Molecular gymnastics

Single-molecule investigations of protein jumping and dna dancing

Ganji, Mahipal

DOI
10.4233/uuid:665728be-2b91-41e7-9a60-2f83f7ee4728

Publication date
2016

Document Version
Final published version

Citation (APA)
Ganji, M. (2016). *Molecular gymnastics: Single-molecule investigations of protein jumping and dna dancing*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:665728be-2b91-41e7-9a60-2f83f7ee4728

MOLECULAR GYMNASTICS:
SINGLE-MOLECULE INVESTIGATIONS OF PROTEIN JUMPING AND DNA DANCING

MAHIPAL GANJI

MOLECULAR GYMNASTICS:
SINGLE-MOLECULE INVESTIGATIONS OF
PROTEIN JUMPING AND DNA DANCING

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 op dinsdag 25 oktober 2016 om 12:00 uur

door

Mahipal Ganji

Master of Science in de Nanobiophysics
Technische Universität Dresden, Germany
geboren te Khajipoor, India.

This dissertation has been approved by the
promotor: Prof. dr. C. Dekker
copromotor: Dr. E.A. Abbondanzieri

Composition of the doctoral committee:

Rector Magnificus	Chairman

Independent members:

Prof. dr. A. Engel	Technische Universiteit Delft
Prof. dr. A. Kapanidis	University of Oxford, Oxford
Prof. dr. D. Rueda	Imperial College London
Prof. dr. E. Peterman	Vrije Universiteit Amsterdam
Dr. G. Bokinsky	Technische Universiteit Delft

Reserve member:

Prof. dr. N. Dekker	Technische Universiteit Delft

Keywords: Single-molecule, fluorescence, reverse transcriptase, strand displacement-synthesis, protein-DNA interactions, protein rebinding, DNA supercoils, plectoneme pinning

Published by: Mahipal Ganji
Cover Image: Dr. Jacob Kerssemakers
Printed by: Gildeprint

Casimir PhD series 2016-29

ISBN 978-90-8593-273-4

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

CONTENTS

1
Introduction

The continuous functioning of an enormous number of molecular machines ensures that information flows at the nanometer scale for various needs of life. Deoxyribonucleic acid (DNA) is central to all life processes because it stores hereditary information and transfers it to the offspring. Human genomic DNA is about 2 meters long but compacted into a micro-meter scale compartment called the nucleus at the center of our cells. Proteins that are work horses of the cellular world play a central role in organisation, replication, repair, and segregation of DNA. In this chapter, several key concepts of life processes at molecular level including central dogma of molecular biology, reverse transcription, and DNA supercoils are briefly discussed. Furthermore, an outline of the theses is presented.

The word “nuclein” was coined by Swiss physician Friedrich Miescher when he discovered an unknown substance (DNA) from leukocytes’ nucleus in the year 1869 and he pioneered the research in extracting DNA in the subsequent years (as reviewed by Ralf Dahm (1)). In the year 1944, McCarty et al confirmed that DNA is the material of inheritance (2). But it was James Watson and Francis Crick who provided the famous double-helical molecular model of DNA in the year 1953 (3), thanks to Rosalind Franklin and Maurice Wilkins for the X-ray crystal structures. The discovery then revolutionised the field of molecular biology and our understanding of life processes at cellular and molecular level.

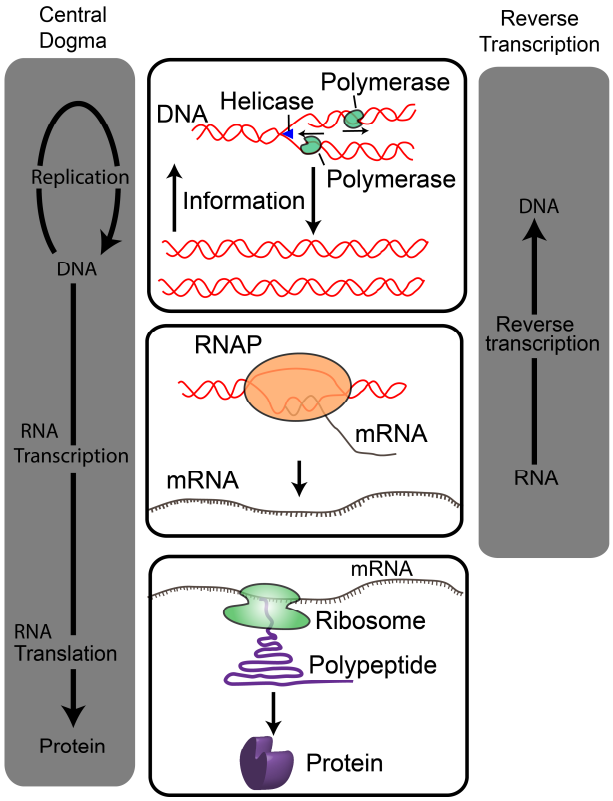


Figure 1.1**: The processes of central dogma of molecular biology and reverse transcription**. Schematics of three essential steps of the central dogma, DNA replication, RNA transcription, and RNA translation, are showed in three boxes. **Top box** -Replisome system consisting of DNA polymerase (green), helicase (blue), and other accessary duplicates DNA during replication. **Middle box** - In RNA transcription, RNA polymerase copies the DNA to mRNA. **Bottom box** -Finally, the copied mRNA will then be translated to polypeptide which will be folded into specific structure to become a functional protein. During reverse transcription RNA genome will be converted to DNA.

1.1 The central dogma of molecular biology

In the year 1958, Francis Crick proposed the central dogma of molecular biology (*4*) in which molecular information flows unidirectional from DNA to RNA and then to protein (Figure 1.1). The information at molecular level is stored in terms of DNA. The building blocks of DNA are thymine, adenine, guanine, and cytosine. Thymine is complementary to adenine and guanine to cytosine in the double-stranded DNA structure. A molecular machine called RNA-polymerase transcribes DNA into RNA. Therefore, RNA is also composed of four building blocks (uracil, adenine, guanine, and cytosine) but slightly different chemical composition than DNA. Unlike the double-stranded structure of DNA, RNA usually exists a single-stranded molecule. But some parts of RNA can self-complement to form the so called secondary structures. Proteins are different polymer molecules compared to DNA and RNA that they are a long chain of amino acids. They are twenty different naturally occurring amino acids each encoded by a three nucleotide RNA during RNA translation. The RNA translation results in a chain of amino acids called polypeptide. But the specific function of a protein is determined by the three-dimensional structure that a polypeptide attains.

1.2 Reverse transcription

Contrary to the central dogma of molecular biology, single-stranded RNA genome of a virus needs to be converted to double-stranded proviral DNA for a successful infection of new cells (Figure 1.1). In fact, there are class of viruses called retrovirus, all of them store their genetic information in terms of RNA. Conversion of RNA genome to double-stranded DNA takes place in the cytoplasm of the host cell. The process of conversion of single-stranded RNA to double-stranded proviral DNA genome is called reverse transcription. An enzyme called reverse transcriptase catalyses the process of reverse transcription. (*5, 6*)

1.2.1 Role of HIV-1 reverse transcriptase in the life-cycle of the virion

The human immunodeficiency virus (HIV) is a lentivirus belongs to the class of retroviruses. The genetic information of retroviruses is stored in single stranded RNA (ssRNA). Once the virus infects the host cells, the ssRNA genome needs to be converted to double-stranded proviral DNA (dsDNA). The conversion of ssRNA to proviral dsDNA is orchestrated mostly by a single enzyme called reverse transcriptase (RT). The process of conversion is called reverse transcription because the genetic information flows opposite to the central dogma of molecular biology.

X-ray crystallographic studies provided an ample amount of structures of RT apo-enzyme and RT co-crystals with nucleic acids and RT inhibitors. Structurally, RT is a hetero-dimer consists of p66 and p51 subunits. While p51 subunit mainly plays a structural role in gripping nucleic acid substrates, p66 subunit carries two catalytic functions- DNA polymerisation and RNase H activity (Figure 1.2A). Reverse transcription starts with in the virus before the infection of a cell where RT uses cellular tRNA as a primer to copy (-)-strand DNA from viral RNA genome (*7*). After synthesising of every few tens of nucleotides, RT switches to RNase H function to make a cut in the copied RNA genome. It is believed that the resultant short fragments will fall off from the newly copied DNA since they are too small to be annealed via base complementarity. But the longer RNA fragments (>7 nucleotides) left intact to DNA will be displaced by RT during the synthesis of complimentary DNA strand (*8-11*). However, it is remaining unclear whether the displacement synthesis is sequence dependent or not. In chapter-3 of this thesis we addressed this problem by testing the ability of RT to displace different GC-rich non-templates. While synthesising (-) strand DNA and RNase H degradation of copied RNA, RT processes two short RNA fragments known as polypurine tracts (PPT) to use as primers for the initiation of (+)-strand DNA synthesis (*12*). Prior to elongation, the PPT is resistant to degradation by RNase H activity of RT. After addition of around twenty nucleotides to the PPT primer, RT makes a cut precisely at the end of the PPT fragment. The fate of left over PPT RNA fragment is not known. In chapter 3 of this thesis we show that RT can displace the PPT fragment efficiently. DNA synthesis then continues until RT reaches the cellular tRNA which was initially used for starting of the reverse transcription. At this moment, RT cleaves off the tRNA leaving a sticky end (*13*). A circular intermediate is generated by annealing this sticky end to the (-)-strand DNA. The completion of reverse transcription requires to displace those RNA fragment resulted from RNase H activity of RT during (-) strand DNA synthesis. Also, RT needs to perform a DNA displacement synthesis to be able to generate full length proviral DNA. In order to perform all these functions RT dynamically switches between DNA-polymerisation and RNase H functions (*14-16*). These characteristics of RT makes this enzyme a good model system to study DNA-protein interactions. Therefore, in chapter-4 we used fluorescently labelled RT for understanding how DNA binding proteins explore different binding configurations on the target site.

1.2.2 Reverse transcriptase of HIV-1 is a molecular gymnast

As described in the previous section, HIV-1 RT is a multifunctional enzyme that performs both DNA polymerisation and RNA degradation during reverse

transcription. The enzyme is also capable of performing strand-displacement synthesis. X-ray crystallographic studies provided static structures of RT binding to DNA/DNA or DNA/RNA substrate but how the enzyme switches between different functions was not clear. Previous studies indicated that the enzyme cannot simultaneously perform both polymerisation and RNA degradation (*17, 18*), meaning that RT must undergo conformational changes in order to perform different functions. In addition, RT does not support primer-extension on RNA primers with an exception for a special RNA sequence known as polypurine tract (PPT). Once the PPT primer is extended by several nucleotides, RT then makes a cut exactly at the end of the PPT. However, it was not clear how RT distinguishes between RNA and DNA primers and how RT uses PPT as a primer initially and cuts later was not clear until before recent single-molecule FRET (smFRET) studies (*14-16*) (Figure 1.2B).

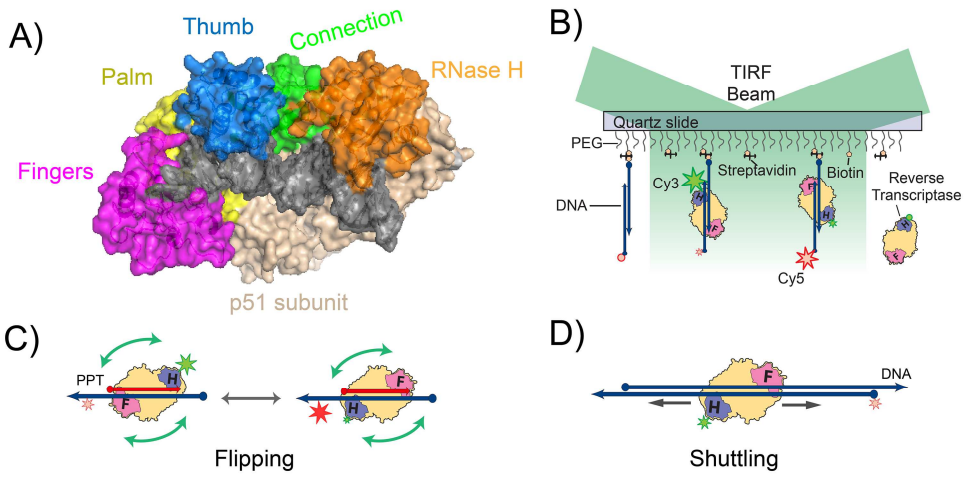


Figure 1.2**: Single-molecule FRET assay revealed that RT flips and shuttles on substrates**. A) Crystal structure of RT in combination with dsDNA substrate. Fingers, palm, thumb, connection, and RNase H subdomains of p66 are represented with pink, yellow, blue, green, orange colors. The p51 subdomain is represented with light brown color. DNA substrate is showed in grey color. B) Single-molecule FRET assay to study dynamic interactions of RT with substrates. C) RT binds on RNA/DNA substrate in two different conformations, and also RT can flip from one orientation to other. D) On long DNA substrates RT can bind on double-stranded region and uses shuttling mechanism to find the target site.

The smFRET studies revealed that RT binds on dsDNA in a polymerase competent mode by gripping the 3'-end of a DNA primer between its fingers and thumb domain as showed in the Figure 1.2. In case of RNA primer, RT binds in flipped orientation compared to the bound state on DNA (Figure 1.2C). On the PPT primer,

RT binds in either polymerisation and RNase H competent mode. Surprisingly, RT was observed to spontaneously flip in its binding orientation on PPT primer in a single binding event (Figure 1.2C). The smFRET studies also revealed that RT is capable of shuttling on long substrates (Figure 1.2D). Flipping in combination with shuttling is advantageous for any enzyme during target searching because it allows the enzyme to explore multiple configurations in order to choose right orientation to perform multiple enzymatic functions in a single encounter (*19, 20*). Flipping transitions offer a unique window that allows to study local interactions of macromolecular complexes. Therefore, we explored the mechanism of flipping at single-molecule level by using RT as a model system.

1.3 DNA supercoiling

Since a good part of the theses reports on supercoiled DNA, in this section we briefly describe about it. DNA is composed of two continuous strands, runs anti-parallel to each other, twist around the helical axis, and kept together by base-pairing and base-stacking forces. The DNA in general can be very long up till few centi-meters but only 2 nm in diameter. This make DNA stiffer in short scales but globally DNA is a semi-flexible polymer. Because of longer length of DNA, during the molecular processes such as replication and transcription, proteins exert torsional stress on DNA (*21-23*). The applied torsional stress results in intertwined DNA structures known as plectonemes in the DNA. Supercoiling plays pivotal role in compaction and organisation of bacterial genome (*24, 25*). For example, a 4.6 Mega-base pairs long circular chromosome of E. coli is packed into a space of ~µm by 1000-fold compaction mainly due to supercoiling (*26*). DNA can exhibit positive supercoiling or negative supercoiling. Over-winding of DNA leads to positive supercoiling and under-winding leads to negative supercoiling.

Supercoiled state of the DNA can be explained using the so called linking number (L_k). The L_k is constant for a topologically closed DNA and is given by sum of writhing and twist numbers (*27, 28*) (Figure 1.3) ($L_k = W_r + T_w$). The number of twists (T_w) in a DNA is equal to the number of helical turns and writhing number (W_r) is the total number of spatial crossings of DNA to itself. For a relaxed DNA, W_r is zero. Since the L_k is conserved for a topologically closed DNA ($dL_k=0$), any change in the T_w, due to action of a protein, change in temperature or salt concentration, will appear as W_r.

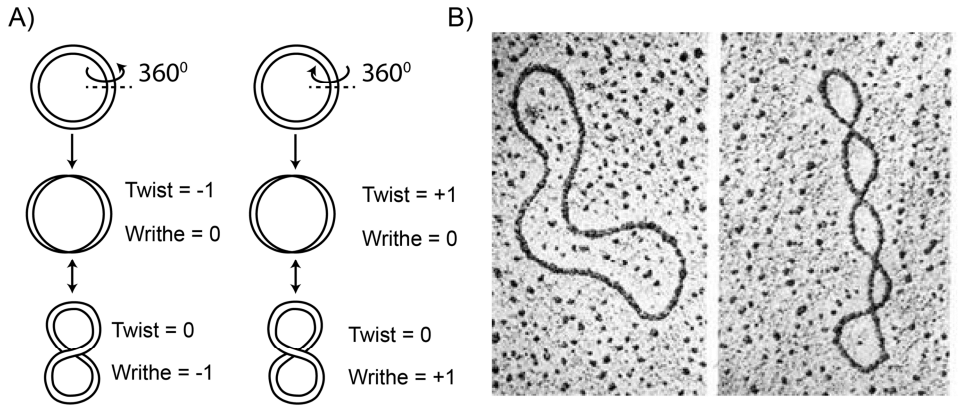


Figure 1.3: **DNA supercoiling**. A) Schematic illustration showing a circular DNA with a twist or writhe. B) Electron microscopy images showing relaxed (left) and supercoiled DNA (right). Image obtained from www.biowiki.ucdavis.edu

DNA plectonemes are dynamic entities: The structure of supercoiled DNA can either be plectonemic (the intertwined DNA) or toroidal. The toroidal structure of DNA is stabilized by binding of nucleosomal proteins. Both atomic-force microscopy and electron microscopy studies revealed that the DNA supercoils assume only plectonemic structures but not toroidal (*27, 29-32*). Single-molecule techniques such as magnetic and optical tweezers have also been used to study the characteristics of supercoiled DNA (*33, 34*). Although, the single-molecule techniques allow to control the density of supercoiling on torsionally constrained DNA, they only provide limited information regarding the characteristics of supercoiled DNA. For instance, the supercoiled state of DNA can only be inferred by the change in length of DNA molecules but not by direct visualization. Recently, plectonemes along supercoiled DNA were visualized by using side-pulling magnetic tweezers in combination with wide-field fluorescence microscopy (*35*). The plectonemes on supercoiled DNA displayed interesting characteristics. Multiple plectonemes could exist on a single supercoiled DNA molecule. Surprisingly, the plectonemes also showed dynamic features such as diffusion and hopping (termination at one position by simultaneous nucleation at a distant position). We wanted to study the plectoneme localization along supercoiled DNA. However, the single-molecule magnetic tweezers technique combined with wide-field magnetic tweezers was turned out be a technically demanding and low-throughput assay. In this theses we developed a novel single-molecule assay to visualize plectonemes on doubly tethered DNA molecules that does not require mechanical manipulation of DNA but only requires a wide-field fluorescence microscopy (chapter 5). Furthermore, by

using this new assay we studied the sequence-dependence localization of plectonemes along supercoiled DNA (chapter 6).

1.4 Outline of this thesis

This thesis presents results from a range of single-molecule studies and bulk biochemical assays. We used well-developed single-molecule FRET assay to study RT-DNA interactions and we also developed a new assay to create plectonemes on doubly tethered DNA for visualizing them in real-time. The doubly tethered DNA assay was then extended to study interactions of Dps, a DNA binding protein from starved cells, with supercoiled DNA and other DNA topologies.

Since we used different techniques in this theses we first described them. **Chapter-2** provides information on the techniques used in this theses. It describes a modified protocol for purification and labelling of RT, a protocol for passivation of quartz surfaces for single-molecule fluorescence studies. The chapter then focuses on the design of fluorescence microscope used in this theses. Two different single-molecule fluorescence assays based on this microscope are described: a single-molecule FRET assay and an intercalation-induced supercoiled DNA (ISD) assay. Other single-molecule assays to study DNA supercoil dynamics are briefly described.

In **chapter-3** of this theses, we investigated the sequence-dependent displacement synthesis by HIV-1 RT using bulk biochemical assay. We tested a range of GC-rich non-templates for displacement synthesis. The displacement synthesis was highly dependent on the GC-content of the DNA. The results also show that RT performs displacement synthesis by actively interacting with the flap of non-template nucleotides.

Single-molecule FRET (smFRET) allows to study distance changes at nanoscale (2-10 nm). In **chapter-4**, we used smFRET to study the dynamics of protein-DNA interactions on the target site. We used fluorescently labelled RT and DNA molecules to study protein flipping at single molecule level. Using this assay we studied the effects of salt and crowding concentrations on the binding of RT to DNA. We then modelled the effects of varied salt and crowding concentrations on the kinetics of flipping. Our results indicate that the non-covalently bound macromolecular complexes undergo of rapid re-bindings.

We then moved to doubly tethered DNA assay where we discovered a simple way of making plectonemes on torsionally immobilized DNA. In **chapter-5** we report a novel single-molecule fluorescence assay, intercalation-induced supercoiled DNA

(ISD) to induce and visualize supercoiled DNA dynamics at single-molecule level. This assay solely depends on the mechano-chemical modification of immobilized DNA molecules. In this assay, we used an intercalating dye to induce and image DNA supercoils. We also show that the mismatched DNA sequences are favourable positions for pinning of plectonemes.

In **chapter-6**, we used the newly developed ISD assay to understand the sequence dependent plectoneme pinning along supercoiled DNA. We first studied DNA constructs with varied GC-content along the length and show that plectoneme pinning is highly dependent on the sequence of DNA. Subsequently, we introduced different length AT-rich segments in a homogenous GC-rich DNA to study if AT-rich sequences pin the plectonemes. We then tested the effects of poly(A)-tracts on the plectoneme localization along the length DNA. Our results clearly indicate that the AT-rich sequences are highly favourable positions for pinning of plectonemes. Our preliminary results show that the multiple poly(A)-tracts along DNA help in localization of DNA.

The ISD assay can also be used to study interactions of proteins and supercoiled DNA. In this line, in **chapter-7**, we visualized the binding of Dps with different topological structures of DNA including supercoils. This assay allowed us to capture new information such as Dps only binds to DNA if two DNA strands are close to each other such as plectonemes or kinks in the DNA. We also observed that Dps is capable of cross-linking two DNA molecules. In addition, Dps-DNA complexes are rich in Dps that they could accommodate excess DNA into them.

1.5 References

1. R. Dahm, Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* **122**, 565 (2008).

2. T. O. Avery, M. C. MacLeod, M. McCarty, Studies on the chemical nature of the substance inducing transformation of Pneumococcal types *J. Exp. Med.* **79**, 22 (1944).

3. W. J.D., C. F.H.C, A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 2 (1953).

4. F. H. C. Crick, The biological replication of macromolecules. *in Symp. Soc. Exp. Biol.* **XII**, (1958).

5. D. Baltimore, RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1209 (1970).

6. H. M. Temin, S. Mizutani, RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**, 1211 (1970).

7. R. Marquet, C. Isel, C. Ehresmann, B. Ehresmann, tRNAs as primer of reverse transcriptases. *Biochimie* **77**, 113 (1995).

8. M. Wisniewski, M. Balakrishnan, C. Palaniappan, P. J. Fay, R. A. Bambara, Unique progressive cleavage mechanism of HIV reverse transcriptase RNase H. *Proceedings of the National Academy of Sciences* **97**, 11978 (2000).

9. J. J. DeStefano, L. M. Mallaber, P. J. Fay, R. A. Bambara, Quantitative analysis of RNA cleavage during RNAdirected DNA synthesis by human immunodeficiency and avian myeloblastosis virus reverse transcriptases. *Nucleic Acids Research* **22**, 3793 (1994).

10. G. M. Fuentes, P. J. Fay, R. A. Bambara, Relationship between Plus Strand DNA Synthesis and Removal of Downstream Segments of RNA by Human Immunodeficiency Virus, Murine Leukemia Virus and Avian Myeloblastoma Virus Reverse Transcriptases. *Nucleic Acids Research* **24**, 1719 (1996).

11. J. J. Champoux, 6 Roles of Ribonuclease H in Reverse Transcription. *Cold Spring Harbor Monograph Archive* **23**, 103 (1993).

12. H. E. Huber, C. C. Richardson, Processing of the primer for plus strand DNA synthesis by human immunodeficiency virus 1 reverse transcriptase. *Journal of Biological Chemistry* **265**, 10565 (1990).

13. J. S. Smith, M. J. Roth, Specificity of human immunodeficiency virus-1 reverse transcriptase-associated ribonuclease H in removal of the minus-strand primer, tRNA(Lys3). *Journal of Biological Chemistry* **267**, 15071 (1992).

14. E. A. Abbondanzieri *et al.*, Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature* **453**, 184 (2008).

15. S. Liu, E. A. Abbondanzieri, J. W. Rausch, S. F. J. L. Grice, X. Zhuang, Slide into Action: Dynamic Shuttling of HIV Reverse Transcriptase on Nucleic Acid Substrates. *Science* **322**, 1092 (2008).

16. S. Liu, B. T. Harada, J. T. Miller, S. F. J. Le Grice, X. Zhuang, Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription. *Nat. Struct. Mol. Biol.* **17**, 1453 (2010).

17. M. Lapkouski, L. Tian, J. T. Miller, S. F. J. Le Grice, W. Yang, Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. *Nat. Struct. Mol. Biol.* **20**, 230 (2013).

18. M. Nowotny *et al.*, Structure of Human RNase H1 Complexed with an RNA/DNA Hybrid: Insight into HIV Reverse Transcription. *Mol. Cell* **28**, 264 (2007).

19. G. Sasnauskas, G. Kostiuk, G. Tamulaitis, V. Siksnys, Target site cleavage by the monomeric restriction enzyme BcnI requires translocation to a random DNA sequence and a switch in enzyme orientation. *Nucleic Acids Res.* **39**, 8844 (2011).

20. G. Sasnauskas *et al.*, A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3′–5′ and 5′–3′ strands by rotating a single active site. *Nucleic Acids Res.* **38**, 2399 (2010).

21. B. J. Peter, C. Ullsperger, H. Hiasa, K. J. Marians, N. R. Cozzarelli, The Structure of Supercoiled Intermediates in DNA Replication. *Cell* **94**, 819 (1998).

22. L. F. Liu, J. C. Wang, Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences* **84**, 7024 (1987).

23. B. ten Heggeler-Bordier, W. Wahli, M. Adrian, A. Stasiak, J. Dubochet, The apical localization of transcribing RNA polymerases on supercoiled DNA prevents their rotation around the template. *The EMBO Journal* **11**, 667 (1992).

24. R. Kavenoff, O. Ryder, Electron microscopy of membrane-associated folded chromosomes of Escherichia coli. *Chromosoma* **55**, 12 (1976).

25. L. Postow, C. D. Hardy, J. Arsuaga, N. R. Cozzarelli, Topological domain structure of the Escherichia coli chromosome. *Genes & Development* **18**, 1766 (2004).

26. T. A. Azam, A. Iwata, A. Nishimura, S. Ueda, A. Ishihama, Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J. Bacteriol.* **181**, 6361 (1999).

27. T. C. Boles, J. H. White, N. R. Cozzarelli, Structure of plectonemically supercoiled DNA. *Journal of Molecular Biology* **213**, 931 (1990).

28. J. H. White, Self-Linking and the Gauss Integral in Higher Dimensions. *American Journal of Mathematics* **91**, 693 (1969).

29. S. D. Levene, C. Donahue, T. C. Boles, N. R. Cozzarelli, Analysis of the structure of dimeric DNA catenanes by electron-microscopy. *Biophysical Journal* **69**, 1036 (1995).

30. J. M. Fogg *et al.*, Exploring writhe in supercoiled minicircle DNA. *Journal of Physics-Condensed Matter* **18**, S145 (2006).

31. R. N. Irobalieva *et al.*, Structural diversity of supercoiled DNA. *Nat Commun* **6**, (2015).

32. J. Bednar *et al.*, The twist, writhe and overall shape of supercoiled DNA change during counterion-induced transition from a loosely to a tightly interwound superhelix- possible implications for DNA-structure *in vivo*. *Journal of Molecular Biology* **235**, 825 (1994).

33. T. R. Strick, J. F. Allemand, D. Bensimon, V. Croquette, Behavior of Supercoiled DNA. *Biophysical Journal* **74**, 2016 (1998).

34. A. La Porta, M. D. Wang, Optical torque wrench: angular trapping, rotation, and torque detection of quartz microparticles. *Phys Rev Lett* **92**, 190801 (2004).

35. M. T. J. van Loenhout, M. V. de Grunt, C. Dekker, Dynamics of DNA Supercoils. *Science* **338**, 94 (2012).

2
Methods

In order to answer the key scientific questions outlined in the introduction, I needed to overcome some important technical hurdles. First and foremost, probing of the flipping dynamics required a donor labeled RT with activity comparable to its wildtype counterpart and acceptor fluorophore labeled oligonucleotide. I modified existing protocols of RT purification in order to maximize the protein purity and labelling efficiency. Secondly, single molecule FRET requires a transparent surfaces capable of rejecting non-specific binding of proteins. I adopted a protocol for passivation of quartz surfaces by PEG coating. Finally, detecting single molecules require high signal-to-noise-ratio (SNR), hence I tested different electromagnetic gains in order to set maximum SNR. At the we describe the single-molecule fluorescence assays for studying protein-DNA interactions and the dynamics of supercoiled DNA.

2.1 Preparation of biological samples

In the first part of this section I will explain methods for obtaining Cy3-labelled RT, Cy5-labelled oligonucleotides, and PEG coated quartz slides for performing single molecule studies. In the second part, the design of two-color fluorescence microscope, optimization of electro-magnetic gain settings of the emCCD, and mapping between two channels of the emCCD will be discussed.

2.1.1 Purification and labelling of RT

We modified existing protocols to maximize the yield of active enzyme and improve the labelling efficiency. By ensuring a majority of RT is both labelled and active, the protocols I developed are crucial in ensuring the single molecule results can be compared effectively to measurements made in bulk.

Growth of cells: The p66 and p51 subunits of reverse transcriptase were expressed and purified separately in *Escherichia coli* strain M15 containing pDMI.1 (expresses *lac* repressor) and pDS56 (expresses p66 of HIV-1 RT) plasmids. The strain was obtained from Le Grice lab (National Cancer Institute, Frederick, USA). This dual plasmid system is resistant to both kanamycin and ampicillin. Consequently, a 25 µg/l kanamycin (Sigma Aldrich) and 100 µg/ml ampicillin (Sigma Aldrich) was added to the cell cultures. The p66 subunit was expressed in *E. coli* strain P6HRT, c-, C561, E478Q and the p51 subunit was expressed in P6HRT51 strain. Cells were first grown on agar plates in both the cases. A colony from agar plate was inoculated in 100 mL of LB medium. The next day, six litres (6x1 litre) of liquid culture was inoculated (OD_{600}=0.1). Cell cultures were grown to mid-log phase (OD_{600} ~0.7) at 37 °C before inducing the expression of protein by adding a 1mM IPTG. After 4 hours of induction, the cells were harvested by centrifuging at 34000x g for 30 minutes. After this step the pellet was either stored at -80 °C or carried to lysis for purification.

Cell lysis: All the purification steps were performed at 4 °C. The pellet was dissolved in 50 mM $NaPO_4$ pH 7.8, 500 mM NaCl with 1mM PMSF and 0.5 mg/ml Lysozyme. The mixture was stirred for 30 minutes in order to remove any clumps. Cells were lysed by French pressing two times at 1 kbar. The lysis was then centrifuged at 34000x g for 45 minutes. The supernatant of the lysis was carried out for further purification. Following procedure explains the purification of p66 subunit but the same applies for p51 subunit as well.

Ni-NTA column purification: 1 M imidazole, pH 7.8 was added to the supernatant to a final concentration of 10 mM. Ni-NTA beads (GE Healthcare) of 3 ml were washed thoroughly with Milli-Q purified water and subsequently with buffer (50 mM $NaPO_4$ pH 7.8, 500 mM NaCl, 10% glycerol, 10 mM imidazole) in a 50 ml Falcon tube. The beads mixture was briefly centrifuged to collect the beads. The supernatant from lysed cells was added to the beads and incubated for around 2 hours. After this step, the beads with proteins were again thoroughly washed by means of centrifugation (2000 rpm for 2 minutes) with 100 ml of 50 mM $NaPO_4$, pH 7.8, 500 mM NaCl, 10% glycerol, 20 mM imidazole, subsequently with a 100 ml of 50 mM $NaPO_4$, pH 6.0, 500 mM NaCl, 10% glycerol, 20 mM imidazole to get rid of non-specifically bound proteins. At this point, p66 is bound to the Ni-NTA beads. Addition of higher concentration of imidazole buffer to the beads will lead to protein elution since imidazole competes with the Ni-His tag interaction (*1*). The beads were then re-suspended in a 7.5 ml of 50 mM $NaPO_4$, pH 6.0, 500 mM NaCl, 10% glycerol, 500 mM imidazole and briefly mixed for 5 minutes before pouring into a 20 ml empty column. After the beads settled on the column, protein was collected by opening the cap. Once the liquid level reached to beads height in the column, 7.5 ml additional buffer was added to the column to collect the remaining protein. The total volume of the sample was 15 ml with 500 mM NaCl final concentration. All the following purification steps were performed on Akta-HPLC system (GE Healthcare).

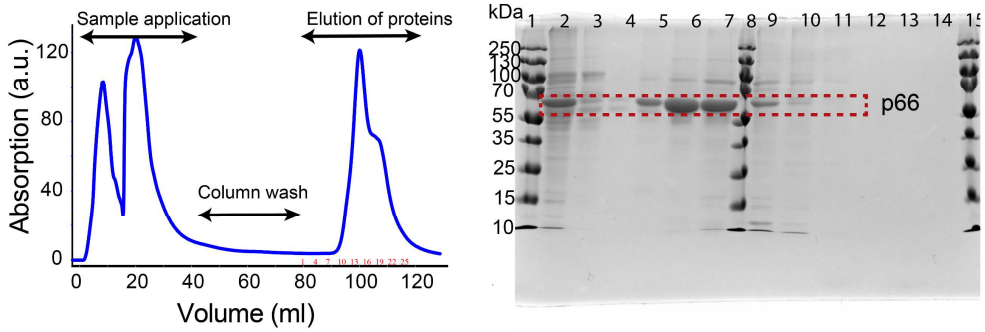


Figure 2.1**: Left- Chromatogram of p66 purification through heparin column.** Curve was obtained from Akta and it represents the absorption of proteins at 280 nm. Different parts of the purification procedures indicated in the figure. The numbers in red color are the numbers of the elution fractions. **Right- Commassie blue staining analysis of different elution fractions**. Lanes 1, 8 and 15 are ladder. Lane 2 was the resultant of Ni-NTA column, lane 3 was the heparin column wash, and lanes 4-7 are fractions 10, 12, 14 and 16, respectively. Lanes 9-11 are fractions 18, 20 and 22 respectively. Lanes 13 and 14 are empty

Heparin-Sepharose column purification: Heparin is a highly negatively charged polypeptide with a similar size to DNA (*2*). Therefore, many DNA binding proteins can bind to heparin as well which makes heparin a useful material for purification. Since p66 itself contains a DNA binding groove, it can bind to heparin and subsequently be eluted by applying a high concentration of salt buffer. For this reason, the protein sample from NiNTA column purification should be diluted to contain lower salt concentration (<75 mM). The protein sample was drop-wise diluted to 100 ml with 50 mM $NaPO_4$, pH 7.0, 10% glycerol. A 5 mL heparin-Sepharose column (GE Healthcare) was equilibrated with 50 ml of water and 50 ml of 50 mM $NaPO_4$, pH 7.0, 50 mM NaCl, 10% glycerol. The sample was loaded at 0.5 ml/min flow rate. The column was then washed with 25 ml of 50 mM $NaPO_4$, pH 7.0, 50 mM NaCl, 1 mM DTT, 10% glycerol. Note that DTT was included in the washing buffer in order to reduce the disulfide bonds that can be formed between p66 molecules. Subsequently, the column was washed with 25 ml of the same buffer but without DTT. Proteins were eluted with a 25 ml gradient of 50 mM $NaPO_4$, pH 7.0, 50 mM NaCl, 10% glycerol to 50 mM $NaPO_4$, pH 7.0, 1 M NaCl, 10% glycerol. 1.5 mL fractions were collected (Figure 2.1-Left). Alternate fractions from the elution peak of the chromatogram were run on a 10% SDS page-gel and analyzed by Coomassie blue staining (Figure 2.1- Right). Fractions from 12-16, where p66 was largely present, were collected and concentrated to 3 ml by the use of centrifugal filters (Millipore 10 kDa MWCO). Since the collected proteins were now in the reduced form and at high concentration, they could be readily labelled with Cy3-mono maleimide dye(*3*). A 1:10 molar ratio of Cy3-monomaleimide (GE Healthcare) was added to the protein sample. The sample was then degassed to remove O_2 and N_2 was flushed in to create an inert condition. The sample was left at room temperature for 2 hours before quenching the labelling reaction by adding a 10 times molar excess of 2-mercaptoethanol. The Cy3-labelled p66 was diluted to 15ml and was then applied on to heparin column again to wash off the excess dye remained after the labelling reaction. The column was washed rigorously with around 50 ml of buffer to get rid of all the free dye. After this step the entire protein was eluted by application 25 ml gradient of 50 mM $NaPO_4$, pH 7.0, 50 mM NaCl, 10% glycerol to 50 mM $NaPO_4$, pH 7.0, 1 M NaCl, 10% glycerol. The entire protein was pooled without any further analysis and prepared for MonoS column purification.

MonoS column purification: Since the MonoS column (GE Health care) is a anion exchange column, it is essential to keep the protein in low salt (<50 mM) buffer otherwise the protein may not bind to the column at all. The resultant protein from the heparin column was concentrated in separate 5 mL fractions with Amicon®

Ultra-15 10K centrifugal filter devices. Each of the 5 ml sample was concentrated to a final volume of 200 µl. Concentrated solute was recovered and diluted (1:25) with 50 mM $NaPO_4$, pH 7.0, 10% glycerol. During the dilution some fraction of protein precipitated. In order to remove precipitate, the whole sample was spun down at 4000 rpm for 30 minutes at 4 °C. The pellet was discarded and the supernatant was used for further purification on the MonoS column. A 1

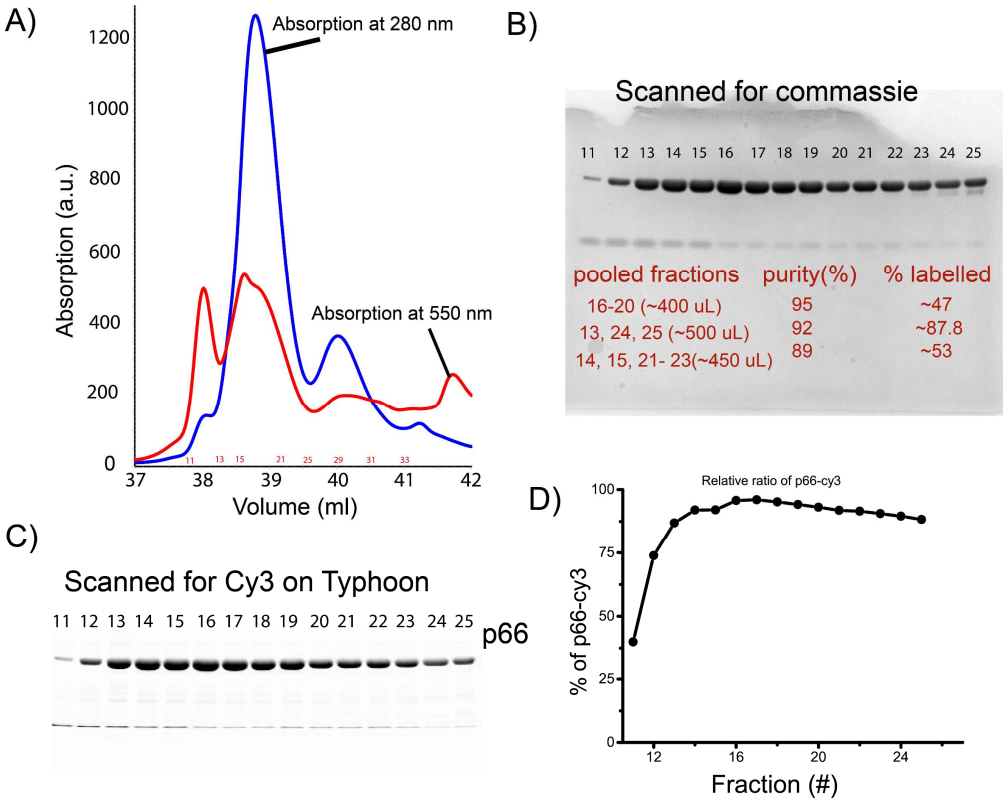


Figure 2.2**: MonoS column purification of Cy3-labelled p66**. Top left: Chromatogram showing the elution of Cy3-labelled p66 from MonoS column. Absorptions at 280 nm (blue) and 550 nm (red), corresponding to protein and Cy3, respectively, were simultaneously tracked during the elution. The numbers in red color on X-axis are the elute fraction number. Each of the fractions were further analyzed by fractionating on a 10% SDS page gel. The gel was first imaged on Typhoon scanner for Cy3 (bottom left). Two bands can be seen in Cy3 scanned gel in which the predominant one belongs to Cy3-labelled p66 and the weaker one is a contaminant protein. The gel was further stained with Commassie blue and was imaged with gel imager (top left). Based on the Cy3 fluorescence of p66 and other proteins observed in Typhoon scanning, an estimation was made for purity of sample and it was plotted as percent of p66 present relative to the contaminant (bottom right). Based on both the Cy3 scanning on Typhoon and Commassie blue staining analysis, the fractions were categorized into three types. The first category (fractions 13-25) contained highest purity (> 95%) with ~

47% labelling, the second (fractions 13,24, and 25) was with 92% purity with ~88 % labelling and the third one (fractions 14, 15, 21-23) was with less purity also with less labelling efficiency. The labelling efficiency was determined by NanoDrop measurements.

ml MonoS column was equilibrated with 10 ml of water and then with 10 ml of 50 mM $NaPO_4$, pH 7.0, 25 mM NaCl, 10% glycerol. Sample is loaded at 0.5 mL/min. The column was then washed with 10 ml of 50 mM $NaPO_4$, pH 7.0, 25 mM NaCl, 10% glycerol at 1 ml/min. Proteins were eluted with a 20 ml gradient of 50 mM $NaPO_4$, pH 7.0, 25 mM NaCl, 10% glycerol to 50 mM $NaPO_4$, pH 7.0, 1 M NaCl, 10% glycerol at 1 ml/min. 250 µl fractions were collected (Figure 2.2). All the fractions were analysed further on SDS-PAGE by commassie staining and for Cy3 label (Figure 2.2). Both commassie and Cy3 scanning showed in Figure 2.2. Fractions 16-20 contained a 47% labelled p66 with less than 5% other detectable proteins. Hence, these fractions were pooled and 20 µl aliquots were shock frozen with liquid N_2 before storing at -80 °C.

The same protocol was followed for the purification of p51 subunit. Since p51 was not labelled, the second round of heparin purification was skipped. Therefore, it was purified in three consecutive purifications: i) Ni-NTA, ii) heparin, and iii) MonoS column.

Preparation of reverse transcriptase: Since the functional reverse transcriptase is a hetero-dimer consisting of p66 and p51 subunits, these separate subunits needed to be combined to re-constitute the enzyme. For this, p66 of 1 µM and 10 molar excess of p51 was prepared in 50 mM Tris-HCl (pH 8.0), 50 mM NaCl and 6 mM $MgCl_2$. The mixture was incubated at 37 °C for 2 hours followed by 4 hours at room temperature. Since the complexes of p66/p66 and p51/p51 homo-dimers are weaker compared to p66/p51 hetero-dimers (*4*) and the concentration of individual subunits is much higher than the dissociation constant ($K_d <$ 10 nM), we expect that the majority of p66 will be in hetero-dimer. Because the concentration of p51 we used was very low compared to the K_d (>100 µM) of p51/p51 homo-dimers and p51 alone shows no binding affinity to DNA, excess p51 does not interfere with heterodimer RT binding to DNA.

Reverse transcriptase activity assay: A primer extension assay was performed in order to confirm the activity of labelled protein. 100 µl of 50 nM of labelled protein was incubated with a 50 nM of 19-bp primer (5'-/Cy5/AC TTA CGT TCT CGA TCA CTA GT-3')/ 36-bp template (5'-ATT AGA TTA GCC CTT CCA GTA CTA GTG ATC GAG ACA GTA AGT GGC GTG GC-3') DNA in a buffer constituting of 50 mM Tris-HCl (pH 8.0), 50 mM NaCl, 6 mM $MgCl_2$, 1 mM DTT and 0.1 mM EDTA at 37 °C for 10 minutes. The primer was

labelled with Cy5 at its 5'-end, which enabled us to track the extension by reverse transcriptase. Initiation of primer extension was done by addition of 250 µM dNTP mixture. The primer extension was allowed for following time points: 0.25, 0.5, 0.75, 1, 2, 3, 5, and 10 minutes. The extension was terminated by mixing a 10 µl of reaction resultant to 10 µl of stop buffer (96% formamide and 20 mM EDTA). The products were loaded on a 10% urea polyacrylamide gel which was pre-run for 1.5 hours at 350 V. The products were separated by electrophoresis at 333V for 1.5 hours. The gel was then scanned on Typhoon scanner (GE healthcare) for visualizing the primer extension (Figure 2.3). The results indicate that Cy3-labelled RT is as active as its unlabeled counterpart, consistent with previous findings (*5*, *6*).

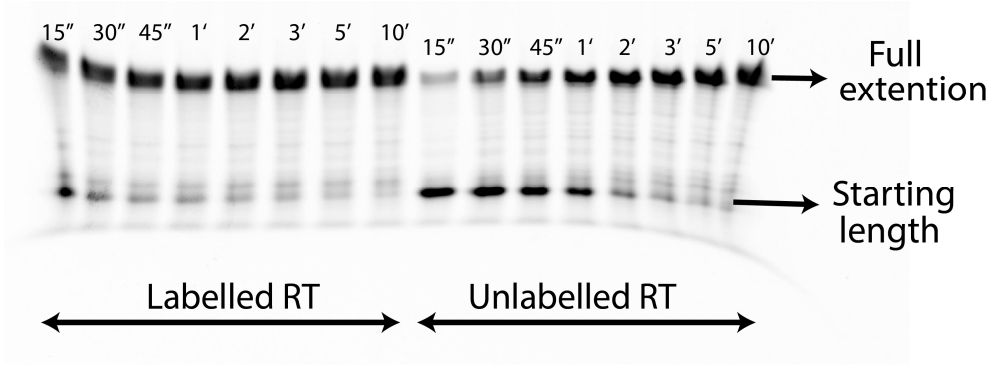


Figure 2.3**: Measuring the activity of RT-Cy3** Activity of the labelled protein in comparison with unlabelled protein (provided by *Le Grice lab*, USA) was determined by primer extension assay. Lower activity of unlabelled protein compared may be because overestimation of its concentration due to impurities

2.1.2 Labelling oligonucleotides for single molecule FRET and primer-extension assay

DNA and RNA substrates were purchased from IDT technologies (IDT; Coralville, IA). A thymine nucleotide was amino modified with C6 linker in order to label oligonucleotides with a fluorophore. Oligonucleotides were dissolved in Milli-Q at 1 mM concentration. Mono-reactive Cy5 NHS ester (GE healthcare) was dissolved in DMSO (Sigma) at 20 mM final concentration. In order to maximize labeling, the reaction was performed in a freshly made solution of 100 mM sodium tetraborate-HCl, (pH 8). The labeling buffer consists of 5 µl of 1mM oligonucleotide, 5 µl of 20 mM Cy5-NHS ester, and 25 µl of 100 mM sodium tetraborate-HCl buffer. The mixture was incubated at room temperature for around 4 hours. Excess non-reactive dye was removed by ethanol precipitation. We usually achieve around 50% of labeling efficiency after first round labelling. By repeating the labeling reaction for

one more time results in around 100% efficiency (Figure 2.4). For single molecule FRET, a 1:10 ratio mixture of the complimentary strands with biotin modifications at their 5'-end were annealed to Cy5-labelled strand by heating at 80 °C for 5 min and allowed to cool to room temperature. The strategy of having excess labeled primers would result in annealing of all the biotinylated primers. Excess labeled primers would not interfere with single molecule experiments since they will be washed away during buffer exchange. A different strategy was adopted for annealing substrates for the strand-displacement synthesis assays since this assay only detects labeled primers. Substrates for strand displacement synthesis were prepared by mixing 1:2:4 ratio of primer-Cy5:template:non-template. The excess template and non-template primers ensure annealing of all the labeled primers. The annealing buffer was 100 mM Tris-HCl (pH 8.0), 1mM EDTA and 100 mM NaCl. The products were stored at -20 °C.

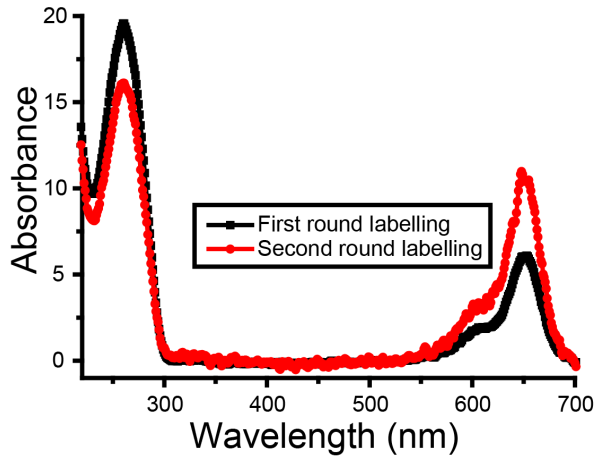


Figure 2.4**: Repetitive labelling of oligonucleotides to get high labelling efficiency**. The absorption spectra of single round labelled oligo-Cy5 (black line) and repeating of labelling for second time (red line). The spectra were recorded on Nanodrop by 10 times diluting the oligo-Cy5 mixture (corrected for dilution).

2.2 PEGylation of quartz slides for fluorescence measurements

Slides cleaning by piranha etching: Macroscopic contaminants on microscope quartz slides (G. Finkenbeiner, Inc.) were removed by scrubbing with detergent. After washing with water, slides were placed in a home-made Teflon holder. Slides were then sonicated in acetone for 20 minutes and subsequently in 1M KOH for 20 minutes with a Milli-Q purified water washing in between. In parallel, cover slips were also etched in 1M KOH for around 2 hours. Since we image quartz slides in

our single molecule experiments, slides need to be as clean as possible. To obtain high surface quality, the slides were etched in piranha solution (3:1 ratio of H_2SO_4:H_2O_2) for around 45 minutes. Since piranha solution is a strong oxidizing agent, it will not only remove the organic matter but it will also result in a highly hydrophilic surface.

