met name van belang daar het experimentele werk dat dit proefschrift omvat (Hoofdstukken 6-9) de toepassing van relatief hoge krachten (15-95 picoNewton) vereiste. Het toenemende parallellisatievermogen ging uiteraard ook gepaard met verregaande instrumentatie-ontwikkelingen aan de opstelling hardware (installatie van betere camera's) en de software (versnellen van de volgalgoritmen). Dit zijn zaken die in dit proefschrift minder letterlijk aan bod komen. Voorts beschrijven de daaropvolgende 4 hoofdstukken (Hoofdstukken 6-9) experimenten die met de geparallelliseerde magnetische pincetopstelling gemeten zijn.

In het kort kunnen hoofdstukken 1 en 2 beschouwd worden als inleidende hoofdstukken voor respectievelijk het veld van de biofysica in het algemeen en de ontwikkelingen, toepassingen en data-analysemethoden met betrekking tot de geparallelliseerde magnetische pincetmetingen in dit proefschrift. Hoofdstuk 3 verkent de mogelijkheden en de potentie van het gebruiken van covalente boei- en verankeringsmethoden en vergelijkt deze met de – doorgaans gebruikte – op elektrostatische antilichaaminteractie gebaseerde verankeringsmethoden. Hoofdstuk 4 is in feite een stap voor stap pragmatische handleiding voor een succesvolle totstandkoming van geparallelliseerde enkel-molecuulmetingen in het hoge krachtenregime. Het presenteert het scala aan DNA/RNA-molecuul ontwerpmogelijkheden en illustreert de methode aan de hand van 2 experimenten: meting van de interactiesterkte tussen het Tus-eiwit en een *Ter*-DNA-sequentie (zoals in detail in hoofdstuk 8 beschreven), alsmede hoe de Large T antigen helicase van

Simian virus 40 de dubbele DNA helix afwikkelt.

Hoofdstuk 5 beschrijft het ontwerp en de implementatie van een temperatuurcontrolesysteem voor de magnetische pincet. Niet alleen is temperatuurstabiliteit is een vereiste voor enkel-molecuulmetingen: het kunnen meten bij temperaturen tussen de 30°C en 40°C maakt het mogelijk om biologische processen te meten die niet eens op kamertemperatuur tot stand kunnen komen. We beschrijven de stabiliteit van ons systeem alsmede het kalibratieproces, en demonstreren alsmede kwantificeren de temperatuurgevoeligheid van een bij constante kracht reversibel op- en ontvouwende DNA haarspeldstructuur.

In hoofdstukken 6 en 7 meten we hoe de RNA-afhankelijke RNA polymerase P2 van bacteriofaag $\phi 6$ een dubbelstrengs RNA-polymeer kopieert (transcribeert). We lezen de snelheid waarmee dit gebeurt uit doordat dit proces gepaard gaat met de transformatie van RNA in de dubbelstrengs helixvorm naar de enkelstrengs lineaire vorm. Doordat de magnetische pincet een constante kracht uitoefent op het RNA – en enkelstrengs RNA een langgerektere vorm aan kan nemen dan dubbelstrengs RNA – gaat dit gepaard met een lengtetoename naarmate er meer enkelstrengs RNA ontstaat. We karakteriseren de polymerase door onder andere te meten bij verschillende krachten en nucleotideconcentraties. We meten voor honderden polymerasetrajecten hoe lang dit enzym erover doet om 10 nucleotiden te transcriberen. Op deze manier brengen we het stochastische (pauzeer)gedrag van de polymerase in detail in kaart en stellen we een kinetisch model op dat twee verschillende pauzetypen onderscheidt. Hoofdstuk 7 belicht een ander aspect van het P2 gedrag: een door diffusie gestuurd proces waarbij de polymerase een aantal stappen op zijn schreden terugkeert (backtracking). We postuleren dat dit relatief zeldzame fenomeen kan leiden tot het ontstaan van een initiatiemogelijkheid voor een nieuwe polymerase en laten zien dat de polymerase pauzedistributiespectra ook zonder kinetisch model gekwantificeerd kunnen worden.