Amino-silanization: Immediately after piranha etching, the slides were transferred to Milli-Q water and subsequently sonicated in methanol for around 15 minutes to remove traces of water molecules. Alongside, amino-silanization mixture (100 ml of methanol, 5 ml of acetic acid and 10 ml of aminopropyltriethoxysilane) was prepared. Slides and cover slips were then incubated in the amino-silane mixture for 20 minutes and then washed with methanol thoroughly. The reaction will result in surfaces with homogenous amino (-NH_2) functional groups which are readily available for attaching any molecules with –NHS ester functional groups.

Passivation of quartz surfaces with PEG: The amino-silanized quartz slides and cover slips were taken out from methanol and rinsed thoroughly with Milli-Q. Methoxy poly(ethylene glycol) succinimidyl valerate (mPEG-SVA) (Laysan Bio, Inc.) was prepared in a freshly prepared 0.1 M sodium bicarbonate (pH 8.5) buffer. Quartz slides were placed on a flat surface on which a 70 µl of 25 mM mPEG-SVA was applied. Cover slips were placed on each slide to sandwich the mPEG-SVA solution. The slide-cover slip sandwiches were then incubated overnight under dark and humid environment. The next morning, slides and cover slips were washed with Milli-Q after disassembling the sandwiches. The PEG coated slides and cover slips were then stored under inert conditions at -80 °C until used.

2.3 Multi-color fluorescence microscope set-up for single molecule studies

Design of the microscope: Studying the interaction of biomolecules with single molecule fluorescence microscopy requires appropriate wavelength laser lines to excite the fluorophores attached on the biological samples. The fluorophores we chose to visualize are YOYO-1 (Molecular Probes), Cy3 (GE healthcare), Sytox Orange (Molecular Probes) and Cy5 (GE healthcare). The emission and excitation spectra of these fluorophores is shown in Figure 2.5. Cy3 and Cy5 have been used to study bio-molecular interactions using single molecule FRET (*5*, *6*). YOYO-1 is a strong DNA intercalating dye with high quantum yield. However, it gives rise to high number of double stranded DNA breaks due to photo-damage. Sytox orange is a very good intercalating stain for visualizing double stranded DNA due to its high

quantum yield (90%) coupled with fast binding and dissociation rates (*7*). The fast kinetic rates not only ensure that reaching binding equilibrium fast but also minimizes the problem of photo-bleaching, thereby avoiding photo-damage to DNA.

A schematic diagram of entire microscope set-up is shown in Figure 2.6. We used 488 nm laser line (Cobolt MLD) to excite YOYO-1, 532 nm laser line (Cobolt Samba) to excite either Cy3 or Sytox orange, and 640 nm laser line (Cobolt MLD) to excite Cy5. Dichroic mirrors were used to combine the lasers. The lasers are always operated at highest laser powers (60 mW for 488nm and 100 mW for the other two). An appropriate laser intensity is chosen by tuning the Acoustic Optic Tunable Filter (AOTF-AA optoelectronics). The AOTF also allows us to alternate between two lasers at high temporal resolution. Alternating between two lasers helps to confirm the presence of the fluorophores while performing FRET measurements (*8*).

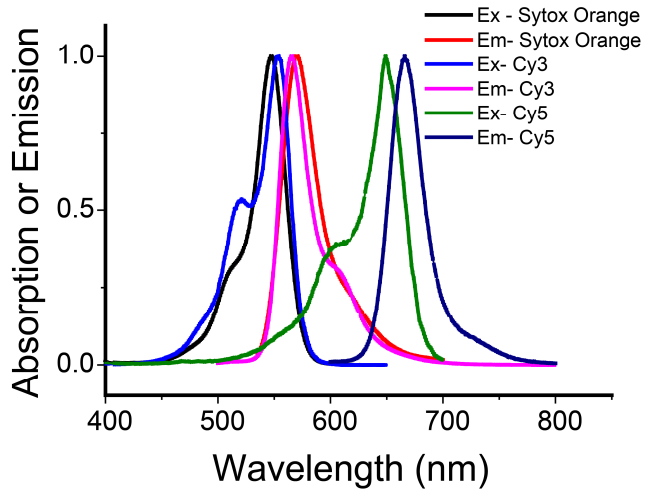


Figure 2.5**:** Normalized excitation and emission spectra of Sytox orange, Cy3 and Cy5

The laser beam from the AOTF is then steered onto a quartz prism at normal angle to get total internal reflection thereby creating an evanescent field (Figure 2.6). The prism and quartz slide were optically coupled by sandwiching immersion oil (Olympus) with appropriate refractive index between them. Coupling of the slide and prism via immersion oil results in creating the evanescent field on the surface of quartz slide. Alternatively, the lasers can be applied in epi-fluorescence mode by focusing at the back-focal plane of the objective through a dichroic cube.

The molecules we want to visualize are immobilized on a PEG passivated quartz surface via biotin-streptavidin-biotin linkage. We immobilize Cy5-labelled DNA on the quartz surface for single molecule FRET and visualize them in prism type TIRF mode. Doubly tethered 21 kb DNA molecules are visualized using Sytox Orange stain in epi-fluorescence mode. A water immersion objective lens (60x Olympus) with numerical aperture of 1.2 is used to collect the emitted fluorescence. The collected fluorescence is then roughly collimated using a pair of lenses (f_1=50 mm and f_2=100 mm) separated by sum of their focal lengths. This configuration also provides an ideal location to narrow the width of field of view by placing a slit at close to the shared focal point of the two lenses. The emission is then separated into two channels by a dichroic mirror (FF635-Di02-Semrock). This dichroic is reflective to light emitted by Cy3 or Sytox orange while it is transparent to the wavelengths emitted by Cy5. Band pass filters at 731/137 nm and 571/72 nm are employed to avoid cross talk between the channels. Additionally, a 550 nm long pass filter also employed in Cy3 emission path to eliminate any direct 532 nm laser light that might pass through to the EMCCD camera (Andor Ixon 897). Finally, another dichroic mirror is employed to combine the emitted fluorescence signals onto the single EMCCD chip. This system allows to display and record emission from Cy3 or Sytox Orange and Cy5 next each other on

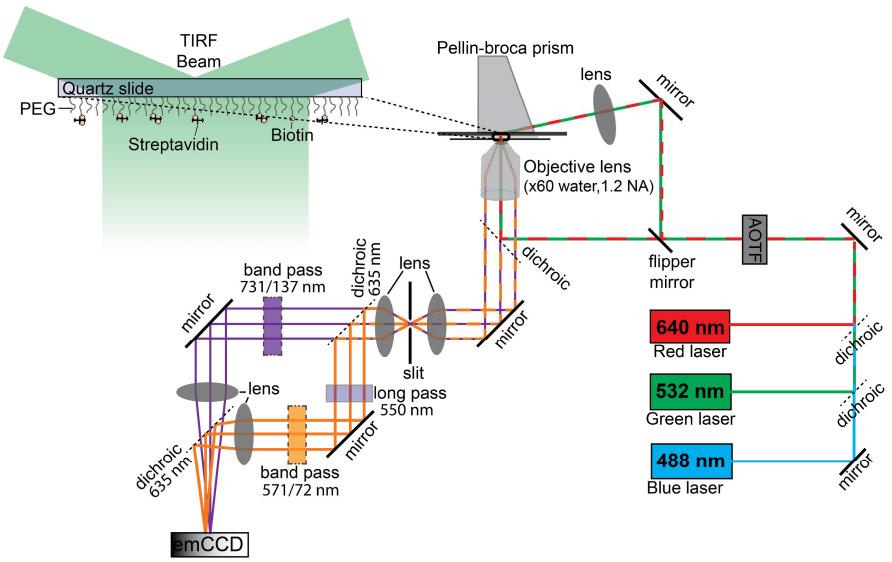


Figure 2.6**:** Fluorescence microscope set-up. The microscope can be operated either in TIRF mode or epi-fluorescence mode.

the EMCCD simultaneously. We chose to use two separate tube lenses to minimize chromatic aberrations. The microscope is operated by a custom LabVIEW routine. The timing of ALEX and EMCCD is done through LabVIEW NIDAQ-mx, using a PCIe-6320 card and a BNC-2120 breakout box (National Instruments). The EMCCD maintained at -90 °C during the data acquisition.

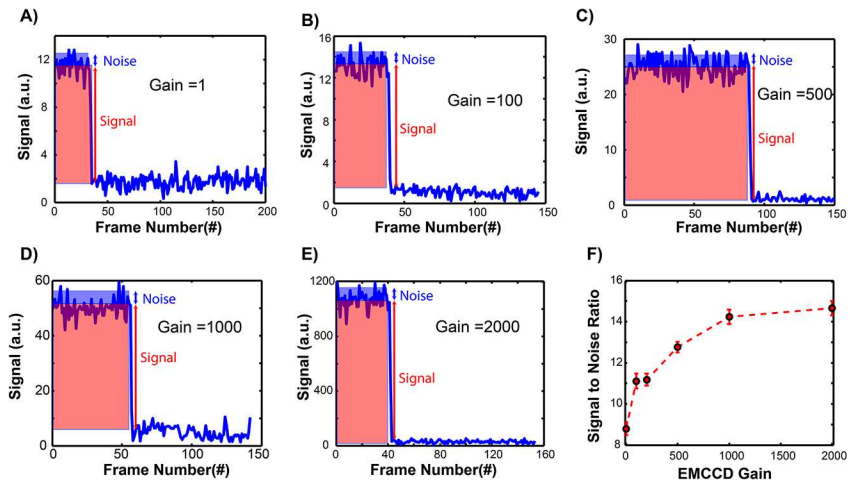


Figure 2.7**: Estimating signal-to-noise ratio from single molecule photo-bleaching traces**. (A-E) Representative photo-bleaching curves for indicated EM-gain. The noise was obtained from the standard deviation in the signal. Strength of the signal was estimated by subtracting the mean of background intensity from the mean Both the signal and noise are indicated in the plots. (F) Plot shows the SNR as a function of applied EMCCD gain. Each SNR was obtained from at least ten individual molecules that photo-bleached during acquisition.

Setting optimal EMCCD gain from single molecule photo-bleaching measurements: Cooling down the CCD to around -100 °C and electron multiplying gain are central to the EMCCD detection of single molecules with high signal-to-noise-ratio (SNR). The lower temperatures of the CCD give rise to a reduction in the noise. The optimal EM-gain for a given EMCCD should be set experimentally because higher gains might lead to decreased SNR. We set to measure single molecule bleaching events in order to measure the optimal EM-gain for obtaining highest signal to ratio. We immobilized Cy5-labelled 19nt single-stranded DNA on a PEGylated quartz slide via biotin-streptavidin-biotin linkage. The molecules were then excited with 640 nm wavelength laser and imaged at 10 frames/second using our prism-type TIRF microscope. By applying high laser power we could photo-bleach some of the Cy5 molecules within the imaging duration. Photo-bleaching of fluorophores is necessary to obtain background fluorescence. The mean intensity of single molecules was

corrected by subtracting the background mean intensity after photo-bleaching. The noise and signal was estimated for each molecule as shown in Figure 2.7A-E. From Figure 2.7F it can be seen that the SNR is reaching to plateau around an EM-gain of 1000. Based on these observations we used a gain of 1000 for the subsequent fluorescence measurements.

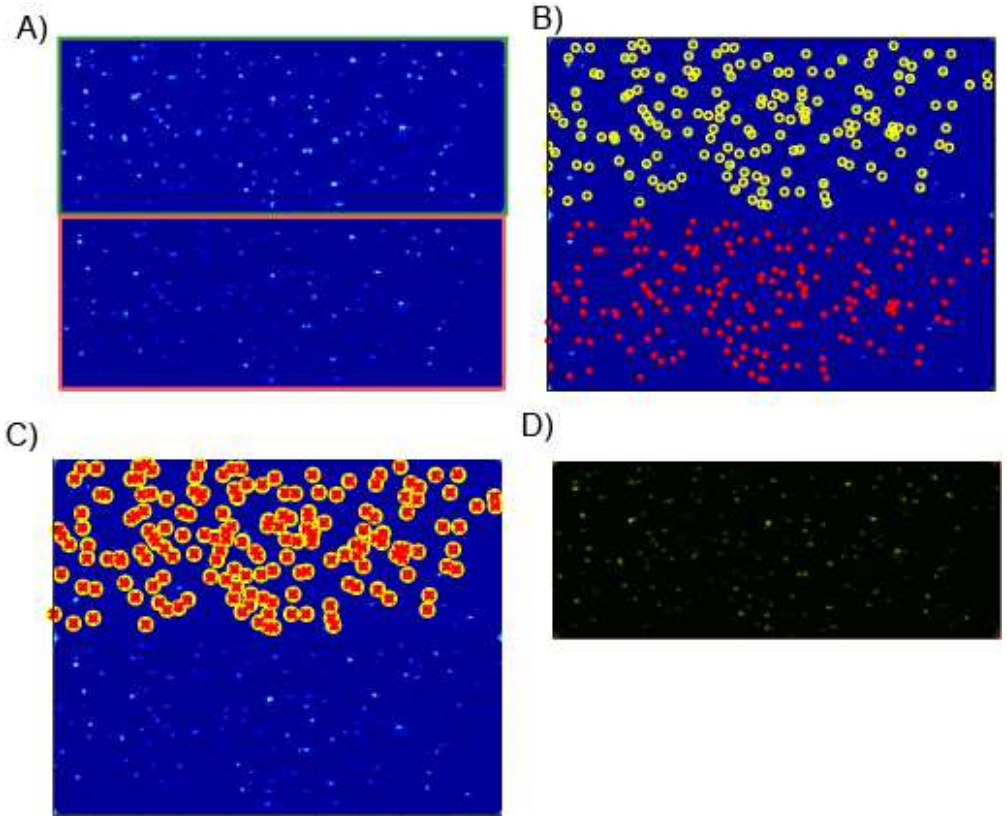


Figure 2.8**: Geometrical mapping between two channels using multi-color Tetraspeck beads. A)** High Density beads obtained by combining twenty sparsely distributed beads images. Two channels are indicated with green and magenta rectangles. **B)** Identifying the centers of beads in both the channels. **C)**Identifying the corresponding beads after mapping between the two channels. Here, the yellow circles are original coordinates identified and the red crosses are identified spots from geometrical mapping. **D)** Overlaid image of two channels after mapping every pixel in the bottom channel to its corresponding pixel in the top channel.

Mapping between two channels using multi-color fluorescent beads: As described in the section 2.3.1, for single molecule FRET, we image the emission from two fluorophores next each other on a single CCD. This way we could follow intensity

of an acceptor and its corresponding donor fluorophore in time but to do this we first need a geometrical mapping of every acceptor and donor molecules. In order to obtain a high precision mapping between two channels we used multi-color Tetraspeck beads (Invitrogen). Since these beads fluoresce in multiple wavelengths, by imaging a sparsely distributed field of view we can identify every fluorescent spot in one channel with its corresponding spot in the second channel. We repeatedly translate the microscope stage for imaging different field of views. By combining multiple fields of views we obtain a high density of paired bead localizations (Figure 2.8A). After this we identify all the spots in the combined image in each of the channels (Figure 2.8B). Using custom written Matlab code we then identify the paired coordinates of the beads in two channel.

These coordinates are then fit to a fifth-order polynomial function to co-localize every bead in the bottom channel to the ones in top channel (Figure 2.8C). The overlaid image of bottom and top channel is shown in Figure 2.8D. The yellow colored beads in overlaid image are indicative of excellent co-localization of green spots in the top channel with red spots in the bottom channel. The root mean squared displacement (RMSD) between original and mapped beads in this case is 0.49 pixels which indicates that we can follow the FRET pair in time with subpixel resolution.

2.4 Fluorescence based single-molecule techniques to probe protein-DNA interactions in real-time

Proteins are workhorses for the cellular world that their function is essential at every step of a cell cycle. For example, as described above, conversion of DNA to RNA is done by a protein called DNA polymerase in co-ordination with few other proteins. Therefore, studying the DNA-protein interactions provide a crucial information in controlling the molecular activities. The scale at which the dynamic protein-DNA interactions happen varies from nano-meter to couple of microns. For instance, a conformational change in a protein bound to DNA may be necessary to switch between two different functions. This kind of interactions require a protein to explore couple of nano-meters distance. On the other hand, some proteins are known to search a target on DNA by one-dimensional diffusion. In this case, a protein has to interact with DNA while translocating for several micro-meters. As a consequence, one needs to adopt different experimental techniques depending on the range of interactions to probe.

Single-molecule FRET as a spectroscopic nano-ruler: If a donor fluorophore and acceptor fluorophore are sufficiently close to each other and the emission spectrum of former one overlaps with the absorption spectrum of the latter one, an energy transfer occurs between the donor to acceptor. In the year 1948, German physicist

Theodor Förster theoretically described the phenomenon of non-radiative energy transfer between two fluorophores (*9*). The mechanism is then called Förster Resonance Energy Transfer (FRET) in honour of his name. Around two decades ago FRET was adopted to study the interactions of single donor and single acceptor fluorophore (*10*). Because FRET is sensitive in the length range of 2-10 nm, it became an effective tool to measure the dynamics biomolecular complexes (*5, 6, 8, 11, 12*). Single-molecule FRET (smFRET) has extensively been used to probe the conformational changes of protein- DNA complexes (Figure 2.9A). The conventional smFRET is limited to capture the distance changes of 2-10 nm between donor and acceptor fluorophores, however, the length range may be pushed further by adopting three-color or four-color FRET (*13*). In chapter 4 of this theses, we extended the smFRET to probe the transition state dynamics of HIV-1 reverse transcriptase. In general, smFRET is insensitive to visualize the corresponding distances of transition states. We circumvent this by modulating the RT-DNA interactions under various salt and macromolecular crowding concentrations. We then modelled the change in the height of transition states to provide an insight into the flipping of RT.

2.5 Stretched DNA to probe DNA-protein interactions

In order to probe the interaction of proteins with DNA on longer length scales, on the order of microns, one has to adopt either doubly-tethered DNA or DNA curtains assay (*14-16*). DNA curtains is a high-throughput technique which allows to visualize numerous nicely aligned DNA molecules in parallel (*17*). To set-up the DNA curtains assay one needs substrates sculpted with nano-structures which is a bottleneck because it requires a lot of optimisation and also an access to cleanroom. In that sense, doubly-tethered DNA assay is less tedious that it requires DNA molecules with biotin labels on both the ends and PEG-biotin coated surface (*15, 18, 19*). In these assays, a long piece of DNA is stretched and attached on a passivated surface using biotin-streptavidin-biotin linkage (Figure 2.9B). If the immobilized DNA is stretched sufficiently (around 75% to the contour length), the DNA can be imaged using total internal reflection fluorescence (TIRF) microscopy. The doubly-tethered DNA assay is a good platform to study protein-DNA interactions. In chapter 7 of this theses, we implemented this assay to study interactions of *E. coli* Dps, DNA-binding proteins from starved cells, with DNA in real-time. Based on the observations we provided new insights into the mechanism of DNA compaction by Dps.

Real-time visualization of supercoiled DNA dynamics: DNA is enormously long compared to the size of a cell that contains it. For example, 0.002 mm sized *Escherichia coli* bacterium contains a $4.7x10^6$ bases (1.6 mm) of DNA in it. Because of this difference, DNA must be folded tightly in order to fit into the cell. Furthermore, the genome should also be well-organised to be able to access any genes of interest out of several thousand genes (*20*). Bacterial genome is known to compact mainly by supercoiling of DNA (*21*). The mechanical properties of DNA are well understood based on single molecule biophysical studies (*22-25*). However, understanding of the organisational dynamics of DNA at molecular level is just started to emerge because of the development of new single-molecule tools (*26-28*). One interesting mechanical property of DNA is that application of twists leads to intertwined coils called plectonemes. The plectonemes were observed to diffuse and hop along DNA (*28*), perhaps to accumulate at certain position on the gene thereby allow or control the gene expression. Further comprehensive understanding of the supercoiled DNA and its interactions with proteins require tools that are both easily adaptable and provide deeper insights into the dynamics of supercoiled DNA. Hence, we developed a novel single-molecule assay to study supercoiled DNA dynamics in real-time. This technique can be readily adaptable by researchers who have access to simple fluorescence microscopy.

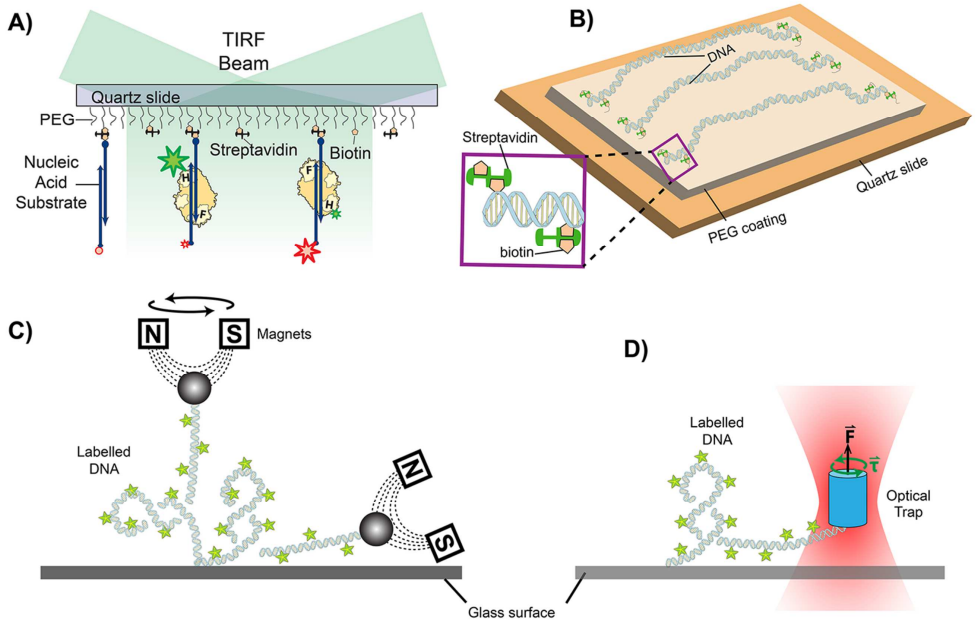


Figure 2.9**: Fluorescence based single-molecules techniques for studying protein-DNA interactions in real-time. A)** Schematic diagram showing single-molecule FRET assay. The

acceptor fluorophore (Cy5) labeled nucleic acid substrates are immobilized on the PEG-coated surface via biotin-streptavidin-biotin linkage. Binding of donor (Cy3) labeled RT on nucleic acid substrate gives a FRET signal. **B)** Doubly-tethered stretched DNA (≥20 kb) on quartz surface. The ends of DNA molecules are immobilized with multiple biotin-streptavidin-biotin linkages for topologically constraining (insert). Plectonemes can be induced on the immobilized DNA molecules by adding an intercalating dye. Binding dynamics of proteins on the immobilized DNA molecules can be visualized in real-time by tracking both the label on protein and staining the DNA. **C)** Side-pulling magnetic tweezers combined with wide-field fluorescence microscopy to study DNA supercoils and protein-DNA interactions. **D)** Combining the optical torque-wrench and fluorescence microscopy for studying protein-DNA interactions.

Intercalation induced Supercoiled DNA (ISD)- a novel single-molecule assay to visualize supercoiled DNA: We developed a novel single-molecule visualization technique based on simple fluorescence microscopy to study the plectonemic DNA dynamics on a doubly-tethered DNA molecule (Figure 2.9B). The plectonemes were induced and visualized in the doubly tethered DNA molecules by using intercalating dye molecules. In chapter 5 of this theses we implemented the Intercalation induced Supercoiled DNA (ISD) technique to show that the plectonemes can be pinned to mispaired bases of DNA. Furthermore, in chapter 6 of this theses we used this technique to show that AT-rich segments along DNA are preferable positions for plectonemes to be localized. Moreover, ISD assay is first of its kind with a potential to directly visualize the interactions of proteins with plectonemic DNA. To this end, in chapter 7 of this theses we attempted to visualize the interaction of Dps with plectonemic DNA. Our results suggest that Dps favours to bind on supercoiled DNA.

2.6 Side-pulling single-molecule force spectroscopy techniques to study DNA plectoneme dynamics

Here we briefly mention about two other potential techniques to study the dynamics of DNA plectonemes and protein-plectonemes interactions. Visualization of plectonemes require DNA to be labelled with fluorescent markers. These assays are built on the widely used single-molecule force spectroscopic techniques. In these techniques, one end of DNA molecule is attached to the surface and the other end to either the magnetic bead or to a quartz cylinder. The plectonemes are induced by applying twists on DNA using either by rotating magnets or rotating the polarization of light, the former one is called “magnetic tweezers” (*24*, *25*) and the latter one as “optical torque-wrench” (*27*, *29*, *30*). Subsequently, the molecule can be brought horizontal for imaging the dynamics of plectonemes along the length of DNA (Figure 2.9C and D). A combination of wide-field fluorescence microscopy and side-pulling single-molecule magnetic tweezers was already implemented to study the dynamics of DNA plectonemes (*28*). Combination of fluorescence with optical

trapping has been implemented to visualize DNA along its length (*31*, *32*). However, visualization of supercoiled DNA dynamics and protein-supercoiled DNA interactions are yet to be realised using optical torque-wrench.

A point to note that these two side-pulling force-spectroscopy techniques allow to precise control of the degree of supercoiling of DNA but both the techniques require a complicated apparatus making them less reachable to wider area of researchers. In this sense, our new assay, the ISD is accessible by most researchers.

2.7 References

1. P. N. Hengen, Purification of His-Tag fusion proteins from Escherichia coli. *Trends in Biochemical Sciences* **20**, 285 (1995).
2. K. Ishii, A. Katase, T. Andoh, N. Seno, Inhibition of topoisomerase I by heparin. *Biochemical and Biophysical Research Communications* **104**, 541 (1982).
3. D. S. Neblock *et al.*, Conjugation and evaluation of 7E3 x P4B6, a chemically crosslinked bispecific F(ab')2 antibody which inhibits platelet aggregation and localizes tissue plasminogen activator to the platelet surface. *Bioconjugate Chemistry* **3**, 126 (1992).
4. R. A. Marko *et al.*, Binding Kinetics and Affinities of Heterodimeric versus Homodimeric HIV-1 Reverse Transcriptase on DNA–DNA Substrates at the Single-Molecule Level. *The Journal of Physical Chemistry B* **117**, 4560 (2013).
5. E. A. Abbondanzieri *et al.*, Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature* **453**, 184 (2008).
6. S. Liu, E. A. Abbondanzieri, J. W. Rausch, S. F. J. L. Grice, X. Zhuang, Slide into Action: Dynamic Shuttling of HIV Reverse Transcriptase on Nucleic Acid Substrates. *Science* **322**, 1092 (2008).
7. A. S. Biebricher *et al.*, The impact of DNA intercalators on DNA and DNA-processing enzymes elucidated through force-dependent binding kinetics. *Nat Commun* **6**, (2015).
8. A. N. Kapanidis *et al.*, Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8936 (2004).
9. V. T. Förster, Zwischenmolekulare Energiewanderung und Fluoreszenz *Annalen der Physik (in German)* **437**, 20 (1948).
10. T. Ha *et al.*, Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences* **93**, 6264 (1996).

11. R. Roy, S. Hohng, T. Ha, A practical guide to single-molecule FRET. *Nat Meth* **5**, 507 (2008).
12. S. J. Holden *et al.*, Defining the Limits of Single-Molecule FRET Resolution in TIRF Microscopy. *Biophys. J.* **99**, 3102 (11/3/, 2010).
13. J. Lee *et al.*, Single-Molecule Four-Color FRET. *Angewandte Chemie International Edition* **49**, 9922 (2010).
14. S. E. Halford, J. F. Marko, How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040 (2004).
15. I. Bonnet *et al.*, Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Res.* **36**, 4118 (2008).
16. S. Redding, E. C. Greene, How do proteins locate specific targets in DNA? *Chemical Physics Letters* **570**, 1 (5/10/, 2013).
17. E. C. Greene, S. Wind, T. Fazio, J. Gorman, M.-L. Visnapuu, in *Methods in Enzymology*, G. W. Nils, Ed. (Academic Press, 2010), vol. Volume 472, pp. 293-315.
18. H. Yardimci, A. B. Loveland, A. M. van Oijen, J. C. Walter, Single-molecule analysis of DNA replication in Xenopus egg extracts. *Methods (San Diego, Calif.)* **57**, 179 (2012).
19. L. Cuculis, Z. Abil, H. Zhao, C. M. Schroeder, Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat Commun* **6**, (2015).
20. D. J. Sherratt, Bacterial Chromosome Dynamics. *Science* **301**, 780 (2003).
21. V. F. Holmes, N. R. Cozzarelli, Closing the ring: Links between SMC proteins and chromosome partitioning, condensation, and supercoiling. *Proceedings of the National Academy of Sciences* **97**, 1322 (2000).
22. J. Lipfert, M. Lee, O. Ordu, J. W. J. Kerssemakers, N. H. Dekker, Magnetic Tweezers for the Measurement of Twist and Torque. (2014).
23. H. Brutzer, N. Luzzietti, D. Klaue, R. Seidel, Energetics at the DNA Supercoiling Transition. *Biophysical Journal* **98**, 1267 (2010).
24. F. Mosconi, J. F. Allemand, D. Bensimon, V. Croquette, Measurement of the Torque on a Single Stretched and Twisted DNA Using Magnetic Tweezers. *Physical Review Letters* **102**, 078301 (2009).
25. T. R. Strick, J. F. Allemand, D. Bensimon, V. Croquette, Behavior of supercoiled DNA. *Biophys. J.* **74**, 2016 (1998).
26. S. Chong, C. Chen, H. Ge, X. S. Xie, Mechanism of Transcriptional Bursting in Bacteria. *Cell* **158**, 314 (2014).

27. J. Ma, L. Bai, M. D. Wang, Transcription Under Torsion. *Science* **340**, 1580 (2013).
28. M. T. J. van Loenhout, M. V. de Grunt, C. Dekker, Dynamics of DNA Supercoils. *Science* **338**, 94 (2012).
29. S. Ha, Y. Ussembayev, R. Janissen, M. van Oene, N. H. Dekker, Fabrication and Surface Functionalization of Highly Birefringent Rutile Particles for Trapping in an Optical Torque Wrench. *Biophys. J.* **108**, 337a (2015).
30. J. Lipfert, M. M. van Oene, M. Lee, F. Pedaci, N. H. Dekker, Torque Spectroscopy for the Study of Rotary Motion in Biological Systems. *Chemical Reviews* **115**, 1449 (2015).
31. S. Lee, S. Hohng, An Optical Trap Combined with Three-Color FRET. *J. Am. Chem. Soc.* **135**, 18260 (2013).
32. M. J. Comstock, T. Ha, Y. R. Chemla, Ultrahigh-resolution optical trap with single-fluorophore sensitivity. *Nat Meth* **8**, 335 (2011).

3
Strand-Displacement Synthesis by HIV-1 Reverse Transcriptase

The process of displacement synthesis, where a DNA polymerase must unwind a nucleic acid duplex without the aid of a helicase, is required for a range of polymerases to properly function in the cell. Viral polymerases, including HIV reverse transcriptase (RT), are known to use displacement synthesis to copy thousands of nucleotides of genetic code. However, the mechanism of displacement synthesis remains unclear. In order to elucidate whether RT actively or passively unwinds downstream duplexes, we measured RT elongation against a series of duplex structures. We find that the both the thermodynamic stability and the backbone content of the duplex influences the rate of DNA replication. We also find that the presence of a nucleic acid flap on the non-template strand increases the efficiency of displacement synthesis. This is the first evidence that RT makes direct contact with the non-template nucleic acid strand. Our results are inconsistent with a passive model of displacement synthesis requiring RT wait for the duplex to melt on its own. Instead, we propose an active mechanism related to some DNA helicases.

M. Ganji, S. F. J. Le Grice, and E. A. Abbondanzieri, *to be submitted*

3.1 Introduction

Many DNA polymerases, while carrying no intrinsic helicase activity, can nevertheless unwind downstream DNA annealed to the template strand through the process of strand displacement synthesis (Figure 3.1). Viral polymerases in particular often rely on displacement synthesis rather than encoding an additional helicase or clamp that must work in concert with the polymerase. Without displacement synthesis, diverse viruses including φ29-bacteriophage (1) and HIV (2,3) would not be able to transcribe.

In retroviruses, several forms of displacement synthesis are orchestrated by the enzyme reverse transcriptase (RT) while converting the single stranded RNA genome into proviral double-stranded DNA. The details of several steps in the HIV-1 reverse transcription process are showed in Supplementary Figure 3.1. RT initiates (-)-strand DNA polymerization by recruiting a cellular tRNA primer to the PBS of viral RNA (4). After initiation, RT must displace secondary structures in the single stranded RNA template to proceed. Simultaneously, RT performs RNase H cleavage on the copied RNA, creating short RNA fragments of size 6-16 nucleotides (5-8). During minus strand synthesis, RT precisely cleaves two short RNA fragments known as polypurine tracts (PPT) to use as primers for the initiation of (+)-strand DNA synthesis (9). After the addition of around 20 DNA nucleotides to the PPT primer, RT makes a cut precisely at the division between the RNA and DNA backbone before proceeding with (+)-strand synthesis. However, the left over PPT RNA fragments intact to the (-)-strand DNA and the fate of this fragment is not known.

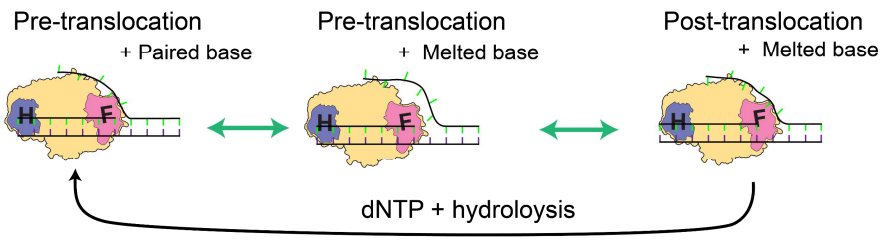


Figure 3.1: Schematic diagram illustrating individual steps in performing strand-displacement synthesis by a DNA polymerase.

DNA (+)-strand synthesis proceeds from the PPT regions and initially requires RT to displace the short RNA oligonucleotides generated by RNase H cleavage. Once RT reaches the 5' end of the (-)-strand template, RT cleaves off the tRNA leaving a sticky end (10). A circular intermediate is generated by annealing this sticky end to the 3'-end of the (-)-strand DNA. RT must then displace DNA in order to extend the 3'-end of the (-)-strand DNA back to the PPT to complete the proviral DNA.

At several critical steps, RT must therefore displace both RNA and DNA in order to copy the viral genome. However, the mechanism of strand displacement remains unclear. Previous studies show RT pauses at regions of strong secondary structure (11-14), suggesting the stability of the downstream duplexes affect the rate of displacement synthesis. RT has also been shown to unwind ~2 nt of downstream RNA even in the presence of dNTP (15,16), implying RT makes specific contacts that stabilize the bound complex during displacement synthesis. In order to elucidate the sequence dependence displacement synthesis, we assayed primer extension and strand displacement synthesis of RT on different substrates with a range of percentage of GC.

3.2 Methods

3.2.1 Oligonucleotides and preparation of displacement constructs

Table 3.1: oligonucleotide used for primer extension, RNA-, and DNA- displacement synthesis assays

Primer /5amMC12/ATTATGGATCTTAGCCACTTT
PPT RNA (non-template) UAAAAGAAAAGGGG
PPT DNA(non-template) TTAAAAGAAAAGGGG
PPT template CCCCTTTTCTTTTAAAAAGTGGCTAAGATCCATAGCTGCCTTGT
13% GC RNA (non-template) AAUUGACUAAUUAAU
13% GC DNA (non-template) AATTGACTAATTAAT
13% GC template ATTAATTAGTCAATTAAAGTGGCTAAGATCCATAGCTGCCTTGT
46% GC RNA (non-template) AUAGCGAUAGCGAUG
46% GC DNA (non-template) ATAGCGATAGCGATG
46% GC template CATCGCTATCGCTATAAAGTGGCTAAGATCCATAGCTGCCTTGT
93% GC RNA (non-template) CGCCCGCUCCCGCCG
93% GC DNA (non-template) CGCCCGCTCCCGCCG
93% GC template CGGCGGGAGCGGGCGAAAGTGGCTAAGATCCATAGCTGCCTTGT

All the oligonucleotides were purchased from IDT technologies (IDT Coraville Inc.) (Table 3.1). All the templates are designed to accommodate the same primer in order

to minimize the effects from RT- primer/template interactions. The primer listed in the Table 3.1 have an amino functional group at its 5'-end that we labelled with Cy5-NHS ester (GE healthcare) as following. Labelling reaction was done in a freshly made 0.1M sodium bicarbonate-HCl, pH 8.5 buffer. The reaction contained an approximately 142 µM primer and 20 times molar excess of Cy5-NHS ester dye. After 4 hours of incubation at room temperature, excess dye was removed by ethanol precipitation. The reaction resulted in less than 50% labelling efficiency. We repeated the labelling reaction second time which resulted in around 100% efficiency.

For primer-extension assay, we mixed 1:2 molar ratio of Cy5-labeled primer (5 µM) to template in TE buffer with 100 mM NaCl and heated the mixture on a heat block to 80 °C. Since we take fluorescence intensity of the primer as a read out for the fraction of extension, we strategically added twice more template to make sure most of the primer annealed to template. The heat block was left at room temperature for cooling the mixture to room temperature over a period of time.

For displacement synthesis assay we added an excess non-template primer to make sure that the most of the primer-template contains a non-template. Therefore we mixed 1:2:4 molar ratio of primer:template:non-template at 5 µM concentration of primer in TE buffer with 100 mM NaCl. The mixture was then heated to 80 °C on heat block and subsequently cooled to room temperature on a heat block for annealing the Cy5-lebeled primer and non-template to template. The resultant was then diluted to 250 nM and stored at -20 °C.

3.2.2 Preparation of HIV-1 reverse transcriptase

Reverse transcriptase of HIV-1 is a heterodimeric multifunctional enzyme. While both the polymerisation and RNase H functions reside in p66 subunit, p51 subunit plays a role in gripping the substrates. Individual subunits were purified as described earlier (17). The p51 subunit was purified in its wild type form. An E478Q mutation was introduced into p66 subunit to abolish the RNase H function (18). The functional heterodimers, final concentration of 1 µM, were reconstituted by mixing 1:5 ratio of p66 and p51 subunits and incubating at 37 °C for 2 hours. Aliquots of 2 µl were stored at -20 °C. An aliquot was left at room temperature for around 2 hours before using. The dissociation constant (K_d <10 nM) of heterodimers is low compared to the homodimers we expect majority of p66 in the form of heterodimer. The concentration of p51 was too low to form homodimers and the p51 alone shows no affinity to DNA.

3.2.3 Primer-extension and displacement synthesis assays

All the primer-extension and displacement synthesis assays were performed in reaction buffer: 50 mM Tris-HCl, pH 8.0, 50 mM KCl, 8 mM $MgCl_2$ at room temperature with indicated concentration of PEG 8K. A 50 µl of reaction mixture containing 10 nM DNA substrates and 20 nM RT were pre-incubated for 5 minutes. The extension or displacement synthesis was initiated by adding 200 µM of dNTPs mixture. Reactions were terminated at 1, 2, 5, 10 and 30 minutes by adding a 10 µl of reaction mixture to 10 µl of 20 mM EDTA and 96% formamide. The reactions were then heated at 90 °C for 5 minutes and immediately chilled on ice in order to denature the DNA products. The products were then fractionated on a 10% urea-polyacrylamide gel. The gel was then imaged for Cy5 fluorescence on Typhoon scanner. The amount of unextended, partially extended, and fully extended primer intensity was analysed using 1D gel analysis software.

3.3 Results

3.3.1 Primer extension by reverse transcriptase on different GC rich templates

First we tested the ability of RT to polyemerise DNA primer on different GC rich templates. We designed four templates which can accommodate the same primer but with different GC content in the extendable region. The first template was selected from poly purine tract (PPT) region of the HIV genome. We designed this substrate such a way that primer was situated right behind the PPT sequence. The other three substrates contain the same sequence in the primer region but the template sequence down stream to the primer was varied. This way we expect RT to bind on the primer-template with equal affinity and then we can study the ability to polymerise different GC sequences. The resuts showed that RT polymerises on general sequences much more efficiently than on the PPT template (Figure 3.2). It was known that RT pauses at certain sequences especially on poly(T) regions. The PPT template indeed contains two thymine quartets seperated by single nucleotide. The first quartet was immediately upstream to 3'-end of the primer where RT was found to be predominantly paused (see lanes 1-4 of Figure 3.2A). Up on over coming the first pause, RT encounters the next strong pauses around 7 and 11 nucleotides upstream to the 3'-end of the primer that can be seen from the lanes1-4 of Figure 3.2A.

We quantified the intensity of the primer bands by separating them into three catageries: (1) unextended (U), (2) partially extended (P), and (3) fully extended (F) fractions. The amount of fully extended primer was calculated as follows

$$Fully\ extended\ fraction = \frac{F}{F+P+U} \qquad (1)$$

Because RT pauses strongly on PPT template, the amount of fully extended primer was linearly increased over time instead of generally observed exponential behaviour (Figure 3.2C- blue curve).

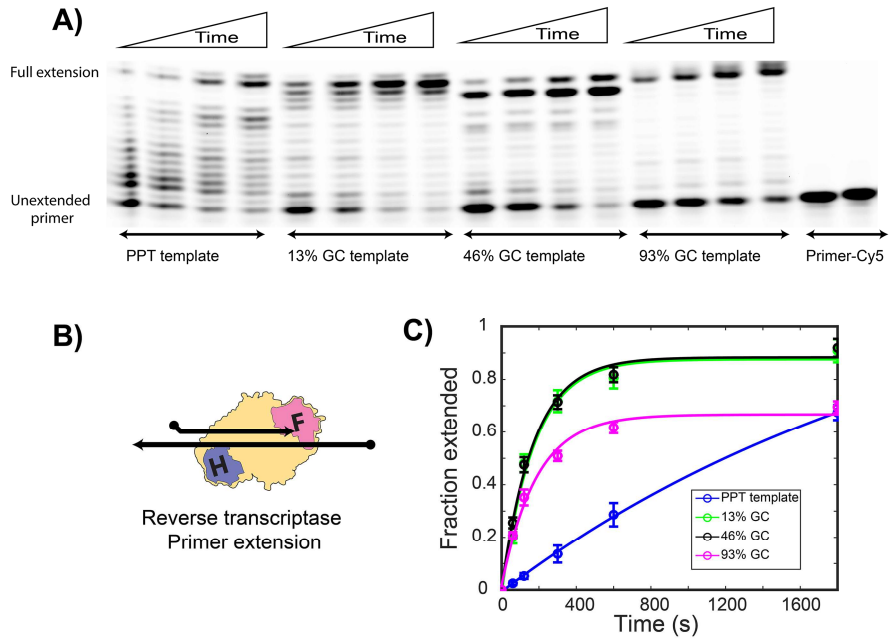


Figure 3.2**: Primer extension by reverse transcriptase is weakly affected by GC content**. Four template sequences were designed annealed to a common primer but had varying GC content in the 15 nt downstream. the same primer but the downstream of primer in the template is designed to have different GC content. **A)** Representative gel showing primer extension by reverse transcriptase. First four lanes represent the extension of primer through PPT region. Lanes 5-8 represent the extension on low GC (13%) template region. Lanes 9-12 represent the extension on 46% GC template. Lanes 13-16 represent the extension on GC rich (93%) template sequence. Lanes 17-20 are un-extended primer. The primer elongation in this gel was terminated at 1, 2, 5, and 10 minutes. **B)** Cartoon showing the RT (yellow) positioned on DNA primer/template (Black arrows) in the polymerase mode. **C)** Plot showing the fraction of fully extended primer over time. Error bars are s.e.m. from at least three experiments. Exponential fits to the fraction extension versus time yielded rate constants of $4x10^{-4}$, $5.7x10^{-3}$,$5.7x10^{-3}$, and $5.3x10^{-3}$ /s.

Contrary to the PPT template, primer extension by RT on other GC rich templates was not much different to each other and the primers were extended exponentially with time (Figure 3.2). There was some fraction of primer that was not extended in the case of 93% GC substrates probably because it was not annealed properly. However, the rate of extension obtained from exponentail fitting was not any different compared to 13% and 46% GC substrates indicating that primer extension proceeded in the same fashion. Next, we investigated if RT can extend the primers in the presence of the upstream RNA fragments resulted from RNase H activity of RT.

3.3.2 RT displacement synthesis through RNA fragments

RNase H activity of HIV-1 RT results in different sized RNA fragments. Some of the fragments will be intact to the DNA by base complementarity if the size of the fragments are longer than 10 nucleotides. We tested if the presence of these RNA fragments impede the DNA polymerisation. We used the same sudstrates used for primer extension assay but with a complimentary RNA fragment upstream to the 3'-end of the primer (see cartoons in Figure 3.3). These substrates allowed us to test the sequence dependent RNA strand displacement synthesis.

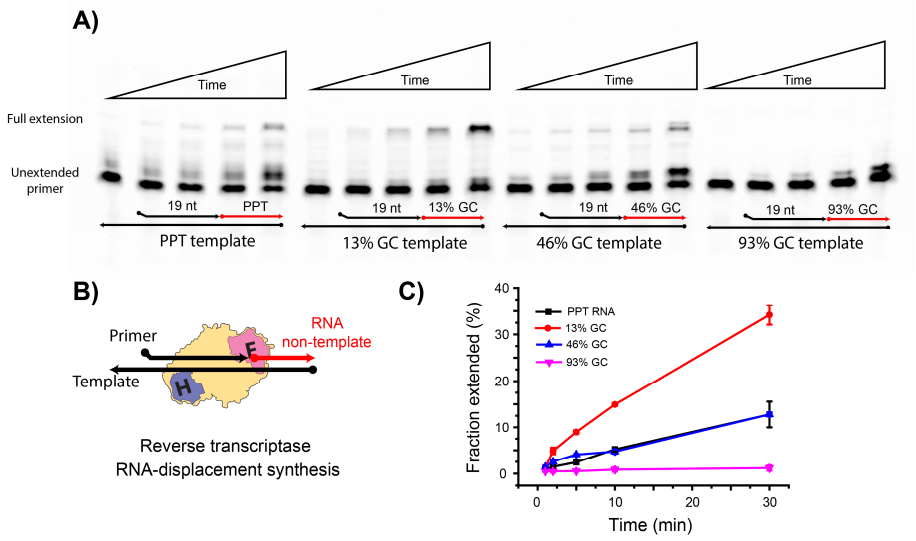


Figure 3.3**: Displacement synthesis through RNA non-templates**. Primer extension by displacement of the non-template PPT RNA, 13% GC RNA, 46% GC RNA and 93% GC RNA. First four lanes represent the displacement synthesis through PPT RNA region. **A)** Representative gel showing the RNA displacement synthesis. Lanes 5-8 represent the displacement synthesis through low GC (13%) non-template RNA region. Lanes 9-12 represent the displacement synthesis through 46% GC non-template RNA. Lanes 13-16

represent the displacement synthesis through GC rich (93%) non-template RNA. The displacement synthesis in each case was terminated at 1, 2, 5, 10, and 30 minutes. **B)** Cartoon showing RT (yellow) bound on DNA primer/template (black arrows) with an up-steam RNA non-template (green arrow). **C)** Plot showing the fraction of fully extended primer by displacing RNA non-template over time. Error bars are s.e.m. from at least three experiments.

The results indicated that the RNA strand displacement synthesis was strongly impeded by the presence of the upstream RNA non-template primer and also dependent on the GC content of the non-template primer to be displaced (Figure 3.3). The displacement of 46% GC and PPT non-template with equal efficiency was some what surprising because we expected PPT RNA to be displaced easily because it contains 33% GC. It should be noted that RT showed a slow primer extension on PPT template,and hence the presence of upstream PPT RNA primer seemed to have a minimal effect compared to the other upstream non- template primers. It could be because PPT RNA exhibit localized anomalous basepairing to DNA (19,20). The 13% GC non-template RNA fragment was easiy displaced compared to the highest GC non-temlate (93%). Presence of 93% GC non-template primer showed to inhibit primer extension beyond 2 nucleotides. We next wanted to probe the displacement synthesis of DNA fragments with identical sequences and compare that to the displacement synthesis of RNA.

3.3.3 RT displacement synthesis through DNA fragments

RT must also perform DNA displacement synthesis during several stages of reverse transcription. Here, we tested the efficiency of RT to displace DNA fragments with

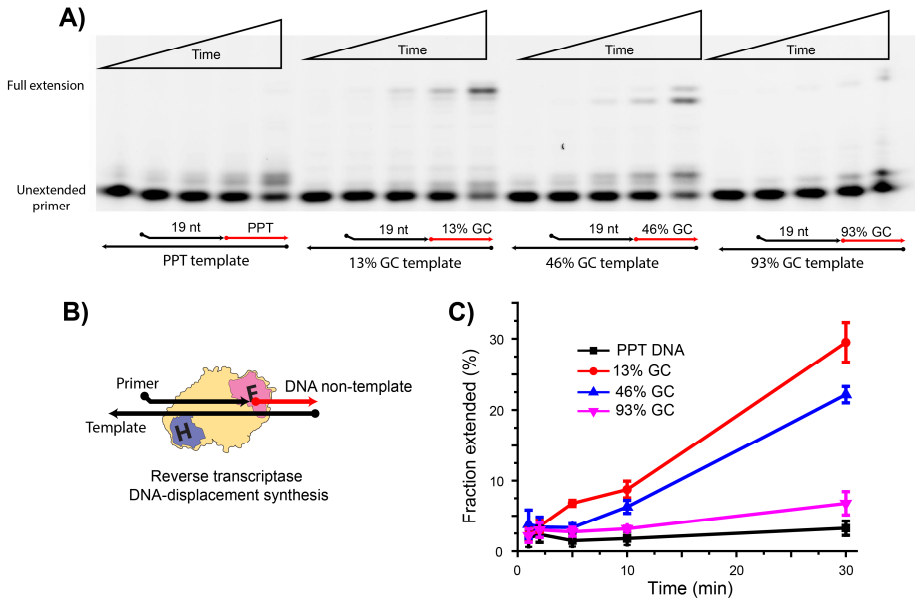


Figure 3.4**: Primer extension by displacement of the non-template PPT DNA, 13% GC DNA, 46% GC DNA and 93% GC DNA**. First four lanes represent the displacement synthesis through PPT DNA region. Lanes 5-8 represent the displacement synthesis through low GC (13%) non-template DNA region. Lanes 9-12 represent the displacement synthesis through 46% GC non-template DNA. Lanes 13-16 represent the displacement synthesis through GC rich (93%) non-template DNA. In each case, the primer elongation was terminated at 1, 2, 5, 10, and 30 minutes. **B)** Cartoon showing RT (yellow) bound on DNA primer/template (black arrows) with an up-steam DNA non-template (red arrow). **C)** Plot showing the fraction of fully extended primer by displacing DNA non-template over time. Error bars are s.e.m. from at least three experiments.

different GC contents. We once again used the same DNA substrates that we used in primer extension and RNA displacement synthesis but with DNA upstream non-templates (see cartoons in Figure 3.4).