Hoofdstuk 8 analyseert de interactie tussen DNA-bindingseiwit Tus en de specifieke DNA sequentie waaraan het bindt *Ter*. Het is een eigenaardig systeem dat een rol speelt bij de beëindiging van DNA replicatie in de *E. coli* bacterium. De eigenaardigheid zit in het feit dat het Tus-*Ter* systeem als een soort diode fungeert: afhankelijk van welke kant de DNA replicatie-eiwitten Tus-*Ter* benaderen, worden de replicatie-eiwitten doorgelaten of juist tegengehouden. Er heerste enige controverse rondom de veronderstelde manier waarop dit eenzijdige tegenhouden door Tus-*Ter* zou werken. De kristalstructuur en biochemische studies suggereren een mechanische werking die veroorzaakt wordt door het met replicatie gepaard gaande afwikkeling van het DNA. Er is echter ook nog een andere, moeilijk uit te sluiten hypothese die suggereert dat een direct contact tussen Tus en de replicatie-eiwitten verantwoordelijk is voor de asymmetrische Tus-*Ter* blokkade. De kracht die de magnetische pincet op moleculen uitoefent kan gebruikt worden om de met DNA replicatie gepaard gaande afwikkeling van de dubbele DNA helix na te bootsen – afwikkeling in afwezigheid van de replicatie-eiwitten dus. Deze experimenten toonden direct aan dat de diodewerking van Tus-*Ter* ook zonder replicatie-eiwitten functioneel is. We karakteriseren het systeem voorts aan de hand van allerlei aan Tus of *Ter* aangebrachte mutaties, en stellen een kinetisch model op waardoor we het aannemelijk maken dat het Tus-*Ter* blokkeringsmechanisme in drie sequentiële stappen totstandkomt.

Hoofdstuk 9 borduurt voort op de in hoofdstuk 8 vergaarde inzichten, echter nu ook met inachtneming van twee andere op Tus-*Ter* gerichte enkel-molecuulstudies waarbij er *wel* replicatie-eiwitten gebruikt zijn. We bespreken de huidige stand van zaken aan de hand van de drie studies – studies die in grote lijnen tot gelijkwaardige conclusies komen, maar in de details elkaar ook hier en daar lijken tegen te spreken. Ook blijkt er sprake te zijn van enige complementariteit: door de in verschillende studies gemaakte observaties te combineren komen we ook tot nieuwe inzichten – zoals een allereerste afschatting van de kracht die het *E. coli* replicatiemechanisme op het DNA uitoefent. Uiteraard wordt er ook een blik op toekomstige uitdagingen met betrekking tot opheldering van dit ogenschijnlijk eenvoudige moleculaire blokkeringsmechanisme.

In Hoofdstuk 10 breng ik mijn persoonlijke conclusies aangaande de geparallelliseerde magnetische pincettechniek te berde, en richt ik me meer op de technische uitdagingen die naar mijn persoonlijke mening voor deze niche van de biofysica in het verschiet liggen. Zo tracht ik ook een aantal richtlijnen op te stellen voor het optimaal ontwerpen van een DNA/RNA-boeisysteem, afhankelijk van de eigenschappen van het te onderzoeken biologische systeem. Ook tracht ik het effect van experimentele ruis op de geanalyseerde data te bespreken. Hierbij dient wel vermeld te worden dat onderzoek naar het effect van ruis nog werk in uitvoering is: het complete effect van ruis op de gemeten polymerase (pauze)distributies begrijpen we nog niet volledig en dient nog geformaliseerd te worden. Tot slot schets ik wat ik zie als potentieel interessante ontwikkelingsrichtingen voor de geparallelliseerde magnetische pincettechniek.

ACKNOWLEDGEMENTS

Congratulations! You have made it through the scientific chapters towards the final part of my thesis. What a relief it must be to finally reach a lighter and more personal note after reading all those science filled pages. Now for a very special note dedicated to those whom without this thesis would not even exist – as well as those whom without it would just have been a lot less fun. Excuse me? This is the first section you are reading? The only section even?! Where has your passion for science gone? Well, fair enough I guess. At least you in this way acknowledge it is people that make science possible! Thanks to the many I have had the pleasure to scientifically interact with over the past four years, I now know so much more than I did back then,[1] for which I am extremely grateful.