Similar to the RNA displacement synthesis, DNA displacement also showed strong dependency on the GC content of the non-template (Figure 3.4A). Quantified results of displacement of non-template DNA fragments showed in Figure 3.4C. Surprisingly, displacement synthesis of PPT DNA was the slowest among all the non-templates tested. The displacement synthesis of PPT DNA did not show more than 3 nucleotides extension and the others were displaced more efficiently comparing to their RNA counter parts (compare Figure 3.3 and Figure 3.4C). The striking difference between RNA and DNA PPT displacement synthesis can be due to the difference in base-pairing with DNA template.

The results clearly indicated that RT was less efficient in displacement synthesis of either RNA or DNA non-template fragments. We then tested the effects of macromolecular crowding on the displacement synthesis of RT.

During the (-) strand DNA synthesis, RT encounters RNA hairpins and during the (+) strand DNA synthesis, RT encounters RNA fragments resulted from RNase H digestion, *in virio*. RT must able to displace these RNA fragments for successful reverse transcription. RT must also perform DNA displacement synthesis for generating proviral DNA for integration into the host-genome. However, the above results indicate that RT was not efficient in displacing the non-template RNA or DNA fragments. In view of this, we tested if the presence of macromolecular crowding enhances the efficiency of displacement synthesis. We tested the 46% GC content substrate for both the RNA and DNA strand displacement synthesis in the presence of 100 mg/ml of PEG 8K. The results showed that RT displaces both RNA and DNA non-templates slightly more efficiently under crowding compared to the non-crowded conditions (Supplementary Figure 3.2). However, the displacement synthesis was still much slower than the primer extension. The higher crowding concentration in the cytoplasm of cells (around 400 mg/ml) may boost the strand displacement synthesis.

During the displacement synthesis there was considerable amount of primer accumulated after extension for three nucleotides (Figure 3.3 and Figure 3.4) probably indicating that the polymerization of first three nucleotides is a limiting step. Because, once the displacement synthesis was above three nucleotides, the synthesis continued to the end of template which might indicate of RT interaction with the 3-nucleotide flap of non-template. To gain insights into the RT interaction with the non-template flap we performed displacement synthesis on a construct that contained three- or six- nucleotide non-template flap (see Figure 3.5B).

3.3.4 Non-template DNA flap enhances displacement synthesis activity

We then probed the dependence of DNA displacement synthesis on the flap length of non-template which RT encounters upon initiation of displacement synthesis. For this we created four different constructs using 38-bp Cy5-labelled DNA primer, 63-bp DNA template, and 15-bp non-template DNA with zero-bp, 3-bp, or 6-bp flap (Figure 3.5B). These constructs were subjected to DNA displacement synthesis by RT and the results were fractionated on urea-PAGE (Figure 3.5A). The primer-extension and displacement synthesis (no flap) was as seen in the previous sections. But the displacement synthesis in the presence of 3- or 6- nucleotide flap was much faster comparing to the zero-bp flap displacement synthesis. This increase in activity

of RT strongly supports the idea that it interacts with the flap of non-templated DNA.

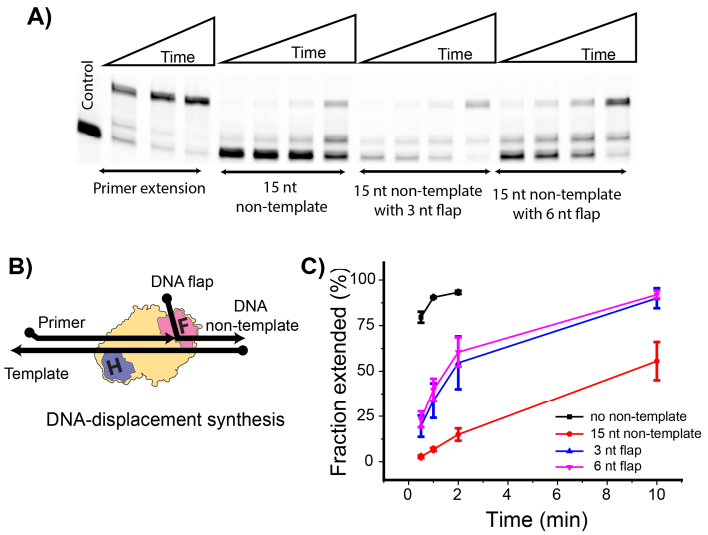


Figure 3.5: **Flap length dependence reverse transcriptase strand displacement synthesis**. A) A representative gel showing the primer extension, displacement synthesis with no flap and with 3- or 6- nucleotide flap. **B)** Cartoon showing RT (yellow) bound on DNA primer/template /non-template (black arrows) with a flap. **C)** Plot showing the fraction of fully extended primer by displacing DNA non-template with a flap over time. Error bars are s.e.m. from at least three experiments.

3.4 Discussion

Viral reverse transcriptase is a unique enzyme that orchestrates different functions in order to convert a single stranded RNA genome into proviral DNA genome. While synthesizing DNA using RNA templates, RT cleaves the copied RNA into short fragments by RNase H activity. The RNase H digestion results in RNA fragments of different lengths ranging from few nucleotides to tens of nucleotide (5,6,21). Moreover, RT specifically process a 15 nucleotide purine rich RNA fragment to be used as a primer to initiate (+)-strand DNA synthesis (9). During the (+)-strand DNA synthesis, RT encounters the remained RNA fragments. Previous research showed that, apart from these explicit enzymatic functions, RT performs strand displacement synthesis (3). RT thus capable of displacing non-templated RNA or DNA fragments.

Our results indicate that the displacement synthesis by RT is very slow compared to the normal DNA polymerization. The displacement synthesis was strongly

dependent on the sequence of the non-template fragment, but the displacement of PPT RNA and DNA were exceptions. Primer extension on the PPT template was anticipated to be slow because RT was known to pause on poly(T) (25). In view of the observation that the primer extension on the PPT template was much slower compared to the primer

Table 3.2**:** Predicted free energy changes for DNA/DNA and RNA/DNA duplex formations at room temperature. These free energy differences were predicted using nearest neighbor approximations (22,23) and corrected for 50 mM NaCl according to an ampirical formula provided in the ref. (24)

Nucleic acid sequences	**% of GC**	**$\Delta G°$(kcal/mol)**	**Fraction extended**
d(AATTGACTAATTAAT)	13	-12.53	29.45
d(ATAGCGATAGCGATG)	46	-18.49	22.15
d(CGCCCGCTCCCGCCG)	93	-28.88	6.78
d(TTAAAAGAAAAGGGG)	33(PPT DNA)	-16.50	3.24
r(AATTGACTAATTAAT)	13	-9.99	34.17
r(ATAGCGATAGCGATG)	46	-17.54	12.78
r(CGCCCGCTCCCGCCG)	93	-25.51	1.17
r(TTAAAAGAAAAGGGG)	33 (PPT RNA)	-17.54	12.77

extension on the other templates, we compared the displacement of PPT RNA and PPT DNA non-templates. While the PPT RNA template was displaced by RT, the PPT DNA non-template was observed to hinder the polymerization beyond three nucleotides. This result reiterates the fact that PPT-RNA/DNA deviates from the normal base pairing as showed by both the X-ray crystal structure and biochemical footprinting studies (19,20). In the absence of this anomalous base pairing interactions, PPT RNA was expected to displace least efficiently because the primer-extension on this template itself was rather slow. This hypothesis was supported by the observation that PPT DNA was the slowest in terms of displacement synthesis. Therefore, our results strongly supporting the view that PPT RNA/DNA assumes a unique structure, in order for selective processing and successive usage as a primer.

We calculated free energy of annealing using the nearest neighbor estimation (22,23) to compare with the displacement synthesis of each substrate (Table 3.2). The predicted free energy changes for the formation of DNA/DNA and RNA/DNA duplexes globally agree with the sequence dependent displacement synthesis albeit PPT RNA and PPT DNA non-templates being the exceptions. Although, the change in the free energy of formation of PPT RNA/DNA and PPT DNA/DNA duplexes were comparable, our data showed that the PPT RNA/DNA was the easiest to

displace while the PPT DNA/DNA was only displaced maximum of three nucleotides. A detectable enhancement in the displacement synthesis of both RNA and DNA was observed in presence of crowding molecules. However, the polymerization reactions were still slower compared to the normal primer extension, perhaps indicating that higher cellular crowding (around 400 mg/ml) would promote the displacement synthesis. Once again, this confirms that the PPT RNA/DNA takes different structure to be selectively recognized by RT to be used as a primer for (+)-strand DNA synthesis. Our results also support that RT interacts with the flap on not-templated DNA actively in order for displacement synthesis.

Taken together these results suggest that in the absence of flap, displacement synthesis is very slow compared to primer-extension and proceeds only if the nucleotides melt due to stochastic reversible denaturation (13,26). However, our results support that as the flap builds up by three nucleotides or more, RT interacts with the flap to fray the next paired nucleotide of non-template strand (15). This melting step is strongly dependent of the base paired nucleotides because the displacement synthesis showed a strong dependence on the sequence of nucleic acids (13). We describe that RT can then translocate one nucleotide forward on the template to be able to ready for binding and hydrolysis of a next cognate dNTP (27). Accordingly, RT continues in cycles of fraying non-template DNA, translocating one nucleotide forward, and recruit and hydrolyze a dNTP to perform the displacement synthesis (Figure 3.1). In addition to this nucleic acid chaperone, the nucleocapsid protein (NC) of HIV, might also assist in accelerating the displacement synthesis (28,29).

In conclusion, we showed that the primer extension with in the PPT region was much slower compared to the other templates due to pausing of RT. The displacement synthesis was observed to be rather sequence dependent and the high GC content non-template fragments were observed to hinder the polymerization reaction (13). Although speculative, it might be that HIV evolved to contain biased AU-content (~58%) in its RNA genome (30) to be effective for strand displacement synthesis. Surprisingly, while RT was able to displace the non-template RNA-PPT, the displacement of non-template DNA-PPT was completely blocked indicating the structural anomaly of RNA-PPT (19). Our observation also indicated that the including crowding molecules enhances the displacement synthesis. Highly crowded cellular environment, where the biochemical reactions happen, together with low GC-content viral genome probably aid RT for successful reverse transcription.

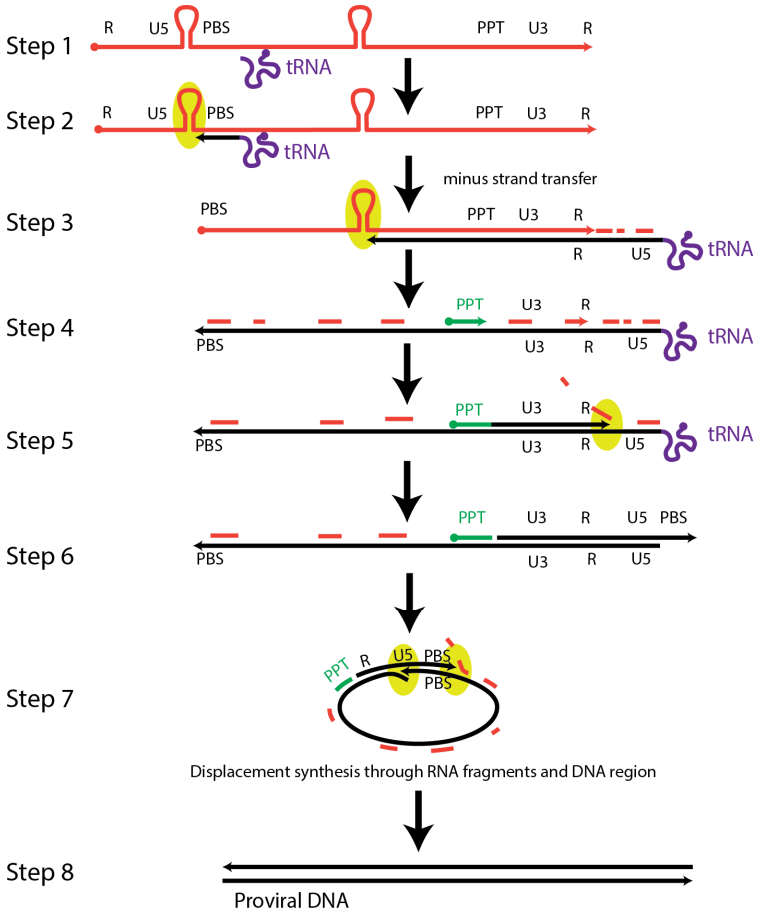
3.5 References

1. Blanco, L., Bernad, A., Lázaro, J.M., Martín, G., Garmendia, C. and Salas, M. (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *Journal of Biological Chemistry*, **264**, 8935-8940.
2. Fuentes, G.M., Rodríguez-Rodríguez, L., Palaniappan, C., Fay, P.J. and Bambara, R.A. (1996) Strand Displacement Synthesis of the Long Terminal Repeats by HIV Reverse Transcriptase. *Journal of Biological Chemistry*, **271**, 1966-1971.
3. Huber, H.E., McCoy, J.M., Seehra, J.S. and Richardson, C.C. (1989) Human immunodeficiency virus 1 reverse transcriptase. Template binding, processivity, strand displacement synthesis, and template switching. *J. Biol. Chem.*, **264**, 4669-4678.
4. Marquet, R., Isel, C., Ehresmann, C. and Ehresmann, B. (1995) tRNAs as primer of reverse transcriptases. *Biochimie*, **77**, 113-124.
5. Wisniewski, M., Balakrishnan, M., Palaniappan, C., Fay, P.J. and Bambara, R.A. (2000) Unique progressive cleavage mechanism of HIV reverse transcriptase RNase H. *Proceedings of the National Academy of Sciences*, **97**, 11978-11983.
6. DeStefano, J.J., Mallaber, L.M., Fay, P.J. and Bambara, R.A. (1994) Quantitative analysis of RNA cleavage during RNAdirected DNA synthesis by human immunodeficiency and avian myeloblastosis virus reverse transcriptases. *Nucleic Acids Research*, **22**, 3793-3800.
7. Fuentes, G.M., Fay, P.J. and Bambara, R.A. (1996) Relationship between Plus Strand DNA Synthesis and Removal of Downstream Segments of RNA by Human Immunodeficiency Virus, Murine Leukemia Virus and Avian Myeloblastoma Virus Reverse Transcriptases. *Nucleic Acids Research*, **24**, 1719-1726.
8. Champoux, J.J. (1993) 6 Roles of Ribonuclease H in Reverse Transcription. *Cold Spring Harbor Monograph Archive*, **23**, 103-117.
9. Huber, H.E. and Richardson, C.C. (1990) Processing of the primer for plus strand DNA synthesis by human immunodeficiency virus 1 reverse transcriptase. *Journal of Biological Chemistry*, **265**, 10565-10573.
10. Smith, J.S. and Roth, M.J. (1992) Specificity of human immunodeficiency virus-1 reverse transcriptase-associated ribonuclease H in removal of the minus-strand primer, tRNA(Lys3). *Journal of Biological Chemistry*, **267**, 15071-15079.
11. Arts, E.J., Li, Z. and Wainberg, M.A. (1995) Analysis of primer extension and the first template switch during human immunodeficiency virus reverse transcription. *J. Biomed. Sci.*, **2**, 314-321.

12. Cirino, N.M., Cameron, C.E., Smith, J.S., Rausch, J.W., Roth, M.J., Benkovic, S.J. and Le Grice, S.F. (1995) Divalent cation modulation of the ribonuclease functions of human immunodeficiency virus reverse transcriptase. *Biochemistry*, **34**, 9936-9943.
13. Kim, S., Schroeder, C.M. and Xie, X.S. (2010) Single-Molecule Study of DNA Polymerization Activity of HIV-1 Reverse Transcriptase on DNA Templates. *J. Mol. Biol.*, **395**, 995-1006.
14. DeStefano, J.J., Buiser, R.G., Mallaber, L.M., Fay, P.J. and Bambara, R.A. (1992) Parameters that influence processive synthesis and site-specific termination by human immunodeficiency virus reverse transcriptase on RNA and DNA templates. *Biochim. Biophys. Acta*, **1131**, 270-280.
15. Whiting, S.H. and Champoux, J.J. (1994) Strand displacement synthesis capability of Moloney murine leukemia virus reverse transcriptase. *J. Virol.*, **68**, 4747-4758.
16. Winshell, J., Paulson, B.A., Buelow, B.D. and Champoux, J.J. (2004) Requirements for DNA Unpairing during Displacement Synthesis by HIV-1 Reverse Transcriptase. *J. Biol. Chem.*, **279**, 52924-52933.
17. Grice, S.F.J.L., Cameron, C.E. and Benkovic, S.J. (1995) Purification and characterization of human immunodeficiency virus type 1 reverse transcriptase. *Methods in Enzymol*, **262**, 5.
18. Schatz, O., Cromme, F.V., Grüninger-Leitch, F. and Grice, S.F.J.L. (1989) Point mutations in conserved amino acid residues within the C-terminal domain of HIV-1 reverse transcriptase specifically repress RNase H function. *FEBS Lett.*, **257**, 4.
19. Kvaratskhelia, M., Budihas, S.R. and Le Grice, S.F.J. (2002) Pre-existing Distortions in Nucleic Acid Structure Aid Polypurine Tract Selection by HIV-1 Reverse Transcriptase. *J. Biol. Chem.*, **277**, 16689-16696.
20. Sarafianos, S.G., Das, K., Tantillo, C., Clark, A.D., Ding, J., Whitcomb, J.M., Boyer, P.L., Hughes, S.H. and Arnold, E. (2001) *Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA*.
21. DeStefano, J.J., Buiser, R.G., Mallaber, L.M., Myers, T.W., Bambara, R.A. and Fay, P.J. (1991) Polymerization and RNase H activities of the reverse transcriptases from avian myeloblastosis, human immunodeficiency, and Moloney murine leukemia viruses are functionally uncoupled. *Journal of Biological Chemistry*, **266**, 7423-7431.
22. Sugimoto, N., Nakano, S.-i., Yoneyama, M. and Honda, K.-i. (1996) Improved Thermodynamic Parameters and Helix Initiation Factor to Predict Stability of DNA Duplexes. *Nucleic Acids Research*, **24**, 4501-4505.
23. Sugimoto, N., Nakano, S.-i., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M. and Sasaki, M. (1995) Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes. *Biochemistry*, **34**, 11211-11216.

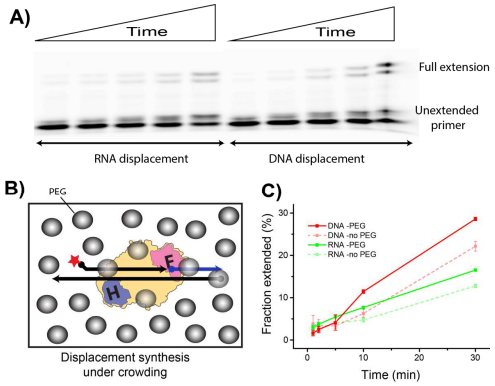
24. John SantaLucia, J. and Hicks, D. (2004) The Thermodynamics of DNA Structural Motifs. *Annual Review of Biophysics and Biomolecular Structure*, **33**, 415-440.
25. Klarmann, G.J., Schauber, C.A. and Preston, B.D. (1993) Template-directed pausing of DNA synthesis by HIV-1 reverse transcriptase during polymerization of HIV-1 sequences in vitro. *Journal of Biological Chemistry*, **268**, 9793-9802.
26. Liu, S., Harada, B.T., Miller, J.T., Le Grice, S.F.J. and Zhuang, X. (2010) Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription. *Nat. Struct. Mol. Biol.*, **17**, 1453-1460.
27. Suo, Z. and Johnson, K.A. (1997) Effect of RNA secondary structure on the kinetics of DNA synthesis catalyzed by HIV-1 reverse transcriptase. *Biochemistry*, **36**, 12459-12467.
28. Cosa, G., Harbron, E.J., Zeng, Y., Liu, H.W., O'Connor, D.B., Eta-Hosokawa, C., Musier-Forsyth, K. and Barbara, P.F. (2004) Secondary structure and secondary structure dynamics of DNA hairpins complexed with HIV-1 NC protein. *Biophys. J.*, **87**, 2759-2767.
29. Levin, J.G., Guo, J., Rouzina, I. and Musier-Forsyth, K. (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog. Nucleic Acid Res. Mol. Biol.*, **80**, 217-286.
30. AC, v.d.K. and B., B. (2012) The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. *Retrovirology*, **9**, 14.

Supplementary Figures



Supplementary Figure 3.1**: Steps in reverse transcription**. **Step 1:** RT of HIV-1 first recruits cellular tRNA at primer binding site (PBS) upon infection of the host cell. **Step 2:** RT then starts to copy the viral RNA to DNA using a tRNA primer. RT also digests the copied RNA into smaller fragments. The smallest RNA fragments (<10 nucleotides) fall off the DNA while the longer ones are remain bound. RT performs RNA displacement synthesis when it encounters a RNA hairpin. Here and in the subsequent steps, displacement synthesis is highlighted with a yellow ellipse **Step 3:** Minus strand transfer occurs: the copied DNA along with tRNA are transferred to the other end of the viral RNA. During this step, RT encounters more RNA hairpins where RT has to perform displacement synthesis **Step 4:** RT copies the remaining RNA genome while digesting the copied RNA into smaller pieces. During these steps, RT specifically processes a short RNA polypurine tract to be able to use as a primer for (+) strand DNA synthesis. **Step 5:** Synthesis of (+) strand DNA begins at the PPT primer. RT displaces those RNA fragments left on the template DNA and it makes a cut exactly at the end of the PPT RNA. **Step 6:** DNA synthesis reaches 5′ end of the (-) strand DNA. RT then cuts off the tRNA leaving a sticky end. **Step 7**: Circularization of the DNA by annealing the PBS sites. RT once again needs to perform strand displacement synthesis through the left over RNA fragments and PPT which will still be intact with the DNA. At this point RT also performs DNA displacement synthesis through U5-region towards R-

region **Step 8:** Linearization of the fully synthesised proviral DNA which is ready to be incorporated into the host cell genome.



Supplementary Figure 3.2**: Displacement of the non-template RNA and DNA under macromolecular crowding.** The reaction buffer contained 100 mg/ml PEG 8K. First five lanes represent the displacement synthesis through 46% GC non-template RNA region. Lanes 6-10 represent the displacement synthesis through 46% GC non-template DNA region. In each case, the primer elongation was terminated at 1, 2, 5, 10, and 30 minutes. **B)** Cartoon showing RT (yellow) bound on DNA primer/template (black arrows) with and up-steam DNA or RNA non-template (blue arrow) under macromolecular crowding (grey spheres). **C)** Plots showing the fraction of fully extended primer by displacing DNA non-template (red) and RNA non-template (green) over time at 100 mg/ml PEG 8K. Plots for the displacement synthesis of DNA (light red) andRNA (light green) are also showed for comparison. Error bars are s.e.m. from three experiments.

4
DNA Binding Proteins Explore Multiple Local Configurations

Finding the target site and associating in a specific orientation are essential tasks for DNA-binding proteins. In order to make the target search process as efficient as possible, proteins should not only rapidly diffuse to the target site but also dynamically explore multiple local configurations before diffusing away. Protein flipping is an example of this second process that has been observed previously, but the underlying mechanism of flipping remains unclear. Here we probed the mechanism of protein flipping at the single molecule level, using HIV-1 reverse transcriptase (RT) as a model system. In order to test the effects of long-range attractive forces on flipping efficiency, we varied the salt concentration and macromolecular crowding conditions. As expected, increased salt concentrations weaken the binding of RT to DNA while increased crowding strengthens the binding. Moreover, when we analyzed the flipping kinetics, i.e. the rate and probability of flipping, at each condition we found that flipping was more efficient when RT bound more strongly. Our data are consistent with a view that DNA bound proteins undergo multiple rapid re-binding events, or short hops, that allow the protein to explore other configurations without completely dissociating from the DNA.

This chapter has been published as M. Ganji, M. Docter, S. F. J. Le Grice, and E. A. Abbondanzieri, DNA binding proteins explore multiple local configurations during docking via rapid rebinding, *Nucleic Acids Res.*, 44, 8376-8384 (2016)

4.1 Introduction

Proteins that non-covalently bind and release nucleic acids must undergo a complex three-dimensional search to find specific targets in the crowded environment of the cell. On distance scales of microns to nanometers, this search on long nucleic acid duplexes can be facilitated by a combination of sliding (continuous attachment) and hopping (short three dimensional excursions) (1-3). However, as a protein approaches its target, it must become precisely oriented relative to the target site within the nucleic acid substrate in order to perform enzymatic reactions. These last steps of the target search involve diffusion over nanometer distances, and therefore occur at microsecond to millisecond time scales that are both too fast to observe directly with many experimental techniques and too slow to easily simulate using all-atom molecular dynamics approaches.

Protein flipping has recently emerged as an important component of the last steps of the target search mechanism and in the subsequent exploration of alternate local configurations. It has been proposed that type II restriction enzymes use a flipping mechanism to cleave complimentary DNA strands during a single encounter with the target sequence (4-6) and the homeodomain transcription factor HoxD9 was observed to flip to bind on two oppositely oriented binding sites within a single DNA molecule (7). Additionally, flipping was directly observed in single molecule complexes of HIV reverse transcriptase (RT) bound to various nucleic acid substrates (8-10). Furthermore, protein flipping was used as a tool to discriminate between nucleoside and non-nucleoside RT inhibitors (11,12). Flipping requires the enzyme to rotate around a vector perpendicular to the axis of the duplex and is distinct from rotation along the nucleic acid helix (13). Because of this, during a flipping event many specific contacts must be disrupted between the protein and nucleic acid and then immediately reformed with the positions of the two nucleic acid strands in the duplex reversed.

We have chosen RT as a model system to understand the mechanism of flipping using the technique of single-molecule FRET (smFRET). The time resolution of smFRET can be pushed as high as a few milliseconds with high illumination intensities (14) but at a cost of increasing the rate of photo-bleaching (15). Instead of attempting to measure the FRET efficiency during the short-lived flipping intermediates, we chose to measure the rate constant of flipping (k_f), dissociation (k_{off}), and binding (k_{on}) of RT on DNA substrates to decipher the mechanism of the transitions. We therefore selected a frame rate and laser intensity that provides relatively high spatio-temporal resolution (~1 nm and 100 msec) while allowing us

to capture the entire interaction of RT with the DNA with minimal photo-bleaching (~10%).

We have taken advantage of two experimental parameters that modify the affinity of RT for DNA: salt concentration and macromolecular crowding. Low salt concentrations preferentially strengthen the stability of bound states containing strong electrostatic interactions, while crowding agents preferentially strengthen the stability of states that form a compact complex. In other DNA polymerases, the bound states are stabilized by a decrease in the salt concentration of the buffer (16,17) or an increase in macromolecular crowding (18-22). However, the magnitude of this effect on the bound states of RT still remains to be experimentally determined. Typical *in vitro* measurements of RT are frequently performed at low salt concentrations (<50 mM) whereas concentrations of ~150 mM better mimic the binding interactions observed *in vivo* (23).

Furthermore, using transition state theory we can correlate changes in k_f and k_{off} with changes in the free energy barriers associated with these transitions. Therefore, by measuring how the rate constants and equilibrium constants associated with binding, flipping, and dissociation change as a function of salt and macromolecular crowding, we can determine the relative stability of each transition state.

For a given bound state (B) and an associated flipped bound state (F), we consider two possible mechanisms for flipping to occur which we define as “tumbling” and “hopping” (Supplementary Figure 4.1). In the tumbling model (Figure 4.1a-left), the enzyme maintains some non-covalent contacts with substrate during the flipping transition and the complex becomes slightly less compact (Supplementary Figure 4.1). The flipping transition by tumbling is therefore distinct from the dissociation transition. Tumbling can be thought of as an analogue to the sliding transition that has been shown to mediate one-dimensional diffusion of proteins on DNA (13,24,25). In the hopping model (Figure 4.1a-right), the enzyme enters a short-lived pseudo-bound intermediate state representing the short diffusive excursions near the DNA that allow flipping or dissociation to occur (Supplementary Figure 4.1). A pseudo-bound state (or ‘effective’ state) has been previously introduced to capture the essential kinetic behavior of diffusing macromolecules undergoing competitor-induced dissociation (26). An analogous hopping transition has been shown to allow some proteins to diffuse along DNA (27,28).

If flipping occurs by tumbling, we expect that RT breaks electrostatic contacts and becomes less compact in the tumbling transition state. We therefore would expect that k_f would decrease as salt concentration decreases or macromolecular crowding

increases. However, if flipping occurs by hopping, k_f will be affected by both the rate at which RT enters into the less compact pseudo-bound state and the probability that the enzyme rebinds from the pseudo-bound state rather than dissociating. If the long-range (0-2 nm) attractive forces between the enzyme and the DNA increase, the rate of entering the pseudo-bound state will decrease but the likelihood of rebinding will increase, making it possible that the net k_f could increase (supplementary materials and Supplementary Figure 4.7). Below we show that changes in k_f and flipping probability (P_{flip}) can be explained by the hopping model but not the tumbling model.

4.2 Materials and Methods

4.2.1 Labeled RT and nucleic acid substrates preparation

Individual subunits of HIV-1 RT were expressed separately as described earlier (29). An E478Q mutation was introduced in p66 subunit to eliminate RNase H activity (30). To allow for site-specific labeling, native cysteines in the p66 subunit were converted to serine and a unique cysteine was introduced at its C-terminus (31). All purification steps were performed at 4 °C. As a first round of purification, p66 was expressed and purified via a self-poured Ni-NTA (GE Healthcare) column. The resultant p66 was applied to a heparin-Sepharose column (GE Healthcare). The enzyme was reduced with DTT on the column and then the column was washed with 50 ml of DTT free sodium phosphate buffer. Reduced p66 was eluted by applying a 75 mM-1M NaCl gradient and purified protein was immediately mixed with Cy3-maleimide dye (GE Healthcare) at a 1:10 protein:dye ratio. The labeling reaction was allowed to continue for ~2 hours at room temperature before quenching with a 10-fold molar excessive of 2-mercaptoethanol. The resultant mixture of p66 and Cy3 was loaded onto the heparin-Sepharose column while tracking Cy3 absorption at 550 nm. The column was washed with sodium phosphate buffer, until free dye was removed. Cy3-labelled p66 was then eluted by applying a salt gradient. Labeled protein was concentrated by mono-S chromatography (GE Healthcare). The reaction resulted in ~50% labeled p66 with >95% purity as observed from absorption and SDS-PAGE analysis, respectively. Functional RT heterodimers of p66 and p51 were prepared by mixing a 1:10 ratio of p66-Cy3:p51. The mixture was incubated at 37 °C for 2 hours and then for 4 hours at room temperature. The dissociation rates of p51/p51 or p66/p66 homodimers are fast enough compared to p66/p51 heterodimer to allow most to exchange within 6 hours (32). Because the concentration of individual subunits is much higher than the dissociation constant (K_d<10 nM) of heterodimer, we expect to have a majority of p66 in the form of heterodimer.

Aliquots were stored at -80 °C until just before the measurements. Since p51 alone has low affinity for DNA, excess of p51 does not interfere with heterodimer RT binding to DNA.

DNA substrates were purchased from IDT technologies (IDT; Coralville, IA). Where applicable, a thymine nucleotide was amino modified with C6 linker, in order to label oligonucleotides with an acceptor fluorophore. Mono-reactive Cy5 (GE healthcare) was attached to an amine group on the nucleic acids as described by the manufacturer. Excess non-reactive dye was removed by ethanol precipitation. Absorption measurements indicate that the labeling efficiency was close to 100%. A 1:10 ratio mixture of the complimentary strands with biotin modifications at their 5'-end were annealed to Cy5-labelled strand by heating at 80 °C for 5 min and allowed to cool to room temperature. The annealing buffer was 100 mM Tris-HCl (pH 8.0), 1mM EDTA and 100 mM NaCl. The product was stored at -20 °C.

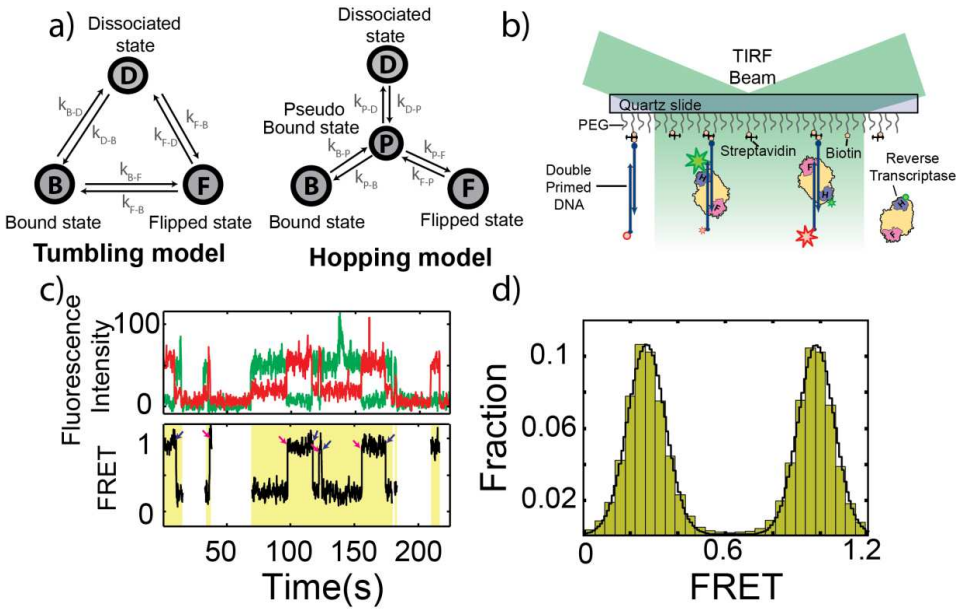


Figure 4.1**: Single-molecule FRET assay for probing RT dynamics. a)** Two possible models describing flipping transition ***left***- tumbling model, a basic three state model in which RT can undergo a transition of flipping or dissociation from a bound state and ***right***- hopping model, a model with a pseudo bound state in which RT must undergo to the transition of flipping or dissociation through the pseudo state. **b)** Schematic diagram of detection of two binding orientations of RT on a double primer DNA substrate. Larger green star and smaller red star represent a low FRET binding mode, larger red star and smaller green star represent a high FRET binding mode of freely diffusing Cy3(green sphere)-labelled RT to surface immobilized Cy5(red sphere)-labelled DNA. **c)** FRET analysis of RT binding to a double-primer DNA substrate. Top: fluorescence time traces of Cy3 (green) and Cy5 (red) under 532

nm laser excitation. Bottom: FRET value calculated over the duration of the binding events. Arrows represent low-to-high (pink) and high-to-low(blue) FRET flipping events. Yellow shaded region identifies the bound state and **d)** FRET distribution histogram of RT binding on 19-bp double primer DNA at 50 mM NaCl. The histogram was constructed from 1534 binding events.

4.2.2 Monitoring the flipping of RT by single-molecule FRET assay

Microscope quartz slides (G. Finkenbeiner, Inc.) and cover slips were passivated with PEG as described (33). Surface passivated slides, together with cover slips, were stored under inert conditions at -80 °C until used.

Flow cells were assembled by sandwiching double-sided tape between a cover slip and a quartz slide. Typically, the volume of the flow cell was ≤10 µl. 100 µl of buffer A (50mM Tris-HCl, pH 8.0, 50 mM NaCl, and 6 mM $MgCl_2$) was applied through the flow cell before applying a 0.2 mg/ml streptavidin in buffer A for 1 minute. Excess streptavidin was removed with 100 µl of buffer A before incubating a 50 pM biotin-DNA substrate. Finally, the flow cell was washed with 100 µl of buffer A to remove unbound DNA. Fluorescence measurements were performed in an imaging buffer consisting of 50mM Tris-HCl, pH 8.0, 6 mM $MgCl_2$, 20 nM of RT-Cy3, 2 mM trolox, 0.2% triton-X100, and an oxygen scavenging system (0.3 mg/ml glucose oxidase and 40 µg/ml catalase with 5% (w/v) glucose as a substrate). Additionally, PEG 8K was used as a crowding agent and NaCl was used to vary the salt concentration. Where needed, the concentrations of NaCl and PEG 8K are mentioned.

4.3 Results

4.3.1 Rapid flipping of RT on a 19-bp double primer dsDNA

Previously, smFRET was used to show that RT spontaneously flips 180° on various nucleic acids (8-10). To study the mechanism of flipping, we immobilized a 19-bp double primed dsDNA (dpdsDNA) (Figure 4.1b) in which both strands can serve as primer or template (see Supplementary Materials for DNA sequences). The symmetric structure of dpdsDNA substrate allows RT to bind in two different orientations (Figure 4.1b), and the two binding orientations should occur with roughly equal probability. We first examined the binding of RT on dpdsDNA using smFRET under low salt concentration (50 mM NaCl) and low crowding (0% PEG 8k) conditions. Time trajectories showed that on this substrate RT bound in distinct states with corresponding FRET values of about 0.3 and 1 (Figure 4.1c).

A histogram of more than 1500 binding events shows that RT spends roughly the same amount of time in each FRET state (Figure 4.1d). Moreover, RT was observed to flip spontaneously between the two binding modes (indicated with arrows in Figure 4.1c). Measuring the binding dynamics allowed us to directly measure the frequency of binding and the binding dwell times, which are direct measures of k_{on} and k_{off}, respectively (Supplementary materials). Using this assay we next sought to explore how salt and crowding conditions influence the stability of RT binding to DNA.

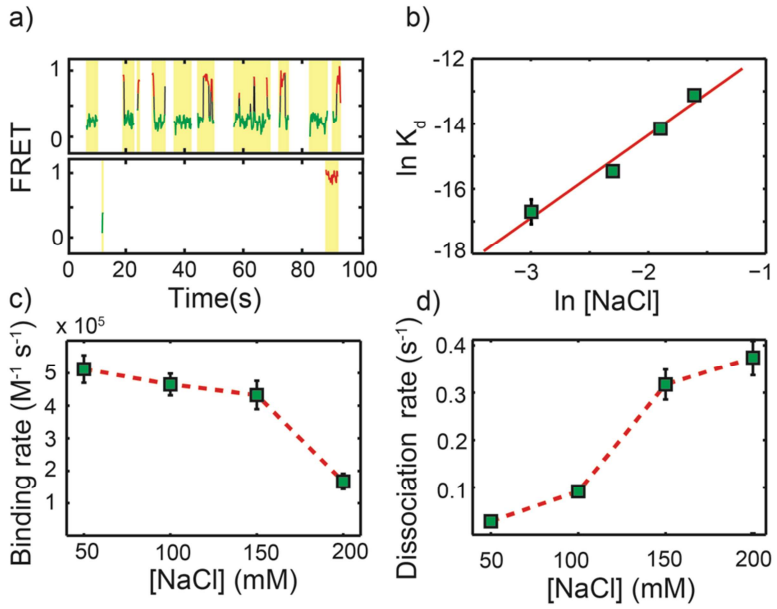


Figure 4.2**: Binding kinetics of RT on DNA as a function of salt concentration. a)** Example FRET traces of RT binding dynamics on dpdsDNA. Top: binding dynamics at 50mM NaCl, bottom: at 200 mM NaCl. Green and red represent a low FRET and high FRET bound states. Black line represents a flipping transition between the two bound states. Yellow shaded region identifies the bound state. **b)** NaCl linkages ($\partial \ln K_d$ versus $\partial \ln[\mathrm{NaCl}]$) for the binding of RT to 19-bp dpdsDNA. The slope of the linear fit (red line) to the data (squares) is 2.55, the thermodynamic net average number of ions released upon RT-DNA complex formation. **c)** Salt concentration dependent binding rate calculated from the frequency of RT binding to DNA over the observation time. **d)** Salt concentration dependent dissociation rate derived from the dwell times of RT binding. Dashed red line is a reference line connecting the data points. Number of binding events analyzed for 50, 100, 150, and 200 mM salt concentrations are 1534, 2700, 4431, and 659, respectively.

4.3.2 RT binds weakly on DNA under increased ionic concentration

The interactions of a protein-nucleic acid complex are in general strongly influenced by the salt concentration of the buffer. Typically, proteins interact weakly with DNA under high salt concentrations (16,22). The ability of RT to bind to DNA under increased salt concentration was first probed using a bulk primer-extension assay (Supplementary materials). We tested the ability of RT to extend a Cy5-labelled 19-bp DNA primer on a 52-bp template under various salt (NaCl) concentrations (50 and 300 mM). This primer-extension assay revealed that the initial bound fraction of DNA molecules decreased with an increased salt concentration (Supplementary Figure 4.2), which suggests that screening of ionic charges weakens the binding of RT to DNA templates.

To complement the ensemble results that RT binds less tightly on DNA under increased salt concentration, and to acquire the individual kinetic parameters of flipping and dissociation, we returned to the smFRET assay, which allows precise determination of the dissociation constant (K_d), k_{on}, k_{off}, and k_f. These quantities were corrected for the effects of photo-bleaching (Supplementary materials).

Using smFRET we next measured binding of RT on DNA under increased salt concentration. RT exhibited increased dwell times with a higher frequency of binding events at 50 mM NaCl compared to a salt concentration relevant to *in vivo* conditions (150-200 mM NaCl) (Figure 4.2a and Supplementary Figure 4.3). Analysis of a large set of single molecule traces (> 500 traces in each test condition) demonstrates that K_d increases with an increase in salt concentration (Figure 4.2b). We also determined the thermodynamic linkage number (L_n) of the interaction which is a measure of binding affinity, given by:

$$L_n = \partial\left(\frac{\ln(K_d)}{\ln[NaCl]}\right) \qquad (1)$$

L_n roughly corresponds to the number of counter ions released during formation of a protein-DNA complex (16,34,35). A linear fit to the linkage plot yielded an L_n of 2.5 (Figure 4.2b). This net release of ions is similar to what was observed for other DNA polymerases (16). The single-molecule results therefore confirm that RT binds weakly under increased salt concentrations.

Next, we estimated the binding and dissociation rates from single-molecule binding events (Supplementary materials) to identify the origin of the change in K_d. k_{off} increased and k_{on} decreased with increased salt concentration (Figure 4.2c and d),

indicating that the transition state corresponding to binding and dissociation shifted relative to both the bound state and the dissociated state. We note that at a salt concentration relevant to *in vivo* conditions, RT binds DNA infrequently and dissociates rapidly once bound. Given that RT copies ~9.8 kilobases of double stranded DNA *in vivo*, we therefore expect cellular conditions stabilize bound RT through some other mechanism.

4.3.3 Impact of macromolecular crowding on binding of RT to DNA

We next investigated if macromolecular crowding is a key factor in determining RT binding stability at a salt concentration that is representative of *in vivo*. We chose a widely-used crowding agent (21,36), PEG 8K, whose radius of gyration is comparable to the length of RT (37). We measured binding dynamics on DNA at 150 mM NaCl under various PEG 8K concentrations using both the bulk primer-extension assay (Supplementary materials) and smFRET. With increasing PEG 8K concentration, the fraction of bound DNA increased as measured by the primer-extension assay. The results suggest that increasing the volume fraction occupied by the PEG 8K molecules leads to an increase in the binding stability of RT on DNA (Supplementary Figure 4.2). Similar increases in stability were observed when other crowding molecules were used (Supplementary Figure 4.2d).

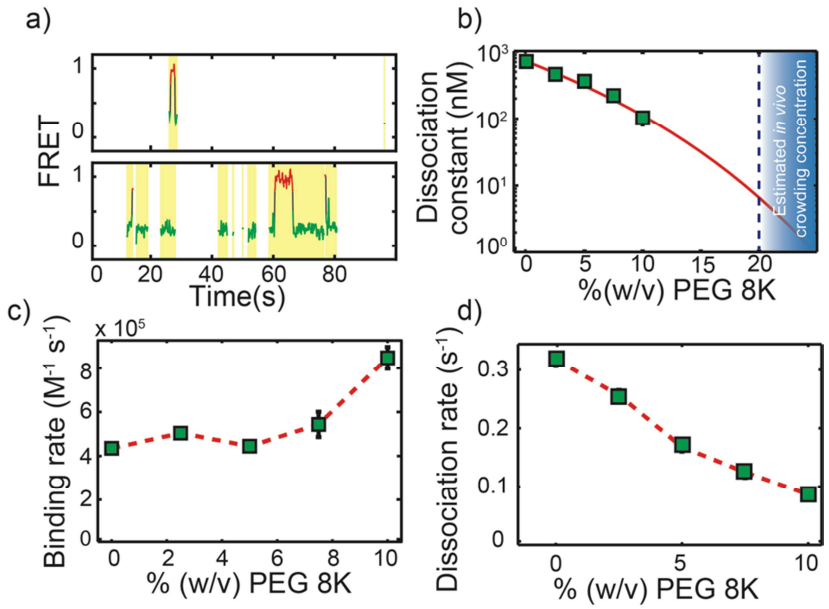


Figure 4.3**: Effects of increased macromolecular crowding on the binding kinetics of RT on dpdsDNA. a)** representative FRET traces of RT binding dynamics on dpdsDNA. Top:

data obtained at 150mM NaCl, bottom: data obtained at 150 mM NaCl and 10% (w/v) PEG 8K. Flipping (black line) of RT between low FRET (green) and high FRET (red) is pronounced under increased crowding. Yellow shaded region identifies the bound state. **b)** Dissociation constant (K_d) of RT binding on DNA under macromolecular crowding. Red curve is a fit to the data (squares) based on scaled particle theory. Blue region is where the concentration of cellular crowding falls in which RT is estimated to have sub-nanomolar binding affinity. **c)** and **d)** Rate of binding and dissociation of RT binding on DNA as a function of PEG 8K at 150 mM NaCl, respectively. Dashed red line is a reference line connecting the data points. Number of binding events analyzed for 0, 2.5, 5, 7.5, and 10 % (w/.v) PEG 8K concentrations are 4431, 3925, 4120, 911, and 1049, respectively.

In order to confirm this relationship, smFRET measurements of RT-DNA interactions in the presence of PEG 8K were performed at 150 mM NaCl. Addition of 10% (w/v) PEG 8K lead to an increase in both the binding frequency and dwell times compared to the dilute conditions (Figure 4.3a,Supplementary Figure 4.3). We found that increasing the volume fraction of PEG 8K to 10% caused an order of magnitude decrease (i.e. tightening) in K_d (Figure 4.3b). In addition, a small but measurable increase in k_{on} of RT (Figure 4.3c) and decrease in k_{off} was observed as crowding increased (Figure 4.3d).

To determine the binding affinity of RT at *in vivo* crowding conditions, we fit the decrease in the K_d of RT-DNA using scaled-particle theory (21,38,39). This theory predicts that the more compact RT-DNA complex is entropically favored over the dissociated RT and DNA as the number of crowding molecules increases. By modelling the DNA, RT, and the RT-DNA complex with simple geometric forms approximating their size and shape (Supplementary materials), we find we can fit the relative decrease in the dissociation constant without any additional free parameters (Figure 4.3b). Extrapolation of this curve to physiological crowding conditions suggests RT binds with a sub-nanomolar dissociation constant *in vivo*. We also measured the effects of crowding at 200 mM NaCl with similar results (Supplementary Figure 4.4). We therefore find that the effects of macromolecular crowding on the dissociation constant of RT can be interpreted straightforwardly from the entropic effects described in scaled particle theory.

Because increased macromolecular crowding also increases the viscosity of solution, we tested the effect of viscosity on the binding of RT to DNA independently. We used high concentrations of D-(+)-dextrose in order to create a similar effective viscosity. We did not observe a significant difference in K_d under increased D-(+)-dextrose concentration (Supplementary Figure 4.5a). However, a decrease in k_{on} and k_{off} was observed (Supplementary Figure 4.6b and c). These results indicate that the increased viscosity due to increased crowing could only have slowed the kinetics of

binding and dissociation slightly, but did not affect the stability of the bound complex.

4.3.4 RT binds stably on DNA with an incoming nucleotide

Both salt concentration and crowding can alter the attraction of RT and DNA even at longer ranges (~1 nm separation). In order to increase in the stability of RT-DNA complexes without stabilizing the transition states, we examined the binding in the presence of cognate nucleotides (dNTP). Previously, an RT/DNA/dNTP ternary complex was shown to have higher stability than RT/DNA alone (8). We therefore prepared a new double primer DNA of 21-bp with a chain terminating nucleotide (dideoxynucleotide) at one of the 3'-priming ends (Figure 4.4a and see Supplementary Information for DNA sequences). In the absence of this nucleotide, the binding kinetics of RT on this substrate was comparable to the standard dpdsDNA (compare Figure 4.2a-bottom and Figure 4.4b-top). The addition of 250 µM dTTP (the next cognate nucleotide for the high FRET orientation) lead to RT binding predominantly in a high FRET state (Figure 4.4b-middle). Since the addition of nucleotides lead to

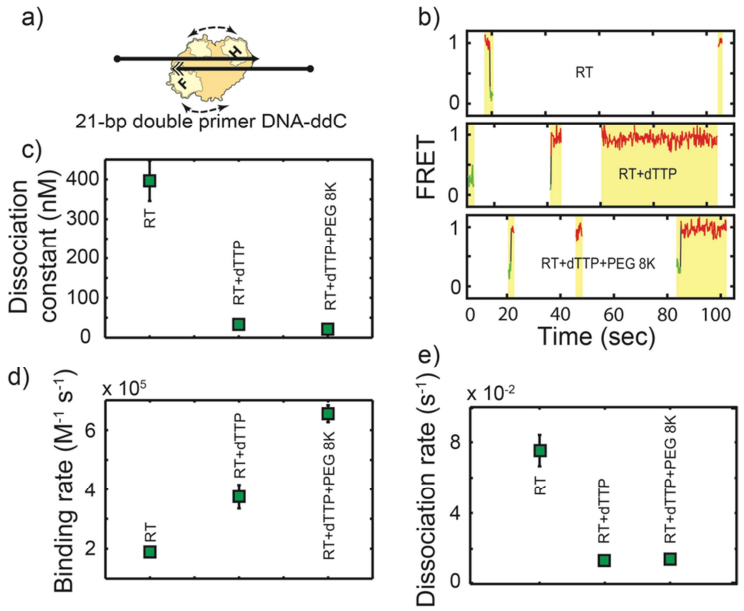


Figure 4.4**: Binding dynamics of RT on 21-bp dpdsDNA with an incoming dNTP. a)** A 21-bp double primer DNA construct with a chain-terminating (ddC) priming end (double arrow) to study the effects of an incoming nucleotide on RT binding kinetics. **b)** Representative FRET traces of RT binding dynamics on 21-bp double primer with a chain terminating dNTP. Top: RT binding under 150 mM NaCl, middle: 150 mM NaCl and 250

µM dTTP and bottom: 150 mM NaCl, 250 µM dTTP and 7.5% PEG 8K, respectively. Yellow shaded region identifies the bound state. **c)** Dissociation constant of RT binding on 21-bp DNA-ddC. **d)** and **e)** Rate of binding and rate of dissociation of RT on 21-bp DNA-ddC. Number of binding events analyzed for RT, RT+dTTP, and RT+dTTP+PEG 8K are 393, 284, and 471, respectively.

stabilization of high FRET state only, we only compared the high FRET data on this substrate. As expected, the presence of cognate nucleotides greatly decreased the effective K_d (Figure 4.4c). This was caused primarily by a decrease in k_{off} rather than an increase in k_{on} (Figure 4.4d and e). We also explored the combined effects of 250 µM dTTP and 7.5% (w/v) PEG 8K, and found that binding was further stabilized. Under these conditions, k_{on} increased significantly and k_{off} decreased significantly compared to RT alone (Figure 4.4). We conclude that short-range interactions primarily change k_{off} rather than k_{on}.

4.3.5 Increased ionic concentration makes flipping less likely

Having established the effect of salt concentration and crowding on the bound states, we next wanted to explore how the flipping transition state responded to these conditions. These measurements can determine the strength of ionic interactions and the degree of compaction in the transition states and any short-lived intermediate states involved in flipping. We again analyzed binding events as a function of salt concentration (Figure 4.2a), but we now extracted the effective k_f , i.e. the number of flipping events per second bound, and P_{flip}, i.e. likelihood of flipping versus dissociation from a high or low FRET event (Supplementary materials). The P_{flip} is determined solely by the ratio of k_f and k_{off}. While increasing salt concentration increased k_{off} , the number of flipping events dropped dramatically (Figure 4.2a). P_{flip} dropped monotonically with increased salt concentration, with a total decrease of >98% from 50 mM to 200 mM (Figure 4.5a). This sharp decrease in P_{flip} can be partly attributed to the increase in k_{off}. However, k_f also dramatically decreased as the salt concentration was increased (Figure 4.5b). As a whole, k_f at 200 mM NaCl decreased by >95% compared to 50 mM NaCl. These results indicate that k_f and k_{off} move in opposite directions in response to salt concentration.

4.3.6 Flipping is more likely in the presence of crowding agents

We next extracted k_f and P_{flip} from binding events measured in the presence of crowding agents (Figure 4.3a). A monotonic increase in P_{flip} was observed as the PEG 8K concentration increased from 0% to 10% (Figure 4.5c). Addition of 10% (w/v) PEG 8K resulted in a 10-fold increase in P_{flip}. We also found that while k_{off} decreased in the presence of crowding agents, k_f consistently increased (Figure

4.5d). Taken together, we once again see that k_f moves in the opposite direction from k_{off} with crowding. It also indicates that flipping occurs frequently at *in vivo* salt and crowding conditions.

4.3.7 Flipping is equally likely but less frequent in the presence of cognate dNTP

Finally, we wished to determine how flipping was influenced when no long range attractive potential exists but the stability of the bound state is altered. We extracted k_f and P_{flip} from high FRET binding in the presence of 250 µM cognate dNTP (Figure 4.4b). While P_{flip} was minimally affected, k_f decreased dramatically (Figure 4.5d and f). This indicates that k_f and the k_{off} scaled together, keeping P_{flip} roughly constant.

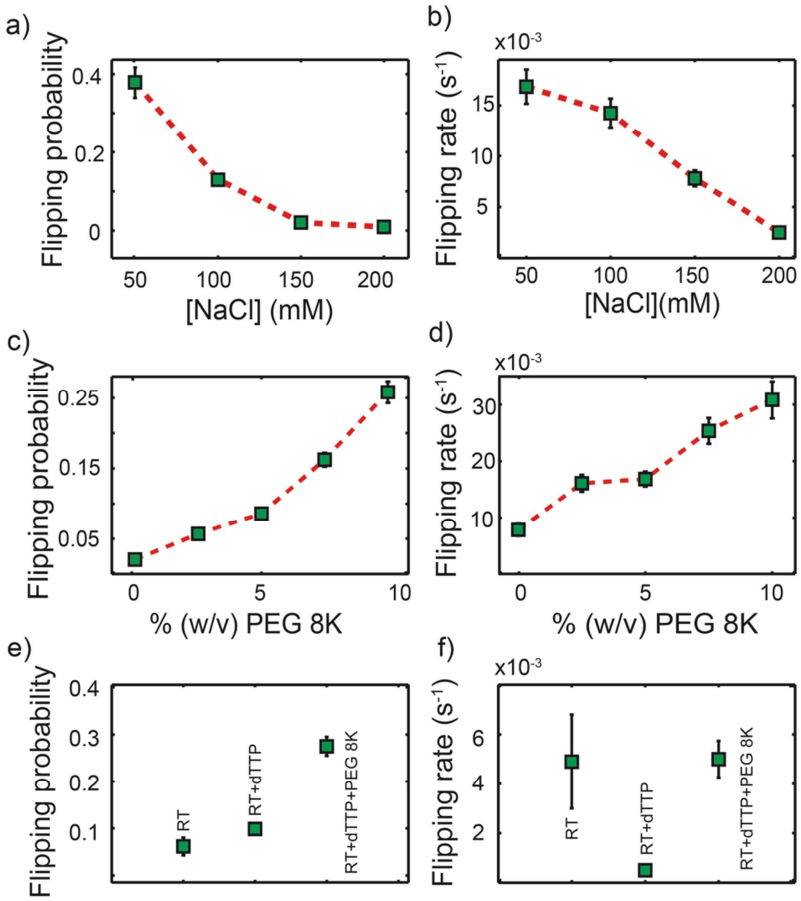


Figure 4.5**: Effects of increased salt concentration and macromolecular crowding on the flipping transition of RT. a)** NaCl dependence of probability of flipping of RT on 19-bp double primer. **b)** NaCl dependence of rate of flipping. **c)** PEG 8K dependence of probability of flipping of RT at 150 mM NaCl. **d)** PEG 8K dependence of rate of flipping of RT at 150 mM NaCl. **e)** Probability of flipping of RT on 21-bp double primer DNA-ddC (Figure 4.4a). **f)** Rate of flipping of RT on 21-bp double primer.

Addition of dTTP and 7.5% PEG 8K further stabilized the binding, but increased P_{flip} and k_f. We conclude from this that short-range stabilization of the RT bound state affects k_f and k_{off} identically. Taken together, these results suggest that long-range forces are likely the cause of the differential effects in k_f and k_{off} observed under different salt and crowding conditions.

4.4 Discussion

Flipping is an important transition, allowing proteins to rapidly find specific targets on nucleic acid substrates. We considered two potential mechanistic models of flipping, "tumbling" and "hopping". A key difference between our models is that k_f competes with k_{off} directly from the bound state in the tumbling model, while hopping requires dissociation into a pseudo-bound state that allows the RT to rebind or fully dissociate (Figure 4.1a).

We describe both the tumbling and hopping models using Markov chains (Figure 4.1a). Each model has exactly three free parameters that can describe the flipping behavior. These free parameters can be related to the experimentally measured parameters (the rate constants of binding, dissociation, and flipping) (Table 4.1). In the case of tumbling, each rate constant from the model corresponds directly to a measured rate constant. In the hopping model, we assumed the pseudo-bound state was too short lived to contribute directly to bound lifetimes. Therefore, the behavior was entirely described by the two rate constants of entering into the pseudo-bound state and the ratio of the rates exiting the pseudo-bound state, i.e. the branching ratio r. The relationships between these three free parameters and the measured rate constants were derived (Supplementary materials and Table 4.1). Based on the derived rate constants, we calculated the associated free energy diagrams for each model at both 50 mM and 200 mM NaCl with 0% PEG 8K (Figure 4.6, orange and black lines). A similar set of diagrams was generated for 150 mM NaCl at 0% and 10% PEG 8K (Supplementary Figure 4.6, black and orange lines). In the hopping model, we have used the same energy well to represent both the bound state and the flipped state, since these states are roughly symmetric on our DNA substrate. We aligned these free energy diagrams so that the energy of the bound state was the same under each buffer condition.

Table 4.1**:** Equations describing probability of flipping and rate of flipping for tumbling and hopping model

	Tumbling model	*Hopping model*
Rate constant of dissociation (k_{off})	$k_{B\to D}$	$k_{B\to P}\left(\frac{1+r}{2+r}\right)$
Rate constant of flipping (k_f)	$k_{B\to F}$	$\left(\frac{k_{B\to P}}{2+r}\right)$
Rate constant of binding rate (k_{on})	$k_{D\to B}$	$\left(\frac{k_{D\to P}}{2+r}\right)$

r is the branching ratio defined by: $r=\frac{k_{P\to D}}{k_{P\to B}}$

If crowding and salt concentration only affected the stability of the bound states, then in both models we would expect k_{off} to scale with k_f and P_{flip} to remain constant, as was observed when cognate nucleotide was added to solution. However, if crowding and salt concentration also influence the stability of the transition states, then the P_{flip} can change as well. Our data reveal that the P_{flip} did not remain constant in these two cases. Instead, decreased salt concentration or increased crowding increased the P_{flip} (Figure 4.5a and c). This shows that the transition states are affected by the changes in the buffer conditions, and the manner in which P_{flip} changes supports the hopping model.

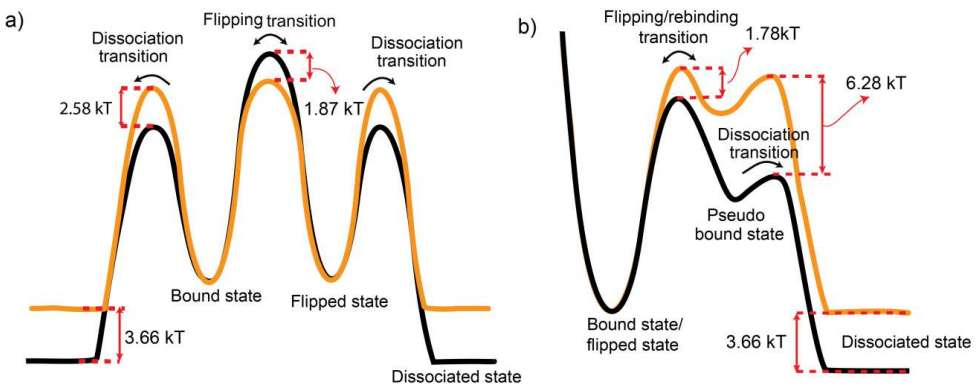


Figure 4.6**: Free energy diagrams of RT flipping kinetics on DNA.** The free energy diagrams were drawn for tumbling model and hopping model based on the kinetic and thermodynamic data obtained under different NaCl concentrations. **a)** Free energy diagram corresponding to the tumbling model of RT binding and flipping kinetics under high salt concentration (black). Decreased salt concentration leads to increase in the free energy

barrier height for dissociation transition and decrease in the barrier height for flipping transition relative to the bound or flipped bound state (orange). **b)** free energy diagram explaining flipping kinetics based on hopping model with an extra pseudo-bound state. In this free energy diagram both the bound state and flipped state are identical. Color convention is the same as in the free energy diagram for tumbling model.

First, we discuss why the tumbling model has difficulties explaining the data. In the tumbling model, k_f is expected to increase with ionic concentration because the tumbling transition state should break some of the ionic contacts that are present between RT and DNA in the bound state (40,41). Instead, we observed that k_f *decreased* as the salt concentration increased. k_f was also expected to decrease with increased crowding because the tumbling transition state should become less compact compared to the bound state. Instead, we observed that k_f *increased* as the volume fraction of PEG 8K increased. If we use a tumbling model to fit our experimental data, we therefore have to conclude that the flipping transition is more destabilized by increased salt concentration relative to the bound state (Figure 4.6a). Conversely, the flipping transition would have to be stabilized by increased crowding (Supplementary Figure 4.7a). This would suggest an unphysical transition state that is both more compact and has more electrostatic contacts between the RT and DNA than the bound state.

Next, we discuss why the hopping model is consistent with our experimental observations. The hopping model assumes that the macromolecular complexes undergo many short-lived diffusive excursions and re-bindings before a complete dissociation can occur. These short-lived excursions are kinetically equivalent to a pathway where RT must move from the bound state to a pseudo-bound state before dissociating (26) (Figure 4.1a). When RT enters the short-lived pseudo-bound state, it has a chance to dissociate or to rebind in either the original orientation or a flipped orientation. For simplicity and because our DNA template is nearly symmetric, we assume there is an equal chance of rebinding in flipped state or in the original orientation. Decreasing salt concentration or increasing crowding should create more long-range attractive forces between RT and DNA (42,43). We expect that these long-range attractive forces should increase the chance of rebinding relative to the chance of dissociation from the pseudo-bound state. To provide support for this assumption, we numerically simulated the Brownian motion of two spherical particles in close proximity. In the presence of an attractive force extending for ~2 nm from each particle, both the number and frequency of diffusive excursions taking the particles more than 7 nm apart (the diameter of RT) before they came into contact again increased (Supplementary Figure 4.7).

Therefore, we expect P_{flip} and k_f to increase as the salt concentration is decreased or crowding is increased, as we observe in the data. Fitting our experimental data to the hopping model, we in fact observe that at lower salt concentrations the rebinding/flipping transition is destabilized by a smaller amount than the dissociation transition (Figure 4.6b). A similar result is derived from the crowding data (Supplementary Figure 4.6b). Therefore, we conclude that flipping proceeds by hopping rather than tumbling.

In addition to this primary conclusion, our research demonstrates that macromolecular crowding plays a major role in determining how strongly RT binds nucleic acids. *In vitro* experiments on RT, both in bulk and at single-molecule level, are frequently performed in dilute solutions with low salt concentration (32,44). This contrasts with the crowded and salty milieu inside cells and viruses, with around 200-400 mg/ml of different macromolecules and effective salt concentrations of ~150 mM (45-49). In line with previous findings on crowding-induced effects on the stability of DNA binding proteins (19,22), we find that increased crowding at salt concentrations relevant to those found *in vivo* can result in comparable binding affinities to those achieved at low salt concentrations under dilute conditions (Figure 4.2 and Figure 4.3). In fact, fitting K_d to scaled-particle theory suggests that RT binds DNA with a sub-nanomolar K_d at physiological crowding conditions. We therefore emphasize that crowding is an important variable to control when trying to predict *in vivo* rates from *in vitro* data, particularly when predicting the effects of antiretroviral drugs on the efficiency of reverse transcription.

Our results have implication for other macromolecular complexes as well. The fact that flipping of RT can be modeled through hopping and rebinding is consistent with the recent observation that macromolecular DNA complexes rebind after entering a pseudo-bound state, allowing the bound lifetime to be shortened by competitive inhibition (26). Indeed, any physical model of the dissociation of two diffusing macromolecules would imply that some rebinding should occur with a finite probability, since the free molecules undergo a random walk. Our observations provide empirical evidence that these short excursions can allow flipping to occur before the molecules rebind.

The highly crowded environment inside the cell provides long-range attractive interactions between many binding macromolecules similar to those provided by our *in vitro* crowding agents, increasing the probability of rebinding and flipping. This increase means that for any encounter between two macromolecules, a wider range of configurations can be explored through rebinding before complete dissociation

occurs. This exploration can make the assembly of macromolecular complexes kinetically more efficient (50,51). Rebinding and flipping also aids enzymes that perform more than one reaction per encounter, such as DNA polymerases that contain an exonuclease site (52), type II restriction enzymes (4-6) or ERK phosphorylation by MEK (53,54). We have demonstrated that smFRET combined with kinetic analysis provides a useful tool for exploring the role rebinding and flipping play in all of these systems.

4.5 References

1. Berg, O.G., Winter, R.B. and Von Hippel, P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929-6948.

2. Berg, O.G. and von Hippel, P.H. (1985) *Annu. Rev. Biophys. Biophys. Chem.*, **14**, 131.

3. Halford, S.E. and Marko, J.F. (2004) How do site‐specific DNA‐binding proteins find their targets? *Nucleic Acids Research*, **32**, 3040-3052.

4. Sasnauskas, G., Kostiuk, G., Tamulaitis, G. and Siksnys, V. (2011) Target site cleavage by the monomeric restriction enzyme BcnI requires translocation to a random DNA sequence and a switch in enzyme orientation. *Nucleic Acids Research*, **39**, 8844-8856.

5. Sasnauskas, G., Zakrys, L., Zaremba, M., Cosstick, R., Gaynor, J.W., Halford, S.E. and Siksnys, V. (2010) A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3′-5′ and 5′-3′ strands by rotating a single active site. *Nucleic Acids Research*, **38**, 2399-2410.

6. Pingoud, A., Wilson, G.G. and Wende, W. (2014) Type II restriction endonucleases—a historical perspective and more. *Nucleic Acids Research*, **42**, 7489-7527.

7. Ryu, K.-S., Tugarinov, V. and Clore, G.M. (2014) Probing the Rate-Limiting Step for Intramolecular Transfer of a Transcription Factor between Specific Sites on the Same DNA Molecule by 15Nz-Exchange NMR Spectroscopy. *Journal of the American Chemical Society*, **136**, 14369-14372.

8. Abbondanzieri, E.A., Bokinsky, G., Rausch, J.W., Zhang, J.X., Le Grice, S.F.J. and Zhuang, X. (2008) Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature*, **453**, 184-189.

9. Liu, S., Abbondanzieri, E.A., Rausch, J.W., Grice, S.F.J.L. and Zhuang, X. (2008) Slide into Action: Dynamic Shuttling of HIV Reverse Transcriptase on Nucleic Acid Substrates. *Science*, **322**, 1092-1097.

10. Liu, S., Harada, B.T., Miller, J.T., Le Grice, S.F.J. and Zhuang, X. (2010) Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription. *Nat Struct Mol Biol*, **17**, 1453-1460.

11. Sharma, K.K., Przybilla, F., Restle, T., Godet, J. and Mély, Y. (2016) FRET-based assay to screen inhibitors of HIV-1 reverse transcriptase and nucleocapsid protein. *Nucleic Acids Research*.

12. Sharma, K.K., Przybilla, F., Restle, T., Boudier, C., Godet, J. and Mély, Y. (2015) Reverse Transcriptase in Action: FRET-Based Assay for Monitoring Flipping and Polymerase Activity in Real Time. *Analytical Chemistry*, **87**, 7690-7697.

13. Blainey, P.C., Luo, G., Kou, S.C., Mangel, W.F., Verdine, G.L., Bagchi, B. and Xie, X.S. (2009) Nonspecifically bound proteins spin while diffusing along DNA. *Nat Struct Mol Biol*, **16**, 1224-1229.

14. Zheng, Q., Juette, M.F., Jockusch, S., Wasserman, M.R., Zhou, Z., Altman, R.B. and Blanchard, S.C. (2014) Ultra-stable organic fluorophores for single-molecule research. *Chemical Society Reviews*, **43**, 1044-1056.

15. Kong, X., Nir, E., Hamadani, K. and Weiss, S. (2007) Photobleaching Pathways in Single-Molecule FRET Experiments. *Journal of the American Chemical Society*, **129**, 4643-4654.

16. Datta, K. and LiCata, V.J. (2003) Salt Dependence of DNA binding by Thermus aquaticusand Escherichia coli DNA Polymerases. *Journal of Biological Chemistry*, **278**, 5694-5701.

17. Akabayov, B., Akabayov, S.R., Lee, S.-J., Tabor, S., Kulczyk, A.W. and Richardson, C.C. (2010) Conformational dynamics of bacteriophage T7 DNA polymerase and its processivity factor, Escherichia coli thioredoxin. *Proceedings of the National Academy of Sciences*, **107**, 15033-15038.

18. Minton, A.P. (1981) Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers*, **20**, 2093-2120.

19. Zimmerman, S.B. and Harrison, B. (1987) Macromolecular Crowding Increases Binding of DNA Polymerase to DNA: An Adaptive Effect. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 1871-1875.

20. Ellis, R.J. (2001) Macromolecular crowding: obvious but underappreciated. *Trends in Biochemical Sciences*, **26**, 597-604.

21. Zhou, H.-X., Rivas, G. and Minton, A.P. (2008) Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences. *Annual Review of Biophysics*, **37**, 24.

22. Akabayov, B., Akabayov, S.R., Lee, S.-J., Wagner, G. and Richardson, C.C. (2013) Impact of macromolecular crowding on DNA replication. *Nat Commun*, **4**, 1615.

23. Kao-Huang, Y., Revzin, A., Butler, A.P., O'Conner, P., Noble, D. and Von Hippel, P.H. (1977) Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: Measurement of DNA-bound Escherichia coli lac repressor in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 5.

24. Blainey, P.C., van Oijen, A.M., Banerjee, A., Verdine, G.L. and Xie, X.S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proceedings of the National Academy of Sciences*, **103**, 5752-5757.

25. Farge, G., Laurens, N., Broekmans, O.D., van den Wildenberg, S.M.J.L., Dekker, L.C.M., Gaspari, M., Gustafsson, C.M., Peterman, E.J.G., Falkenberg, M. and Wuite, G.J.L. (2012) Protein sliding and DNA denaturation are essential for DNA organization by human mitochondrial transcription factor A. *Nat Commun*, **3**, 1013.

26. Paramanathan, T., Reeves, D., Friedman, L.J., Kondev, J. and Gelles, J. (2014) A general mechanism for competitor-induced dissociation of molecular complexes. *Nat Commun*, **5**.

27. Bonnet, I., Biebricher, A., Porté, P.-L., Loverdo, C., Bénichou, O., Voituriez, R., Escudé, C., Wende, W., Pingoud, A. and Desbiolles, P. (2008) Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Research*, **36**, 4118-4127.

28. Gowers, D.M., Wilson, G.G. and Halford, S.E. (2005) Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 15883-15888.

29. Grice, S.F.J.L., Cameron, C.E. and Benkovic, S.J. (1995) Purification and characterization of human immunodeficiency virus type 1 reverse transcriptase. *Methods in Enzymol*, **262**, 5.

30. Schatz, O., Cromme, F.V., Grüninger-Leitch, F. and Grice, S.F.J.L. (1989) Point mutations in conserved amino acid residues within the C-terminal domain of HIV-1 reverse transcriptase specifically repress RNase H function. *FEBS Lett.*, **257**, 4.

31. Rausch, J.W., Sathyanarayana, B. K., B., M. K. and Grice, S.F.J.L. (2000) Probing Contacts between the Ribonuclease H Domain of HIV-1 Reverse Transcriptase and Nucleic Acid by Site-specific Photocross-linking. *J. Biol. Chem.*, **275**, 8.

32. Marko, R.A., Liu, H.-W., Ablenas, C.J., Ehteshami, M., Götte, M. and Cosa, G. (2013) Binding Kinetics and Affinities of Heterodimeric versus Homodimeric HIV-1 Reverse Transcriptase on DNA–DNA Substrates at the Single-Molecule Level. *The Journal of Physical Chemistry B*, **117**, 4560-4567.

33. Chandradoss, S.D., Haagsma, A.C., Lee, Y.K., Hwang, J.H., Nam, J.M. and Joo, C. (2014) Surface Passivation for Single-molecule Protein Studies. *J. Vis. Exp.*

34. Wyman Jr, J. (1964) In C.B. Anfinsen, M. L. A. J. T. E. and Frederic, M. R. (eds.), *Advances in Protein Chemistry*. Academic Press, Vol. Volume 19, pp. 223-286.

35. Record Jr, M.T., Ha, J.-H. and Fisher, M.A. (1991) In Robert, T. S. (ed.), *Methods in Enzymology*. Academic Press, Vol. Volume 208, pp. 291-343.

36. Dupuis, N.F., Holmstrom, E.D. and Nesbitt, D.J. (2014) Molecular-crowding effects on single-molecule RNA folding/unfolding thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 8464-8469.

37. Thiyagarajan, P., Chaiko, D.J. and Hjelm, R.P. (1995) A Neutron Scattering Study of Poly(ethylene glycol) in Electrolyte Solutions. *Macromolecules*, **28**, 7730-7736.

38. Reiss, H., Frisch, H.L. and Lebowitz, J.L. (1959) Statistical Mechanics of Rigid Spheres. *The Journal of Chemical Physics*, **31**, 369-380.

39. Lebowitz, J.L., Helfand, E. and Praestgaard, E. (1965) Scaled Particle Theory of Fluid Mixtures. *The Journal of Chemical Physics*, **43**, 774-779.

40. Ding, J., Das, K., Hsiou, Y., Sarafianos, S.G., Clark Jr, A.D., Jacobo-Molina, A., Tantillo, C., Hughes, S.H. and Arnold, E. (1998) Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody fab fragment at 2.8 Å resolution1. *Journal of Molecular Biology*, **284**, 1095-1111.

41. Sarafianos, S.G., Das, K., Tantillo, C., Clark, A.D., Ding, J., Whitcomb, J.M., Boyer, P.L., Hughes, S.H. and Arnold, E. (2001) Crystal structure of HIV‐1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *The EMBO Journal*, **20**, 1449-1461.

42. Vologodskii, A. and Cozzarelli, N. (1995) Modeling of long-range electrostatic interactions in DNA. *Biopolymers*, **35**, 289-296.

43. Yodh, A.G., Lin, K., Crocker, J.C., Dinsmore, A.D., Verma, R. and Kaplan, P.D. (2001) Entropically driven self–assembly and interaction in suspension. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **359**, 921-937.

44. Schauer, G.D., Huber, K.D., Leuba, S.H. and Sluis-Cremer, N. (2014) Mechanism of allosteric inhibition of HIV-1 reverse transcriptase revealed by single-molecule and ensemble fluorescence. *Nucleic Acids Research*, **42**, 11687-11696.

45. Fulton, A.B. (1982) How crowded is the cytoplasm? *Cell*, **30**, 3.

46. Spitzer, J. (2011) From Water and Ions to Crowded Biomacromolecules: In Vivo Structuring of a Prokaryotic Cell. *Microbiology and Molecular Biology Reviews*, **75**, 491-506.

47. del Álamo, M., Rivas, G. and Mateu, M.G. (2005) Effect of Macromolecular Crowding Agents on Human Immunodeficiency Virus Type 1 Capsid Protein Assembly In Vitro. *Journal of Virology*, **79**, 14271-14281.

48. Briggs, J.A.G., Simon, M.N., Gross, I., Krausslich, H.-G., Fuller, S.D., Vogt, V.M. and Johnson, M.C. (2004) The stoichiometry of Gag protein in HIV-1. *Nat Struct Mol Biol*, **11**, 672-675.

49. Benjamin, J., Ganser-Pornillos, B.K., Tivol, W.F., Sundquist, W.I. and Jensen, G.J. (2005) Three-dimensional Structure of HIV-1 Virus-like Particles by Electron Cryotomography. *Journal of Molecular Biology*, **346**, 577-588.

50. Fange, D., Berg, O.G., Sjöberg, P. and Elf, J. (2010) Stochastic reaction-diffusion kinetics in the microscopic limit. *Proceedings of the National Academy of Sciences*, **107**, 19820-19825.

51. Lomholt, M.A., Zaid, I.M. and Metzler, R. (2007) Subdiffusion and Weak Ergodicity Breaking in the Presence of a Reactive Boundary. *Physical Review Letters*, **98**, 200603.

52. Wuite, G.J.L., Smith, S.B., Young, M., Keller, D. and Bustamante, C. (2000) Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature*, **404**, 103-106.

53. Takahashi, K., Tănase-Nicola, S. and ten Wolde, P.R. (2010) Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proceedings of the National Academy of Sciences*, **107**, 2473-2478.

54. Aoki, K., Takahashi, K., Kaizu, K. and Matsuda, M. (2013) A Quantitative Model of ERK MAP Kinase Phosphorylation in Crowded Media. *Sci. Rep.*, **3**.

Supplementary Materials

Sequences of DNA molecules used in the single-molecule FRET

Double primer double-stranded DNA (dpdsDNA) sequences:

Primer 1: 5′-/CGG TGC CAC GCC ACT TAC TGT CTC GAT CAC TAG T-3′

Primer 2: 5′-/5Biotin/ATT AGC CCT TCC AGT ACT AGT GAT CGA GAC AGT A -3′ Double primer double-stranded DNA with a chain terminating 3′-end sequences:

Primer 1: 5′-/CGG TGC CAC GCC ACT TAC TGT CTC GAT CAC TAG TAC-3′

Primer 2: 5′-/5Biotin/ATT AGC CCT TCC AGT ACT AGT GAT CGA GAC AGT A -3′

Amino functionalized cytosine base (indicated with green color in primer 1) was labelled with Cy5 fluorophore and biotin label on 5′-end of primer 2 was used for immobilization od duplex DNA on streptavidin coated surface. Complementary bases for DNA hybridization are identified with red color.

Total internal reflection fluorescence microscopy

A schematic illustration of the home built prism based TIRF microscopy set-up is shown in Supplementary Figure 4.8. We combine two colour lasers (532 nm Cobolt Samba, and 642 Cobolt MLD) with a dichroic mirror (FF635-Di01-Semrock). The combined beam is put into an AOTF (AA optoelectronics) to allow fast alternating laser excitation (ALEX) between the two colours (1). This is required to track two labels (Cy3-labelled RT and Cy5-labelled DNA). The beam is then focused into a prism at a nearly normal incidence, making it easier to keep both coloured beams co-localized at the image plane. The total internal reflection occurs at the quartz-water interface, allowing TIRF illumination of our sample. With a 60x water immersion objective (Olympus UPLSAPO, 1.2 NA), we image the sample onto a narrow slit (Thorlabs) to remove light originating from outside a narrow band in the image plane. We then split the path into the two emission colours. Having the slit and separate paths for each colour allow us to display and record both colours next to each other on an EMCCD (Andor Ixon 897) simultaneously. The microscope is operated by a custom LabVIEW routine. The timing of ALEX and EMCCD is done through LabVIEW NIDAQ-mx, using a PCIe-6320 card and a BNC-2120 breakout box (National Instruments). In a typical experiment we record around 2000-2500 frames with a 100 msec frame rate. The first frame is obtained under the excitation wavelength of 640 nm, followed by 530 nm illumination for the next 20 frames. Alternating between the green and red illumination makes it possible to localize all DNA substrates on the surface and allows us to identify any photo-bleaching of the Cy5 (1). ImageJ (http://imagej.nih.gov/ij/) was employed to identify all the Cy5

spots corresponding to DNA. The data was further treated with a custom written Matlab (Mathworks Inc.) routine to filter the background and for further analysis of the data. A mapping function between the green and red channels was obtained by the repeated imaging and translation of a sparse field of Tetraspeck beads (Invitrogen), identifying the paired coordinates of the red and green spots, and fitting the data to a fifth-order polynomial function. Mapping was needed for co-localizing Cy5-labelled DNA spots in red channel with the corresponding Cy3-labelled RT fluorescence in green channel. Regions of interest (5x5 pixels) centered on the Cy5 spots were cropped and the corresponding fluorescent intensities were plotted in time. The dwell times and FRET values were derived from this data. Development of the code was inspired by previous works (2,3).

Activity assay for Cy3-labelled RT

We performed a primer extension assay to confirm the activity of RT after labelling with Cy3. The reaction buffer is composed of 50 mM Tris-HCl, pH 7.8, 50 mM KCl, 8 mM $MgCl_2$, 1 mM DTT 0.2 mM EDTA. Cy5-labelled, 21 nucleotide primer (20 nM) was annealed to 50 nucleotide template. The primer/template was then pre-incubated with Cy3-labelled RT (1 µM) in the reaction buffer for 10 minutes before adding dNTP mixture to initiate the polymerization. The primer extension was quenched at various time points by adding a mixture of 500 mM EDTA and 90% (v/v) formamide. The products were heated at 90 $^{\circ}$C for 5 minutes and were separated on a 10% urea-polyacrylamide (Bio-Rad) gel. Intensities of fluorescence bands from extended primers were measured with a Typhoon scanner (GE healthcare). The intensity of bands were analyzed using 1D gel image analysis. The labelled enzyme showed similar activity to the unlabelled RT (data not shown). This agrees with previous results that show Cy3-labelled RT to have primer extension activity comparable to its unlabelled counterpart (4).

Single turnover primer-extension assay to study the effects of increased salt and crowding concentrations

Effects of increased ionic concentration and macromolecular crowding were measured by primer extension assay. The assay is same as the activity assay described above except that it also included a heparin trap. The heparin was added to the reaction along with the dNTP mixture to ensure a single run of nucleotide incorporation followed by dissociation (5). In order to estimate the initial bound DNA fraction we quenched the reaction at three time points (30 sec, 2 min, and 5 min). These samples were separated on a urea-PAGE gel as described earlier (Supplementary Figure 4.2a). The amount of primer extension was measured at each

time point and extrapolated to Y-axis. The Y-intercept was used to estimate the initial bound fraction of DNA. The assay was repeated under varying salt concentrations (50-300 mM) and different crowding agents (PEG 8K and Ficoll 70K from 0-10% (w/v)).

A representative gel shown in supplementary figure 3a was used to examine the effects of increased macromolecular crowding at 150 mM NaCl. Increased crowding not only increased the initial bound fraction of DNA, but also increased the processivity of RT (Supplementary Figure 4.3b). Under increased salt concentrations, RT was found to interact weakly with DNA (Supplementary Figure 4.2c). At physiologically relevant salt concentrations (150 mM of NaCl), the bound fraction of DNA increased by more than 50% as the crowding fraction increased to 10% (Supplementary Figure 4.2d).

Analysis of flipping kinetics, correcting for the photo-bleaching of fluorophores and derivation of kinetic parameters for tumbling and hopping models

Binding of Cy3-labelled RT on dpdsDNA resulted in an increase of fluorescence above background and FRET values of either 0.3 or 1 for low FRET and high FRET orientations (Figure 4.1d). If a binding transition occurs between an unbound state and a bound state with a FRET value of below 0.65, we mark the subsequent frames as a low FRET event, while if the FRET is above 0.65 the frames are marked as a high FRET event. A history-dependent threshold was applied to identify flipping events, using a FRET cut-off of 0.6 to identify transitions from high-FRET to low-FRET events and a cut-off of 0.7 to identify transitions from low-FRET to high-FRET events. Additionally, we only counted flipping transitions between states that lasted for longer than two frames each, to avoid false positives (Supplementary Figure 4.9). In this way, we could obtain all the binding dwell times corresponding to low FRET binding events and high FRET binding events, as well as the total number of flipping transitions. We calculated the binding rate, dissociation rate, and the flipping rate based on the total number of binding events, total number of flipping events, and the total observation time of the DNA molecules.

The photo-bleaching rate of Cy3 and Cy5 was calculated in a separate experiment. RT-Cy3 or DNA-Cy5 were separately immobilized on a quartz slide via anti-his-antibody or biotin-streptavidin, respectively. The immobilized molecules were continuously recorded under the excitation of their respective laser-lines and recorded at 10 frames per second. The amount of time that individual Cy3 or Cy5 molecules fluoresced before entering into a permanent dark state was measured. This data was fit with an exponential decay function to obtain the characteristic time

constant of photo-bleaching k_{bleach} for both the Cy3-labelled RT and Cy5-labelled DNA (Supplementary Figure 4.10). The bleaching rate of Cy5-DNA ($k_{bleach,Cy5}$) was found to be 0.0093 sec^{-1} while the bleaching rate of Cy3-RT ($k_{bleach,Cy3}$) was found to be 0.027 sec^{-1}. The higher bleaching rate of Cy3 was partly caused by the local environment near the protein, since Cy3 attached to DNA bleached more slowly (data not shown). The measured bleaching rate of Cy3 on RT was used to correct the dissociation rate (k_{off}), binding rate (k_{on}), rate of flipping (k_{flip}) and probability of flipping (P_{flip}) as described by the equations below. Here we use an asterisk (*) to indicate parameters that have been corrected for the effects of photo-bleaching.

$$k_{off,low} = \frac{N_{b_l} - N_{f_l}}{T_{bound,low}}$$

$$k^*_{off,low} = k_{off,low} - (k_{bleach,Cy3} \cdot (1 - E_l))$$

$$k_{off,high} = \frac{N_{b_h} - N_{f_h}}{T_{bound,high}}$$

$$k^*_{off,high} = k_{off,high} - (k_{bleach,Cy3} \cdot (1 - E_l) + k_{bleach,Cy5} \cdot E_h)$$

$$k^*_{off} = \left(\frac{1}{2}\right)(k^*_{off,high} + k^*_{off,high}) \qquad (1)$$

Note that we assume that the bleaching rate of Cy3 scales as $(1 - E_{FRET})$, since high FRET will lower the chance of entering the triplet state, while the bleaching rate of Cy5 scales with the FRET efficiency, since this amounts to a modulation of the excitation intensity of Cy5. Here, $E_l = 0.3$ and $E_h = 0.98$, the average low FRET and high FRET values observed on dpdsDNA. Because $k_{bleach,Cy3} \cdot (1 - E_l) << k_{bleach,Cy5} \cdot E_l$, we did not include the effects of Cy5 photo-bleaching in the low FRET state.

$$k^*_{on,low} = \frac{N_{b_l}}{\left(T_{tot} - \frac{N_{b_l}}{k^*_{off,low}} - \frac{N_{b_h}}{k^*_{off,high}}\right)[RT]}$$

$$k^*_{on,high} = \frac{N_{b_h}}{\left(T_{tot} - \frac{N_{b_l}}{k^*_{off,low}} - \frac{N_{b_h}}{k^*_{off,high}}\right)[RT]}$$

$$k^*_{on} = \left(\frac{1}{2}\right)(\mathrm{k}^*_{on,low} + \mathrm{k}^*_{on,high}) \qquad (2)$$

$$k^*_{flip} = \left(\frac{1}{2}\right)(\mathrm{N}_{f_l} \cdot \mathrm{k}^*_{off,low} + N_{f_h} \cdot \mathrm{k}^*_{off,high})$$

$$\text{and } P^*_{flip} = \left(\frac{1}{2}\right)\left(\frac{N_{f_l}}{N_{b_l}} + \frac{N_{f_h}}{N_{b_h}}\right)$$

where,

N_{b_l} -total number of low FRET binding events observed

N_{b_h} - total number of high FRET binding events observed

N_{f_l} - total number of observed low-to-high FRET flipping events

N_{f_h} - total number of observed high-to-low FRET flipping events

T_{tot} - total observation time of all the molecules

$T_{bound,low}$ - sum of the binding dwell times of all the low FRET events

$T_{bound,high}$ - sum of the binding dwell times of all the high FRET events

$[RT]$ - *concentration of reverse transcriptase*

An examples FRET trace is shown in Supplementary Figure 4.9 in which all of the above parameters are defined.

These measured rates can then be used to derive the rates in each model. The tumbling model contains 6 rate constants: $k_{D\to F}, k_{D\to B}, k_{F\to B}, k_{B\to F}, k_{F\to D}$ and $k_{B\to D}$ (see Figure 4.1a). Because we assume the bound state and the flipped state to be symmetric, we conclude that $k_{D\to F} = k_{D\to B}, k_{F\to B} = k_{B\to F},$ and $k_{F\to D} = k_{B\to D}$. Therefore, the tumbling model can be completely described using three kinetic free parameters, $k_{D\to B}, k_{B\to F}$ and $k_{B\to D}$. These parameters have a direct correlation to the measured rates, namely:

$$k_{off}^{*} = k_{B\to D}$$

$$k_{on}^{*} = k_{D\to B}$$

$$k_{flip}^{*} = k_{B\to F}$$

$$K_d = \frac{k_{B\to D}}{k_{D\to B}}$$

The hopping model also contains 6 rate constants: $k_{D\to P}, k_{P\to D}, k_{P\to B}, k_{B\to P}, k_{P\to F}$ and $k_{F\to P}$. Again because of the symmetry of the flipped state and the bound state, we assume that $k_{P\to F} = k_{P\to B}$, and $k_{F\to P} = k_{B\to P}$. We further assume that the pseudo-bound state is too short lived to contribute to the bound lifetimes. This is supported by the observation that flipping appears instantaneous at 100 msec frame rates with no intermediate FRET states observed during the transition. Therefore we can only measure the ratio $r = \frac{k_{P\to D}}{k_{P\to B}}$ rather than the absolute values of the rates out of the pseudo-bound state. Because of these two assumptions, we can describe the hopping model with three kinetic free parameters, namely $k_{D\to P}, k_{B\to P}$ and $r = \frac{k_{P\to D}}{k_{P\to B}}$. We can relate these free parameters to the measured rate constants by first defining the probabilities of flipping, rebinding, or dissociating from the pseudo-bound state:

$$P_{flip}^{pseudo} = P_{bind}^{pseudo} = \frac{k_{P\to F}}{k_{P\to F} + k_{P\to B} + k_{P\to D}} = \frac{1}{2+r}$$

$$P_{dissoc.}^{pseudo} = \frac{k_{P\to D}}{k_{P\to F} + k_{P\to B} + k_{P\to D}} = \frac{r}{2+r}$$

The measured rate constants are therefore related to the hopping model parameters by:

$$k_{off}^{*} = k_{B\to P} P_{dissoc.}^{pseudo} = k_{B\to P} \frac{r}{2+r}$$

$$k_{on}^{*} = k_{D\to P} P_{bind}^{pseudo} = k_{D\to P} \frac{1}{2+r}$$

$$k_{flip}^{*} = k_{B\to P} P_{flip}^{pseudo} = k_{B\to P} \frac{1}{2+r}$$

$$K_d = \frac{k_{B\to P}}{k_{P\to B}} \frac{k_{P\to D}}{k_{D\to P}} = \frac{k_{B\to P}}{k_{D\to P}} \cdot r$$

We also can relate the measured probability of flipping, P_{flip}^{*}, to r by considering the conditional probability that the enzyme flipped given that it did not revert back to the previous bound state:

$$P_{flip}^{*} = \frac{P_{flip}^{pseudo}}{1 - P_{bind}^{pseudo}} = \frac{1}{1+r}$$

Using this relationship, we can write the kinetic free parameters of the hopping model as functions of the measured rate constants:

$$r = \frac{k_{off}^{*}}{k_{flip}^{*}}$$

$$k_{B\to P} = k_{flip}^{*} \left(2 + \frac{k_{off}^{*}}{k_{flip}^{*}} \right)$$

$$k_{D\to P} = k_{on}^{*}\left(2+\frac{k_{off}^{*}}{k_{flip}^{*}}\right)$$

Using these rate constants, we can now calculate the free energy differences in each model (Figure 4.5) and identify how buffer conditions affect these free energies. For the tumbling model, the free energy of binding (ΔG_B) , the height of the dissociation barrier $(\Delta G_{B\to D}^{\ddagger})$, and the height of the flipping barrier $(\Delta G_{B\to F}^{\ddagger})$ shift by the following amounts as the salt concentration increases from 50 to 200 mM:

$$\Delta\Delta G_B^{\ddagger} = kT\cdot\log\left(\frac{K_d^{50}}{K_d^{200}}\right)$$

$$\Delta\Delta G_{B\to D}^{\ddagger} = kT\cdot\log\left(\frac{k_{on}^{50}}{k_{on}^{200}}\right)$$

$$\Delta\Delta G_{B\to F}^{\ddagger} = kT\cdot\log\left(\frac{k_{flip}^{50}}{k_{flip}^{200}}\right)$$

In the case of hopping model, we can specify the free energy of binding (ΔG_B), the height of the barrier from the bound state into the pseudo-bound state $(\Delta G_{B\to P}^{\ddagger})$, and the height of the barrier from the pseudo-bound state to the dissociated state $(\Delta G_{P\to D}^{\ddagger})$,

$$\Delta\Delta G_B^{\ddagger} = kT\cdot\log\left(\frac{r_{50}}{r_{200}}\frac{k_{B\to P}^{50}}{k_{B\to P}^{200}}\frac{k_{D\to P}^{200}}{k_{D\to P}^{50}}\right)$$

$$\Delta\Delta G_{B\to P}^{\ddagger} = kT\cdot\log\left(\frac{k_{B\to P}^{50}}{k_{B\to P}^{200}}\right)$$

$$\Delta\Delta G_{P\to D}^{\ddagger} = kT\cdot\log\left(\frac{r_{50}}{r_{200}}\right)+\Delta\Delta G_{B\to P}^{\ddagger}$$

The same equations are applied to calculate free energy diagram for crowding data (Supplementary Figure 4.7)

Fitting the dissociation constant of RT/DNA complex under macromolecular crowding to scaled-particle theory

Scaled particle theory was originally developed by Lebowitz and co-workers to estimate the free energy of cavity formation in a given liquid solution (6,7), and later extended to understand the effects of macromolecular crowding on the equilibrium of binding of non-interacting hard spheres (8). Recently, this theory was applied to explain experimentally observed macromolecular crowding effects (9,10). We fit our experimental data (Figure 4.3b and Supplementary Figure 4.5) using scaled particle theory (8). For this fit we modelled the DNA as a cylinder of radius 1 nm and length of 6.33 nm (19 bp) and RT is a sphere of radius 3.5 nm. The bound complex of RT-DNA was modelled as a sphere whose volume is equal to the sum of the volumes of the DNA cylinder and RT sphere. The PEG 8K was then modelled as a sphere of radius 1.5 nm based on neutron scattering results (11). Under these assumptions, the model closely predicts the observed shift in the dissociation constant without any additional free parameters (Figure 4.3b and Supplementary Figure 4.5). Because cellular crowding fractions are typically around 20-30 %(w/v), we extended the range of the model to >20 % (w/v) PEG 8K. It is evident from this curve (Figure 4.3b) that RT is predicted to bind DNA with a sub-nanomolar affinity under cellular and viral crowding conditions.

Brownian dynamics simulations

We performed Brownian dynamics simulations to understand the influence of attractive forces on the rate of rebinding and/or flipping between two spherical particles of radii 1.5 nm (DNA) and 3.5 nm (RT). The simulation starts at the bound state where the two particles are in contact, i.e. separated by sum of the radii. During the simulation we allowed the particles to diffuse by Brownian motion under the influence of an exponentially decaying attractive force between them. The distance between centres of two particles was tracked. If the surfaces of the molecules move away from each other by more than 7 nm (the diameter of RT) before rebinding, then rotational diffusion should allow the spheres to potentially reorient ($< \theta^2 > = \sim 180°$). We therefore counted these events as rebinding that could potentially lead to flipping. The total amount of time the particles spent before drifting apart by more than 12 nm without rebinding was used to calculate the bound time. The simulations were performed for 400 particles each for 100,000 time steps with a fixed maximum

force and different screening distances, which simulates different salt concentrations in our experiments. Simulations were also performed with a fixed distance parameter and different maximum attraction forces to reflect the effects of macromolecular crowing. Using the total bound time and number of rebinding events that could lead to flipping, we then calculated an average rate of rebinding. The results showed that the rate of rebind increases either with the increased attraction distance and force (Supplementary Figure 4.7). Hence, these simulations confirm that the increased long-range attractive forces arising from either decreased salt concentration or increased macromolecular crowding enhance the rate of rebinding.

REFERENCES

1. Kapanidis, A.N., Lee, N.K., Laurence, T.A., Doose, S., Margeat, E. and Weiss, S. (2004) Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U.S.A.*, 101, 8936-8941.

2. Holden, S.J., Uphoff, S., Hohlbein, J., Yadin, D., Le Reste, L., Britton, O.J. and Kapanidis, A.N. (2010) Defining the Limits of Single-Molecule FRET Resolution in TIRF Microscopy. *Biophysical Journal*, 99, 3102-3111.

3. Ha, T., Zhuang, X., Kim, H.D., Orr, J.W., Williamson, J.R. and Chu, S. (1999) Ligand-induced conformational changes observed in single RNA molecules. *Proceedings of the National Academy of Sciences*, 96, 9077-9082.

4. Abbondanzieri, E.A., Bokinsky, G., Rausch, J.W., Zhang, J.X., Le Grice, S.F.J. and Zhuang, X. (2008) Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature*, 453, 184-189.

5. Xu, H.-T., Quan, Y., Schader, S.M., Oliveira, M., Bar-Magen, T. and Wainberg, M.A. (2010) The M230L Nonnucleoside Reverse Transcriptase Inhibitor Resistance Mutation in HIV-1 Reverse Transcriptase Impairs Enzymatic Function and Viral Replicative Capacity. *Antimicrobial Agents and Chemotherapy*, 54, 2401-2408.

6. Lebowitz, J.L., Helfand, E. and Praestgaard, E. (1965) Scaled Particle Theory of Fluid Mixtures. *The Journal of Chemical Physics*, 43, 774-779.

7. Reiss, H., Frisch, H.L. and Lebowitz, J.L. (1959) Statistical Mechanics of Rigid Spheres. *The Journal of Chemical Physics*, 31, 369-380.