Of course Nynke, you are the first person I want to extend my gratitude to. Thank you for making this whole adventure possible to begin with. Where are the days filled mostly with you piling up ready-for-grading statistical physics homework sets on my desk? Over the past four years I have gotten to know you as a hard-working professional, someone who's standards set for others are almost as high as those you set for yourself. Of the many things I have learnt from you, I enjoy the acquired writing/story building skills the most every day. Thank you for always being very open to me about your line of reasoning as to why you make certain decisions – this has also been a valuable learning experience. Thank you for always supporting me, for trusting me to learn to make my own judgements, for your generosity when it comes to conferences and lab trips, as well as your amenability to deliberate a certain (research) approach. As I grew into my PhD I feel we both learned to know exactly how we could count on each other, this made us a very efficient team. Thank you.

David, of course you are next in line. My mentor from day 1, understood what you were saying from day 2, and a lifelong friend from day 3. Thank you for all you have taught me, for all the instrumentation developments you have pushed forward over the years (which without I would still be measuring one hairpin at a time – in true single-molecule fashion though!), for sharing your passion for doing research, for all the bad jokes, good advice, and countless other things that got me where I am now. We were a particularly fantastic team during the long P2 months (during which I at some point lost my training wheels), grabbing Terabytes of data day and night, I will not forget. Miraculously, you were able to do this all – and become a PI(!) – while reading *L'équipe* for 15 hours a day, incredible. Naturally, it is an absolute honor to have you as one of my paranymphs.

Martin! *De facto* my other PI – as well as my advisor and friend. Thanks for letting me barge into your office unannounced on countless occasions to pick your gifted, mathe-

---

[1] Even though knowing more makes you increasingly aware of all the things you do not know.

matically inclined brain. I have learnt a lot from you, both scientifically as on a personal level.

I would also like to thank my committee members for the efforts you have made in reading my thesis and pointing out the various inconsistencies the draft version still contained. Thank you to the TU Delft based members Cees, Marileen, Martin, Nynke – and I would especially like to thank the external members Gijs Wuite, Peter McGlynn, Vincent Croquette and Nicholas Dixon for respectively traveling (all the way) to Delft from Amsterdam, York, Paris, and even across 9 timezones from Wollongong, Australia for my defense. I am honored!

A special word of thanks to Prof. Dixon – Nick – not only because you are clearly not afraid of traveling great distances, but because you have been our close collaborator ever since my very first days in Delft. Back in those days it was not at all clear what kind of results would emerge from single-molecule studies of this peculiar system called Tus-*Ter*, or if placing precious protein samples into the hands of an unexperienced grad student of a biophysics group was even a good idea in the first place. Thank you Nick, for your trust, the continuous support, the many hours of scientific discussions over Skype, and sharing your vast body of knowledge with me. In addition, you have also taught me that science can be a very passionate enterprise, as different kinds of experimental assays can yield opposing views. Yet what I have also learned from you is that these opposing views are often the driving force towards a better scientific understanding. Thank you.

Where would BN be without the professional support from its secretaries and support staff? That's right, nowhere! Dijana, thanks so much for all that you have done for me ever since Texel, and for the joking around about Sinterklaas of course. Emmylou, you are seriously cool. Thanks for the Cuba advice, sharing the latest ins and outs, and all the Kavli events you have (co)-organized. Amanda, thanks for all the newsletters and your help. Thank you Jolijn and Chantal for making the secretaries' office generally a great place to go for a chat. I am sorry that Christmas trees keep on falling apart before Christmas, I guess trees are not what they used to be. A big thank you also to Esther, Marije, Liset, Angela and others for all the energy you invest in BN: you have fostered the growth of this department.

This thesis would not be the same if it were not for the world class technical support I have enjoyed. Bronwen, then Theo, and later Andrea: it is not entirely unlikely that I would – in a final act of desperation – have resorted to trying to grow constructs by planting oligo's in soil and watering them, if it were not for you. Thank you so much. Thank you Jelle for the great instruments and bespoke parts you have made for me. Thank you also to Susanne, Ilja, Jacco, Serge, Erwin, Anna, Inge, Margreet, Jaap, Sasha, and Roland.