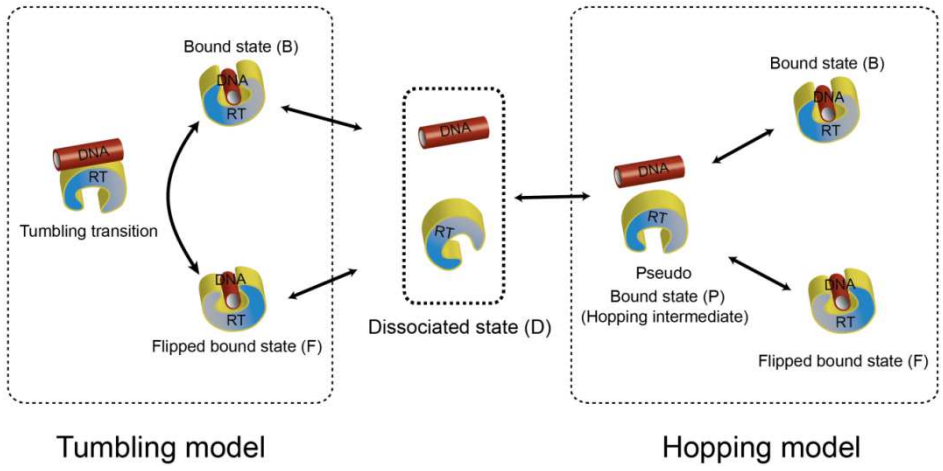
8. Zhou, H.-X., Rivas, G. and Minton, A.P. (2008) Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences. *Annual Review of Biophysics*, 37, 24.

9. Aoki, K., Takahashi, K., Kaizu, K. and Matsuda, M. (2013) A Quantitative Model of ERK MAP Kinase Phosphorylation in Crowded Media. *Sci. Rep.*, 3.

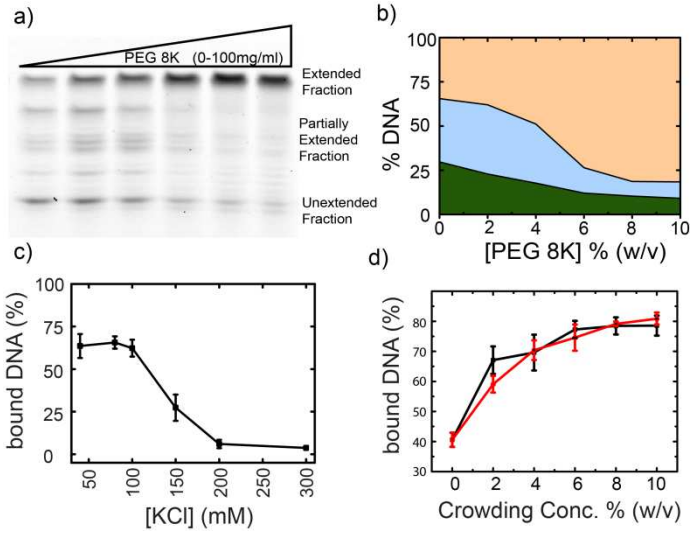
10. Dupuis, N.F., Holmstrom, E.D. and Nesbitt, D.J. (2014) Molecular-crowding effects on single-molecule RNA folding/unfolding thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U.S.A.*, 111, 8464-8469.

11. Thiyagarajan, P., Chaiko, D.J. and Hjelm, R.P. (1995) A Neutron Scattering Study of Poly(ethylene glycol) in Electrolyte Solutions. *Macromolecules*, 28, 7730-7736.

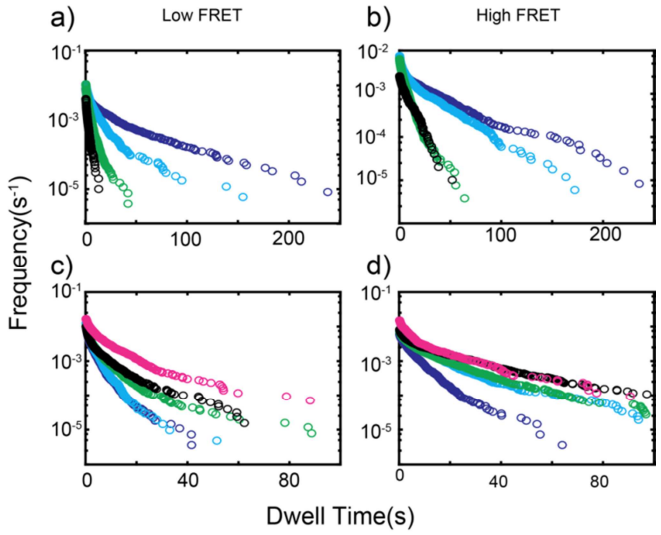
Supplemental figure legends



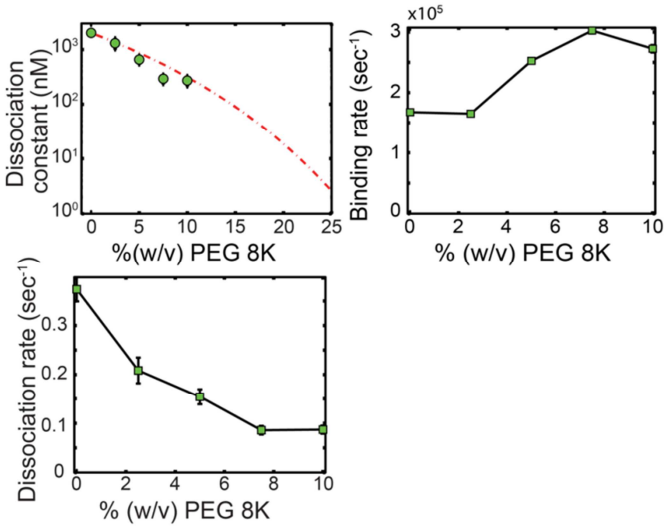
Supplementary Figure 4.1**: Schematic diagram representing the pathways of tumbling and hopping mechanism of flipping**. The DNA is represented with a cylinder (red) and RT with a "U"-shape (yellow-blue). In the case of tumbling model (left box), RT first binds DNA from a dissociated state (middle box) in one of the two bound states and flipping between the two states happens by tumbling over the DNA. In the case of hopping model (right box), RT and DNA from dissociated state (middle box) first enters into pseudo bound state. The pseudo-bound state is an intermediate between bound state, flipped state, and dissociation state. Flipping happens if the enzyme just entered into pseudo-bound state from its one of the bound states, rotates 180° and binds to the other bound state.



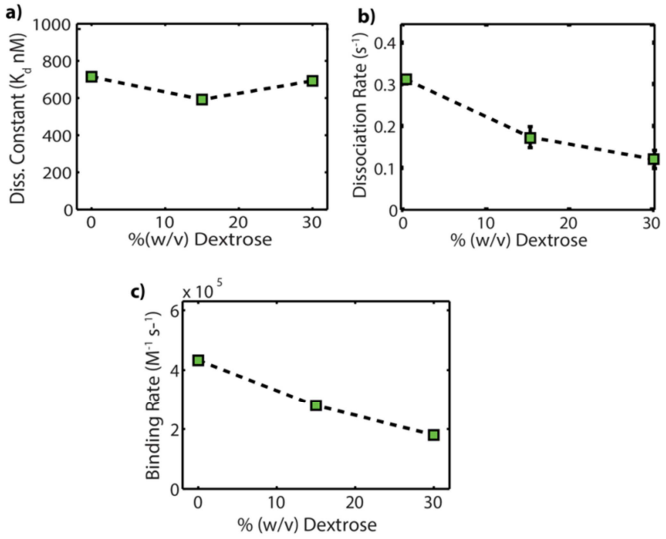
Supplementary Figure 4.2**: Single round primer extension assay under increased salt and macromolecular crowding**. **a)** A representative 10% urea-polyacrylamide gel of a primer extension assay in the presence of a heparin trap. The assay was conducted at 40 nM RT and 150 nM DNA-Cy5 with 2 mg/ml heparin and varying concentrations of PEG 8K (0, 2, 4, 6, 8, and 10 % (w/v) PEG 8K from left to right lane). **b)** Summed fractional intensity of different DNA fractions at increasing PEG 8K concentrations. Orange: fully extended DNA fraction, blue: partially extended DNA fraction and green: un-extended DNA fraction. **c)** Percent of extended DNA at increased KCl concentrations. **d)** The percent of extended DNA under increased crowding concentrations at 150 mM KCl. Black line represents PEG 8K data and red line represents Ficoll 70K data.



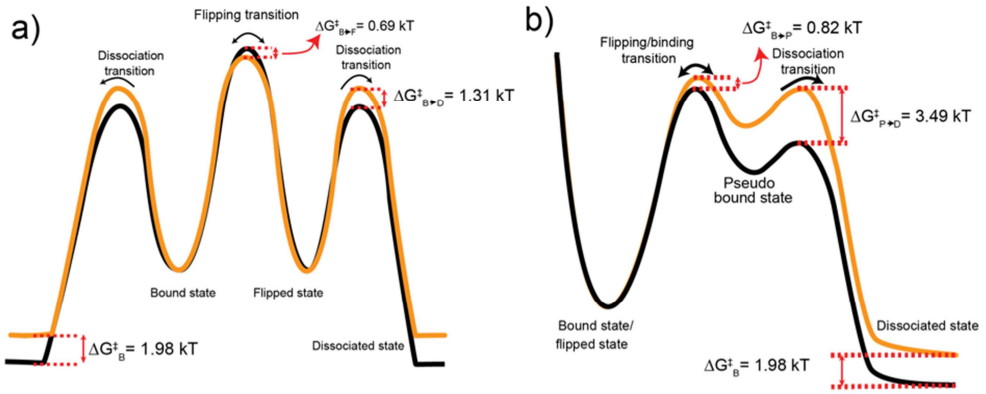
Supplementary Figure 4.3: Survival frequency of RT binding on double primer DNA at varying salt and PEG 8K concentrations extracted from single-molecule FRET measurements. a) Low FRET and b) high FRET frequency of RT-DNA complexes with lifetimes greater than or equal to the indicated dwell time at 50 mM (blue), 100 mM (cyan), 150mM (green) and 200 mM (black) NaCl concentration. c) Low FRET and d) high FRET frequency of RT-DNA complexes with lifetimes greater than or equal to the indicated dwell time at 0 % (blue), 2.5 % (cyan), 5 % (green), 7.5 % (black) and 10 % (magenta) (w/v) PEG 8K at 150 mM NaCl.



Supplementary Figure 4.4**: Binding dynamics of RT on dpdsDNA at 200 mM NaCl under increased macromolecular crowding**. **a)** Dissociation constant of binding of RT to dpdsDNA under increased concentration of PEG 8K. The dashed red line is a fit of scaled particle theory to the data (green circles) obtained at 200 mM NaCl. The fitting was done in the same way as in Figure 4.2b. **b)** and **c)** Rate of binding and rate of dissociation of RT under increased crowding concentration. Number of binding events analysed at 0, 2.5, 5, 7.5, and 10 % (w/v) PEG 8K were 659, 681, 903, 1156, and 1421, respectively.

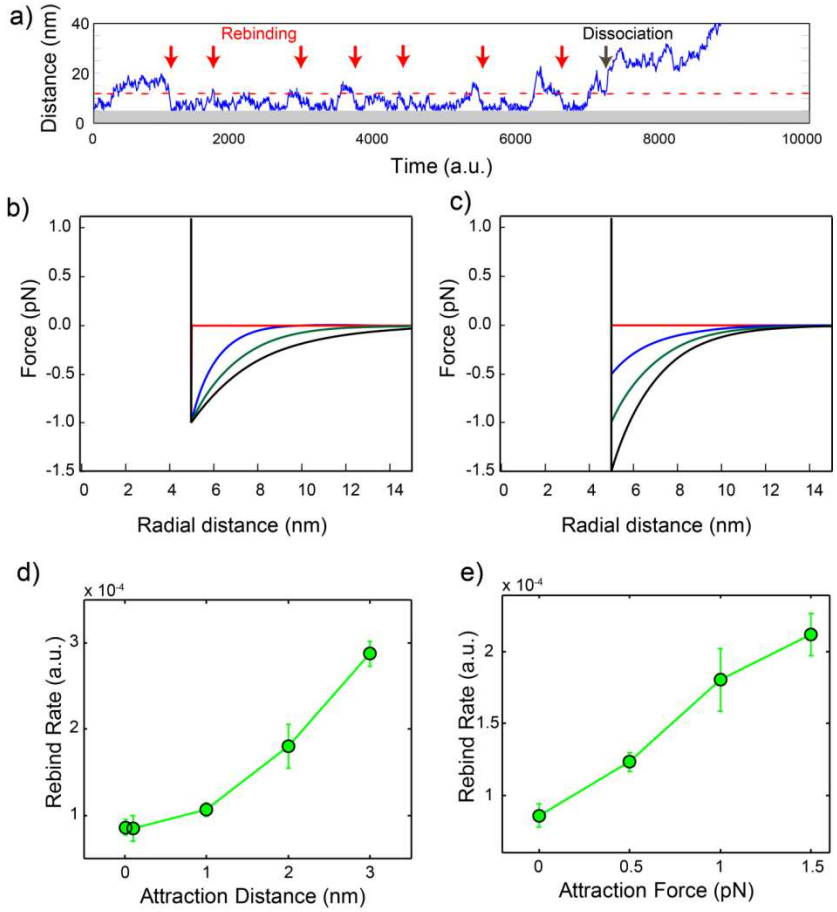


Supplementary Figure 4.5: **Binding of Cy3-labelled RT on dpdsDNA at 150 mM NaCl under high concentrations of dextrose**. **a)** Dissociation constant of RT binding to DNA is relatively constant under increased dextrose concentration. **b)** The rate of dissociation of RT is decreased under increased dextrose concentration. **c)** The rate of binding of RT is also decreased due to increased dextrose concentration. Number of binding events analysed at 0, 15, and 30 % (w/v) Dextrose are 4431, 478, and 375, respectively.



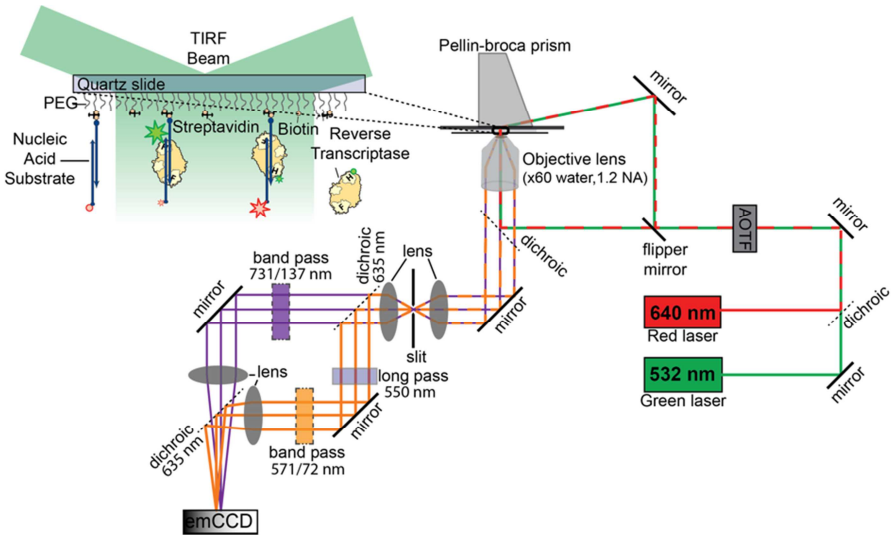
Supplementary Figure 4.6**: Free energy diagrams of RT flipping kinetics on DNA under increased crowding.** The free energy diagrams were drawn for tumbling model and hopping model based on the kinetic and thermodynamic data obtained under different crowding concentrations. **a)** Free energy diagram corresponding to the tumbling model of RT binding and flipping kinetics under dilute conditions at 150 mM NaCl (black). Increased crowding concentration leads to increase in the free energy barrier height for dissociation transition and decrease in the barrier height for flipping transition relative to the bound or flipped bound state (orange). **b)** Free energy diagram explaining flipping kinetics based on hopping model with an extra pseudo-bound state. In this free energy diagram both the bound state and

flipped state are identical. Color convention is the same as in the free energy diagram for the tumbling model.

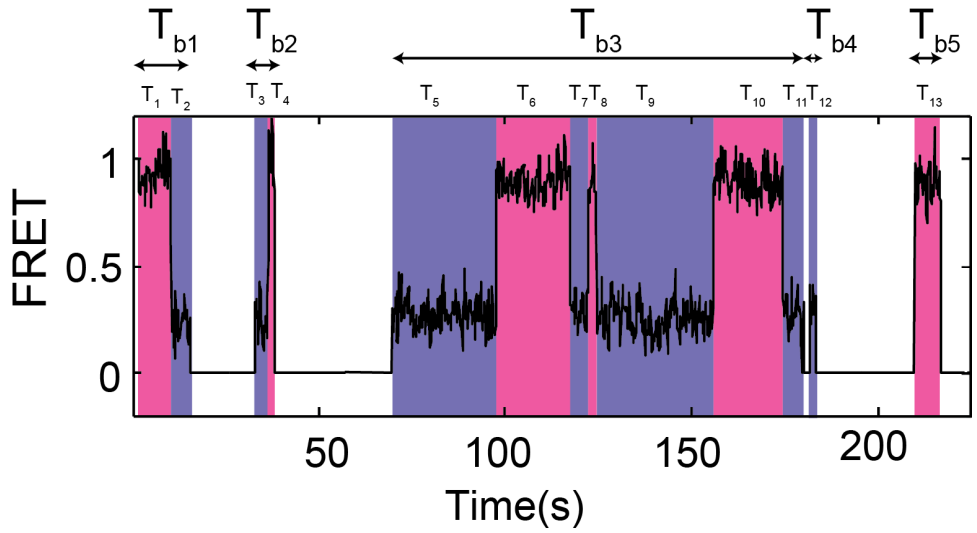


Supplementary Figure 4.7**: Brownian dynamics simulations of two diffusing particles held under an attractive force field. a)** Representative time trace showing the distance between centres of two particles undergoing Brownian diffusion with exponentially decaying attractive force of with a maximum value of 1 pN and a distance parameter of 3 nm between them. Grey region represents the minimum distance (5 nm) between centres of two particles before they come into contact. During the simulation rebinding events were tallied when the distance returned to 5 nm after reaching 12 nm (red arrows). We assume that the two particles are completely dissociated if the two particles become separated more than 12 nm and do not rebind (grey arrow). **b)** Curves showing exponentially decaying force fields as a function of distance between the centres of each particle. The screening distance was varied to simulate different salt concentrations. The distance parameter for the red, blue, green, and black curves are 0, 1, 2, and 3 nm, respectively and the maximum force was 1 pN in all the cases. **c)** Curves showing exponentially decaying force fields with different maximum forces, meant to simulate the effects of crowding. The maximum forces for red, blue, green, and black curves are 0, 0.5, 1, and 1.5 pN, respectively and the distance parameter was 2 nm for

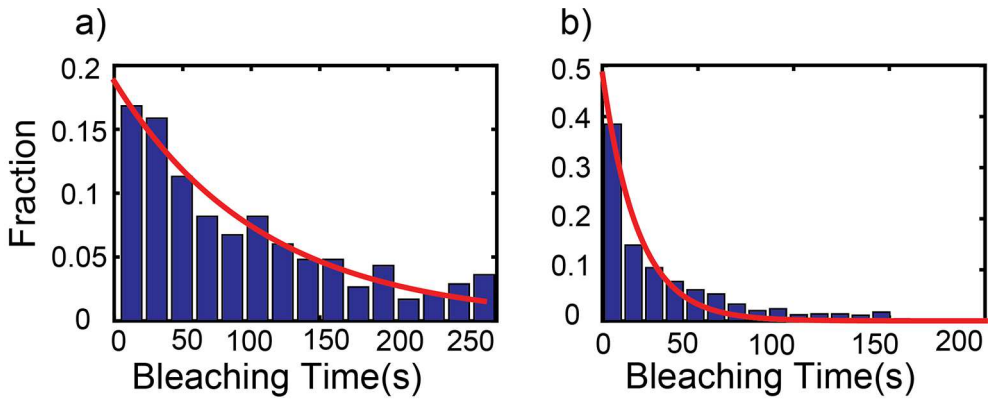
all the curves. **d) and e)** The rate of rebinding of two particles was calculated by taking the ratio of number of rebinding events and total amount of bound time for 1600 simulations for each point. In **(d)** the attraction distance was varied with a maximum force of 1 pN, while in **(d)** the maximum force was varied with a constant distance parameter of 2 nm.



Supplementary Figure 4.8**: Schematic diagram of the prism-type total internal reflection fluorescence microscopy.** The excitation lasers are focused through a Pellin-broca prism to create a broad TIRF excitation region in the image plane. Freely diffusing RT (yellow) binds to immobilized DNA (blue) producing bursts of fluorescence. This fluorescence is imaged through an objective onto a mechanical slit. The light is then separated into two channels by a dichroic mirror and focused on two halves of an emCCD detector.



Supplementary Figure 4.9**: Kinetics of flipping of reverse transcriptase on dpdsDNA**. The single molecule FRET time trace contains five binding events labelled as T_{b1}, $T_{b2,}$ T_{b3}, T_{b4} and T_{b5}. Each binding event further divided into either low FRET, high FRET or combination of both with single/multiple flipping transitions. The sub-states are labelled as T_1, T_2, ……., T_{13}. Pink and blue colors denote for low and high FRET binding events, respectively. If the pink and blue are not separated by white color then a flipping event is scored. Eight flipping events can be seen due to the fact that the transition between two sub-states (between T_1 and T_2, T_3 and T_4, T_5 and T_6, T_6 and T_7, T_7 and T_8, T_8 and T_9, T_9 and T_{10}, T_{10} and T_{11}) happened instantaneously. T_{12} and T_{13} are two binding events without a flipping event. In this example trace, during the total observation time of 225 sec, there are seven low FRET events ($N_{b_l} = 7$) with total low FRET bound time $T_{bound,low}$ of T_2+T_3+T_5+T_7+T_9+T_{11}+T_{12}, six high FRET events ($N_{b_h} = 6$) with total high FRET bound time ($T_{bound,high}$) of T_1+T_4+T_6+T_8+T_{10}+T_{13}, four low FRET to high FRET flipping transitions ($N_{f_l} = 4; T_3 \rightarrow T_4, T_5 \rightarrow T_6, T_7 \rightarrow T_8, and\ T_9 \rightarrow T_{10}$), four high FRET to low FRET flipping transitions ($N_{f_h} = 4; T_1 \rightarrow T_2, T_6 \rightarrow T_7, T_8 \rightarrow T_9, and\ T_{10} \rightarrow T_{11}$) and the total bound time is T_{b1}+T_{b2}+T_{b3}+T_{b4} +T_{b5}=150s.



Supplementary Figure 4.10**: Rate of photo-bleaching of RT-Cy3 and DNA-Cy5**. Histograms of photo-bleaching events of **a)** Cy5 labelled DNA primer and **b)** Cy3 labelled on RT obtained from single molecule fluorescence analysis. Red curves are exponential fits to the histograms.

5
Single-Molecule Fluorescence Assay to Study DNA Supercoil Dynamics

DNA supercoiling crucially affects cellular processes such as DNA replication, gene expression, and chromatin organization. However, mechanistic understanding of DNA supercoiling and the related DNA-processing enzymes has remained limited, mainly due to the lack of convenient experimental tools to probe these phenomena. Here, we report a novel high-throughput single-molecule assay for real-time visualization of supercoiled DNA molecules, named ISD (Intercalation-induced Supercoiling of DNA). We use an intercalating dye to induce supercoiling of surface-attached DNA molecules as well as to visualize coiled-loop structures (i.e. plectonemes) formed on DNA. The technique is solely based on epifluorescence microscopy and requires no mechanical manipulation of the DNA molecules. This new assay allows to track positions and sizes of individual plectonemes and characterize their position-dependent dynamics such as nucleation, termination and diffusion. We describe the ISD technique and demonstrate its potential by establishing that plectonemes are pinned to a local 10-nucleotide long mispaired sequence along a double-stranded DNA molecule.

This chapter has been published as M. Ganji*, S.H. Kim*, J. van der Torre, E. Abbondanzieri, and C. Dekker, Intercalation-based single-molecule fluorescence assay to study DNA supercoil dynamics, Nano Letters, 16, 4699−4707 (2016)

5.1 Introduction

DNA supercoiling plays a vital role in most cellular processes including DNA replication and gene expression.[1-6] Failure to control the degree of supercoiling can be lethal.[7] Indeed, the supercoiling state of genomic DNA is precisely controlled over the course of the cell cycle and is subject to change in response to environmental factors such as temperature jumps or starvation.[8, 9] Electron-microscopy studies revealed that the *Escherichia coli* genome is organized into 50-100 topological domains with a length of 20-100 kb, where each domain allows for independent control of the degree of supercoiling of its DNA segment.[10, 11] Most regions of the DNA are maintained in a negatively supercoiled state for easier access of the DNA bases to proteins, but positively supercoiled domains are also present.[12] Interestingly, thermophilic bacterial species keep most DNA in a positively supercoiled state to avoid excessive denaturation of DNA at the high temperature of their environment.[13, 14]

Supercoiling induces significant alterations in the three-dimensional structure of DNA. For example, the formation of a type of DNA supercoil called plectonemes, where the DNA helix is coiled onto itself (cf. Figure 5.1D), brings distant DNA segments in close proximity, which subsequently alters the activity of DNA-processing proteins such as transcription-regulatory proteins or site-specific recombinases.[15-19] These enzymatic activities in turn can change the supercoiling state of DNA transiently or permanently[5, 20]. Also, local defects in the B-form DNA structure such as kinks or bubbles have been theorized to affect the locations of plectonemes.[21]

Despite the ubiquitous importance of DNA supercoiling, a mechanistic understanding of the role of supercoiled DNA in various cellular processes has remained limited. This is mainly due to the lack of a convenient experimental platform that allows real-time *in situ* visualization of the DNA structure under different degrees of supercoiling. Indirect measurements, such as biochemical assays of enzymatic activity on supercoiled DNA, hardly allow for an unambiguous interpretation of the DNA structure. And direct measurements of DNA structure using electron-microscopy and atomic-force microscopy only provide static images of DNA plasmids with limited control of the supercoiling state.[22-30] Force-spectroscopy techniques such as magnetic tweezers and optical torque wrenches have been developed for mechanical measurements on supercoiled DNA with precise control of the tension and torque applied.[31-33] However, most of these single-molecule techniques merely measure the end-to-end extension of the DNA which

provides only limited structural information. Real-time visualization of the dynamics of plectonemes within a single DNA molecule has only recently been achieved by combining fluorescence microscopy and magnetic tweezers in a side-pulling geometry.[34] This technique provided a powerful demonstration of the feasibility to visualize and study the DNA supercoiling, but the fairly complicated instrumentation and sample preparation and the low throughput of the method have prevented it from becoming widely accessible.

Here, we report a novel high-throughput single-molecule assay to control and visualize DNA supercoils using only a conventional fluorescence microscope. We use intercalating dyes to induce supercoils within a linear DNA molecule that is bound to a surface at its two ends in such a way that it is torsionally constrained. Intercalation of dye molecules between the DNA bases results in a local change of the rise and twist of the B-form DNA[35-37], which globally induces twist to the DNA molecule.[38] Indeed, we show that this simple approach makes it possible to precisely control the degree of supercoiling of the DNA by changing the concentration of the intercalating dye. Therefore, there is no need for direct mechanical manipulation on the DNA, which requires a more complicated experimental measurement apparatus.[34] Moreover, visualization of the plectonemes that are induced by the coiling of the DNA is directly accessible by measuring the fluorescence intensity of the intercalating dye. We name this new assay Intercalation-induced Supercoiling of DNA (ISD). Using this assay, we measured the size and position of individual plectonemes along DNA to characterize their position-dependent dynamics such as nucleation, termination, and diffusion. We then applied this new assay to study a pinning effect of plectonemes at a local single-stranded region of mismatched bases.

5.2 Results and discussion

To visualize supercoiled DNA (Figure 5.1), we prepared 20kb-long linear DNA with multiple biotins labeled at its end regions (~500 bp at each of the ends). To determine the direction of each DNA molecule, we further labeled one of the DNA ends with Cy5 which can be identified under 640 nm laser illumination (Supplementary Information,). We flowed the DNA molecules into a microfluidic channel in which the surface is coated with streptavidin (Figure 5.1A and Supplementary Figure 5.1A). Under constant flow, one end of the DNA molecule first binds to the surface, causing the DNA molecule to be linearly stretched along the flow (Figure 5.1B). Subsequently, the other end of the DNA binds to the surface (Figure 5.1C).[39, 40] We set the flow rate to obtain a DNA extension of around 65-70% of its B-form contour length (Supplementary Figure 5.2A). Unbound DNA

molecules were washed off before the injection of a buffer containing 30 nM of the intercalating dye Sytox Orange (SxO). We chose the monomeric cyanine nucleic-acid stain SxO for inducing and visualizing DNA supercoils for a number of reasons: First, the high fluorescence quantum yield of the dye ($\Phi = 0.9$) provides a high signal-to-noise-ratio in imaging at our 532 nm excitation. Second, SxO has relatively high binding and dissociation rates[41] which ensures that equilibrium is achieved immediately after buffer exchange and which also minimizes the number of photo-bleached dye molecules bound to DNA (as those will disassociate and get replaced). Third, the dye exhibits a more than 500-fold fluorescence enhancement upon binding to DNA, minimizing the fluorescence background from the free dyes in the solution. Upon flushing the buffer containing SxO into the flow cell, dye molecules bind to DNA by intercalating between bases, resulting in local unwinding of helical structure. Because each end of the DNA has multiple biotin-streptavidin linkages, the torsional stress induced by the intercalation of the dye molecules is not be relaxed and instead accumulates to induce positive supercoiling in the form of plectonemes (Figure 5.1D).

We employed dual-color epifluorescence microscopy in order to visualize the tethered DNA molecule and the plectonemes induced in it (Supplementary Figure 5.1). Shown in the Figure 5.1E is a representative fluorescence image with eight stretched DNA molecules. Note that the DNA molecules display a homogeneous fluorescence intensity with additional bright fluorescent local spots, which are consistent with the formation of local plectonemes (Figure 5.1E-G, red arrows) [cf. the very similar images in a previous report of DNA supercoils].[34] These spots disappeared suddenly and irreversibly over time under prolonged laser illumination, presumably because of the photo-induced appearance of a local nick along the DNA molecule. After the photo-induced nicking, the intensity profile of the DNA molecule became very homogeneous along the length of the DNA, showing no detectable sequence-dependence in the binding of SxO dye (Supplementary Figure 5.2B). Note that these fluorescent spots display the exact same behavior as the local plectonemes observed with side-pulling magnetic tweezers.[34] To further confirm that the observed spots were indeed plectonemes, we performed two additional experiments. First, we visualized DNA molecules which were pre-treated with a nicking enzyme. As expected, we did not observe any bright spots on these nicked DNA molecules. Second, we directly visualized plectonemes emerging from DNA exhibiting bright spots using flow stretching. We selected a DNA molecule which was not oriented parallel to the direction of flow (Figure 5.1E-G and Supplementary Movie 1). Before applying flow, the molecule was linear and showed the characteristic bright spots (Figure 5.1F). During the flow, under an oblique angle, a

branch in the DNA molecule emerged in the direction of flow (Figure 5.1G), resulting in a Y-shaped molecule. The downstream-flow-aligned branch of the DNA exhibited a higher fluorescence intensity compared to the other branches, indicating this branch was a plectoneme containing two DNA strands wound around each other. After prolonged exposure to

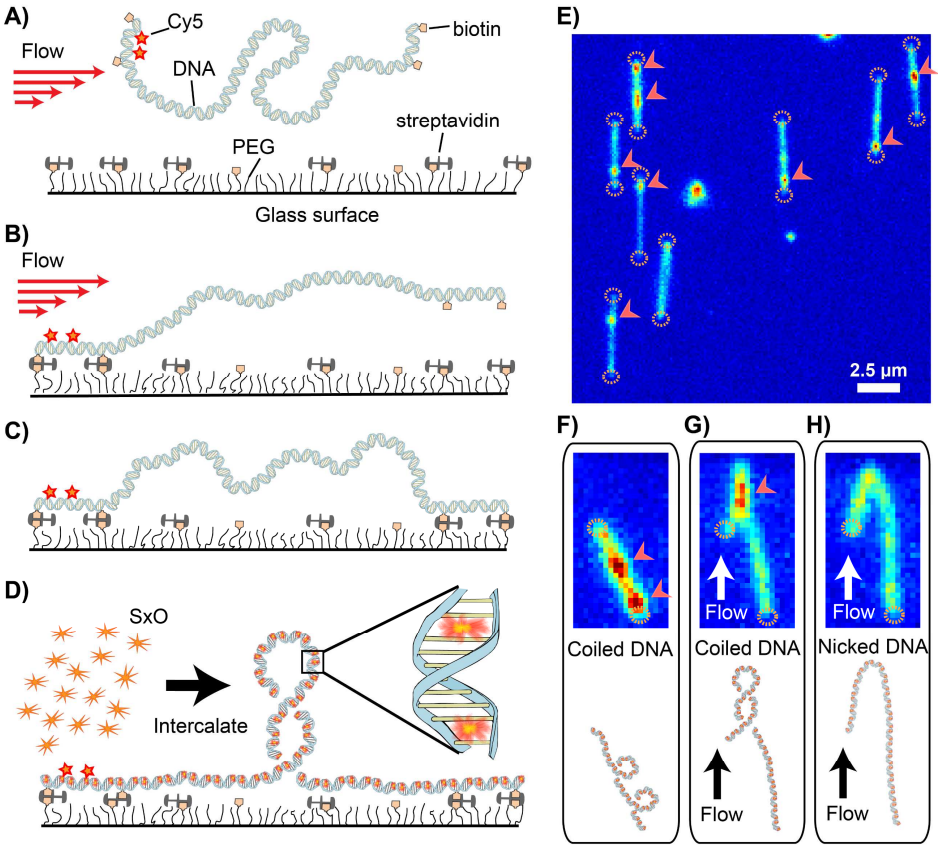


Figure 5.1. Visualization of supercoiled DNA induced by intercalating dyes. (A-D) Schematics of the preparation of doubly-tethered, supercoiled DNA. (A) Biotinylated-DNA is applied under flow to a surface containing immobilized biotin-streptavidin complexes. Cy5 (red star) fluorophores identify the DNA orientation. (B) One of the DNA ends binds to the surface causing the DNA molecule to be stretched by the flow. (C) Binding of the remaining DNA end locks the molecule in an extended conformation. Multiple streptavidin-biotin links at each DNA end ensure the DNA is torsionally constrained. (D) Binding of the intercalating Sytox orange dyes (SxO, orange stars) unwinds the DNA and generates torsional stress that lead to plectonemes in the DNA. (E) Example of a fluorescence image of SxO stained DNA obtained with the epifluorescence microscope. Among the eight stretched DNA molecules in the field of view, seven showed plectonemes (indicated by arrows).

Dashed circles represent DNA anchor points. (F) Image of a single DNA molecule with multiple plectonemes (red arrows) observed without flow. A schematic of DNA molecule is drawn at the bottom for visual guide. (G) Visualization of the same DNA molecule under flow shows a single plectoneme (red arrow) stretched in the direction of flow. (H) After photo-induced nicking, the plectonemes vanish and the DNA adopts a J-shaped structure under flow.

the excitation light, the plectonemic region suddenly unwound, likely due to photo-induced nicking of the DNA, creating a J-shaped DNA molecule with a uniform intensity (Figure 5.1H). These observations confirm that the bright fluorescent spots that we observed were indeed DNA plectonemes.

Having established that SxO can be used to form plectonemes, we next used magnetic tweezers[42] to examine if the intercalation of SxO dyes modifies the mechanical properties of DNA. We prepared DNA molecules identical to those used above except that the biotin labels at the Cy5-end were replaced with digoxigenin. We then tethered the digoxigenin-labeled ends of the DNA molecules to a glass surface coated with anti-digoxigenin and attached the biotin-labeled DNA ends to magnetic beads coated with streptavidin. Then, by using a pair of magnets, we observed the mechanical response of the DNA molecules under a range of tensions and torques in the presence of SxO.

We first examined the end-to-end length change of a torsionally unconstrained (i.e. nicked) DNA at different applied forces in the presence or absence of SxO (Figure 5.2A). Force-extension curves from torsionally unconstrained DNA are shown in Figure 5.2B. In the force range of 1-5 pN, we observed an increase in the end-to-end distance of the DNA as we increased the concentration of SxO. For example, at 3 pN force and 30 nM SxO, we observed 21% increase in the DNA extension. Importantly, upon normalizing the end-to-end length of the DNA with its maximum length at 5pN, the force-extension curves of the DNA for all SxO concentrations overlapped with each other (Figure 5.2B, inset), implying that the intercalation of the dye did not significantly alter the mechanical properties of the DNA other than slightly extending the contour length. When we flushed out the flow cell with a SxO-free buffer, the force-extension curve nearly reverted back to that of the bare DNA, confirming that the dye binding is reversible (Figure 5.2B, grey dashed line).

We also measured force-extension curves for torsionally constrained DNA (Figure 5.2C). In the absence of SxO, no difference was observed between torsionally unconstrained and constrained DNA, indicating the DNA was initially torsionally relaxed (Figure 5.2B and D, black curves). However, upon addition of SxO, the torsionally constrained DNA exhibited a different behavior (Figure 5.2D) compared

to the torsionally unconstrained DNA (Figure 5.2B). As we increased the concentration of SxO, the end-to-end extension of the torsionally constrained DNA systematically decreased due to the formation of plectonemes. This empirical force-extension relationship measured with magnetic tweezers also allows to estimate the tension within the DNA in ISD measurements (Supplementary Figure 5.2A). We then measured the extension-rotation curves of DNA molecules at 3 pN under different SxO concentrations (Supplementary Figure 5.2C). When we unwound the torsionally

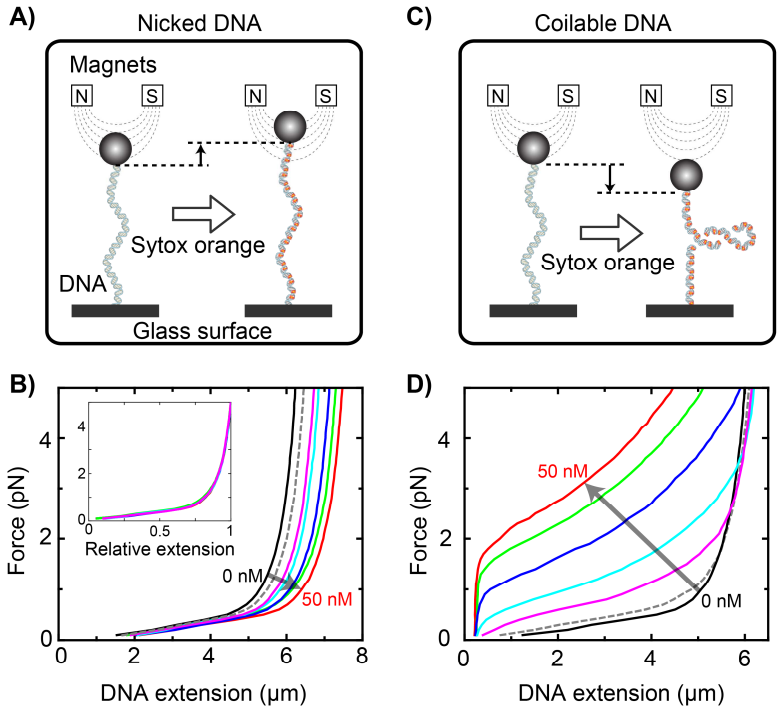


Figure 5.2. DNA force-extension curves from single-molecule magnetic tweezers assay. (A) Schematic diagram showing lengthening of nicked DNA molecule, held in magnetic field created by a pair of magnets, upon binding of SxO. (B) Force-extension curves of nicked DNA measured at various SxO concentrations (black, pink, cyan, blue, green, and red: 0, 5, 10, 15, 30, and 50 nM. The dotted grey curve was obtained after washing off 50 nM SxO with SxO free buffer). Inset shows the same DNA extension curves normalized to their maximum length. (C) Schematic diagram shows the intercalating dye-induced plectoneme formation on torsionally constrained DNA, resulting in the decrease in the DNA extension. (D) Force-extension curves of the torsional constrained DNA (color scheme for the SxO concentrations is the same as in (B)).

constrained DNA, the end-to-end length linearly increased until it reached to a plateau where all the dye-induced plectonemes were relaxed. Importantly, the slopes within the linearly increasing regions of the rotation curves (i.e., on the right side of Supplementary Figure 5.2C) remained the same regardless of the SxO concentration, implying that the size and curvature of the loops in plectonemes did not change significantly upon binding of SxO in our buffer conditions. From the measured rotation curves, we estimate the number of coils (writhing number) put in plectonemic DNA due to binding of SxO. For instance, at a DNA stretching of 3 pN with 30 nM SxO concentration the number of turns induced into the DNA was 31. Thus, by using force-rotation curves from single-molecule magnetic tweezers as reference, we are able to estimate the number of coils applied on each DNA molecule in our ISD assay.

Having successfully characterized the degree of dye-induced supercoiling, we return to the ISD assay to investigate the dynamics of individual plectonemes. We recorded fluorescence movies of the DNA dynamics with 100 ms exposure time per frame (Figure 5.3A). The fluorescence images of the same DNA molecules were also measured after photo-induced nicking of the DNA to confirm that the plectonemes disappeared (Figure 5.3B). We extracted fluorescence intensity profiles along individual DNA molecules for every frame to build intensity kymographs (Supplementary Figure 5.3A-B). For quantitative analysis, the fluorescence intensity was converted to DNA density by mapping the intensity profiles of the supercoiled DNA onto the intensity profile of the corresponding torsionally relaxed molecule (see Supplementary Figure 5.3C-E, and Material and Methods for detail). A characteristic example of the converted DNA density kymograph is shown in Figure 5.3C.

Next, we applied a threshold algorithm to the DNA density kymograph to track the position and size of individual plectonemes (Figure 5.3D and Supplementary Figure 5.3F). Plectonemes appearing in consecutive frames were considered to be continuous if they appeared within 3 pixels (~360 nm or ~1 kb) of each other. In order to reduce false-positives due to noise, we only included plectonemes that were present for at least two consecutive frames. In this way, diffusing plectonemes (Figure 5.3E) could be tracked over time. If plectonemes in consecutive frames were separated by more than 3 pixels, we assumed that the existing plectoneme terminated and a new plectoneme had nucleated at a different position (Figure 5.3F). We marked the position of first appearance of each plectoneme as the nucleation point and the last position as the terminating point (Figure 5.3G). In Figure 5.3C (right panel), we plotted the tracked position of the individual plectonemes on the DNA density kymograph. Note that the trajectories of the plectonemes in Figure

5.3D were plotted as a function of the genomic position along the DNA (i.e. in units of kilobase pairs). We also obtained the mean-square-displacement (MSD) of plectonemes from the individual time-trajectories of plectoneme displacement, to yield a diffusion coefficient D = 4.8 kb^2/s (0.13 μm^2/s), from the slope of the MSD versus time (Supplementary Figure 5.4). The size of individual plectonemes was determined from the fluorescence intensity by summing up the excess amount of DNA above the average DNA density in a local region.

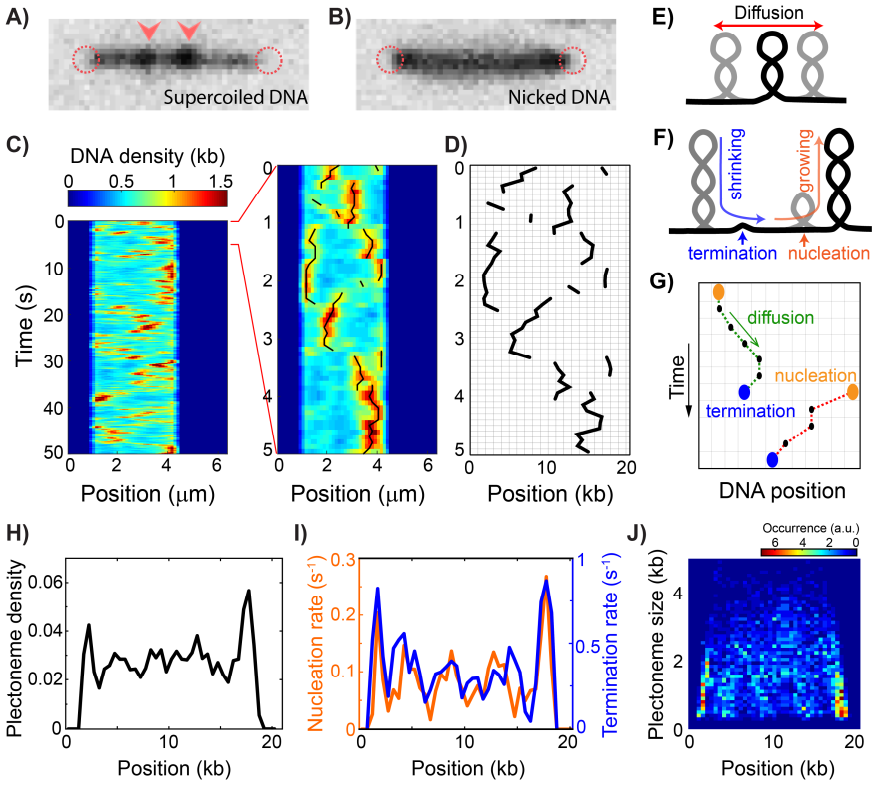


Figure 5.3. Real-time observation of DNA plectoneme dynamics. (A) Fluorescence image of a supercoiled DNA with multiple plectonemes (red arrows). Dotted circles indicate DNA anchoring positions on surface (B) Fluorescence intensity image of the same molecule as in (A) after nicking. (C) Representative DNA density kymograph of a supercoiled DNA molecule. Close up view of the first 5 second is shown in the right panel. Overlaid black curves show the positions of individual plectonemes identified by a threshold algorithm. (D) The plectoneme positions are plotted in DNA base pair space (i.e. in the unit of kb). (E) Schematic diagram showing the diffusion of a plectoneme. (F) Diagram showing the nucleation and termination processes coupled with the plectoneme growth and shrinkage. (G)

Schematic DNA plectoneme kymograph showing nucleation, diffusion, and termination of plectonemes. Two plectonemes (green and red) are presented. Presence of plectoneme was represented with dots. Nucleation and termination events were shown with orange and blue dots, respectively. (H-I) The measured plectoneme density (H), the nucleation and termination rates (I), and the position-dependent plectoneme size distributions from the single DNA molecule presented in (C).

Based on the determined position and size of individual plectonemes, we constructed a probability density function for the plectonemes, hereafter referred to as the plectoneme density, versus the genomic position along the DNA. The plectoneme density was obtained by building a normalized histogram of the genomic positions of individual plectonemes collected from all time frames (Figure 5.3H, see Material and Methods for details). We also counted the total number of nucleation and termination events at each position of the DNA to obtain position-dependent nucleation and termination rates (Figure 5.3I). The observed size of plectonemes varied between 150 bp to 4 kb (Figure 5.3J), in good agreement with our previous study.[34]

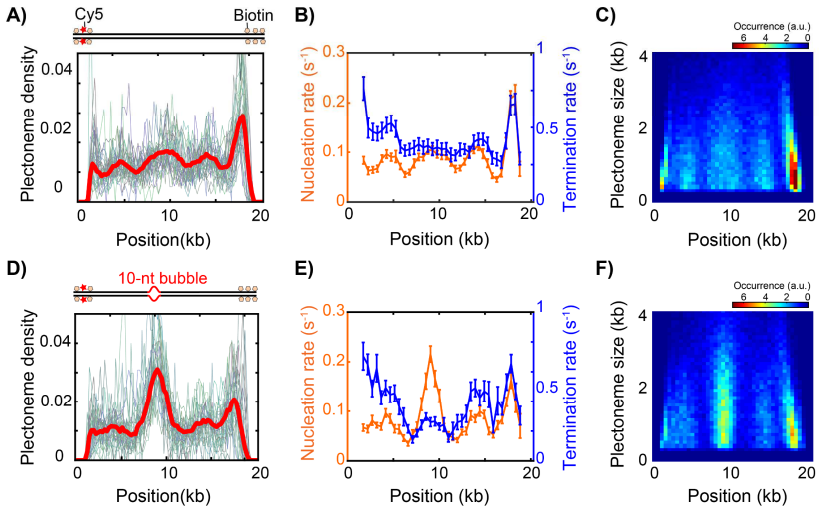


Figure 5.4. DNA sequence-dependent pinning of plectonemes. (A) Plectoneme densities of 46 identical DNA molecules (thin lines) and their average (red thick line). The schematic of the DNA construct is shown above (red star and orange pentagons represent Cy5 and biotin, respectively). (B) Averaged nucleation (orange) and termination (blue) rates observed. (Error bars are sem) (C) Averaged position-dependent plectoneme size distributions. (D) Plectoneme densities of 47 identical DNA molecules (thin lines) that contain a local 10-nt mismatched sequence located at 8.8 kb away from the Cy5 labeled end as illustrated in the top panel. Red thick line is the averaged plectoneme density of the individuals. (E) Averaged

nucleation (orange) and termination (blue) rates observed for the 10-nt mismatched sequence. (F) Averaged position-dependent plectoneme size distributions of the 10-nt mismatched sequences.

Interestingly, the observed plectoneme density and the nucleation and termination rates showed position-dependent variations (Figure 5.3G-J). As inhomogeneity in the formation of plectonemes could be due to the particular molecule, we measured many (N=46) identical DNA molecules and averaged the results (see Figure 5.4A-C). The inhomogeneity was still observed in the averages of the plectoneme density (Figure 5.4A, thick red line), the nucleation and termination rates (Figure 5.4B), and the position-dependent plectoneme size distribution (Figure 5.4C). Similarly, dynamical properties of plectoneme such as the life time and local diffusion constant were observed to be position dependent (Supplementary Figure 5.5). This demonstrates that the plectoneme properties reflect features encoded within the DNA sequence.[43] Interestingly, we observed a strong peak near one end of the DNA (at ~18 kb) in the plectoneme density, nucleation and termination rates, and plectoneme size distributions (Figure 5.4A-C). Because there was no such a peak at the other end of the DNA, the strong peak was likely not an artifact of the DNA-surface linkage and instead reflects some property of the local sequence. Interestingly, when we measured the plectonemes formed on negatively supercoiled DNA molecules (see Material and Methods for the details of preparation), we obtained somewhat different patterns in the plectoneme density and the nucleation and termination rates (Supplementary Figure 5.6). While the current paper reports the new methodology to observe DNA supercoils, this sequence dependence is of obvious importance and will be studied in full detail in future work.

In cells, DNA plectonemes can potentially pin their position to local structures such as DNA-bound proteins or defects in the DNA. We used the new ISD technique to address one such example. As nucleation of a plectoneme requires a large bending of a local region of DNA, a locally flexible DNA region would significantly promote the preferential formation and growth of a plectoneme at this particular site as it would require less energy to bend the DNA into the tip of a plectoneme.[44] To examine if a flexible DNA region indeed serves as a nucleation site that localizes a plectoneme, we inserted a 10-nucleotide mismatched sequence (i.e. a local region where the sequences of both strands do not allow base pairing) at a position 8.8-kb away from the Cy5-end of the DNA (Figure 5.4D). The measured plectoneme density then showed a pronounced peak at the position of this mismatch sequence, confirming that the flexible DNA bubble pins plectonemes at this position. As expected, the plectonemes nucleated more frequently at the mismatch site while the termination rate was lowered at this position (Figure 5.4E). Plectonemes also grew

larger at the mismatch site (Figure 5.4F and Supplementary Figure 5.5G). Furthermore, plectonemes were observed to diffuse towards the mismatch site from both sides and survived longer near the mismatched bubble (Supplementary Figure 5.5B and F). We observed the same pinning effect by a mispaired bubble using a side-pulling magnetic tweezers assay where the DNA was visualized using covalently attached dyes (Supplementary Figure 5.7), as described previously.[34] This control measurement rules out that the plectoneme localization observed in Figure 5.4D-F was an artifact from the intercalation of SxO to DNA. The strong localization effect caused by a DNA mismatch bubble suggests that DNA enzymes that induce local DNA melting, such as RNA polymerase, may also localize a plectonemic structure at the location of the enzyme on the DNA.

5.3 Conclusions

Summing up, we have established ISD as a novel high-throughput technique that allows real-time observation of multiple supercoiled DNA molecules simultaneously using a conventional fluorescence microscopy with a simple sample preparation. The first data taken with ISD demonstrates that the visualization of individual plectonemes in a supercoiled DNA can identify sequence-dependent features of the DNA plectoneme dynamics, such as plectoneme pinning by a 10-nucleotide DNA bubble. The throughput of the current assay can be further improved by patterned immobilization of the stretched DNA to visualize hundreds of molecules simultaneously.[45] Additionally, the current spatial resolution of our fluorescence microscopy is limited by the optical diffraction limit and does not allow to distinguish two plectonemes in close proximity (<350nm, <1kb). This resolution could be improved by combining ISD with recent advances of super-resolution microscopy such as STED.[46, 47] Furthermore, ISD is capable of detecting other DNA processes such as protein-mediated loop formation and DNA condensation by utilizing multi-color fluorescence imaging, making the technique more broadly applicable to different biological systems. By significantly lowering the barrier for visualizing dynamic plectoneme structures, ISD allows researchers to explore how DNA structure is influenced by DNA sequence and enzymatic activity on supercoiled DNA.

5.4 Materials and methods

Preparation of DNA: DNA molecules were made by using PCR, oligonucleotide hybridization, digestion and DNA ligation (see Supplementary Text 1 and Supplementary Table 1-2 for detail). Each DNA fragment was digested to the

desired length using the restriction enzyme BsaI which leaves an overhang outside of the recognition site giving us the possibility to specifically ligate different fragments together. The digested DNA fragments were then ligated using T4 DNA ligase (Promega) to make 20kb-long DNA molecules. For surface immobilization, we further extended the DNA molecules at both ends with 500-bp long DNA fragments containing biotinylated dUTP (biotin-16-dUTP, Roche). We also incorporated aminoallyl-dUTP-Cy5 (Jena Bioscience) to one of the biotin-DNA fragments to distinguish between the different ends. The desired DNA molecule was gel purified.

Dual-color epifluorescence microscopy: A schematic of the custom-made epifluorescence microscopy setup is shown in Supplementary Figure 5.1. We combined two lasers (532 nm Cobolt Samba and 640 nm Cobolt MLD, Cobolt AB) using a dichroic mirror and sent them through an acousto-optic tunable filter (AOTF, AA Opto electronic) for a programmable switching of the two lasers lines. The lasers were then focused at the back-focal plane of an objective lens (60x UPLSAPO, NA 1.2, water immersion, Olympus) for the wide-field, epifluorescence-mode illumination. Fluorescence emission from the sample was collected by the same objective lens and separated from the laser beams by using a dichroic mirror (Di01-R405/488/543/635, Semrock). The fluorescence signal was focused by a lens (f_1=50 mm) and spatially filtered by passing through a slit. The fluorescence was then collimated again by using another lens (f_2=100 mm) before the spectral separation of SxO and Cy5 by a dichroic mirror (FF635-Di02, Semrock). Band pass filters at 731/137 nm (FF01-731/137, Semrock) and 571/72 nm (FF01-571/72, Semrock) were employed to avoid cross talk between the channels. Additionally, a 550 nm long pass filter (FEL0550, Thorlabs) was placed in the SxO emission path to eliminate the back scattered 532 nm laser light. Finally, two lenses (f = 200 mm) and a dichroic mirror (FF635-Di02, Semrock) were used to combine and image the fluorescence signals onto an EMCCD camera (Ixon 897, Andor). The AOTF and the EMCCD was synchronized for the alternative laser excitation by using a PCIe-6320 card and a BNC-2120 breakout box (National Instruments) and controlled by a home-built LabVIEW software. The EMCCD was operated at -90 °C and at an em-gain of 1000 for data acquisition.

Intercalation-induced Supercoiling of DNA (ISD): A flow cell for DNA immobilization and buffer exchange was prepared by sandwiching a quartz slide and a glass coverslip with double-sided tape. Holes in the quartz slide serves as an inlet and outlet for buffer exchange (Supplementary Figure 5.1B). Typically, a flow channel holds 10 µl of solution. The inner surface of the flow channel was

passivated with polyethlyleneglycol (PEG) to suppress nonspecific binding of SxO and 2% (wt/wt) of the PEG molecules were functionalized with biotin so that the biotinylated DNA could be immobilized on the surface via biotin-streptavidin linkage.[48]

For immobilization of DNA, we first flowed 20 µl of 0.1 mg/ml streptavidin into a flow cell and incubated for 1 min. After washing excess streptavidin that did not bind to the surface, we introduced 30 µl of 1~10 pM of biotinylated-DNA molecules at a flow rate of 50 µl/min (Figure 5.1A). Immediately after the flow, we further flowed 200 µl of a washing buffer (40 mM Tris-HCl, pH 8.0, 20 mM NaCl, 0.4 mM EDTA) at the same flow rate to ensure stretching and tethering of the other end of the DNA to the surface (Figure 5.1C). We typically obtained a stretch of around 60-65% of the contour length of the B-form DNA (Supplementary Figure 5.2A).

To induce supercoiling of the tethered DNA, we flowed in 30 nM Sytox orange (S11368, Thermo Fisher) in an imaging buffer consisting of 40 mM Tris-HCl, pH 8.0, 20 mM NaCl, 0.4 mM EDTA. We also included an oxygen scavenging system consisting of 2 mM trolox, 40 µg/ml glucose oxidase, 17 µg/ml catalase, and 5% (wt/v) D-dextrose in the imaging buffer for minimization of photo-bleaching of the fluorophores.

To prepare negatively supercoiled DNA, we first immobilized DNA to the surface in the presence of high concentration of SxO (150 nM), and subsequently reduced the dye concentration to 30 nM for the measurements. The subsequent release of pre-bound SxO dyes after immobilization of the DNA results in negative supercoiling of the DNA.

Magnetic tweezers: The magnetic tweezers apparatus is similar to the one reported previously.[49] A flow cell was prepared by sandwiching two polystyrene-coated cover slips with a parafilm as a spacer. The flow cell was then incubated with 1 mg/ml antidigoxigenin in a buffer carrying 20 mM Tris-HCl, pH 8.0, 20 mM NaCl, 5 mM EDTA for 15 minutes. Alongside, streptavidin-coated magnetic beads (2.8 µm diameter, M270, Thermo Fisher) were washed with buffer A (20 mM TrisHCl pH8.0, 200 mM NaCl, 5 mM EDTA, 0.25% (v/v) Tween20) and incubated with the 21-kb DNA functionalized with biotin at one end and digoxigenin at the other end for 15 min. After incubation, unbound DNA was washed with buffer A and the mixture was flowed into the flow cell and incubated for 15 minutes for surface immobilization. The flow cell was washed with excess amount of buffer A until all the unbound beads were removed. After the washing, we introduced the same measurement buffer which we used for the fluorescence measurement with indicated

concentration of SxO to record force-extension curves of DNA. The force-extension curves were measured by approaching the magnet to the DNA tethered surface at a speed of 0.1 pN/s.

Side-pulling magnetic tweezers combined with fluorescence microscopy: The details of the side-pulling magnetic tweezers assay are described in our previous report.[34] Briefly, a 20 kb long DNA functionalized with biotin and digoxigenin at each end was labeled with Cy3 using a commercial nucleic acid labeling reagent (Label IT, MIR3625, Mirus Bio).[50] The labeled DNA was incubated with a streptavidin-coated magnetic beads (1 µm, New England Biolabs) in buffer A for 15 min. The bead/DNA mixture was washed and reconstituted in buffer A to remove unbound DNA molecules and flowed in a square-hollowed glass capillary (8270, VitroCom) pretreated with polystyrene (0.1% in toluene, brief immersing) and antidigoxigenin (1mg/ml, 15 min incubation). After 15 min incubation of the bead/DNA mixture, the capillary was washed with excess amount of buffer A to remove unbound bead/DNA. Then a imaging buffer carrying 20 mM TrisHCl pH 7.5, 2.5 mM dihydroxybenzoic acid, 50 nM Protocatechuate 3,4-Dioxygenase, 1 mM Trolox, 100 µM N-Propyl gallate, and 0.1 % (v/v) Tween20.

The side-pulling magnetic tweezers setup combined with epifluorescence microscopy was built around an inverted microscope (Axiovert 200M, Zeiss) equipped with a 532 nm Laser (Sapphire, Coherent) and a scientific-grade CMOS camera (Neo, Andor).[34] The surface immobilized DNA was first stretched upwards by placing a pair of magnet on top of the capillary to apply rotations to the magnetic bead. After applying a desired number of rotations, the bead was pulled from the side by approaching another magnet horizontally, followed by removal of the top magnets. The tension applied to the DNA was calibrated by measuring the power spectrum of the position fluctuation of the bead.[51] The fluorescence from the labeled DNA were collected by an objective lens (NA1.4, 63x, Oil immersion, Plan-Apochromat, Zeiss) and filtered through a bandpass filter (FF01-585/40, Semrock) to remove back scattered laser light. Images were acquired with 20 ms exposure time.

Image analysis: Fluorescence images were analyzed using custom written software in Matlab (Mathworks, available upon request). The first 10 frames of the measured fluorescence images were averaged and used to determine the end position of individual DNA molecules. After selection of a molecule, the images with 640 nm illumination at the same field of view were loaded and used to identify the direction of the DNA molecule. Then, fluorescence intensity profiles of the DNA along its

longitudinal direction were calculated by summing up the intensities from 11 nearby pixels that lied perpendicular to the DNA at each position. The background intensity of the image was determined by taking the median value of the pixels around the DNA. The background intensity was subtracted from the intensity profile which was then normalized to compensate for photo-bleaching of SxO. After analyzing all the frames, the normalized intensity profiles were aligned to build an intensity kymograph (Supplementary Figure 5.3C).

DNA density estimation from fluorescence intensity: Although peaks in the intensity profile indicate the position of individual plectonemes in real space, it cannot be read as the position in the base pair space because of the nonlinearity caused by presence of other plectonemes which put more DNA into certain pixels (Supplementary Figure 5.3. Image analysis and position/size determination of individual plectonemes. (A) Representative fluorescence images of a supercoiled DNA (top) and the same DNA molecule (bottom) after torsional relaxation due the formation of a nick. (B) Schematic of a supercoiled DNA and its fluorescence intensity profile (top). Two plectonemes (red and green) are shown with its positions in the intensity profile. The DNA that was contained in the two plectonemes is also shown on the torsionally relaxed DNA (bottom). (C) Fluorescence intensity kymograph measured from the supercoiled DNA molecule. (D) Schematic showing the conversion procedure from the raw data to the DNA density. Due to a slightly non-homogeneous illumination profile and varying detection efficiencies in the image area, the intensity of each pixel cannot be directly converted to the amount of DNA. To convert the measured fluorescence intensity to amount of DNA in each pixel, i.e. DNA density, we first obtained fluorescence intensities of each region of the DNA by measuring the intensity profile of the same DNA molecule after nicking. Note that, for the nicked-DNA case, the DNA is homogeneously distributed among the pixels so that the pixel position *can* be directly converted to the base-pair position. Hence, by comparing the area of the left-hand side of each pixel in the coiled DNA intensity profile, i.e., from the cumulative intensity, to that of the nicked DNA, we can determine the amount of DNA in number of base pairs that is contained in each pixel for the coiled DNA. (E) Reconstructed DNA density kymograph that is obtained from the intensity kymograph in C by using the conversion procedure in D. (F) For each frame in the DNA density kymograph, a threshold is applied to detect the position and size of individual plectonemes. (G) Example of position-time trajectories of plectonemes determined from the threshold algorithm.B). Moreover, because of the illumination profile and detection efficiency in the field of view is not exactly the same for each pixel, the observed intensity values at each position of the DNA cannot be directly converted to the amount of

DNA. To calculate the amount of DNA per pixel (DNA density) in the intensity profiles of the supercoiled DNA, we map it into the intensity profile of the same DNA molecule measured after relaxation of torsional stress due to photo-induced nicking (Supplementary Figure 5.3D). We then obtain a kymograph of DNA density (Supplementary Figure 5.3E).

Plectoneme detection and analysis: A threshold algorithm was applied to the DNA density profile to determine the position and size of individual plectonemes. First, the background DNA density was determined by taking global median value of all the DNA density profiles in a kymograph. Then the threshold level was set to be 25% above the background DNA density (Supplementary Figure 5.3F). Among the peaks detected above the threshold, only those that were observed longer than two consecutive time frames were selected as plectoneme. The size of the plectoneme was determined by summing up 5 pixels around each peak. If a plectoneme was found more than 3 pixels away in the next time frame, we counted it as nucleation of new plectoneme at that pixel and simultaneously termination of the plectoneme in the previous time frame (Figure 5.3F). The threshold algorithm applied to the DNA density kymograph determined the positions of each plectoneme in real-space, i.e. pixel position. The pixel positions of each plectoneme were then converted to the genomic positions (base pair position) by summing up of the DNA densities in the left-side pixels of the plectoneme. Once all the plectonemes were identified, we calculated the probability of finding a plectoneme (plectoneme density) at each position along the DNA in base pair space, by counting the number of frames occupied by plectonemes at each position over time. We also calculated the nucleation (termination) rate by counting the number of nucleation (termination) events divided by the sum of time spent without (with) a plectoneme present at that location before the nucleation (termination) of that plectoneme.

5.5 References

1. Wang, J. C. *Annu. Rev. Biochem.* **1985,** 54, (1), 665-697.

2. Liu, L. F.; Wang, J. C. *Proc. Natl. Acad. Sci. USA* **1987,** 84, (20), 7024-7027.

3. Hatfield, G. W.; Benham, C. J. *Annu. Rev. Genet.* **2002,** 36, (1), 175-203.

4. Sherratt, D. J. *Science* **2003,** 301, (5634), 780-785.

5. Chong, S.; Chen, C.; Ge, H.; Xie, X. S. *Cell* **2014,** 158, (2), 314-326.

6. Roca, J. *Chromosoma* **2011,** 120, (4), 323-334.

7. Zechiedrich, E. L.; Khodursky, A. B.; Cozzarelli, N. R. *Genes Dev.* **1997,** 11, (19), 2580-2592.

8. Goldstein, E.; Drlica, K. *Proc. Natl. Acad. Sci. USA* **1984,** 81, (13), 4046-4050.

9. Dorman, C. J. *Trends Microbiol.* **1996,** 4, (6), 214-6.

10. Kavenoff, R.; Ryder, O. *Chromosoma* **1976,** 55, 12.

11. Postow, L.; Hardy, C. D.; Arsuaga, J.; Cozzarelli, N. R. *Genes Dev.* **2004,** 18, (14), 1766-1779.

12. Naughton, C.; Avlonitis, N.; Corless, S.; Prendergast, J. G.; Mati, I. K.; Eijk, P. P.; Cockroft, S. L.; Bradley, M.; Ylstra, B.; Gilbert, N. *Nat. Struct. Mol. Biol.* **2013,** 20, (3), 387-395.

13. Perugino, G.; Valenti, A.; D'Amaro, A.; Rossi, M.; Ciaramella, M. *Biochem. Soc. Trans.* **2009,** 37, (1), 69-73.

14. Forterre, P.; Bergerat, A.; Lopex-Garcia, P. *FEMS Microbiol Rev.* **1996,** 18, (2-3), 237-248.

15. Lavelle, C. *Nat. Struct. Mol. Biol.* **2008,** 15, (2), 123-125.

16. Kouzine, F.; Gupta, A.; Baranello, L.; Wojtowicz, D.; Ben-Aissa, K.; Liu, J.; Przytycka, T. M.; Levens, D. *Nat. Struct. Mol. Biol.* **2013,** 20, (3), 396-403.

17. Baranello, L.; Levens, D.; Gupta, A.; Kouzine, F. *Biochim. Biophys. Acta* **2012,** 1819, (7), 632-8.

18. Kimura, K.; Hirano, T. *Cell* 90, (4), 625-634.

19. De Vlaminck, I.; Vidic, I.; van Loenhout, M. T. J.; Kanaar, R.; Lebbink, J. H. G.; Dekker, C. *Nucleic Acids Res.* **2010,** 38, (12), 4133-4142.

20. Ma, J.; Bai, L.; Wang, M. D. *Science* **2013,** 340, (6140), 1580-1583.

21. Brutzer, H.; Luzzietti, N.; Klaue, D.; Seidel, R. *Biophys. J.* **2010,** 98, (7), 1267-1276.

22. Lyubchenko, Y. L.; Shlyakhtenko, L. S. *Proc. Natl. Acad. Sci. USA* **1997,** 94, (2), 496-501.

23. Irobalieva, R. N.; Fogg, J. M.; Catanese Jr, D. J.; Sutthibutpong, T.; Chen, M.; Barker, A. K.; Ludtke, S. J.; Harris, S. A.; Schmid, M. F.; Chiu, W.; Zechiedrich, L. *Nat. Commun.* **2015,** 6.

24. Adrian, M.; ten Heggeler-Bordier, B.; Wahli, W.; Stasiak, A. Z.; Stasiak, A.; Dubochet, J. *EMBO J.* **1990,** 9, (13), 4551-4.

25. Amzallag, A.; Vaillant, C.; Jacob, M.; Unser, M.; Bednar, J.; Kahn, J. D.; Dubochet, J.; Stasiak, A.; Maddocks, J. H. *Nucleic Acids Res.* **2006,** 34, (18), e125.

26. Bednar, J.; Furrer, P.; Stasiak, A.; Dubochet, J.; Egelman, E. H.; Bates, A. D. *J. Mol. Biol.* **1994,** 235, (3), 825-47.

27. Boles, T. C.; White, J. H.; Cozzarelli, N. R. *J. Mol. Biol.* **1990,** 213, (4), 931-951.

28. Cherny, D. I.; Jovin, T. M. *J. Mol. Biol.* **2001,** 313, (2), 295-307.

29. Levene, S. D.; Donahue, C.; Boles, T. C.; Cozzarelli, N. R. *Biophys. J.* **1995,** 69, (3), 1036-1045.

30. Fogg, J. M.; Kolmakova, N.; Rees, I.; Magonov, S.; Hansma, H.; Perona, J. J.; Zechiedrich, E. L. *J. Phys. Condens. Matter* **2006,** 18, (14), S145-S159.

31. La Porta, A.; Wang, M. D. *Phys. Rev. Lett.* **2004,** 92, (19), 190801.

32. Strick, T. R.; Allemand, J. F.; Bensimon, D.; Bensimon, A.; Croquette, V. *Science* **1996,** 271, (5257), 1835-7.

33. Mosconi, F.; Allemand, J. F.; Bensimon, D.; Croquette, V. *Phys. Rev. Lett.* **2009,** 102, (7), 078301.

34. van Loenhout, M. T. J.; de Grunt, M. V.; Dekker, C. *Science* **2012,** 338, (6103), 94-97.

35. Gago, F. *Methods* **1998,** 14, (3), 277-292.

36. Lee, J. S.; Morgan, A. R. *Nucleic Acids Res.* **1978,** 5, (7), 2425-2439.

37. Yao, J.; Lowary, P. T.; Widom, J. *Proc. Natl. Acad. Sci. U S A* **1993,** 90, (20), 9364-8.

38. Lipfert, J.; Klijnhout, S.; Dekker, N. H. *Nucleic Acids Res.* **2010,** 38, (20), 7122-32.

39. Yardimci, H.; Loveland, A. B.; van Oijen, A. M.; Walter, J. C. *Methods* **2012,** 57, (2), 179-186.

40. Graneli, A.; Yeykal, C. C.; Prasad, T. K.; Greene, E. C. *Langmuir : the ACS journal of surfaces and colloids* **2006,** 22, (1), 292-9.

41. Biebricher, A. S.; Heller, I.; Roijmans, R. F. H.; Hoekstra, T. P.; Peterman, E. J. G.; Wuite, G. J. L. *Nat. Commun.* **2015,** 6.

42. Strick, T. R.; Allemand, J. F.; Bensimon, D.; Croquette, V. *Biophys. J.* **1998,** 74, (4), 2016-2028.

43. Medalion, S.; Rabin, Y. *J. Chem. Phys.* **2016,** 144, (13), 135101.

44. Matek, C.; Ouldridge, T. E.; Doye, J. P.; Louis, A. A. *Sci. Rep.* **2015,** 5, 7655.

45. Gorman, J.; Fazio, T.; Wang, F.; Wind, S.; Greene, E. C. *Langmuir : the ACS journal of surfaces and colloids* **2010,** 26, (2), 1372-1379.

46. Willig, K. I.; Kellner, R. R.; Medda, R.; Hein, B.; Jakobs, S.; Hell, S. W. *Nat. Methods* **2006,** 3, (9), 721-3.

47. Heller, I.; Sitters, G.; Broekmans, O. D.; Biebricher, A. S.; Wuite, G. J.; Peterman, E. J. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2014,** 15, (4), 727-33.

48. Chandradoss, S. D.; Haagsma, A. C.; Lee, Y. K.; Hwang, J. H.; Nam, J. M.; Joo, C. *J. Vis. Exp.* **2014**, (86).

49. R, V.; JSJ, S.; A, L.; C, D. *PloS one* **2012,** 7, (9).

50. Slattum, P. S.; Loomis, A. G.; Machnik, K. J.; Watt, M.-A.; Duzeski, J. L.; Budker, V. G.; Wolff, J. A.; Hagstrom, J. E. *Mol. Ther.* **2003,** 8, (2), 255-263.

51. van Loenhout, M. T.; Kerssemakers, J. W.; De Vlaminck, I.; Dekker, C. *Biophys. J.* **2012,** 102, (10), 2362-71.

Supplementary Text 1

DNA preparation

To make the bubble construct reported in Figure 5.4, we digested 5 DNA fragments with BsaI and ligated these together using T4 DNA ligase. By using BsaI as a restriction enzyme we could make non-palindromic, specific 4-nucleotide overhangs that gave us the possibility to specifically ligate the fragments together in a known way. Two of the DNA fragments were made by PCRing ~8 and ~11kb. Since this gave a low DNA yield, we cloned these DNA fragments into pCR-XL-topo vector, using the TOPO® XL PCR Cloning Kit (Thermo Fisher).

The biotin-handle was made via a PCR on pBlueScriptIISK and using taq DNA polymerase in the presence of biotin-16- dUTP with the forward primer GACCGAGATAGGGTTGAGTG and Reverse primer TTTTTTTTTTGGTCTCTCCAGCTGGCGTTACCCAACTTAATCGCC.

A ~11.2kb fragment was made through a PCR on unmethylated Lambda DNA (promega) using Phusion DNA polymerase (Thermo Fisher). With the forward primer TTTTTTGGTCTCACTGGCAGGAACAGGGAATGC and reverse primer CGCAGTCCTGTCTGCCCAGG.

To make the bubble we hybridized the following two oligonucleotides with a 10nucleotide mismatch: Oligonucleotide 1: TTTTTTGGTCTCTACGCGCCTCTGCCTTAATCAGCATCTCCTGCGCGGTAT CAGGACGACCAATATCCAGCACCGCATCCCACATGGATTTGAATGCCCG CGCAGTCCTGTCTGCCCAGCGAGAGACCAAAAAA

And oligonucleotide 2: TTTTTTGGTCTCTCGCTGGGCAGACAGGACTGCGCGGGCATTCAAATCCA TGTGGGATGCGGCAGACCCGCGTGGTCGTCCTGATACCGCGCAGGAGAT GCTGATTAAGGCAGAGGCGCGTAGAGACCAAAAAA.

A ~8.3kb fragment was made through a PCR on unmethylated Lambda DNA (promega) using Phusion DNA polymerase (Thermo Fisher) with the forward primer: TTTTTTGGTCTCTGCGTATAAGAAAGCAGACGACATCTGG and Reverse primer: TTTTTTGGTCTCCATACACGGTGATGGTCCCGG.

The biotin-Cy5 handle was made via a PCR on pBlueScriptIISK and using taq DNA polymerase in the presence of biotin-16- dUTP and Aminoallyl-dUTP-Cy5 with the forward primer GACCGAGATAGGGTTGAGTG and reverse primer TTTTTTTTTTGGTCTCTGTATCTGGCGTTACCCAACTTAATCGCC.

The flat DNA construct was made in a similar way, with the following DNA fragments. The biotin-handle was made via a PCR on pBlueScriptIISK and using taq DNA polymerase in the presence of biotin-16- dUTP with the forward primer GACCGAGATAGGGTTGAGTG and reverse primer TTTTTTTTTTGGTCTCTCCAGCTGGCGTTACCCAACTTAATCGCC.

A ~11.2kb fragment was made through a PCR on Unmethylated Lambda DNA (promega) using Phusion DNA polymerase (Thermo Fisher) with the Forward primer TTTTTTGGTCTCACTGGCAGGAACAGGGAATGC and reverse primer TTTTTTGGTCTCTACGCGCGTGCCCATGTTCTCTTTCAG.

(promega) using Phusion DNA polymerase (Thermo Fisher) with the Forward primer TTTTTTGGTCTCTGCGTATAAGAAAGCAGACGACATCTGG and reverse primer TTTTTTGGTCTCCATACACGGTGATGGTCCCGG.

The biotin-Cy5 handle was made via a PCR on pBlueScriptIISK and using taq DNA polymerase in the presence of biotin-16-dUTP and Aminoallyl-dUTP-Cy5 with the forward primer GACCGAGATAGGGTTGAGTG and reverse primer TTTTTTTTTTGGTCTCTGTATCTGGCGTTACCCAACTTAATCGCC.

A ~8.3kb fragment was made through a PCR on unmethylated Lambda DNA

DNA sequences used in this study are summarized in Supplementary Table 1 and 2.

Supplementary Table 1

Bubble construct			
DNA fragment	Primers/oligonucleotide	Template	PCR or hybridization
biotin-handle	GACCGAGATAGGGTTGAGTG	pBlueScriptIISK +	PCR (Gotaq), Bio-11-dUTP
	TTTTTTTTTTGGTCTCTCCAGCTGGCGTTACCCAACTTAATCGCC		
~11,2kb fragment	TTTTTTGGTCTCACTGGCAGGAACAGGGAATGC	Unmethylated Lambda DNA	PCR (phusion)
	CGCAGTCCTGTCTGCCCAGG		
Bubble	TTTTTTGGTCTCTACGCGCCTCTGCCTTAATCAGCATCTCCTGCGCGGTATCAGGACGACCAATATCCAGCACCGCATCCCACATGGATTTGAATGCCCGCGCAGTCCTGTCTGCCCAGCGAGAGACCAAAAAA		Hybridization
	TTTTTTGGTCTCTCGCTGGGCAGACAGGACTGCGCGGGCATTCAAATCCATGTGGGATGCGGCAGACCCGCGTGGTCGTCCTGATA		

	CCGCGCAGGAGATGCTGATTA AGGCAGAGGCGCGTAGAGACC AAAAAA		
~8.3kb fragment	TTTTTTGGTCTCTGCGTATAAG AAAGCAGACGACATCTGG	Unmethylated Lambda DNA	PCR (phusion)
	TTTTTTGGTCTCCATACACGGT GATGGTCCCGG		
biotin-Cy5 handle	GACCGAGATAGGGTTGAGTG	pBlueScriptII SK +	PCR (taq), Bio-11-dUTP + Aminoallyl-dUTP-Cy5
	TTTTTTTTTTGGTCTCTGTATCT GGCGTTACCCAACTTAATCGCC		

Supplementary Table 2

Flat construct			
DNA fragment	Primers/oligonucleotide	Template	PCR or hybridization
biotin-handle	GACCGAGATAGGGTTGAGT G	pBlueScriptII SK +	PCR (taq), Bio-11-dUTP
	TTTTTTTTTTGGTCTCTCCA GCTGGCGTTACCCAACTTA ATCGCC		
~11,2kb fragment	TTTTTTGGTCTCACTGGCA GGAACAGGGAATGC	Unmethylated Lambda DNA	PCR (phusion)
	TTTTTTGGTCTCTACGCGCG		

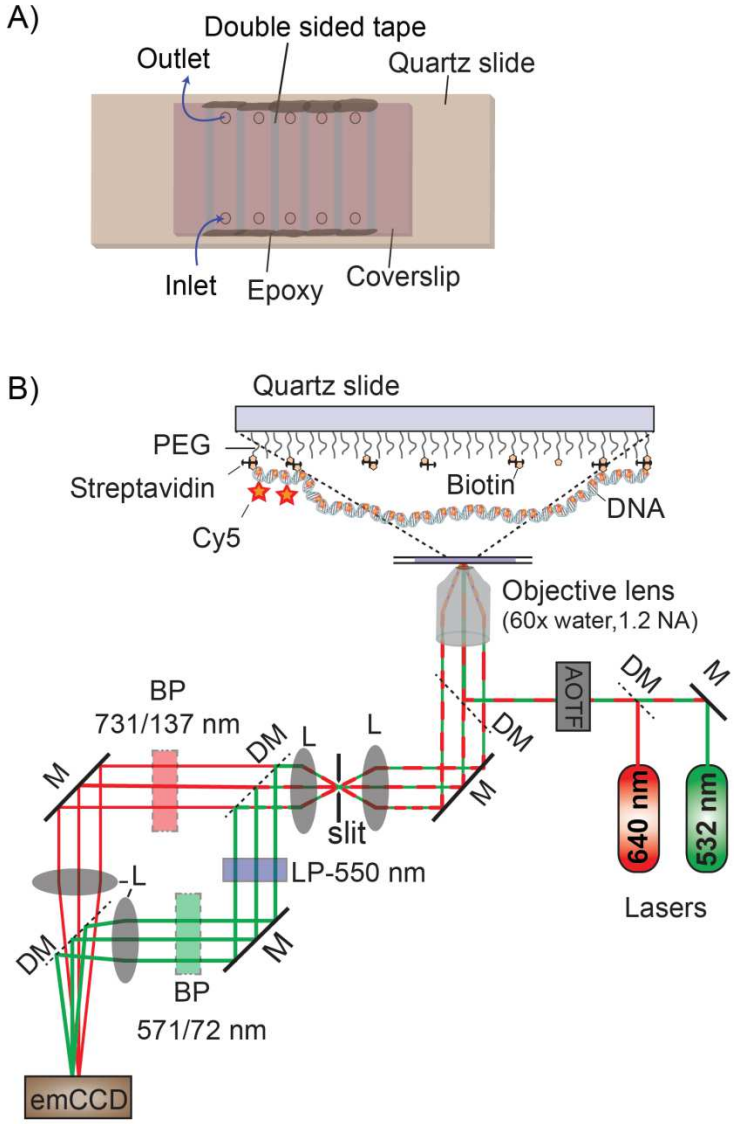
	TGCCCATGTTCTCTTTCAG		
~8.3kb fragment	TTTTTTGGTCTCTGCGTATAAGAAAGCAGACGACATCTGG	Unmethylated Lambda DNA	PCR (phusion)
	TTTTTTGGTCTCCATACACGGTGATGGTCCCGG		
biotin-Cy5 handle	GACCGAGATAGGGTTGAGTG	pBlueScriptII SK +	PCR (Gotaq), Bio-11-dUTP + Aminoallyl-dUTP-Cy5
	TTTTTTTTTTGGTCTCTGTATCTGGCGTTACCCAACTTAATCGCC		

Supplementary movie legend:

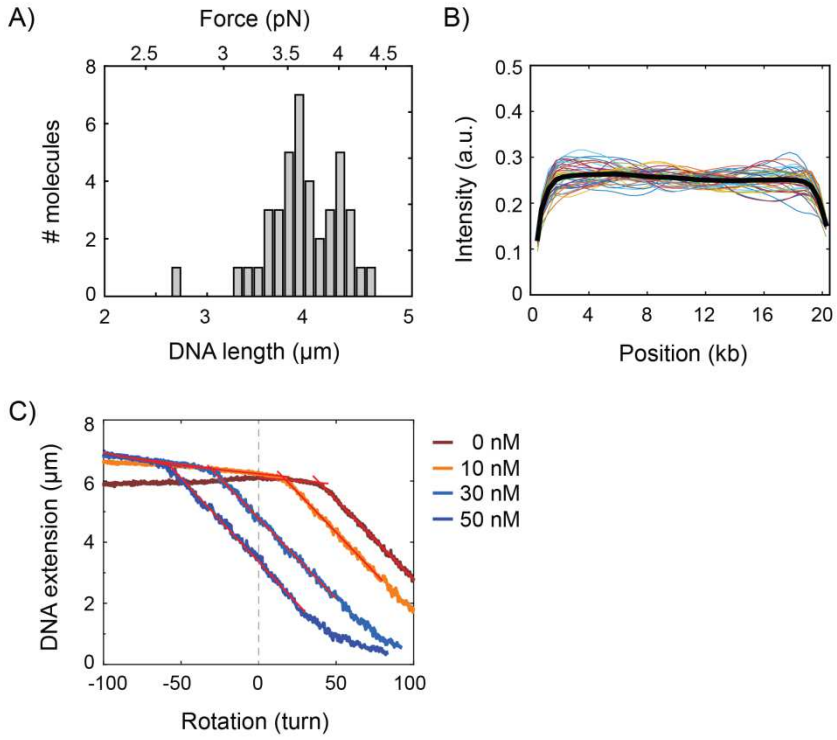
Movie S1:

Movie can be found on the following link: http://goo.gl/sLPKb1. Direct visualization of DNA plectoneme by the application of flow. Movie starts with an immobilized DNA molecule under zero flow featuring dynamic characteristic bright spots indicative of coiled DNA. Flow is briefly applied towards the bottom of the frame, followed by an upward direction. During the flow, the DNA adopted a Y-shaped configuration with brighter fluorescence intensity along the stem pointing in the direction of the flow. This indicates the stem is a single plectoneme. Finally, the molecule abruptly becomes J-shaped, likely due to a photo-induced single-stranded break in the DNA relaxing the torsional stress. Frames from the same movie are shown in Figure 5.1F-H.

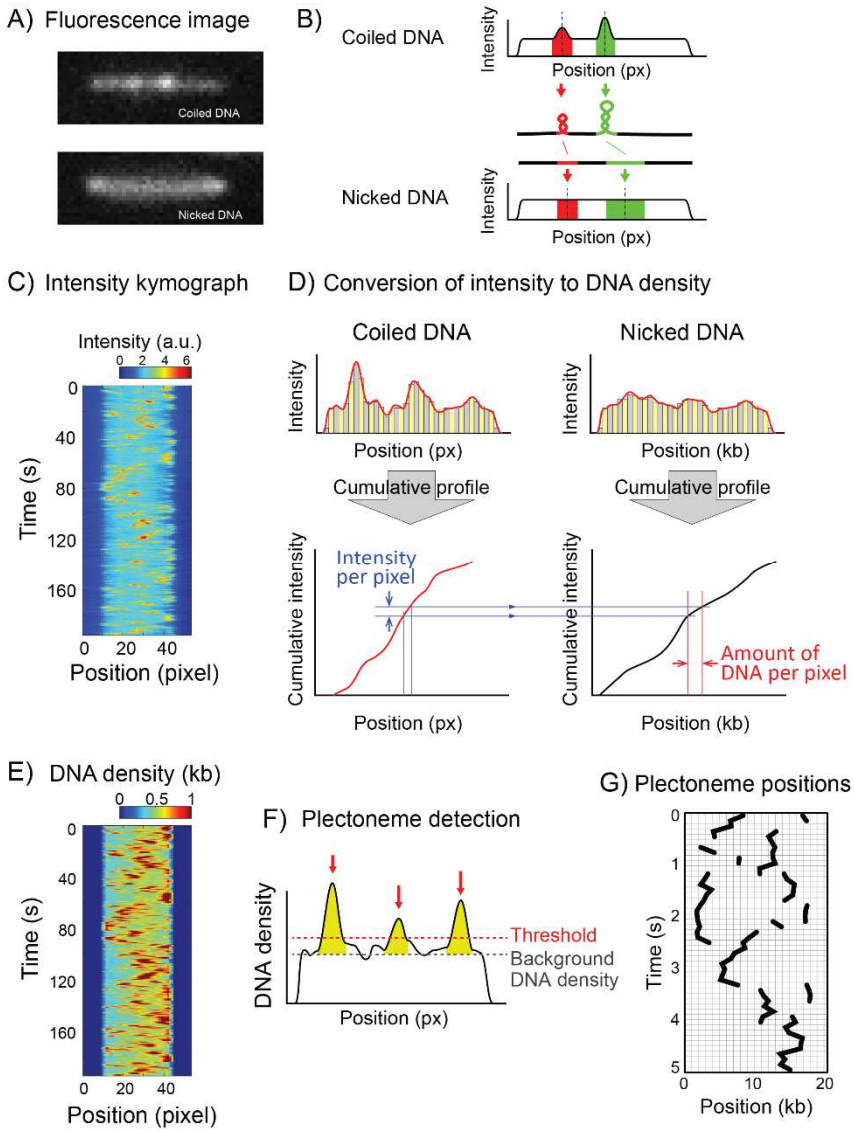
Supplementary Figures



Supplementary Figure 5.1**.** Schematic of the dual-color epi-fluorescence microscopy and the sample chamber (A) Microfluidic sample chamber prepared by sandwiching a double-sided tape between quartz slide and coverslip. Holes drilled in the quartz glass slide served as an inlet and outlet for buffer exchanging. (B) Dual-color epi-fluorescence microscopy consisting of two laser sources (532nm and 640 nm) and an EMCCD. (M: Mirror, DM: Dichroic Mirror, L: lens, LP: Long Pass filter, BP: Band Pass filter, AOTF: Acousto-Optical Tunable Filter)

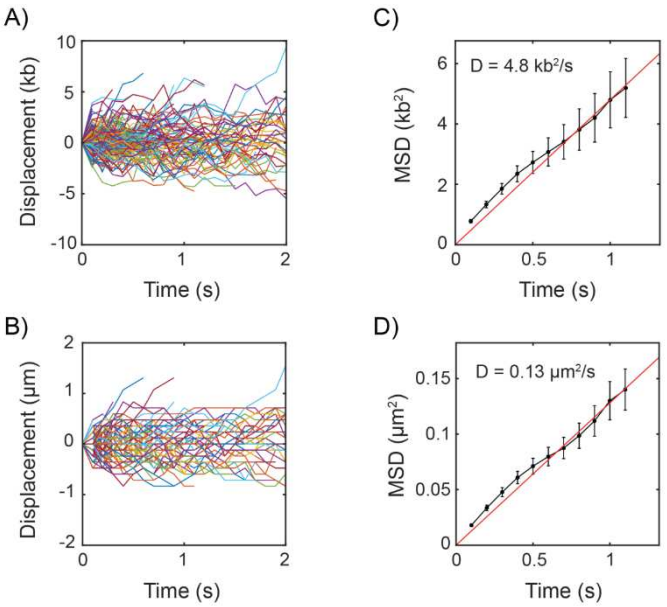


Supplementary Figure 5.2**.** Estimation of force and number of coils induced by SxO. (A) Length distributions of the DNA molecules shown in Figure 5.1E. Using the force-extension relationship in Figure 5.2D, the length was converted to force (top axis). (B) Fluorescence intensity profile of torsionally relaxed DNA molecules measured in the presence of SxO. Thin lines: individual molecules (N=46); Thick black line: averaged intensity profile. (C) Rotation curves measured at different concentrations of the SxO. Red lines are linear fits of the plateau and slope regions. Intercepts of the plateau and slope regions indicate the point where plectonemes start to form. Slopes determined from the linear fits are -0.052, -0.053, -0.053, -0.055 µm/turn for 0, 10, 30 50 nM of SxO, respectively.

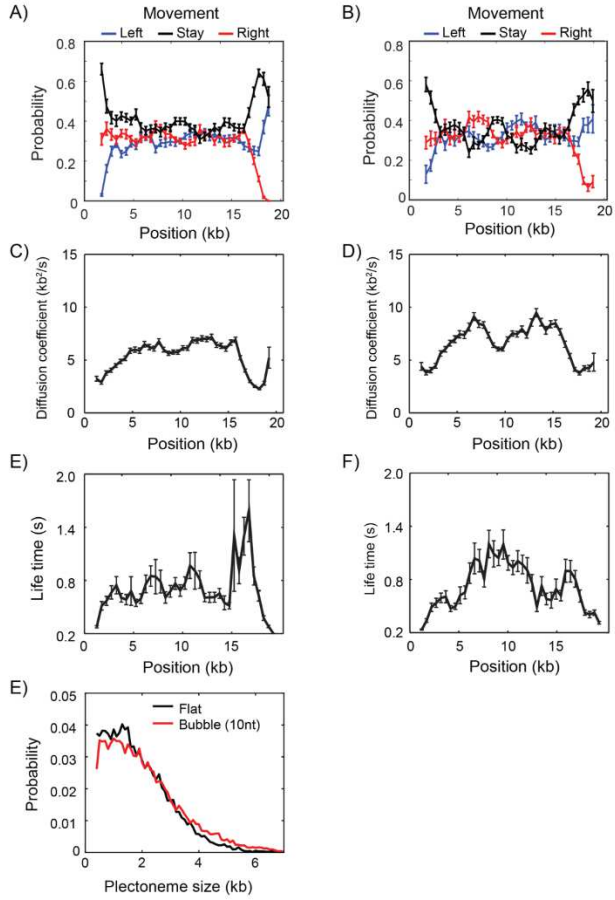


Supplementary Figure 5.3. Image analysis and position/size determination of individual plectonemes. (A) Representative fluorescence images of a supercoiled DNA (top) and the same DNA molecule (bottom) after torsional relaxation due the formation of a nick. (B) Schematic of a supercoiled DNA and its fluorescence intensity profile (top). Two plectonemes (red and green) are shown with its positions in the intensity profile. The DNA that was contained in the two plectonemes is also shown on the torsionally relaxed DNA (bottom). (C) Fluorescence intensity kymograph measured from the supercoiled DNA molecule. (D) Schematic showing the conversion procedure from the raw data to the DNA density. Due to a slightly non-homogeneous illumination profile and varying detection efficiencies in the image area, the intensity of each pixel cannot be directly converted to the amount of DNA. To convert the measured fluorescence intensity to amount of DNA in each

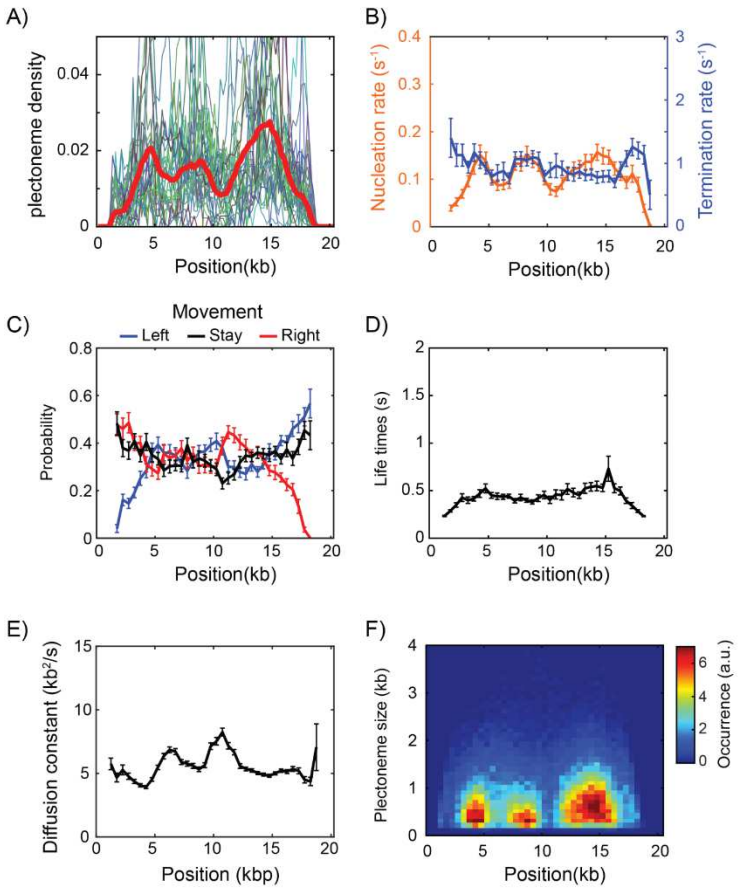
pixel, i.e. DNA density, we first obtained fluorescence intensities of each region of the DNA by measuring the intensity profile of the same DNA molecule after nicking. Note that, for the nicked-DNA case, the DNA is homogeneously distributed among the pixels so that the pixel position *can* be directly converted to the base-pair position. Hence, by comparing the area of the left-hand side of each pixel in the coiled DNA intensity profile, i.e., from the cumulative intensity, to that of the nicked DNA, we can determine the amount of DNA in number of base pairs that is contained in each pixel for the coiled DNA. (E) Reconstructed DNA density kymograph that is obtained from the intensity kymograph in C by using the conversion procedure in D. (F) For each frame in the DNA density kymograph, a threshold is applied to detect the position and size of individual plectonemes. (G) Example of position-time trajectories of plectonemes determined from the threshold algorithm.



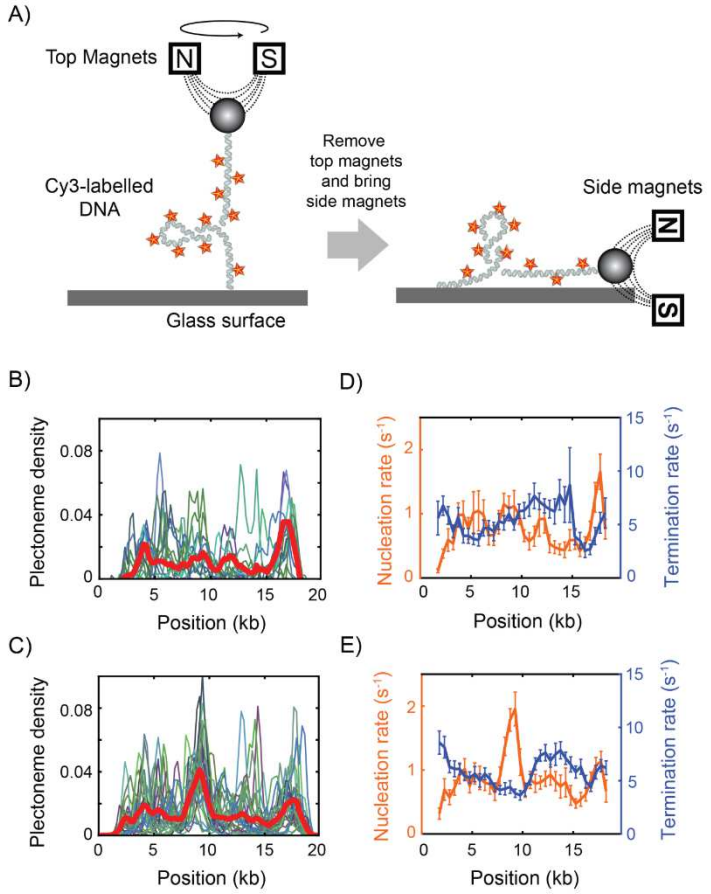
Supplementary Figure 5.4**.** Analysis of individual plectoneme dynamics. (A-B) Displacement-time trajectories of individual plectonemes (N=294) in base-pair space (A) and in real-space (B). (C-D) Mean-square-displacement (MSD) in base-pair space (C) and in real-space (D) obtained from the corresponding time trajectories as in (A-B). The diffusion coefficients (*D*) were determined from the slopes of the linear fits to the MSD (red lines) plots.



Supplementary Figure 5.5**.** Comparison of the position-dependent plectoneme properties with and without 10-nucleotide mismatch sequence. (A-B) Probability of the plectoneme movement along the DNA without (A) and with (B) the 10-nucleotide mismatch sequence. If a plectoneme is moved at least 500bp leftward/rightward in the next frame (i.e. within 100 ms), we regarded the plectoneme as moved to 'Left/Right'. If the distance was smaller than 500bp, we counted it as 'Stay'. The window size of 500bp and the time span of 100 ms is empirically chosen based on the spatial resolution and signal-to-noise ratio of our microscope. Errors are standard errors of the mean. (C-D) Position-dependent diffusion coefficients measured on the DNA without (C) and with (D) 10-nulceotide mismatch sequence. To obtain these local diffusion coefficients (*D*), we first calculated plectoneme displacements (*dx*) at each position of the DNA for a time interval of two successive frames (*dt*=100 ms). We then used following equation to obtain the diffusion coefficient: $<dx^2>=2Ddt$. Errors are standard errors of the mean. (E-F) Averaged life times of plectonemes plotted versus the position where the plectoneme was measured. (E): without and (F) with 10-nucleotide mismatch sequence. Errors are standard errors of the mean. (G) Size distribution of the plectonemes detected at the mismatched sequence (red). For comparison, the size distribution at the same region of the perfectly matched sequence was plotted together (black). Slightly larger plectonemes are observed at the mismatch sequence.