Jelmer, where would I be without you? First of all, it would have been a lot less fun making jokes on David in D012 on my own. Besides this, I would probably be tracking beads by eye if it weren't for you. You're a first class whizz kid – and good company in the Klooster. The short line of communication (i.e. between experimentalist and software developer) we had was, in my opinion, a true symbiosis. Your work was absolutely pivotal in paving the way for all the work published in 2015 and onwards – thus also this thesis. Thank you.

Cees, also a paragraph dedicated to you. Though I did not end up in your lab (I do admit that I would have had a hard time hiring myself back then as well), it was an absolute pleasure to learn from you the traits of (Kavli) column writing – sculpting the text into a clear message like a sculptor crafts his statue. We also had some other interesting exchanges, where you showed to always be up for a debate. When I arrived here, BN still had to fully enter its exponential growth phase, and look where we are now. Our large, vibrant, socially active department would not be here if it were not for you. BN is one-of-a-kind and your brainchild, you can be proud of that.

My thanks goes out to the students who have been brave enough to embark on a project with me, or in some cases have been partially supervised by me. Every supervising experience was a learning experience for me as well. Thanks Kollol (where did those bugs come from?), Tom and Christina, Ivana – greatness is your destiny, and Max. A special thank you to Jordi for all your efforts in building the temperature control system with me. Thanks for being a native speaker of several programming languages by the way, I am curious what life has in store for you. Thanks Jurjen for so rigorously putting the temperature control system to the test during your project.