Supplementary Figure 5.6. Plectonemes on negatively supercoiled DNA molecules. (A) Plectoneme densities obtained from negatively supercoiled DNA molecules (thin lines, N=42) and their average (red thick line). (B) Averaged nucleation (orange) and termination (blue) rates observed. Errors are standard errors of the mean. (C) Probability of the plectoneme movement along DNA. Errors are standard errors of the mean (D) Averaged life times of plectonemes plotted versus the position where the plectoneme was measured. Errors are standard errors of the mean. (E) Position-dependent diffusion coefficients. Errors are standard errors of the mean. (F) Averaged position-dependent plectoneme size distributions.



Supplementary Figure 5.7**.** Plectoneme formation and dynamics observed with the side-pulling magnetic tweezers combined with fluorescence microscopy. (A) A schematic diagram of the side-pulling magnetic tweezers. (Left) Fluorescently labeled DNA molecules is immobilized on the glass surface at its one end. The other end of the DNA is bound to a magnetic bead which is pulled and rotated by using a pair of magnets above. (Right) After applying rotations to the DNA, the top magnets are removed while another magnet is brought from the side. The side magnets stretch the DNA on the surface for fluorescence imaging. (B-C) Individual (thin lines) and averaged (red thick line) plectoneme densities measured on (B) the perfectly matched sequence as in Figure 5.4A, and (C) the 10-nt mismatched sequence as in Figure 5.4B. (D-E) Nucleation and termination rates measured on the perfectly matched sequence and (D) the 10-nt mismatched sequence (E).

6
Sequence-Dependent Plectoneme Pinning Along Supercoiled DNA

DNA supercoiling plays a vital role in cellular processes such as replication, transcription and chromosome organisation. For instance, supercoils generated during the transcription of a gene effect the expression of distant genes by deformation of three-dimensional structure of promoter or activator regions of that gene, suggesting an accumulation of plectonemes at certain sequences of DNA than the others. However, no such sequences have been identified to pin the plectonemes due to unavailability of a technique to probe. Using the technique described in the chapter-5, we studied the sequence-specific organisation of plectonemes on a supercoiled DNA at single-molecule level. For this we tested the effects of GC-content of DNA in localization of plectonemes to regions along DNA. Our results indicate that the low GC-content regions are very favourable positions for localization of plectonemes along supercoiled DNA. In addition, preliminary results suggest that the poly(A) tracts play a role in formation and localization of plectonemes.

S. H. Kim*, M. Ganji*, J. van der Torre, E. Abbondanzieri, and C. Dekker, to be submitted

6.1 Introduction

Organization of the genome is central to every cellular process because some parts of the genome are more demanding for biological needs (*1*). Supercoiling of DNA is the main driving force in condensation of bacterial genome leading to around 1000-fold compaction. Early electron microscopy studies revealed that the 4.6 mega bases long *Escherichia coli* chromosome is organized into 50-100 topological domains of which length varies between 20 to 100 kb (*2, 3*) in which DNA is condensed by supercoiling. In addition, supercoils exhibit transiently due to the molecular processes such as DNA replication and RNA transcription (*4-7*). Both *in vitro* and *in vivo* studies revealed that RNA transcription resulted in building up of positive supercoils ahead of RNA polymerase and negative supercoils behind of it (*6-9*). In *E. coli*, these plectonemes appeared to hamper RNA transcription. The transiently generated torsional stress was eventually removed by the activity of Topo I and gyrase resulted in the transcription bursting *(6)*. Moreover, the supercoiling state of DNA leads to alterations in the three-dimensional structure of DNA which in turn participate in regulating the transcription because these deformations allow for specific protein-DNA interactions (*9-11*). Thus, DNA supercoils play a vital role during every cellular process.

Understanding of plectoneme organization along the supercoiled DNA has been limited, mainly due to unavailability of experimental techniques to probe. Although, single-molecule techniques such as magnetic tweezers provide useful information by measuring end-to-end distance of a supercoiled DNA, these techniques do not permit to measure the position and the dynamics of the plectoneme (*8, 12-15*). In order to directly visualize the dynamics of plectonemes in real time, we developed a fluorescence based technique, Intercalation-induced Supercoiling of DNA (ISD) (see chapter-5) (*16*). In this chapter, we used ISD to understand the sequence dependent dynamic localization of DNA supercoils.

In chapter-5, we found that a 10-nucleotide mismatch localizes plectonemes along supercoiled DNA. In this chapter, we selectively altered the DNA sequences in order to understand the sequence dependent localization of plectonemes at single-molecule level. We first tested three different DNA sequences: 1) a sequence with relatively homogeneous GC-content (designated as FlatDNA), 2) a sequence with large fluctuation in GC-content (FluctDNA), and 3) an AT-rich DNA segment in the

middle of FlatDNA (PeakA-FlatDNA). Along this line, we then created five different AT-rich segments and inserted them in the middle of FlatDNA to test the plectoneme localization dependency on the length of AT- rich DNA segments. These DNA constructs allowed us to understand how the increased length of AT-rich regions modulate localization of plectonemes. We also analysed the dependence of individual kinetic parameters such as rate of nucleation, rate of termination, and movement of plectonemes on the AT-peak length along DNA.

In FlatDNA construct we identified a DNA segment (not AT-rich) that but pins the plectonemes. Analysis of the sequence indicated that there are several poly(A) tracts (up to 8-nucleotides in length) within this region. We then specifically removed the poly(A) tracts in this region and found that the poly(A) tracts were playing a key role in formation and localization of plectonemes.

6.2 Methods and Materials

6.2.1 Preparation of DNA

Relatively homogeneous GC-content DNA plasmids of length ~8. kb and 11 kb were expressed and purified from *E. coli* cells. Two DNA molecules of length 500 bp were synthesized by PCR in the presence of biotin-16-dUTP (Roche). We also incorporated aminoallyl-dUTP-Cy5 (Jena Bioscience) in one of the molecules. Each of the fragments were then ligated to one of the 8 kb and 11 kb DNA molecules using T4 DNA ligase (promega) after endonuclease digestion of plasmids. These fragments were digested to the desired fragments using different restriction enzymes and the fragments were gel purified. After purification of samples, the molecules were then ligated to obtain a ~21 kb DNA. While the biotins at both the ends help in surface tethering of DNA, the Cy5 at one of the ends help in distinguishing between the different ends. DNA molecule of different length AT-rich segments were inserted in the between 8 kb and 11 kb DNA fragments in order to obtain the AT-peak inserted flat DNA constructs. The varying GC-content DNA construct was expressed and purified as a 21 kb plasmid. The plasmid was then digested with restriction enzyme and biotin- and/or Cy5-DNA molecules (500 bp) were ligated at each end for surface immobilization.

Intensity profile of a supercoiled DNA is compared to the intensity profile of corresponding nicked DNA to determine the positions of plectonemes along the length of DNA. This information was further converted to obtain plectoneme density (see methods and materials of chapter-5 for more details).

In this chapter, we also calculated the plectoneme occupancy fraction in addition to plectoneme density. First, we estimated the total size of DNA in each plectoneme (showed with red color in Figure 6.1) in a given time frame. Then, the plectoneme occupancy fraction for each position along DNA was obtained by dividing the total number of occupied frames with total number of observed frames.

6.2.2 Determining the plectoneme density and occupancy along DNA

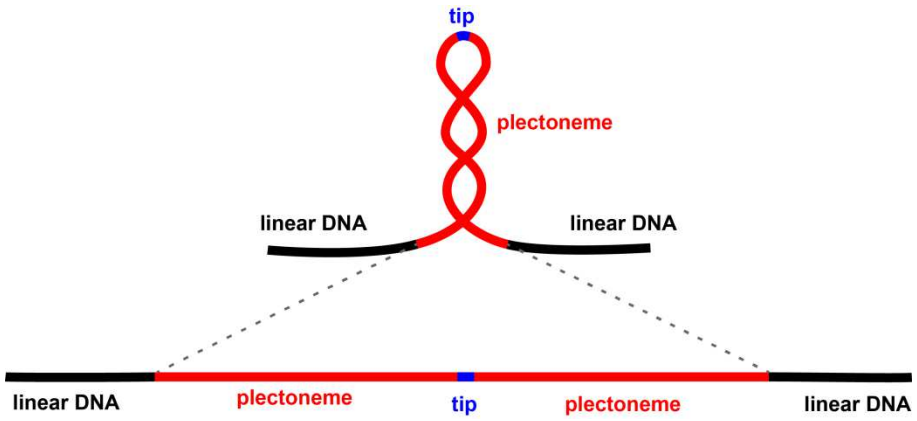


Figure 6.1**: Schematic diagram to show localization of plectonemes on linear DNA.** Plectonemic DNA (showed with red) was stretched to its equivalent linear length to locate the tip position of the plectoneme on linear DNA. Only tip position was accounted for calculating plectoneme density. Entire length of the plectonemic DNA was considered for calculating the plectoneme occupancy fraction along DNA.

6.3 Results

6.3.1 Plectoneme density pattern along DNA depends on the sequence of DNA

We wanted to study if the formation of the plectonemes and their dynamics vary with the DNA sequence. First, we tested three different DNA constructs with different GC compositions. The first DNA construct contained relatively homogeneous GC bases (~55%, FlatDNA, Figure 6.2A). The second DNA construct has a large fluctuation in the GC-content, varying between 30-60% (FluctDNA, Figure 6.2B). Third construct was the modified version of FlatDNA in which we inserted a 3 kb AT-rich DNA segment in the middle of the FlatDNA (9kb inside from the Cy5-end, PeakA-FlatDNA, Figure 6.2C).

We obtained the plectoneme occupancy profile from the double-tethered DNA molecules as described in chapter-5. The plectoneme occupancies of FlatDNA from 40 individual molecules together with their average are showed in Figure 6.2D. Clearly, plectonemes preferentially localize to certain positions along the length of the DNA (a peak at around 17.5 kb). However, the GC-content of the FlatDNA does not contain any corresponding feature (compare Figure 6.2A and D). Here we presented the plectoneme occupancy pattern for which we considered the size of the plectoneme (see Figure 6.1). We also measured the probability of plectoneme formation (plectoneme density) along the DNA by taking the position of the plectoneme at its tip. The results are presented in Supplementary Figure 6.2 which are in agreement with plectoneme occupancy pattern showed in Figure 6.2.

When we measured FluctDNA, the plectoneme occupancy and plectoneme density pattern was completely different than that of the FlatDNA (Figure 6.2E and Supplementary Figure 6.2). We observed two peaks in the plectoneme occupancy pattern. Interestingly, the valley in the plectoneme occupancy appeared at ~10kb where there is a high GC-rich segment (~60% GC). On the other hand, the region around at 12.5kb, where a peak appeared in the plectoneme occupancy pattern, turned out to be AT-rich. Likewise, the first peak around 3 kb also correlates the corresponding dip in GC-content. However, the low GC region between 5 kb-9 kb in FluctDNA was not observed to localize plectonemes. These results suggest that AT-rich regions correlate with the observed peaks in the plectoneme occupancies. In fact, when we plotted correlation between GC-content and plectoneme density along FluctDNA, we observed some correlation between the positions of AT-rich segments and plectoneme density (Supplementary Figure 6.4).

In order to further confirm the localization of plectonemes to AT-rich regions in the above results, we measured the plectoneme occupancy of PeakA-FlatDNA. We strategically placed a 3 kb AT-rich segment in the FlatDNA (i.e. the PeakA-FlatDNA) at around 9 kb where we did not observe any considerable preference for plectonemes. The measured plectoneme occupancy of 20 different molecules showed that the plectonemes preferentially localize at the AT-rich segment (Figure 6.3F), indicating that AT-rich region is a preferential position for plectoneme localization.

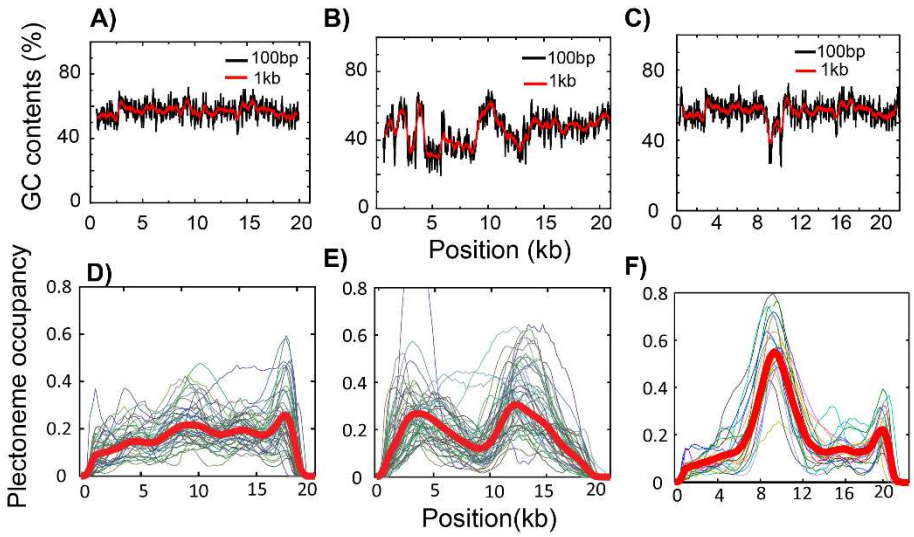


Figure 6.2**: Plectoneme density pattern on different DNA constructs. A)** Distribution of GC content along the length of FlatDNA binned to 100 bp (black) and 1000 bp (red). **B)** DNA construct (FluctDNA) with large fluctuations in GC content along its length. **C)** PeakA-FlatDNA construct with a 3 kb AT rich segment in the middle of FlatDNA. **D), E)**, and **F)** Thin lines are plectoneme density patterns measured with (D) FlatDNA, (E) FluctDNA and (F) PeakA-FlatDNA constructs. We measured 40, 120, and 20 molecules for FlatDNA, FluctDNA and PeakA-FlatDNA, respectively. Red thick lines represent average of the plectoneme densities from individual DNA molecules.

The DNA stain, i.e., binding of SxO, did not show any specificity for either AT-rich or GC-rich regions along DNA as was observed by very homogeneous intensity profile of DNA (Supplementary Figure 6.1), indicating that the observed localization of plectonemes was not influenced by binding of SxO. We also did not see any considerable alterations in the mechanical properties of DNA due to the binding of SxO (chapter-5).

6.3.2 Role of nucleation, termination and diffusion in localization of plectonemes

Our single-molecule fluorescence assay allows us to measure not only plectoneme occupancy pattern but also the kinetics of individual plectonemes. Using the same data used for plectoneme occupancy, we calculated the kinetic parameters such as the rate of nucleation and rate of termination along the length of DNA molecules. We also obtained the probability of plectoneme movement toward left, right or stay at given position along the length of DNA.

Nucleation of a plectoneme happens due to an appearance of a new plectoneme by simultaneous uncoiling of a plectoneme elsewhere (Figure 6.3A). On the other hand, a termination happens if a plectoneme disappears due to absorption of writhe by other faraway plectonemes. Note that the plectoneme nucleation and termination do not necessarily interrelated because an existing plectoneme may partially uncoil for nucleation of new plectoneme. Similarly, termination of a plectoneme may occur by transferring the writhe into an existing plectoneme.

We wanted to understand whether the plectoneme localization depends on the rate of nucleation and termination. The rates of nucleation and termination for FlatDNA are showed in Figure 6.3C. The results indicated that the rates of nucleation and termination show a peak at around 17.5 kb where plectoneme density also showed a peak (Figure 6.3D). On the contrary, termination rate was highest and nucleation rate was lowest from 1-5 kb where the plectoneme density was also the lowest. In line with this result, the nucleation rate of FluctDNA also showed peaks where the plectoneme density showed peaks (compare Figure 6.3D and Figure 6.2E). Interestingly, PeakA-FlatDNA also showed a predominant peak in the rate of nucleation in combination with a predominant dip in the rate of termination at around 10 kb (Figure 6.3E). Note that the plectoneme density also showed a predominant peak around this region (Figure 6.2F). Together, these results indicate that the higher rate of plectoneme nucleation is one of the reasons that assist in localization of the plectonemes.

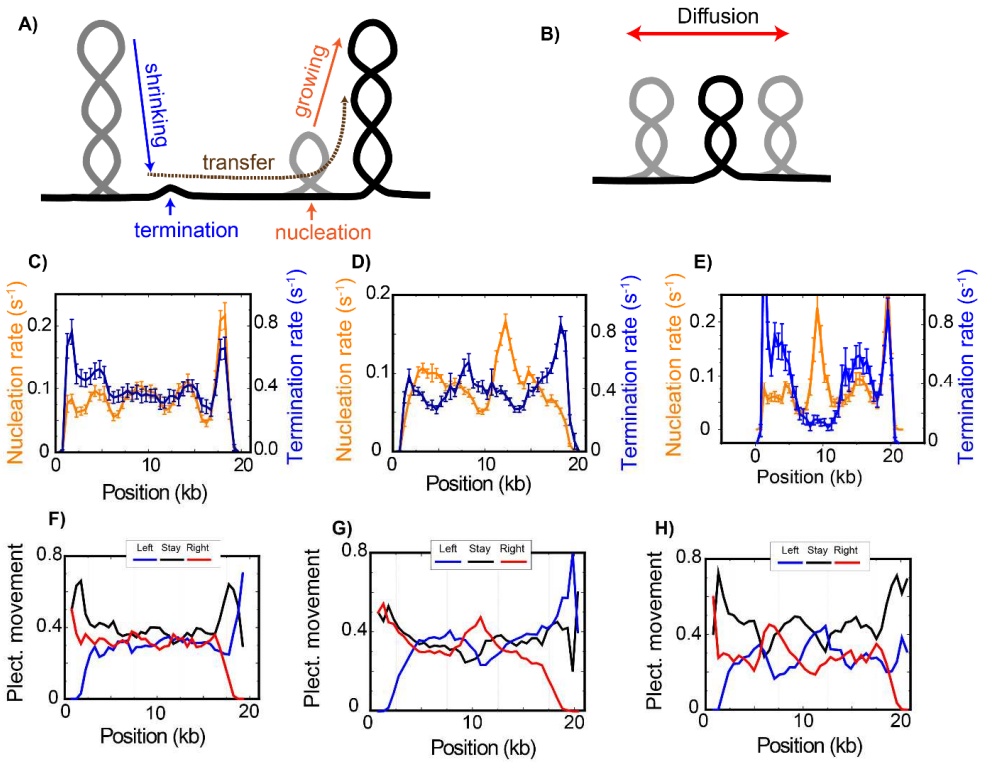


Figure 6.3**: Kinetics of DNA plectonemes- nucleation, termination and diffusion. A)** Schematic of DNA plectoneme nucleation and termination. A plectoneme grows after nucleation on DNA by simultaneous shrinking of another, faraway plectoneme which may eventually terminate. **B)** Schematic showing diffusion of plectoneme. **C), D), and E)** Rates of nucleation (orange) and termination (blue) of plectonemes along the length of DNA on FlatDNA, FluctDNA, and PeakA-FlatDNA. **F), G), and H)** Diffusion of plectonemes on FlatDNA, FluctDNA, and PeakA-FlatDNA.

Diffusion of plectoneme along DNA is another dynamic aspect of plectonemes (*17*) (Figure 6.3B). We analysed our single molecule data to test if the diffusion also contributes to the localization of plectonemes to different regions of DNA. At a given position along the DNA, a plectoneme can move either to right or left side, or stay at the same position. The movement of plectonemes along the FlatDNA were approximately constant (Figure 6.3F). In the case of FluctDNA and PeakA-FlatDNA, the plectonemes showed a higher tendency to both staying constant and moving inward from both the directions at positions where the localization was seen (Figure 6.3G and H).

There is a possibility that the DNA stain (i.e. SxO) might be binding favourably to either the plectonemic DNA or linear DNA which could lead to an artefact in

observed plectoneme localizations. In the case of such possible artefact we would expect unequal nucleation or termination events along DNA, meaning that the system is away from the thermodynamic equilibrium. In order rule out this we counted the total number of nucleation and termination events at every position of DNA during the full course of measurements. The results showed that the total number of nucleation and termination events are equal to each other at every position of DNA, indicating that the system was under thermodynamic equilibrium (Supplementary Figure 6.3).

Taken together, these results indicate both the plectoneme nucleation/termination processes and the diffusion processes contribute to the plectoneme localization.

6.3.3 Plectoneme localization relates to length of AT-rich region

In order to further establish the sequence dependent localization of plectonemes and test if localization of plectonemes is enhanced with the increased length of AT-rich segments, we introduced different length AT-rich segments into the FlatDNA (Figure 6.4A). We prepared five different AT-rich segments of size 0.25 kb (Peak 1), 0.5 kb (Peak 2), 1 kb (Peak 3), 3 kb (Peak 4), and 3.9 kb (Peak 5) and introduced them into FlatDNA at 8.5 kb away from Cy5 end (Figure 6.4A). The distributions of GC content along the length of DNA with AT-Peaks are showed in Figure 6.4B.

The plectoneme occupancy pattern for each of the AT-peaks showed that the localization of plectonemes were strongly dependent on the size of AT-peak (Figure 6.4C). The height of the peak in the plectoneme density at the position where AT-peaks were introduced increased with the size of AT-peaks, indicating that AT-rich segments of DNA are favourable positions for plectonemes. The rate of nucleation followed the same pattern as plectoneme density (Figure 6.4D) such that plectonemes nucleated more frequently at the position where AT-rich segments present. Interestingly, the rate of termination increased elsewhere than the position where the AT-peaks were present (Figure 6.4E).

The average plectoneme movement on DNA showed that the plectonemes move into the AT-rich segments (Figure 6.4F-J). The movement of plectonemes into the AT-peaks increased with the length of the AT-peaks. Moreover, the plectonemes also showed higher tendency to stay within the AT-peaks and it was dependent on the length of the AT-peak.

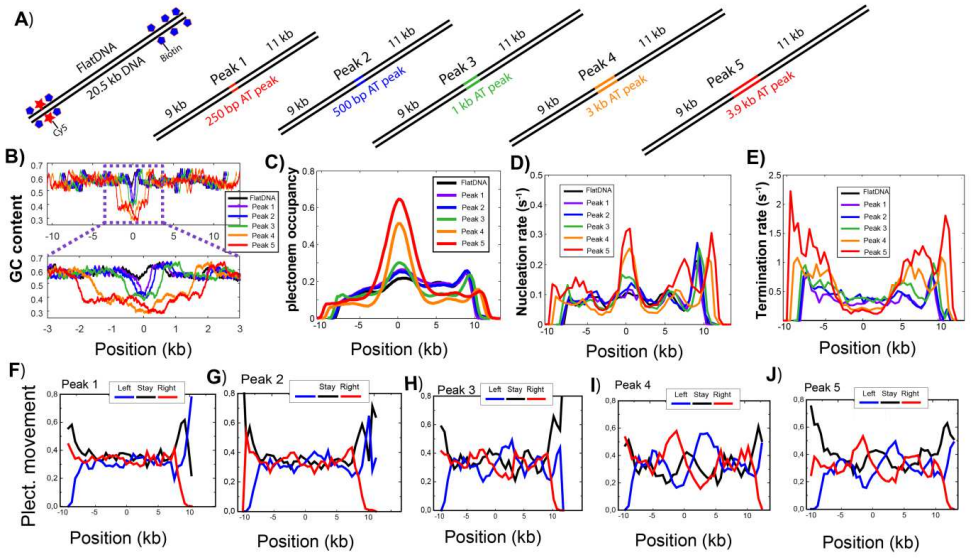


Figure 6.4**: localization of plectoneme by AT-rich sequences**. **A)** Schematic diagrams of DNA constructs with the indicated length of the AT-segment introduced into FlatDNA. **B)** GC content of FlatDNA with 3.9 kb AT-segment. The color lines in the centre represent the length and position of each AT-segment (red- 0.25 kb, orange-0.5 kb, green-1 kb, cyan-3 kb, and blue-4 kb). **C)**, **D)**, and **E)** Plectoneme density (C), the rate of nucleation (D), and the rate of termination (D) versus position of DNA for all five AT peaks together with FlatDNA. The color convention is the same as in (B) and FlatDNA showed in grey color. **F-J)** Probabilities of plectoneme movement along DNA with (F) 250 bp (peak 1), (G) 500 bp (peak 2), (H) 1 kb (peak 3), (I) 3 kb (peak 4) and (J) 3.9 kb (peak 5) AT peak–FlatDNA molecule. Blue curve represents the probability of a plectoneme movement to left side, black represents the plectonemes that do not move in position, and red represent movement to right side.

Altogether, these results suggest that nucleation of plectonemes happen more frequent at AT-rich regions and the AT-rich regions are favourable positions for plectonemes to stay longer which resulted in a decrease in the rate of termination. The plectonemes also showed higher tendency to diffuse in from elsewhere.

Next, we analysed the size of plectonemes for each of the five AT-rich segment inserted constructs. We expected the size of plectonemes to increase with the length of the AT-peak as plectonemes strongly localize to the AT-peaks. For this, we built a 2-D histogram of plectonemes with their sizes and positions (Figure 6.5A-F). The results showed a consistent increase in the size of plectonemes with the size of the

AT-peaks at the position where the AT-peaks were inserted (between red lines in Figure 6.5A-F).

We then looked in to the fraction number of plectonemes at AT-rich region compared to the throughout the length of DNA constructs. The size of plectonemes that appears around the AT-rich inserts increased with the width of the AT-peak (Figure 6.5G). These results indicate that the size of the plectoneme indeed grows with increased length of the AT-rich segment.

6.3.4 Poly(A) tracts assist for localization of plectonemes

The above results strongly indicate that the AT-rich regions are preferential positions for localization of plectonemes. However, the reason behind the peak in the plectoneme density pattern of FlatDNA around 17.5 kb is not clear. Upon looking into the sequence in this region, we found that the sequence with in this region contained multiple poly(A) tracts which could be helping in pinning the plectonemes. We then studied the effect of the poly(A) tracts on plectoneme pinning by inserting this known sequence with and without the tracts in the centre of FlatDNA (around 9 kb from Cy5-end).

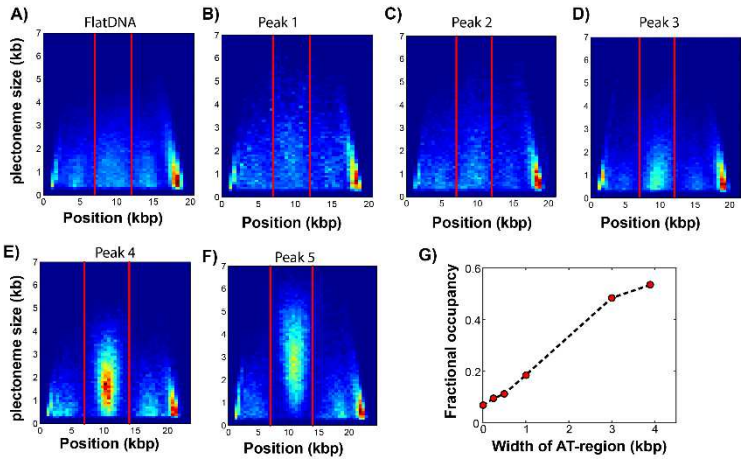


Figure 6.5**: Plectoneme size distributions on DNA molecules with different-sized AT-peaks**. **A)** Position-dependent size distribution of plectonemes with FlatDNA. The of plectonemes number is represented with color, blue represents the least number and the dark red represents highest number. **B)-F)** Plectonemes size distribution on Peak 1, Peak 2, Peak 3, Peak 4 and Peak 5. The color convention is the same as in (A). **G)** Plectoneme occupancy distribution found near the AT-rich regions in (A) to (F). Red vertical lines in (A)-(F) show the AT-rich regions.

First we copied a 1 kb DNA from 17 kb-18 kb region of FlatDNA where three poly(A) tracts were present (Supplementary Information). We then mutated few nucleotides in the poly(A) tracts in order to get rid of them.

We then measured the plectoneme occupancy on each of the constructs. We observed a strong peak in the plectoneme occupancy pattern around 9 kb for the 1 kb DNA insert with poly(A) tracts (orange and red curves-Figure 6.6A) compared to the FlatDNA (black curve-Figure 6.6A). The control construct in which we inserted the same 1 kb DNA in the middle of FlatDNA but without any poly(A) showed no preference for plectonemes. In line with these results, the rate of nucleation showed a peak at the region where we inserted DNA with the poly(A) tracts but not on the construct without the poly(A) tracts. The rate of termination was also showed an agreement with these results. Taken together, these results indicate that the poly(A) tracts assist in the pinning of plectonemes.

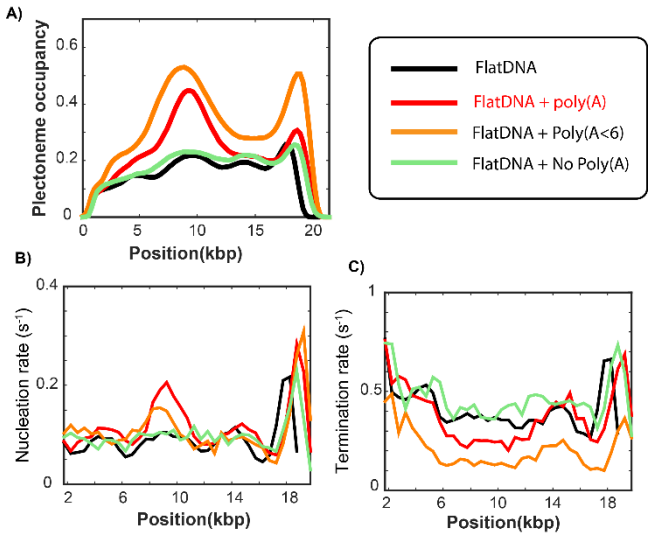


Figure 6.6: **Plectonemes pinning by poly(A) tracts along supercoiled DNA**. A), B), and C) Plectoneme occupancy, the rate of nucleation, and the rate of termination pattern of FlatDNA, FlatDNA with poly(A) tracts, FlatDNA with poly(A<6) tracts, and FlatDNA with no poly(A) tracts. A legend is presented to identify each of the curves represented.

6.4 Discussion

Plectoneme, a coiled loop structure of DNA, forms when a topologically constrained DNA is situated under torsional stress. The plectoneme diffuse along a linear DNA

which helps in bringing two faraway sites close to each other. In addition, a plectoneme disappears from one position, which is accompanied by appearance of another plectoneme at a distant position (*17*). In the current study, by using intercalation-based single-molecule fluorescence assay, we showed that the dynamics of plectonemes directly correlated to the sequence of DNA sequence and confirming that the AT-rich DNA segments are favourable positions for the localization of plectonemes.

By measuring the plectonemes on three different DNA constructs, flat GC content DNA, fluctuating GC content DNA, and flat GC content DNA with an AT-rich segment in the middle, we observed an anti-correlation between plectoneme density and the GC-content of the DNA constructs. Furthermore, we tested AT-rich segments with different lengths and compositions to confirm that the AT-rich regions are favourable positions for the formation of plectonemes.

Our single-molecule technique is capable to measure the nucleation and termination rates, and plectonemes movement along DNA. The results indicated that the rate of nucleation increased within the AT-rich segments without much affecting outside the AT-peaks and this increase was strongly dependent on the length of AT-rich segment. While the rate of termination was suppressed strongly at AT-rich segments, it increased outside the AT-peaks along DNA. Plectonemes were also observed to move towards the AT-rich segments. In addition, we also observe that the size of the plectonemes increased with increased length of AT-rich segments. These results synergetically lead to the observed plectoneme density patterns.

Although, our results clearly indicate that the plectonemes prefer to localize at AT-rich regions along DNA, the appeared peak in the plectoneme density on FlatDNA around 17.5 kb was not understood. Since poly(A)-tracts are known to induce an intrinsic bend in DNA molecules (*18-21*), we analysed the sequence of FlatDNA to identify the positions of poly(A)-tracts. We found that there are few poly(A)-tracts present around this region (Supplementary Information). Our preliminary results indicated that the poly(A)-tracts assist in pinning of plectonemes (Figure 6.6). However, further research is needed to ascertain the effects of poly(A) tracts on plectoneme localization. For example, if poly(A)-tracts really play a role, DNA construct with one or few poly(A)-tracts inserted in to the mutated DNA construct, should pin the plectonemes. This well controlled measurement will facilitate to state whether the poly(A)-tracts are favourable position for plectoneme localization or not. Therefore, our future research will be focused to understand the role of poly(A)-tracts on the localization of plectonemes.

6.5 References

1. D. J. Sherratt, Bacterial Chromosome Dynamics. *Science* 301, 780 (2003).

2. R. Kavenoff, O. Ryder, Electron microscopy of membrane-associated folded chromosomes of Escherichia coli. *Chromosoma (Berl.)* 55, 12 (1976).

3. R. Kavenoff, B. C. Bowen, Electron microscopy of membrane-free folded chromosomes from Escherichia coil *Chromosoma (Berl.)* 59, 12 (1976).

4. J. Roca, The torsional state of DNA within the chromosome. *Chromosoma* 120, 323 (2011).

5. J. C. Wang, DNA Topoisomerases. *Annual Review of Biochemistry* 54, 665 (1985).

6. S. Chong, C. Chen, H. Ge, X. S. Xie, Mechanism of Transcriptional Bursting in Bacteria. *Cell* 158, 314 (2014).

7. L. F. Liu, J. C. Wang, Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences* 84, 7024 (1987).

8. J. Ma, L. Bai, M. D. Wang, Transcription Under Torsion. *Science* 340, 1580 (2013).

9. F. Kouzine *et al.*, Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat Struct Mol Biol* 20, 396 (2013).

10. C. Lavelle, DNA torsional stress propagates through chromatin fiber and participates in transcriptional regulation. *Nat. Struct. Mol. Biol.* 15, 123 (2008).

11. L. Baranello, D. Levens, A. Gupta, F. Kouzine, The importance of being supercoiled: How DNA mechanics regulate dynamic processes. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1819, 632 (2012).

12. T. R. Strick, J. F. Allemand, D. Bensimon, V. Croquette, Behavior of supercoiled DNA. *Biophys. J.* 74, 2016 (Apr, 1998).

13. J. Lipfert, M. Lee, O. Ordu, J. W. J. Kerssemakers, N. H. Dekker, Magnetic Tweezers for the Measurement of Twist and Torque. (2014).

14. F. Mosconi, J. F. Allemand, D. Bensimon, V. Croquette, Measurement of the Torque on a Single Stretched and Twisted DNA Using Magnetic Tweezers. *Physical Review Letters* 102, 078301 (2009).

15. H. Brutzer, N. Luzzietti, D. Klaue, R. Seidel, Energetics at the DNA Supercoiling Transition. *Biophysical Journal* 98, 1267 (2010).

16. M. Ganji, S. H. Kim, J. van der Torre, E. Abbondanzieri, C. Dekker, Intercalation-Based Single-Molecule Fluorescence Assay To Study DNA Supercoil Dynamics. *Nano Lett.* 16, 4699 (2016/07/13, 2016).

17. M. T. J. van Loenhout, M. V. de Grunt, C. Dekker, Dynamics of DNA Supercoils. *Science* 338, 94 (2012).

18. H. C. M. Nelson, J. T. Finch, B. F. Luisi, A. Klug, The structure of an oligo(dA)·oligo(dT) tract and its biological implications. *Nature* 330, 221 (11/19/print, 1987).

19. C. Rivetti, C. Walker, C. Bustamante, Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility1. *J. Mol. Biol.* 280, 41 (7/3/, 1998).

20. K. Tóth, V. Sauermann, J. Langowski, DNA Curvature in Solution Measured by Fluorescence Resonance Energy Transfer. *Biochemistry* 37, 8173 (1998/06/01, 1998).

21. A. Brunet *et al.*, Probing a label-free local bend in DNA by single molecule tethered particle motion. *Nucleic Acids Res.*, (March 12, 2015, 2015).

Supplementary Information:

1 kb DNA with several poly(A) tracts from FlatDNA (17-18 kb region) was copied to the centre of it. Subsequently, the nucleotides in these tracts were mutated to get rid of the poly(A) tracts step-by-step. The original poly(A) tracts were highlighted with yellow color and the mutated nucleotides were showed with red font. These DNA molecules were purchased from IDT technologies (gene blocks).

1) **FlatDNA + poly(A)**
   original 1kb DNA sequence from FlatDNA (17.5 kb to 18.5 kb)

CCGCTACGAAATGCGCGTATGGGGATGGGGGCCGGGTGAGGAAAGCTG
GCTGATTGACCGGCAGATTATTATGGGCCGCCACGACGATGAACAGACG
CTGCTGCGTGTGGATGAGGCCATCAATAAAACCTATACCCGCCGGAATG
GTGCAGAAATGTCGATATCCCGTATCTGCTGGGATACTGGCGGGATTGA
CCCGACCATTGTGTATGAACGCTCGAAAAAACATGGGCTGTTCCGGGTG
ATCCCCATTAAAGGGGCATCCGTCTACGGAAAGCCGGTGGCCAGCATGC
CACGTAAGCGAAACAAAAACGGGGTTTACCTTACCGAAATCGGTACGGA
TACCGCGAAAGAGCAGATTTATAACCGCTTCACACTGACGCCGGAAGGG
GATGAACCGCTTCCCGGTGCCGTTCACTTCCCGAATAACCCGGATATTTT
TGATCTGACCGAAGCGCAGCAGCTGACTGCTGAAGAGCAGGTCGAAAAA
TGGGTGGATGGCAGGAAAAAAATACTGTGGGACAGCAAAAAGCGACGC
AATGAGGCACTCGACTGCTTCGTTTATGCGCTGGCGGCGCTGCGCATCAG
TATTTCCCGCTGGCAGCTGGATCTCAGTGCGCTGCTGGCGAGCCTGCAGG

AAGAGGATGGTGCAGCAACCAACAAGAAAACACTGGCAGATTACGCCC
GTGCCTTATCCGGAGAGGATGAATGACGCGACAGGAAGAACTTGCCGCT
GCCCGTGCGGCACTGCATGACCTGATGACAGGTAAACGGGTGGCAACAG
TACAGAAAGACGGACGAAGGGTGGAGTTTACGGCCACTTCCGTGTCTGA
CCTGAAAAAATATATTGCAGAGCTGGAAGTGCAGACCGGCATGACACAG
CGACGCAGGGGACCTGCAGGATTTTATGTATGAAAACGCCCACCATTCC
CACCCTTCTGGGGCCGGACGGCATGACATCGCTGCGCGAATATGCCGGT
TATCACGGCGGTGGCAGCGGA

2 FlatDNA + Poly(A<6)

Those poly(A) tracts that are longer than six nucleotides were mutated. The mutated nucleotides were showed in red font in the yellow highlighted region.

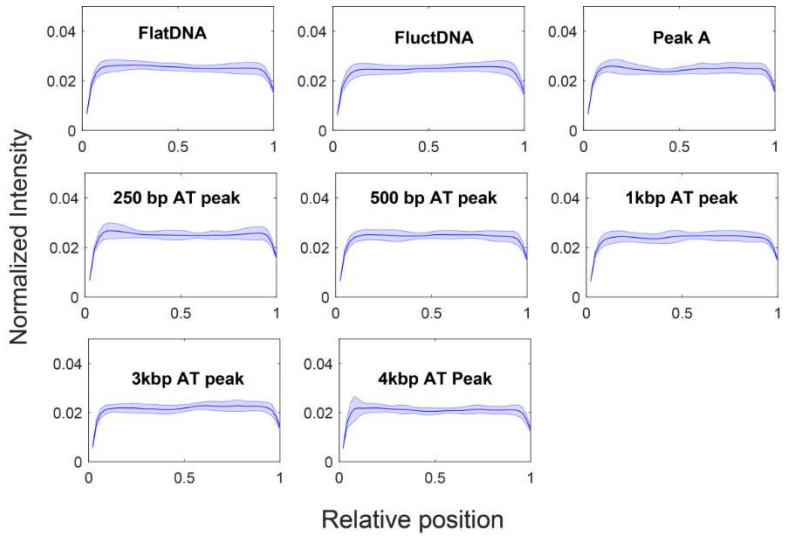
CCGCTACGAAATGCGCGTATGGGGATGGGGGCCGGGTGAGGAAAGCTG
GCTGATTGACCGGCAGATTATTATGGGCCGCCACGACGATGAACAGACG
CTGCTGCGTGTGGATGAGGCCATCAATAAAACCTATACCCGCCGGAATG
GTGCAGAAATGTCGATATCCCGTATCTGCTGGGATACTGGCGGGATTGA
CCCGACCATTGTGTATGAACGCTCGAAGCAACATGGGCTGTTCCGGGTG
ATCCCCATTAAAGGGGCATCCGTCTACGGAAAGCCGGTGGCCAGCATGC
CACGTAAGCGAAACAAAAACGGGGTTTACCTTACCGAAATCGGTACGGA
TACCGCGAAAGAGCAGATTTATAACCGCTTCACACTGACGCCGGAAGGG
GATGAACCGCTTCCCGGTGCCGTTCACTTCCCGAATAACCCGGATATTTT
TGATCTGACCGAAGCGCAGCAGCTGACTGCTGAAGAGCAGGTCGAAAAA
TGGGTGGATGGCAGGAGCAACATACTGTGGGACAGCAAAAAGCGACGC
AATGAGGCACTCGACTGCTTCGTTTATGCGCTGGCGGCGCTGCGCATCAG
TATTTCCCGCTGGCAGCTGGATCTCAGTGCGCTGCTGGCGAGCCTGCAGG
AAGAGGATGGTGCAGCAACCAACAAGAGCACACTGGCAGATTACGCCC
GTGCCTTATCCGGAGAGGATGAATGACGCGACAGGAAGAACTTGCCGCT
GCCCGTGCGGCACTGCATGACCTGATGACAGGTAAACGGGTGGCAACAG
TACAGAAAGACGGACGAAGGGTGGAGTTTACGGCCACTTCCGTGTCTGA
CCTGAGCACATATATTGCAGAGCTGGAAGTGCAGACCGGCATGACACAG
CGACGCAGGGGACCTGCAGGATCCTATGTATGAAAACGCCCACCATTCC
CACCCTTCTGGGGCCGGACGGCATGACATCGCTGCGCGAATATGCCGGT
TATCACGGCGGTGGCAGCGGA

3 FlatDNA + No Poly

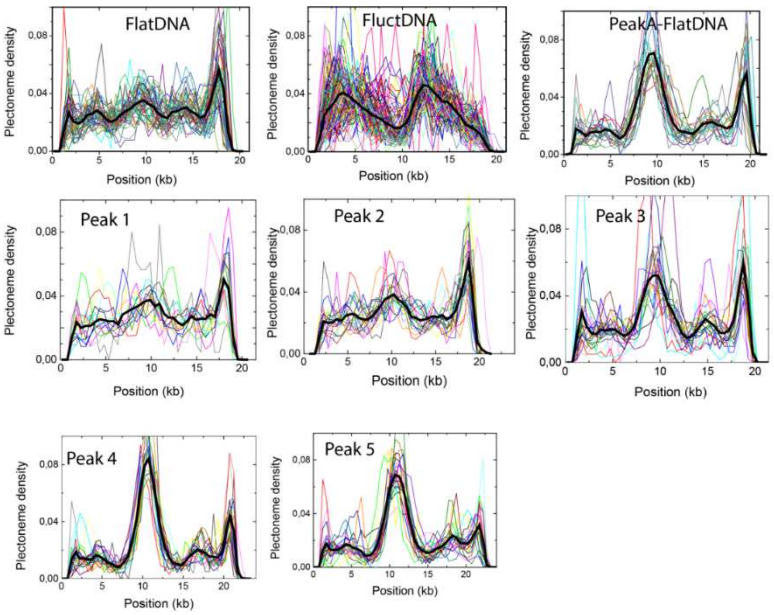
Those poly(A) tracts that are longer than four nucleotides were mutated. The mutated nucleotides were showed in red font in the yellow highlighted region.

CCGCTACGAAATGCGCGTATGGGGATGGGGGCCGGGTGAGGAAAGCTG
GCTGATTGACCGGCAGATTATTATGGGCCGCCACGACGATGAACAGACG
CTGCTGCGTGTGGATGAGGCCATCAATACGACCTATACCCGCCGGAATG
GTGCAGAAATGTCGATATCCCGTATCTGCTGGGATACTGGCGGGATTGA
CCCGACCATTGTGTATGAACGCTCGAAGCAACATGGGCTGTTCCGGGTG
ATCCCCATTAAAGGGGCATCCGTCTACGGAAAGCCGGTGGCCAGCATGC
CACGTAAGCGAAACAACGACGGGGTTTACCTTACCGAAATCGGTACGGA
TACCGCGAAAGAGCAGATTTATAACCGCTTCACACTGACGCCGGAAGGG
GATGAACCGCTTCCCGGTGCCGTTCACTTCCCGAATAACCCGGATATTGC
TGATCTGACCGAAGCGCAGCAGCTGACTGCTGAAGAGCAGGTCGAGCAA
TGGGTGGATGGCAGGAGCAACATACTGTGGGACAGCAACGAGCGACGC
AATGAGGCACTCGACTGCTTCGTTTATGCGCTGGCGGCGCTGCGCATCAG
TATTTCCCGCTGGCAGCTGGATCTCAGTGCGCTGCTGGCGAGCCTGCAGG
AAGAGGATGGTGCAGCAACCAACAAGAGCACACTGGCAGATTACGCCC
GTGCCTTATCCGGAGAGGATGAATGACGCGACAGGAAGAACTTGCCGCT
GCCCGTGCGGCACTGCATGACCTGATGACAGGTAAACGGGTGGCAACAG
TACAGAAAGACGGACGAAGGGTGGAGTTTACGGCCACTTCCGTGTCTGA
CCTGAGCACATATATTGCAGAGCTGGAAGTGCAGACCGGCATGACACAG
CGACGCAGGGGACCTGCAGGATCCTATGTATGAGCACGCCCACCATTCC
CACCCTTCTGGGGCCGGACGGCATGACATCGCTGCGCGAATATGCCGGT
TATCACGGCGGTGGCAGCGGA

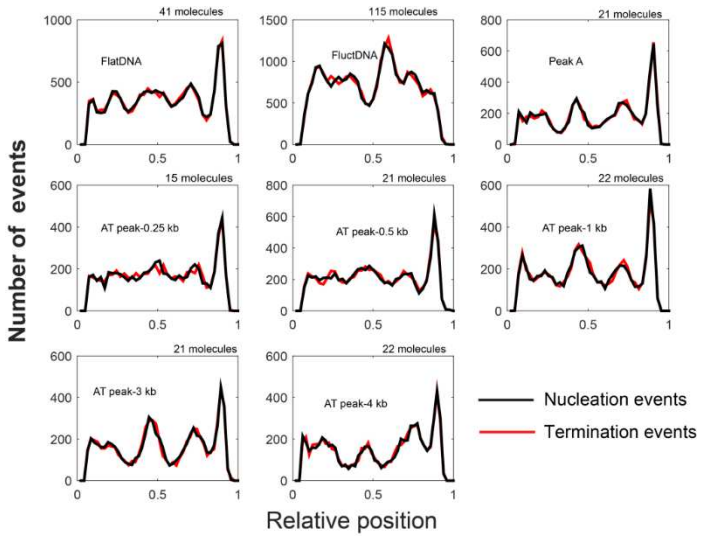
Supplementary Figures



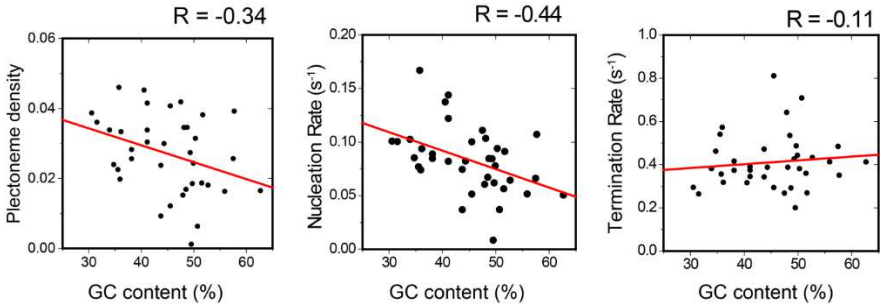
Supplementary Figure 6.1**: Nicked DNA intensity profile upon labelling with Sytox orange.** The name of the DNA construct is indicated in the corresponding plot. Light blue color indicates standrd deviation and thick blue line indicates the mean of the data.



Supplementary Figure 6.2**: Plectoneme density plots.** The name of the DNA construct is indicated in the corresponding plot. Thin color lines indicate plectoneme density from individual molecules and thick black line indicates the mean of the data.



Supplementary Figure 6.3**: Total number of observed nucleation and termination events.** The name of the DNA construct and number of analysed molecules in each case are indicated in the corresponding plot.



Supplementary Figure 6.4**:** Correlation plots between the GC content and (left) the plectoneme density, (middle) the nucleation rate and (right) the termination rate of FluctDNA.

7
Interaction of Dps with Different DNA Topologies

DNA-binding protein from starved cells (Dps) is known to bind DNA cooperatively and compact DNA by forming a dense DNA-Dps biocrystal. In this chapter, we used Intercalation-induced Supercoiled DNA (ISD) technique (see chapter 5) to visualize the interactions of Dps with supercoiled DNA and other DNA topologies. In ISD, we obtain doubly tethered supercoiled DNA, doubly tethered linear DNA, and singly tethered DNA molecules. This versatile technique allowed us to test the activity of Dps on all these different DNA conformations under the same buffer condition. Our results clearly indicate that binding of Dps requires two DNA strands close to each other such as plectonemic DNA and Dps shows no affinity to stretched linear DNA. We also found that Dps can mediate the cross-linking of two DNA molecules. Additionally, we found that Dps-DNA assembly can accommodate extra DNA indicating that Dps-Dps interactions are also present in the assembly. Based on these observations we provide a model describing formation of DNA-Dps complexes.

M. Ganji, N. Vtyurina, M. Docter, E. Abbondanzieri, to be submitted

7.1 Introduction

Supercoiling of the DNA molecule is a major force driving the compaction of bacterial genome (1-4). DNA supercoiling occurs in a topologically closed DNA when the helix undergoes over- or under- winding due to increase in writhe. While over-winding leads to positive supercoiling, under-winding leads to negative supercoiling. Additionally, supercoiled DNA may form three possible structures: a plectoneme, a toroid, or a combination of both. Plectonemes are more common in nature and take part in shaping most bacterial chromosomes.

Bacterial genomes consist of a single circular DNA molecule that is negatively supercoiled contributing to around thousand-fold compaction of the chromosome inside the cell (5). Additionally, the chromosome is constrained by different nucleoid associated proteins (NAPs) into topological domains of 10-100 kbp size (1,2). By binding to multiple DNA sites, NAPs stabilize particular supercoil-dependent structures of the DNA, and also create different local conformations through twisting, bridging, and DNA bending (1).

Despite the complex bacterial genome packaging, regulatory sites of the DNA that are essential for life processes must be made accessible to the transcription and replication machineries. Because of the interplay between supercoiling and genome packaging, any large variations in the supercoiled state of DNA can affect gene expression with severe consequences for the bacterium (3). Therefore, DNA supercoiling should be recognized as a vitally important component of chromosome packaging and the regulation of the cellular processes.

It was found that the extent to which each of the NAPs contributes to DNA supercoiling depends on the growth phase and physiological state of the cell (4,5). This feature makes the level of DNA supercoiling very dynamic and environmentally regulated (4,6). For instance, in response to environmental stresses, the physiological states of the bacterial cell change for the sake of survival and adaptation. One of the responses includes an upregulation of one of the NAPs, Dps, that compacts DNA yet keeps the genome accessible for life sustaining processes like transcription. Another response includes a change in DNA supercoiling state and, subsequently, was found to influence expression of genes (7-9).

As both DNA supercoiling and Dps contribute to the physiological response of bacteria on the environmental changes, a detailed understanding of interplay between DNA supercoiling and Dps protein is required. These findings can shed a light on the mechanisms that modulate transcription during challenging conditions

for bacterial survival and provide insights on overall metabolic regulation. We therefore used an assay developed in the chapter-5 to examine how Dps-DNA complexes are influenced by supercoiling and different DNA topology. The assay allows to visualize DNA in curved or plectonemic topologies during binding of Dps. The assay demonstrates that Dps preferentially binds to plectonemes and enhances the compaction of DNA.

In a previous report, Dps was observed to bind and compact torsionally relaxed DNA through a cooperative Ising mechanism (10). In this chapter, we focus on the affinity of Dps for other DNA conformations, paying particular attention to plectonemic DNA as one of the DNA supercoiling states. Using Intercalation-induced Supercoiling of DNA (ISD) assay (11), we tested the preferences of Dps in binding to different DNA shapes. Additionally, we show that preformed Dps-DNA complexes are able to engulf more DNA and accommodate it inside these complexes. Thus, we explore how preformed supercoiled DNA state influences Dps activity. At the end, we summarize our observations with a model describing binding of Dps to DNA.

7.2 Results and discussion

7.2.1 Preparation of different DNA conformations

To visualize the various DNA conformations and subsequent Dps binding at single-molecule level, we utilized an assay developed for studying supercoiled DNA (see chapter-5 and also ref. (11)). We introduced a linear 20.6 kbp DNA biotinylated at both the ends (Figure 7.1A) into a streptavidin coated flow cell at a flow rate of 30 µl/min. After one end of the DNA attached to the surface the molecules became stretched in the direction of applied flow. While flow continued, the other end of the DNA also attached to the surface at different location. Thus, we obtained DNA molecules doubly tethered to the surface. We also obtained molecules that were tethered at one-end only due to lacking of second biotin handle. We visualized the immobilized DNA molecules using epi-fluorescence microscopy. We choose the intercalating dye SYTOX orange (SxO) for DNA labeling due to the high quantum yield (Φ = 0.9) coupled with high binding and unbinding rates (12) that ensured a rapid equilibrium of DNA-dye complexes and reduced the photo-induced damage of DNA molecules.

We observed spontaneous formation of plectonemes on some fraction of doubly tethered DNA molecules (Figure 7.1B) due to intercalation by SxO. Binding of SxO would result in decrease in the sum of number of helical turns of the DNA (T_w) due

to partial unwinding of the of two neighboring bases by an angle of around 30° as well as additional separation of around ~0.3 nm (13). For a topologically closed DNA molecule, these effects were compensated by the increase in the number of spatial crossings of DNA to itself, known as writhe (W_r) due to conservation of linking number ($L_k = T_w + W_r$) (14,15). The increase in W_r corresponds to formation of a plectoneme. Upon an application of flow, plectonemic DNA stretched in the direction of the flow but remained intertwined.

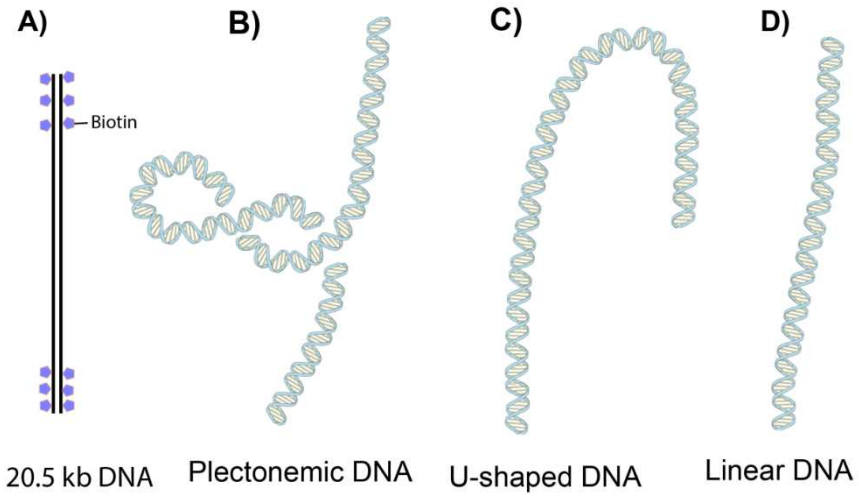


Figure 7.1**. Design of the DNA construct and different DNA topologies tested for the activity of Dps.** *(A)* Cartoon of the DNA construct with biotin handles *(blue pentagons)* at both ends designed for the fluorescence assay. *(B-D)* Schematics of different DNA shapes: (B) plectonemic DNA obtained using ISD technique, (C) inverted J-shaped DNA obtained by application of flow on doubly tethered DNA, and (D) linear DNA obtained by application of flow on singly tethered DNA.

We also found that some of the DNA molecules attached to the surface became J-shaped upon application of flow (Figure 7.1C). We conclude that these molecules were nicked or tethered via only a single biotin, allowing them to torsionally relax. The relative location of the two DNA attachments points also varied among the molecules, creating a range of broader and narrower J-shaped DNA under flow. In the same field of view, we also observed singly tethered DNA molecules that were stretched in the direction of applied flow (Figure 7.1D), because, these molecules were lacking one of the biotin handles.

We stretched DNA by application of flow and examined Dps binding to different conformations of DNA molecules: DNA plectonemes, J-shape and linear DNA molecules.

7.2.2 Binging of Dps to plectonemic, J-shaped and stretched DNA

In order to visualize the binding of Dps to DNA of different conformations, we flowed in Cy5-labelled Dps into the flow cell while recording the fluorescence of SxO from immobilized DNA. We flowed in 200 nM of Cy5-labelled Dps in a buffer containing 40 mM Tris-HCl, 50 mM NaCl, pH 7.3, 5% PEG 8K. During the flow the central part of the DNA molecules adapted one of three conformations: plectonemic DNA, J-shape, and linear DNA **(**Figure 7.1B, C and D).

The results indicated of Dps binding to DNA (frame 50-275) and compacting J-shaped DNA faster than linear DNA but slower than plectonemic DNA. Qualitatively the distinction in time of the DNA compaction by Dps for various DNA conformations can be seen if we focus on the snap-shots of DNA molecules at different time points **(**Figure 7.2**)**. The binding of Dps to J-shape DNA was detected at frame 175 **(**Figure 7.2A), while Dps binding to plectonemic DNA was detected already at frame 125 (Figure 7.2C). At last stretched DNA was compacted by Dps when the flow was reduced almost completely at frame 275 (Figure 7.2B). We also found that Dps was capable of cross-linking two different DNA molecules. This cross-linking can be seen from the snap-shots provided in the figure 2D where two DNA molecules initially not linked to each other. In these frames, while the first molecule is doubly tethered and exhibiting plectonemes, the second one was singly tethered. During the flow the singly tethered DNA molecule was stretched (frames 25-100 Fig.2D) over the doubly tethered molecule and got cross-linked due to binding of Dps (starting from frame 125 in Fig.2D). More examples of cross-linking DNA molecules are provided in the Supplementary Figure 7.1.

By the time the flow was completely stopped (frames after 275), we did not observe any fluctuations across the DNA length that are typical to a doubly tethered DNA without Dps, meaning all the DNA molecules were tightly compacted by Dps. The DNA-Dps complexes can be readily seen as a bright fluorescence spots on image obtained by overlapping fluorescence from DNA and Dps channels **(**Figure 7.3A). Most strikingly, Dps was not detected along the length of the stretched DNA except of the binding site where it nucleated, stabilized and compacted DNA.

We then performed quantitative analysis of the distribution of Dps along DNA molecules. An example molecule for analysis is indicated with a rectangle in the

Figure 7.3A. The average fluorescence intensities of DNA (SxO) and Dps (Cy5) were calculated along the length of the DNA-Dps complexes and plotted them together for comparison (Figure 7.3B). This result showed that Dps does not show any affinity to linear part of the DNA apart from the position where it made a complex with DNA.

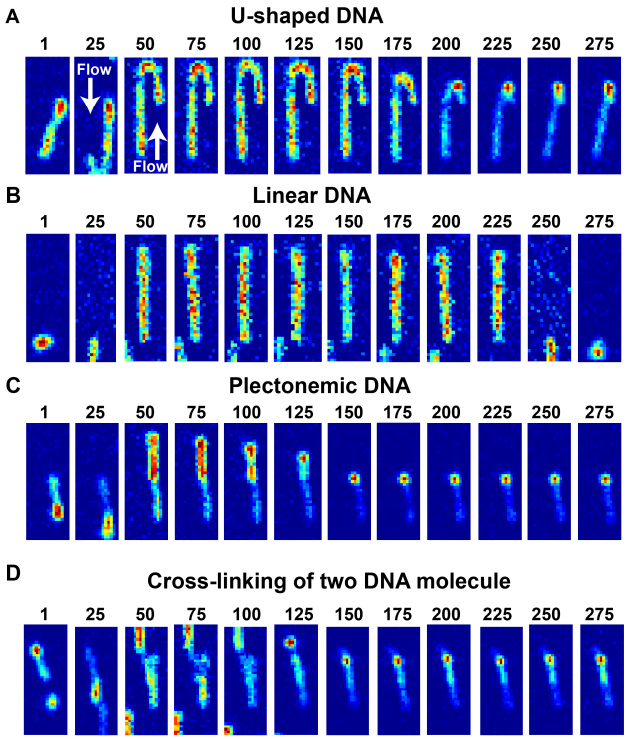


Figure 7.2**. Real-time observation of Dps interacting with U-shape and plectonemic DNA.** All the images were acquired at 100 msec intervals simultaneously for all molecules. The frame number is indicated on each snap-shot. ***(A)*** Snap-shots showing DNA compaction due to binding of Dps to U-shaped DNA formed during the flow as both ends were anchored to the surface. Full compaction of the DNA occurred at frame number 225. ***(B)*** Snap-shots showing DNA compaction due to binding of Dps to linear DNA formed during the flow as one end was anchored to the surface. Full compaction of the DNA occurred at frame number 275. ***(C)*** Snap-shots DNA compaction due to binding of Dps to plectonemic DNA. Plectonemes can be seen as bright fluorescent spots in frames 50-100. ***(D)*** Dps mediated cross-linking of DNA molecules. A singly tethered DNA molecule and a doubly tethered DNA exhibiting plectonemes free from each other before arrival of Dps. The two molecules were cross-linked by binding of Dps.

Therefore, this fluorescence assay not only facilitated to observe DNA compaction mediated by Dps but also provided insights into nucleation of Dps on DNA of various conformational states in real time. Overall, our results strongly suggest that Dps binds and makes a complex

with DNA only if two DNA strands are close to each other such as plectonemic, kinked or flexible DNA. Two interesting findings of this assay are that Dps can mediate cross-linking of DNA molecules and it shows no affinity for stretched linear DNA.

7.2.3 Dps-DNA complex accommodates excess of DNA

Above we observed and quantified the process of Dps-DNA complex formation. However, it remains unclear whether Dps-DNA assemblies have Dps molecules proportional to the amount of wrapped DNA. Here we hypothesized that stable Dps-DNA complexes can form with various ratios of Dps and DNA. Thus, we tested if the preformed Dps-DNA complexes can bind an additional DNA in the absence of free Dps.

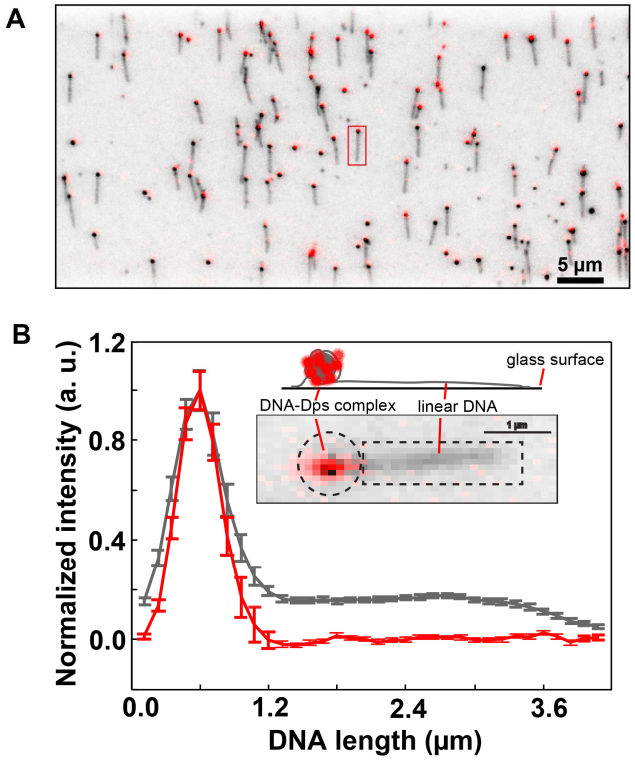


Figure 7.3**. Distribution of Dps molecules along stretched DNA.** ***(A)*** Overlaid fluorescence signals obtained from DNA labelled with SYTOX orange *(comet like structures in green color)* and Cy5-labelled Dps *(bright spots in red color)* after the formation of Dps-DNA complexes. An example of stretched DNA molecule with Dps-DNA complex formed on one sight *(shown in rectangular)*. ***(B)*** Plot showing the intensity distribution of Cy5-Dps *(red)* along labelled DNA with SYTOX orange *(green)* obtained by averaging the intensity of Cy5 and SxO from 90 different structures. Inset represents an example molecule from ***(A)*** illustrating direction of pixels used in the plot..

Before adding free DNA to the flow cell containing preformed Dps-DNA complexes, we exchanged buffer in the flow cell (~50 times the volume of the flow cell) to make sure that there is no free Dps diffusing in the solution. After this step Dps-DNA complexes remained completely stable and in compacted conformation. Then, while recording, we added 100 µl of ~100 fM of 20.6 kbp DNA plasmid labelled with in imaging buffer together with SxO.

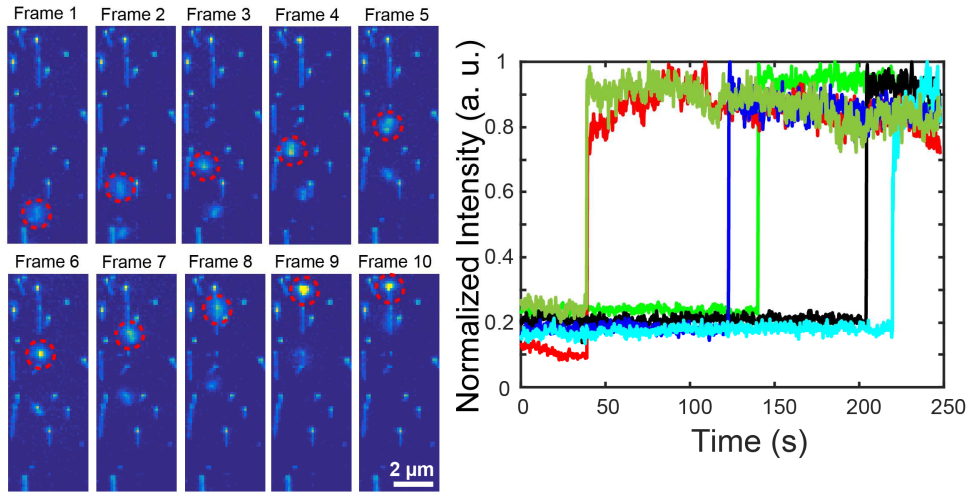


Figure 7.4**. Binding of DNA plasmid to preformed Dps-DNA assembly.** ***(A)*** DNA plasmid *(shown in red circle)* freely moving towards preformed Dps-DNA assembly *(frame 1-8)* and permanently binding to it *(frame 9 and 10)*. Each frame is a snapshot taken at a 200 msec interval. ***(B)*** Normalized time traces of six events when a single step sharp increase in fluorescence intensity of the DNA-Dps complex indicated binding of extra DNA plasmid.

Before binding to the Dps-DNA complex, plasmid DNA remained flexible with no Dps bound (shown in red circle in Figure 7.4A). A difference in fluorescence intensities detected for Dps-DNA complexes before and after addition of plasmid DNA demonstrates that some of complexes bound freely moving DNA plasmid and incorporated it inside of them. Therefore, on the plotted time trace of the intensity of those Dps-DNA complexes that engulfed additional DNA plasmids, a sharp intensity increase is observed (Figure 7.4B). These time traces with single step fluorescent increment confirmed the binding of plasmid DNA inside of the preformed Dps-DNA complexes.

The ability of a fixed amount of Dps to integrate various amounts of DNA into a complex suggests that rather than adopting a single rigid structure, the Dps-DNA complex is flexible and can adopt multiple geometries. This might allow enzymes

moving along the DNA, such as RNAP, to actively rearrange Dps-DNA complexes, providing access to hidden sequences.

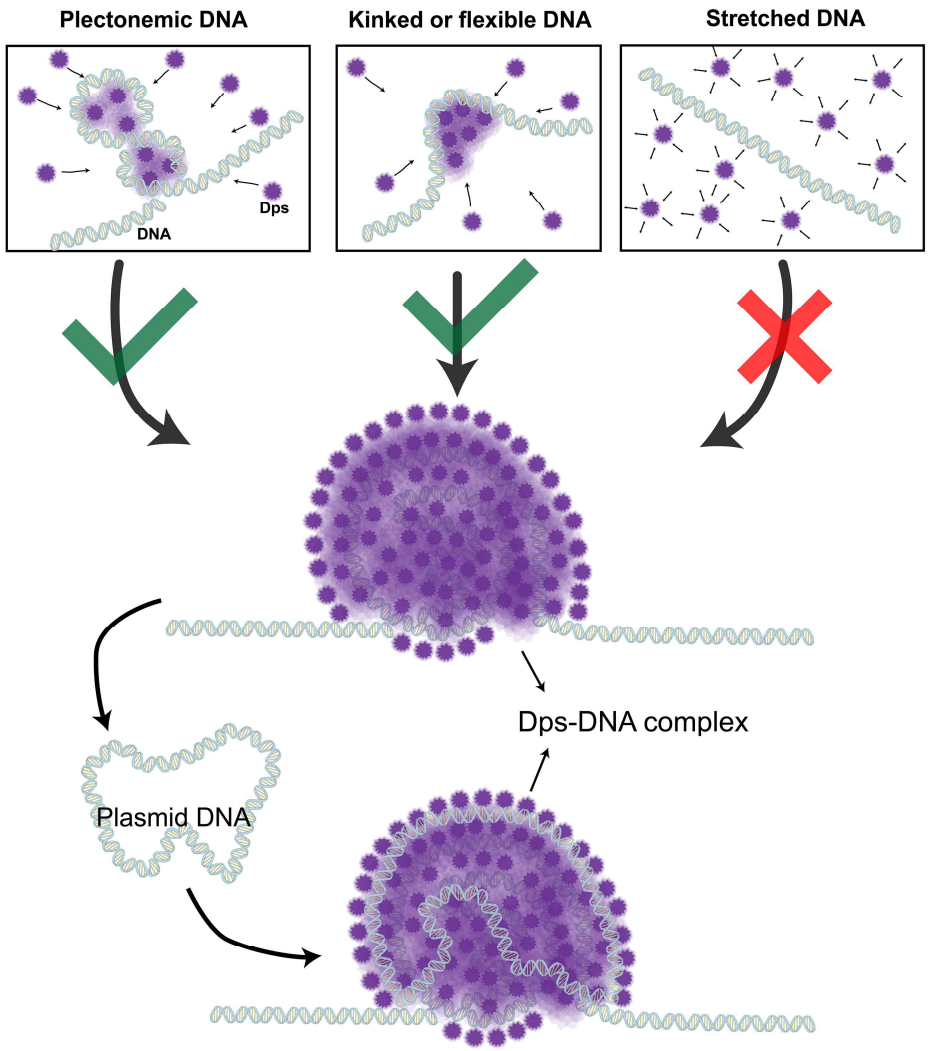


Figure 7.5**. Schematic diagram showing nucleation and formation of Dps-DNA complex**. Plectonemic DNA or kinked/flexible DNA acts as nucleation point for formation Dps-DNA Dps does not bind to an inflexible or taut DNA. If the DNA is held at constant force, formation of Dps-DNA complex is hindered. Dps-Dps interactions are prevalent in the initial Dps-DNA complex. An excess DNA binds to preformed Dps-DNA complex.

7.3 Conclusions

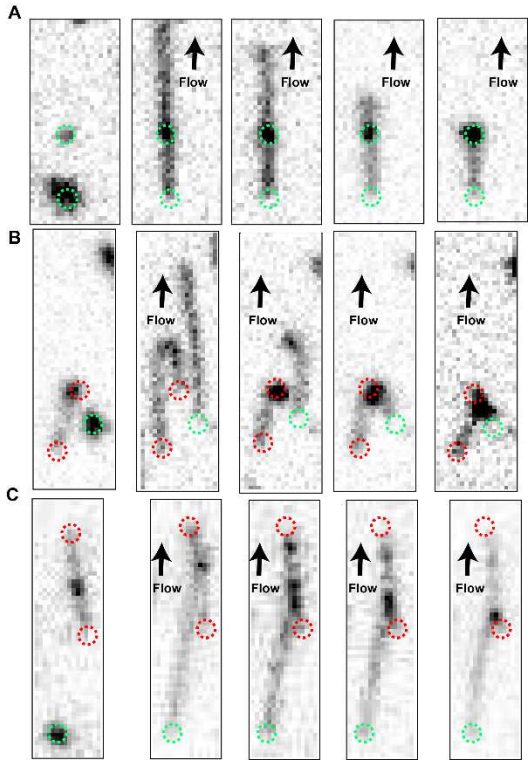
In this chapter we investigated how different DNA conformations and supercoiling states affect Dps binding, nucleation and subsequent DNA compaction. Also we looked at influence of Dps on DNA coiling and pre-coiled DNA. Utilizing ISD assay, we showed that Dps has no affinity for stretched DNA and requires two DNA strands to be close to each other in order for nucleation (Figure 7.5). We considered three different DNA shapes and found that as closer the DNA strands are, as faster Dps mediates DNA compaction. Dps shows highest affinity to plectonemic DNA, then to bent or flexible DNA, and least to linear DNA but only after its released from tension. Binding of Dps leads, therefore, to Dps-DNA complex formation with some Dps-Dps interactions. Additionally, the preformed Dps-DNA complexes can accommodate excess of free DNA due to Dps-Dps interactions with the complexes.

7.4 References

1. Dillon, S.C. and Dorman, C.J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Micro*, **8**, 185-195.
2. Dorman, C.J. (2014) Function of Nucleoid-Associated Proteins in Chromosome Structuring and Transcriptional Regulation. *J. Mol. Microbiol. Biotechnol.*, **24**, 316-331.
3. Peter, B.J., Arsuaga, J., Breier, A.M., Khodursky, A.B., Brown, P.O. and Cozzarelli, N.R. (2004) Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. *Genome Biol*, **5**.
4. Goldstein, E. and Drlica, K. (1984) Regulation of Bacterial-DNA Supercoiling - Plasmid Linking Numbers Vary with Growth Temperature. *P Natl Acad Sci-Biol*, **81**, 4046-4050.
5. Azam, T.A., Iwata, A., Nishimura, A., Ueda, S. and Ishihama, A. (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J. Bacteriol.*, **181**, 6361-6370.
6. Higgins, C.F., Dorman, C.J., Stirling, D.A., Waddell, L., Booth, I.R., May, G. and Bremer, E. (1988) A physiological role for DNA supercoiling in the osmotic regulation of gene expression in S. typhimurium and E. coli. *Cell*, **52**, 569-584.
7. Dorman, C.J. (1996) Flexible response: DNA supercoiling, transcription and bacterial adaptation to environmental stress. *Trends Microbiol.*, **4**, 214-216.
8. Higgins, N.P. (2007) Under DNA stress, gyrase makes the sign of the cross. *Nat. Struct. Mol. Biol.*, **14**, 256-258.

9. Rui, S. and Tse-Dinh, Y.C. (2003) Topoisomerase function during bacterial responses to environmental challenge. *Front Biosci*, **8**, D256-D263.
10. Vtyurina, N.N., Dulin, D., Docter, M.W., Meyer, A.S., Dekker, N.H. and Abbondanzieri, E.A. (2016) Hysteresis in DNA compaction by Dps is described by an Ising model. *Proceedings of the National Academy of Sciences*, **113**, 4982-4987.
11. Ganji, M., Kim, S.H., van der Torre, J., Abbondanzieri, E. and Dekker, C. (2016) Intercalation-Based Single-Molecule Fluorescence Assay To Study DNA Supercoil Dynamics. *Nano Lett.*, **16**, 4699-4707.
12. Biebricher, A.S., Heller, I., Roijmans, R.F.H., Hoekstra, T.P., Peterman, E.J.G. and Wuite, G.J.L. (2015) The impact of DNA intercalators on DNA and DNA-processing enzymes elucidated through force-dependent binding kinetics. *Nat Commun*, **6**.
13. Lee, J.S. and Morgan, A.R. (1978) A rapid method for the measurement of the unwinding angle of intercalating agents and the superhelix density of circular DNAs. *Nucleic Acids Res.*, **5**, 2425-2439.
14. White, J.H. (1969) Self-Linking and the Gauss Integral in Higher Dimensions. *American Journal of Mathematics*, **91**, 693-728.
15. Strick, T.R., Allemand, J.F., Bensimon, D. and Croquette, V. (1998) Behavior of supercoiled DNA. *Biophys. J.*, **74**, 2016-2028.

Supplementary Figure



Supplementary Figure 7.1**: Dps mediated DNA Cross-linking:** (A) Two singly tethered DNA molecules (B and C) a singly tethered and a doubly tethered DNA molecule get stretched during the application of flow and cross-linked due to formation of Dps-DNA assembly. First frame in each of the series show DNA molecules obtained before application of flow and the subsequent frames were obtained under application flow. Intermediates of the compaction can be seen in frames 3 and 4 in each of the series. Green dotted circles indicate the position of singly tethered DNA molecules and the red dotted circles indicate the positions of doubly tethered DNA molecules.

Summary

For the past two decades single-molecule techniques has been helped to reveal an unprecedented detail into the complex molecular processes. Single-molecule FRET (smFRET) is a fluorescence based technique that allows to visualize interactions of two molecules with high spatio-temporal resolution. In this theses smFRET was implemented to study the dynamics of protein-DNA complexes by using HIV-1 reverse transcriptase (RT) as a model system. We also developed a new technique to introduce plectonemes into doubly tethered DNA molecules by exploiting the properties of intercalating dyes. The new technique allows to measure the dynamics of individual plectonemes. We implemented the technique to understand the sequence dependent plectoneme pinning. This technique opens up new possibilities to study interactions of proteins with supercoiled DNA.

In **chapter 2** we modified the protocol to purify and label p66 subunit of HIV-1 reverse transcriptase. Instead of labelling proteins in presence of reducing agent, we first reduced proteins on the column and washed off the reducing agent before reacting with Cy3-maleimide. The labelled protein was then recovered from excess Cy3 molecules by column purification. Chapter 2 also describes about our wide-field fluorescence microscope set-up and different single-molecule fluorescence techniques.

HIV-1 RT is one of the DNA polymerases that performs displacement synthesis without the help of a helicase protein. In **chapter 3** we used primer-extension assays to study the effect of DNA sequence on the displacement synthesis by RT. We found that both the DNA and RNA displacement synthesis are highly dependent on the sequence of backbone. In fact, RT hardly showed any displacement synthesis on DNA construct with 93% GC content. Moreover, the results indicated that RT actively interacts with the flap of non-template DNA in order for displacement synthesis.

Previous single-molecule research on HIV-1 RT showed that RT switches between two binding modes instantaneously on the substrates. However, the underlying mechanism of flipping was not known. Although related phenomenon such as the sliding of proteins on DNA have been visualized at the single-molecule level and modelled, flipping transitions are difficult to observe because they occur on comparatively fast timescales (less than 1 ms) and are limited to nanometre scale rearrangements of the molecules. In **chapter 4** we therefore developed a single

molecule FRET assay to probe the mechanism of flipping which provides a high spatio-temporal resolution. We considered two different models for flipping of protein, a diffusive hop or a continuously bound tumbling transition. In the tumbling model, protein maintains contacts with DNA during the transition state of flipping and in the hopping model protein enters into the pseudo-bound state in order for either flipping or rebinding. In order to differentiate between the two models, we probed the interactions of RT and DNA under varied salt and crowding concentrations and then measured the relative stability of transition state under each condition. The results showed that the increased salt concentration decreased the affinity of RT to DNA, similar to many other proteins. Furthermore, we found that binding affinity of RT increases under increased macromolecular crowding. The effective dissociation constant of RT to DNA was then fit to the scaled-particle theory. This fitting predicted that RT binds to DNA with sub-nanomolar affinity, *in vivo*. We then applied the changes in rate of flipping, rate of binding, and rate of dissociation due to increased salt and crowding to both tumbling and hopping models. While the hopping model was consistent with our observations, the tumbling model made unphysical predictions about the transition state.

DNA supercoiling effects many essential cellular processes such as DNA replication and RNA transcription, but a detailed mechanistic understanding of DNA supercoiling has remained limited mainly due to the lack of a convenient experimental tool. Existing techniques, force-based single-molecule spectroscopy techniques provide limited information regarding the DNA supercoiling. In view of this we developed a novel single-molecule fluorescence assay to visualize individual plectonemes on doubly tethered DNA (**chapter 5**). We used intercalating dyes for inducing plectonemes on topologically closed DNA molecules and named the assay as Intercalation-induced Supercoiled DNA (ISD). This assay allowed us to follow position and size of individual plectonemes in real-time leading to detect position-dependent plectoneme dynamics. Using ISD we showed that the plectonemes localize to mispaired DNA sequences along supercoiled DNA, emphasizing the potential of the current assay to study sequence dependent plectoneme dynamics.

In **chapter 6** we studied the sequence-dependent plectoneme localization along supercoiled DNA using ISD assay. We wanted to test if the fluctuations in GC-density along DNA has any influence on the localization of plectonemes. For this we prepared a DNA construct with relatively homogeneous GC-content and then subsequently we inserted different length of AT-rich regions in the middle of it. These constructs allowed us to test the influence of GC-density on localization of plectoneme along supercoiled DNA. When we measured the plectoneme density on

these DNA constructs, the results clearly indicated that plectonemes prefer to pin at AT-rich regions than GC-rich regions. We also identified that poly(A)-tracts assist in pinning the plectonemes even on positions where the GC-density is comparable to the flanking sequences.

Since DNA supercoiling is prevalent, activity of every DNA binding protein is expected to be different on supercoiled DNA than relaxed DNA. In **chapter 7** we used ISD assay to test the activity of Dps, DNA binding proteins from starved cells, on supercoiled DNA. Our results revealed new insights into the binding of Dps to DNA. We found that Dps only binds to DNA if two strands are close to each other such as in plectoneme structures but Dps does not show any affinity to stretched DNA. Our results indicated that Dps-DNA complexes are rich in Dps that the complexes can accommodate excess DNA. Moreover, we found that Dps is capable of cross-linking two DNA molecules.

Samenvatting

In de afgelopen twee decennia hebben enkel-molecuul technieken bijgedragen aan de ontdekking van complexe moleculaire processen. Enkel-molecuul FRET (smFRET) is een op fluorescentie gebaseerde techniek waarmee het mogelijk is de interactie tussen twee moleculen met hoge spatiotemporele resolutie te visualiseren. In deze thesis is smFRET geïmplementeerd, met behulp van HIV-1 reverse transcriptase (RT) als modelsysteem, om de dynamica van eiwit-DNA complexen te bestuderen. Ook hebben we een nieuwe techniek ontwikkeld om zogenoemde plectoniemen te introduceren in dubbel-verankerd DNA, door gebruik te maken van de eigenschappen van intercalerende kleurstoffen. Deze nieuwe techniek biedt de mogelijkheid om de dynamica van individuele plectoniemen te meten. We hebben deze techniek gebruikt om sequentie-afhankelijke lokalisatie van plectoniemen te begrijpen. Deze techniek biedt een nieuw arsenaal aan mogelijkheden voor het bestuderen van interacties tussen eiwitten en supercoiled DNA.

In **hoofdstuk 2** hebben we het protocol gemodificeerd om de p66-subunit van HIV-1 RT te purificeren en te labelen. In plaats van de eiwitten te labelen in de aanwezigheid van een reducerende stof, hebben wij eerst de eiwitten gereduceerd in een kolom en de reducerende stof weggespoeld alvorens de reactie met Cy3-maleimide te laten plaatsvinden. De gelabelde eiwitmoleculen zijn toen gewonnen uit het Cy3 overschot door kolompurificatie. **Hoofdstuk 2** beschrijft ook onze wijd-veld fluorescentie microscopie-opstelling en verschillende enkel-molecuul fluorescentietechnieken.

HIV-1 RT is één van de DNA polymerasen die verplaatsingssynthese uitvoert zonder hulp van een helicase eiwit. In **hoofdstuk 3** hebben we primer-extensie assays uitgevoerd om het effect van de DNA-sequentie op de verplaatsingssynthese door RT te bestuderen. We vonden dat zowel DNA- als RNA-verplaatsingssynthese in hoge mate afhankelijk zijn van de sequentie van de DNA-ruggegraat. In feite liet RT nauwelijks verplaatsingssynthese zien op een DNA-construct met 93% GC-gehalte. Daarnaast gaven de resultaten aan dat RT actief inwerkt op de klep van het niet-sjabloon DNA bij de verplaatsingssynthese.

Eerdere enkel-molecuul studies van HIV-1 RT lieten zien dat RT instantaan schakelt tussen twee bindingsmodes op de substraten. Echter, het onderliggende mechanisme van toestandswisseling is niet bekend. Hoewel verwante fenomenen, zoals het

glijden van een eiwit over DNA, gevisualiseerd en gemodelleerd zijn op enkel-molecuul niveau, zijn toestandswisselingen moeilijk te observeren omdat ze zich op relatief korte tijdsschalen van minder dan 1 ms voordoen en gelimiteerd zijn tot de herschikking van moleculen op de nanometer schaal. In **hoofdstuk 4** hebben we daarom een enkel-molecuul FRET assay ontwikkeld om het toestandswisselingsmechanisme te onderzoeken met hoge spatiotemporele resolutie.

We hebben gekeken naar twee verschillende modellen voor eiwit-toestandswisseling: het sprongenmodel en het tuimelmodel. In het tuimelmodel houdt het eiwit contact met het DNA gedurende de transitietoestand van de toestandswisseling en in het sprongenmodel betreedt het eiwit een pseudo-gebonden toestand tijdens de toestandswisseling. Om onderscheid te kunnen maken tussen de twee modellen, hebben we de interacties tussen RT en DNA bestudeerd onder verschillende zoutconcentraties en crowdingcondities en hebben we de relatieve stabiliteit van de transitie gemeten onder deze gegeven condities. De resultaten lieten zien dat bij een verhoogde zoutconcentratie de DNA-bindingsaffiniteit van RT bindt lager wordt, zoals dat ook bij andere eiwitten het geval is. Daarnaast hebben we gevonden dat de bindingsaffiniteit van RT toeneemt naarmate de dichtheid van crowdingmoleculen toeneemt. De effectieve dissociatieconstante van RT aan DNA hebben we gefit aan een geschaalde-deeltjestheorie. Deze fit voorspelde dat RT *in vivo* aan DNA bindt met sub-nanomolaire affiniteit. Vervolgens hebben we veranderingen in de mate van toestandswisseling, binding en dissociatie door de verhoogde zoutconcentratie en crowding toegepast op zowel het tuimel- als het sprongenmodel. Waar het sprongenmodel consistent was met onze observaties, maakte het tuimelmodel niet-fysische voorspellingen over de transitie.

DNA-supercoiling heeft effect op vele essentiële cellulaire processen, zoals DNA-replicatie en RNA-transcriptie, maar een gedetailleerd, mechanistisch begrip van DNA-supercoiling is nog niet voorhanden. Dit komt voornamelijk door het ontbreken van geschikte experimentele technieken. Huidige enkel-molecuul krachtspectroscopietechnieken bieden beperkte informatie over DNA-supercoiling. Daarom hebben we een nieuw enkel-molecuul fluorescentie assay ontwikkelt om individuele plectoniemen te induceren op dubbel-verankerde DNA-moleculen (**hoofdstuk 5**). We hebben intercalerende kleurstoffen gebruikt om plectoniemen te induceren op topologisch gesloten DNA-moleculen en hebben het assay Intercalation-induced Supercoiled DNA (ISD) genoemd. Dit assay stelde ons in staat de positie en grootte van individuele plectoniemen in real-time te volgen, resulterend in positie-afhankelijke plectoniem-dynamica. Met behulp van ISD hebben wij laten zien dat plectoniemen zich lokaliseren bij verkeerd gepaarde DNA-sequenties langs

het supercoiled DNA, wat de mogelijkheden van het assay onderstreept om sequentie-afhankelijke plectoniem-dynamica te bestuderen.

In **hoofdstuk 6** hebben we de sequentie-afhankelijke plectoniem-lokalisatie langs supercoiled DNA bestudeerd door gebruik van een ISD assay. We wilden testen of fluctuaties in GC-dichtheid langs het DNA invloed heeft op de lokalisatie van de plectoniemen. Hiervoor hebben we een DNA-construct geprepareerd met een relatief homogene GC-inhoud en hebben we AT-rijke strengen van verschillende lengtes in het midden ingevoegd. Met deze constructen konden we de invloed van GC-dichtheid te testen op de lokalisatie van plectoniemen op supercoiled DNA. Toen we de plectoniem-dichtheid op deze DNA-constructen gemeten hadden, lieten de resultaten zien dat de plectoniemen de vookeur hebben om zich aan AT-rijke regionen te binden, in plaats van aan de GC-rijke regionen. Ook hebben we geobserveerd dat poly(A)-tracts helpen bij het binden van plectoniemen, zelfs op posities waar de GC-dichtheid vergelijkbaar is met de aanliggende sequenties.

Omdat DNA-supercoiling veel voorkomt in het genoom, is het te verwachten dat de activiteit van een DNA-bindend eiwit anders is op supercoiled DNA dan op gerelaxeerd DNA. In **hoofdstuk 7** hebben we een ISD assay gebruikt om de activiteit van Dps (DNA-bindende eiwitten uit verhongerde cellen) te meten op supercoiled DNA. Onze resultaten geven nieuwe inzichten in de binding van Dps aan DNA. We vonden dat Dps alleen bindt aan DNA als twee DNA-strengen dicht bij elkaar zijn, zoals in plectoniemen het geval is. Dps liet daarentegen geen affiniteit zien voor gerelaxeerd DNA. Onze resultaten lieten daarnaast zien dat Dps-DNA complexen rijk zijn aan Dps, zodat ze een overschot aan DNA kunnen accommoderen. Verder vonden we dat Dps de capaciteit heeft om twee DNA-moleculen te crosslinken.

Acknowledgments

Writing this part of the thesis made me to realise that I'm reaching a major milestone in my life. I had a great motivation and support from a lot of people in this feat. I had a great pleasure meeting a lot of new people during the last four and half years. I got an extraordinary support from all these people was. I think this is a right place to acknowledge each and every.

First and foremost, I would like to thank my supervisor Elio for providing me an opportunity to pursue PhD in his lab and for being an inspirational scientist. I cannot imagine that I would have finished my PhD if it was not under his supervision. Elio, starting from my PhD, you guided me in right path whenever I was losing my focus, you nurtured me well as a scientist, you encouraged me to explore every direction I wanted, you always valued my thoughts. Elio, you always had time whenever I feel I'm in the cloud, then I could knock your door, and I would only return with a lot of confidence and motivation. You also supported me to attend conferences in the USA where I could meet eminent scientists. Not only in the lab but I also enjoyed our fun-filled group outings and dinners at your place. I would like to express my gratitude for everything.

The next person I would like to thank is, my promoter, Cees. Thanks to the graduate school for choosing you as my promoter. I'm glad that you are not just my promoter on paper, but I also got a chance to work directly with you. I still remember our first scientific meeting that's when I understood why it's so much exciting to discuss science with you. Meetings with you are filled with exciting discussions and insightful thoughts. You like to think about big and bold questions. I'm extremely delighted to join your group for my next endeavour as a postdoc.

I would like to thank all my committee members, Prof. Andreas Engel, Dr. Greg Bokinsky, Prof. Erwin Peterman, Prof. Achillefs Kapanidis, and Prof. David Rueda for reading and judging of my theses.

I would like to thank my officemates Maarten, Andrew, and Bojk from my whole heart. I cannot ask for better officemates. You guys made my PhD life pleasant. Maarten, what an amazing person you are. Maarten, you always had time for fun conversations, Friday evening beers, fussball, translating all those letters from Dutch, and what not. Bojk, thank you for all the discussions both scientific and non-scientific. Andrew, I always had hard time understanding your accent but I could still laugh on those never ending jokes.

I also like to thank the other members of Abbondanzieri lab. Natalia, my one and only lab-mate, for being such an organised person which helped to keep the stack of lab materials and we could talk each other with graduate school requirements. Thanks for coming all the way to India to attend my wedding. Thanks for all those wonderful house-warming and cooling parties. I wish you good luck in finding a job. I had a great wetlab training from Anna, including protein purification, labelling, and single-molecule experiments. Thank you Anna for the wonderful training. The microscopy and programming expert, Maargreet Docter, trained me in the related areas. Thank you very much Maargreet for your help.

The contributions of awesome “teamplectoneme” (thank you Maarten for the naming), Sung Hyun and Jaco, to my PhD theses is well appreciated. Jaco, you would never say no to our ever growing wish list of DNA constructs. You would only smile for our crazy ideas and say it is doable. Thank you very much Jaco for so much positivity. Sung Hyun, you were almost like my unofficial daily supervisor. You thought me to be critical. We could have endless discussions about our experiments and interpretation of our data almost every day. Our discussions would not end at lab and would continue while we have a drink or two in the city center. Thank you very much for the collaboration and friendship. I wish you all the best for all your future endeavours.

I also want to thank several people that I met during my master’s studies in Germany. My master’s supervisor Marten Mkandawire has always encouraged and guided me to pursue research. His friendly nature made me feel home in during my studies in Germany. Equally affectionate people are Rocco and Mandy. I had so much fun during my time in Germany mainly because of you guys. Thanks a lot.

I would like to express my gratitude to Chirlmin Joo for allowing me to use equipment in his lab and also for suggestions regarding microscopy and single-molecule data analysis. Thank you for the great feedback on my manuscript. I also want to thank the other members of Chirlmin lab: Luuk, Malwina, Mohammad (fussball legend), Stanley, and Jetty, discussions with you guys were very necessary for successful establishment of my single-molecule assays. Jetty and Luuk, good luck with your last steps of your PhD. I also like to thank Thijs and Misha. Misha, thank you for those discussions on target search mechanism of protein.

I like to thank the BN social animals and my friends whom with I could have everyday lunch and/or a drink or two in every Friday evening: Adithya, Maarten, Bojk, Jorine, Viktotia, Sergii, Daniel, Stephanie, Yaroon, Sumit, Jacob, Antony, Fede, Wayne, Siddharth, Laura, Kuba, Yoones, Fabai and Misha. You guys are

really awesome and make BN a lively place to work. Thank you guys for the friendship and nice time. Adi and Kuba, good luck with the last steps of your PhD. Antony and Kasper, thank you for translating summary of theses in to Dutch. Sergii, thank you for the tips in handling word document.

I feel grateful that I interacted with eminent scientists during my time at TU Delft. I like to thank Andreas Engel for being patient in explaining about AFM and electron-microscopy amongst many other interesting discussions. Nynke Dekker for giving me opportunity to be a teaching assistant for statistical physics which helped to shape my Matlab skills. Marileen Dogterom for the cooperation in the extension of my contract. Christophe Danelon, such a down-to-earth person you are, thank you for friendly conversations. Martin Depken, thank you for your support. Anne Meyer, thank you for the conversations and multi-group outings. Greg Bokinsky, thanks for accepting to be part of my committee. Timon Idema, Bertus Beaumont, Marie-Eve Aubin-Tam and Hyun Youk for inspiring scientific talks and friendly conversations. You are all inspired me as a researcher.

Friendly conversations and assistance from the non-scientific staff of BN, Dijana, Amanda, Jolijn, Emmylou, Marije, Jelle, Dimitri, and Esther is well appreciated.

I would also like to thank other members of BN: Richard, Orkide, Allard, Alicia, Jonas, Ilyong, Sam, Peter, Humberto, Roy, Seungkyu, Eugen, Sriram, Pauline, Artur, Afke, Ruben, Afshin, Tania, Nuria, Louis, Maurits, Mathijs, Nicole, Vanessa, Jochem, Da Wei, Mathia, Benjamin, and Fabrizio. Thank you guys for the friendship. Victor, Dominik, and Roland, thank you for friendly conversations. Victor, good luck with finishing up your PhD.

The conversations, assistance, and, guidance from the other technical staff of BN, Pawel, Erwin, Esenguel, Andrea, Sacha, Ilja, Theo, Marek, Jacob, and Ange are also appreciated whole heartedly. You guys were always there to help me with any of the technical issues. Theo, my special thanks to you for being so kind for making those DNA constructs which lead to whole new line of research during my theses. Jacob, thank you very much for the awesome and appealing cover design.

I also like to thank people who are not part of BN anymore but shared some quality time during their stay here: Felix, Calin and Tessa, David, Jan Lipfert, Marijn, Charl, Rifka, Zongbo, Jelmer, Huong, Tim, Magnus, Gautam, and Michela.

I also would like to thank a few people outside BN. Thank you Sathis AP, you are very warm and inspiring person. I had pleasure playing cricket at Concordia where I interacted with lot of friendly fellows. Playing cricket helped me to relieve stress

and regenerate my energy. I want to thank all the members of the club, especially, JP, Jan Willem, Srijith, Sathis, Darwin, Nauman, Tiggy, Shukla, Casper, Manu, Anand, Sid, Kamal, Pradyumna, and Vik.

Last but not least, I want to thank my family. I want to dedicate this thesis to my late parents. Saying thanks is not enough for the kind of support I get from my brothers. They are the best brothers in the world. Their unconditional love and support encouraged me to pursue my dreams. Without their support I would not have come so far. Both my wife and son get at least half the credit in achieving my PhD. I cannot ask for a better family. Your support during this time is tremendous. However late I reach home due to work, you only invite me with a smile. I think I'm fortunate to have such a great family.

Curriculum Vitae

Mahipal Ganji

10-08-1984	Born in Khajipoor, India
1999	Secondary school certificate Govt. High School, Vemulawada, India
1999-2001	Higher secondary school Jawaharlal Nehru Junior college, Vemulawada, India
2003-2006	Bachelor of science SRR Govt. Degree college, Karimnagar, India
2006-2008	Master of science Nizam college, Hyderabad, India
2009-2011	Master of science TU Dresden, Germany
2012-2016	Doctor of Philosophy TU Delft, the Netherlands Title: "Molecular Gymnastics: Single-Molecule Investigation of Protein Jumping and DNA Dancing" Promoter: Prof. dr. Cees Dekker Co-promoter: Dr. Elio Abbondanzieri

List of Publications

1. **M. Ganji**, S. F. J. Le Grice, and E. A. Abbondanzieri, “Sequence-dependent strand displacement synthesis of HIV-1 reverse transcriptase”, To be submitted

2. **M. Ganji**, M. Docter, S. F. J. Le Grice, and E. A. Abbondanzieri, “DNA binding proteins explore multiple local configurations during docking via rapid rebinding”, *Nucleic Acids Res*. doi: 10.1093/nar/gkw666 (2016)

3. **M. Ganji***, S.H. Kim*, J. van der Torre, E. Abbondanzieri‡, and C. Dekker‡, “Intercalation-based single-molecule fluorescence assay to study DNA supercoil dynamics”, *Nano Lett.*, doi: 10.1021/acs.nanolett.6b02213 (2016)

4. S.H. Kim*, **M. Ganji***, J. van der Torre, E. Abbondanzieri‡, and C. Dekker‡, “Sequence-dependent plectoneme pinning along supercoiled DNA”, To be submitted

5. **M. Ganji**, N. Vtyurina, and E. A. Abbondanzieri, “Visualization of Dps interaction with different DNA topologies”, To be submitted

*-Denotes equal contribution