A big thank you to the BN social crew: Jan, thanks for your friend- and mentorship and endless supply of wisdom & anecdotes from day one to this very day. I am sure you are a fantastic PI. What about our index? Or Oktoberfest? / Francesco, you are one of the "big three", just so you know / Jacob, thanks for all the existential discussions, they are exactly what one needs every now and then. The cover of this thesis already shows how highly I think of your artistic skills, in fact both in science and painting. Thank you. / Iwijn, who would have known that I would meet the most competitive Flemish guy ever, in Delft? / Marijn “super-confidence” van Loenhout, thanks for your hospitality in Vancouver / Calin, thanks for all the laughs, best of luck in Cali / Charl, that was one hell of a wedding! / Laura Dickinson, watch out for those spinners / Michela “Regina” De Martino, grazie mille! signed off: vostro umile servitore / Jetty, thanks for being at the center of social organization back in the days, best of luck with the final stretch! / Christophe, you are an inspiring scientist with superb office dance moves, thank you for those countless lunches together / Yaron, thanks for bringing a conscience to the department – and Nelson to the parties / Laura, your smooth Latin American vibes are exactly what a department like ours needs / Tessa, thanks for joining us occasionally on the drill / Behrouz, the gifted savior of dwell time analysis, best of luck! / Roy destroy Lionfit, that was a memorable breakfast at the Mariott, glad you un-left [sic] the lab / Sumit: the level of absurdist banter reached new heights upon your arrival, I tip my hat at you / The Gregs: Greg "Schnanolab" Schneider / Greg “go USA!” Bokinski / Greg "proud father of Olaf the progidy novelist" Pawlik / Adi, our favorite Santa, never a dull party without you, stay safe though. / Fabai, the artist on many different length scales / Yoones Kabiri Kabiri, how’s it going? Thanks for all the laughter and letting the hallway know when-ever you enter or exit / Fede, melancholy suits you well, you and I both know that in a way the future of BN depends on you / Louis, always nice conversations that end up you sending me too complex papers, stay clear of canals though / Steef, I doubt I will ever find a skiing partner as fine as you / Mathijs, suuuup? Living the postdoc dream, respect! / Daniel V, your relentless positivism is admirable / Dan Burnham, thanks for all the tweezers discussions and your British sense of humor / Mathia, didn't we have an agreement? Ik ga niet meer voor je hakken trouwens! / Sophie, love those dance moves, thanks for sharing & teaching / Keegstra, you e-mail terrorist! / Thijs, thanks for D012 laughs back in the days / Kuba, don't skip leg day, take care of your students, and good luck with your challenging project / Elio, thanks for sharing your astute observations at the fora / Vincent Lorent, in many ways the quintessential Gentleman, thank you for bringing that flair to the department in a way only you could have done / Eugen, love that humor of yours, probably because you are merely speaking the truth / Yera, you are a gifted man, sometimes a lesser man like myself can be a downright nuisance, I agree! / Natalia, thanks for all those parties and your fantastic hospitality in Moscow! / Richard, the king of surface chemistry / Siddarth, don't be cocky man / Sergii, biology is messy, keep on making those great pictures though / Misha, bridging the gap between theory and experiment, both literally and socially: respect! / Aysen, let's meet up at burning man / Sung Hyun, best of luck in Korea / Jorine, like what you've done with your hair / Nicole bangelijk Imholtz: allez komaan, ge kunt het! / Malwina, let the good times roll in Oregon / Luuk, call me if you can think of a CRISPR experiment that involves kitesurfing / Mo, the king of the pitch and a regular guest at the FvB, thanks for your friendship / Bertus, thanks for being such an inspiring speaker / Essengül "Essie!"/ Simon Lindhoud, mijn mede-Wageninger in Delft / Ruben, the current driving force behind Real RKC / Everyone of Real RKC, you know who you are / Afshin, thanks for those soccer moments / Seunkyu, our man with a cleanroom plan / Humberto, worthy replacement of Maarten at the foosball table / Chirlmin and Marie-Eve, I enjoyed interacting with you with the forum organization / Rianda, just keep those cakes a comin' / Magnus, the most suavely dressed scientist ever / Sam, replicator of tweezers setups – and hopefully soon DNA, toi-toi-toi! / Pauline, I wish you all the happiness in life, especially now you have found your chubby Frenchman! / Mariana, take care of those twins / Wayne, better start graduating / and the list just goes on and on. Thank you Timon, Andreas, Anne, Hyun, Millie, Nika, Zhongbo, Chaya, Florian, Sriram, Peter, Mihela, Artur, Peter, Belen, Aafke, Amelie, Nikola (NvD), Toivo, Eve, Zoreh, Gautam, Tim, Meghan, Xander, Lilliane – from QT: Alessandro, drop by in the new building as well, BN needs you, Bas Hensen, Julia, Gabriele / QN: Sal Bosman, the other servitore, never a dull conversation – Rifka, Allard, Simon van Vliet, Ammeret, Victor, Vanessa, Dominik, Da, Karsten, Marek, Maurits, Jonás, Fabrizio, Núria, Alicia, Huong, Orkide, Stanley, Iljong, Johannes, and many more.

F060, the office of all offices. Andrew, Maarten, Mahipal, what an absolute pleasure it was to get to know you guys this way. Secretly claiming Jelmer's desk after he left was definitely one of the better choices I have made during my PhD. Whether it was for discussing science, personal matters, cracking jokes or playing foosball: you guys were always there. Ganji: best of luck with the final stretch, almost there! I really enjoy our endless single-molecule biophysics conversations – and the fact that you have a keen interest in stocking up for spontaneous coffee table drinks. Andrew, thanks for sharing your humor and sincerity – and for organizing the 2012 halloween party of course. I have a lot of respect for the life choices you are making now, I will definitely come and visit. Maarten, amazing how you always maintain your stoic and calm presence, no matter what. Respect!

My dearest FvB buddies David and Felix! What great times we had in our bachelor flat! Those Delft days were a true highlight during that phase of a phd when having results is not a given. Great to share the good and bad times with such splendid flat mates. And one of them even turned out to be a master chef and vinologist/virologist! David, thanks to you our place became the BN social hub and the obvious place for big dinners and sunday roasts, thank you! Felix, thanks for bringing in those plain yet much appreciated Dutch sandwiches, and for being my laughing companion when we were yet once again unpleasantly awakened by the smell of steamed salmon in the early morning. I do hope you guys appreciated me dragging in a 500 kg piano and the music I played for you. Of course I will never forget the steam cooker and the 'healthy' cooked-to-death broccoli it produced, as well as our cosy get-together on the Boeroestraat, that place had character alright.

Drill time! Drill sergeant Lipfert, privates Hol and Dulin: drop down and give me 50! Go, Go, Go! Thanks for getting me in shape guys, would not have wanted to see myself rolling out of BN as well. Jan, thanks for your teaching us the endless variations in cross fit exercises, as well as sharing the interesting girl names they have. Felix, my running partner in crime, remember: no matter how much more muscular the others are, we will always be able to outrun them. Come to think of it, we actually beat the buff guys at their own pull-up game. David, thanks for introducing the French jiggle and your skills at the pull-up bar, this will always put a smile on my face.

My dearest social bitches Viktorija, Helena, Maarten and Pawel! When the days turn dark and the end is near, it is you guys I can count on for laughter, banter, and beer. And messages, that is. Just checked: we are well on our way to 10000 – that only in a couple months time, awesome! Viktorija, thanks for bringing in that dark shade of humor so typical of eastern Europe, fantastic personal accounts, and that much needed sassiness: strong independent woman who don't need no man! Helena, sup? Thanks for being a Latin diva – and for sharing those mouthwatering enchiladas and that awesome mole of course. Maarten, thanks for your down to earth touch and – last but not least – Pawel a.k.a. Dr. P: what can I say? I will try not to disappoint you, though we all know that is an impossibility. Our group would not have been the same without your astute observations. I guess those are the perks that come with having a PhD degree.

Cleo, I am still glad Jan made us swap costumes on that infamous 2012 BN halloween party. What can I say? I guess I just love wearing mini-skirts and a spider for a hat. Dank je voor de geweldige tijd die we gehad hebben, al je steun en geduld tijdens wat het centrale gedeelte van mijn promotietijd was, ondanks het feit dat de dingen nou eenmaal anders gelopen zijn. Je hebt van mij een vrolijker, positiever mens gemaakt.

Martin & Olja, thanks for the great times in Rotterdam, your friendship and hospitality through the good times and the bad. I will miss having you guys as fellow citizens! Till, my main man from Aachen/Oxford/London, thanks for your friendship ever since the Oxford days. Nienke, bedankt voor de Californische gastvrijheid, Melina, if you ever read this: I am sorry! I did not *actually* kill Felix! Voor de rest wil ik ook mijn Senaat, Club en Oud-Huis bedanken voor de mooie momenten van afgelopen jaren – de spaarzame momenten waarop ik even aan de wetenschappelijke scene kon ontsnappen, jullie weten wat ik bedoel. Op naar nog vele jaren!

Mam, bedankt voor je niet-aflatende steun afgelopen jaren. Pap, Avery, thanks for the support from overseas. Bell, je uit eergevoel geboren onwil om het pad van de minste weerstand te kiezen vind ik inspirerend, ik mag graag denken dat wij die eigenschap delen. Op naar nog vele sportieve avonturen! Britt & Esmée, bedankt voor jullie gastvrijheid en gezelligheid in Breda! Bianca, de stabiele factor van de familie en mijn paranimf, wat een eer! Luuk & Nel, wie had gedacht dat die studieweken op de zolder van de Zandberglaan hierin zouden uitmonden?

With that, a lot of writing has come to an end. BN, thanks again, I have had a blast. Thanks to you my time here as a graduate student has truly been the best of my life so far.

Bojk
Delft, March 2016

CURRICULUM VITÆ

BOJAN ARTHUR BERGHUIS

20-05-1983	Born in Breda, The Netherlands.

EDUCATION

1996–2002	Secundary education Koninklijk Atheneum Brasschaat Brasschaat, Belgium
2002–2008	B.Sc. Molecular Sciences Wageningen University, The Netherlands
2008–2010	M.Sc. Molecular Sciences (*cum laude*) Wageningen University, The Netherlands *Thesis:* Exploring the N-terminal domain of CP29 with ultrafast fluorescence spectroscopy *Promotor:* Prof. dr. Herbert van Amerongen
2009–2010	Prof. dr. Anthony Watts' Lab Department of Biochemistry Oxford University, United Kingdom
2010	Internship Akzo Nobel Chemicals Deventer, The Netherlands
2011	Consultant at YSE Kirkman Company, Baarn, The Netherlands
2011–2016	Ph.D. Department of Bionanoscience Delft University of Technology, The Netherlands *Thesis:* Enhanced interrogation of DNA roadblocks and RNA polymerases *Promotor:* Prof. dr. Nynke H. Dekker

LIST OF PUBLICATIONS

10. **B.A. Berghuis**, M. Depken, N.H. Dekker *et al.* What is all this fuss about Tus? That and more on non-Gaussian distribution of replication fork velocities. *In preparation*

9. D. Dulin, Z. Yu, T.J. Cui, **B.A. Berghuis**, M. Depken, N.H. Dekker. Real-time observation of a replicative helicase loading on single-stranded DNA. *In preparation*

8. **B.A. Berghuis**, M. Koeber, T. van Laar, N.H. Dekker. High-throughput, high-force probing of DNA-protein interactions with magnetic tweezers. *Submitted*

7. D. Dulin, I. Vilfan, **B.A. Berghuis**, M. Poranen, M. Depken, N.H. Dekker. Backtracking behavior in viral RNA-dependent RNA polymerase provides the basis for a second initiation site. *Nucleic Acids Research* (2015)

6. D. Dulin, **B.A. Berghuis**, M. Depken, N.H. Dekker. Untangling reaction pathways through modern approaches to high-throughput single-molecule force-spectroscopy experiments. *Current Opinion in Structural Biology* **34**, 116–122 (2015).

5. **B.A. Berghuis**, D. Dulin, Z-Q. Xu, T. van Laar, B. Cross, R. Janissen, S. Jergic, N.E. Dixon, M. Depken, N.H. Dekker. Strand separation establishes a sustained lock at the Tus–*Ter* replication fork barrier. *Nature Chemical Biology* **11**, 579–585 (2015).

4. D. Dulin, I. Vilfan, **B.A. Berghuis**, S. Hage, D. Bamford, M. Poranen, M. Depken, N.H. Dekker. Elongation-competent pauses govern the fidelity of a viral RNA-dependent RNA polymerase. *Cell Reports* **10**, 1–10 (2015).

3. Z. Yu, D. Dulin, J.P. Cnossen, M. Kober, M.M. van Oene, O. Ordu, **B.A. Berghuis**, T. Hensgens, J. Lipfert, and N.H. Dekker. A force calibration standard for magnetic tweezers. *Review of Scientific Instruments* **85**, 123114 (2014).

2. R. Janissen, **B.A. Berghuis**, D. Dulin, M. Wink, T. van Laar, N.H. Dekker. Invincible DNA tethers: covalent DNA anchoring for enhanced temporal and force stability in magnetic tweezers experiments *Nucleic Acids Research* **42**, e137 (2014).

1. **B.A. Berghuis**, R.B. Spruijt, R. Koehorst, A. van Hoek, S.P. Laptenok, B van Oort, H. van Amerongen. Exploring the structure of the N-terminal domain of CP29 with ultrafast fluorescence spectroscopy. *European Biophysics Journal* **39**, 631–638 (2010).

In recent decades, single-molecule techniques have become a valuable addition to existing bulk assays to study biological processes. The direct access to reaction kinetics has proven to provide crucial insights into the stochastic behavior and transient dynamics of individual molecules. Within this context, a challenge is to efficiently build up statistics in order to correctly place the occurrence of individual events in perspective of the total distribution of occurrences. A single-molecule technique well equipped for this challenge is magnetic tweezers.

This dissertation encompasses the single-molecule magnetic tweezers studies that set out to dissect the lock mechanism of E.coli replication termination protein Tus with its cognate DNA-binding sequence Ter, as well as the transcriptional activity of viral RNA-dependent RNA polymerase P2 of bacteriophage ϕ6 as it progresses along an RNA template. Both studies capitalize on the recent developments in throughput enhancement that was obtained for single-molecule magnetic tweezers measurements.

Casimir PhD series 2016-08
ISBN: 978-90-8593-252-9