Technische Universität München
TUM School of Natural Sciences

Replication with a twist: Single-molecule studies of the replisome under topological strain

Thomas M. Retzer

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Martin Zacharias

Prüfende der Dissertation:

1. Prof. Dr. Karl Duderstadt
2. Prof. Dr. Matthias Rief

Die Dissertation wurde am 06.09.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 23.10.2024 angenommen.

Thomas Retzer:

Replication with a twist: Single-molecule studies of the replisome under topological strain

Excerpts of this thesis have appeared in the following publication:

Nadia M. Huisjes*, Thomas M. Retzer*, Matthias J. Scherr, Rohit Agarwal, Lional Rajappa, Barbara Safaric, Anita Minnen, Karl E. Duderstadt (2022). **Mars, a molecule archive suite for reproducible analysis and reporting of single-molecule properties from bioimages.** eLife 11:e75899.

*equal contribution

What I cannot create, I do not understand.

Richard Feynman

Table of content

Zusammenfassung

Die Doppelhelix Struktur der DNA stellt während der Replikation erhebliche topologische Herausforderungen dar. Wenn die zu kopierende DNA entwickelt wird um Tochterstränge zu synthetisieren, sammelt sich verdrillte DNA vor der Replikationsgabel an und bildet eine Barriere, wenn sie nicht aufgelöst wird. Gleichzeitig bildet die neugebildete DNA hinter der Gabelung negative Superspiralisierung. Topoisomerasen, Enzyme, die für DNA-Topologie verantwortlich sind, arbeiten daran diese Probleme während der Replikation, Transkription und DNA-Reparatur zu lösen. Jüngste Studien deuten jedoch darauf hin, dass die Anzahl der Topoisomerasen möglicherweise nicht ausreicht, um mit der hohen Geschwindigkeit des bakteriellen Replisoms Schritt zu halten, was bedeutet, dass diese Enzyme anfallende Supercoils nur teilweise auflösen können. Außerdem setzt die stark verdichtete Chromosomenstruktur mit ihrer komplexen Organisation und ihren topologischen Domänen der Topoisomerase-Aktivität räumliche Grenzen. Dies wiederum deutet darauf hin, dass das Replisom selbst über intrinsische Mechanismen zur Bewältigung von topologischem Stress verfügen könnte. Ein möglicher Mechanismus ist die Gabelrotation, bei der sich das Replisom dreht, um topologische Belastungen hinter sich zu lassen, was zur Bildung zusätzlicher Verdrillung der Tochterstränge führt, die später während der DNA-Segregation aufgelöst werden. Die Gabelrotation wurde jedoch nur indirekt durch den Nachweis überschüssiger verdrillter Tochterstränge nach verringerter Topoisomerase-Aktivität beobachtet. In dieser Arbeit haben wir ein modulares DNA-Toolkit entwickelt, um verschiedene topologische Szenarien in konstituierten DNA-Replikationsexperimenten mit Hilfe der Einzelmolekül-Fluoreszenzmikroskopie nachzubilden. Unsere Ergebnisse zeigen, dass topologische Barrieren vor der Replikationsgabel zu Pausen und einer leicht verringerten Replikationsrate führen. Trotz dieser Hindernisse setzt das Replisom die Replikation mit nahezu gleicher Geschwindigkeit fort. Wir beobachteten dynamische Veränderungen des Strangprodukts, insbesondere unter topologischem Stress, was uns zu weiteren Untersuchungen veranlasste. Um diese Dynamik besser zu verstehen, haben wir einen neuartigen Ansatz für die Replikationsexperimente entwickelt, welcher eine räumliche Auflösung des Folgestrangs und des Leitstrangs während der laufenden Synthese ermöglicht. Unsere Beobachtungen deuten darauf hin, dass sich das Replisom während der Replikation mit hoher Geschwindigkeit drehen kann, was dazu beiträgt topologische Spannungen zu bewältigen. Diese Rotation kann zu einem Anstieg der Zahl der Polymerasen in der Nähe des Replisoms führen, was auf potenzielle Neustart-Ereignisse oder DNA-Schaden hinweist. Um die Analyse von Hunderten von Einzelmolekül-Ereignissen zu erleichtern, entwickelten wir des Weiteren ein

Softwarepaket namens Mars (Molecule Archive Suite) entwickelt. Diese Open-Source-Plattform bietet einen flexiblen Arbeitsablauf für die reproduzierbare Analyse aus Bildern abgeleiteten Eigenschaften und könnte als universelle Lösung für die Erstellung von Analysepipelines zur Gewinnung quantitativer Informationen aus komplexen Einzelmoleküldatensätzen dienen. Zusammenfassend lässt sich sagen, dass wir ein Experiment entwickelt haben, um topologische Barrieren bei der Einzelmolekülreplikation zu untersuchen. Unser neuer Imaging Ansatz lieferte neue Einblicke in die Dynamik der Gabelrotation. Abschließend lässt sich sagen, dass die Gabelrotation zum ersten Mal direkt beobachtet wurde, was uns dem Verständnis des komplexen Prozesses der DNA-Replikation näherbringt.

Abstract

The double helical structure of DNA presents significant topological challenges during replication. As the parental DNA is unwound to synthesize daughter strands, overwound DNA accumulates ahead of the replication fork, creating a barrier if left unresolved. Simultaneously, underwound DNA behind the fork forms precatenanes. Topoisomerases, enzymes that manage DNA topology, work to resolve these issues during replication, transcription, and DNA repair. However, recent studies suggest that the number of topoisomerases may not be sufficient to keep pace with the rapid speed of the bacterial replisome, implying that these enzymes can only partially resolve accumulating supercoils. Additionally, the highly compacted chromosomal structure, with its complex organization and topological domains, imposes spatial limitations on topoisomerase activity. This suggests that the replisome itself may have intrinsic mechanisms to cope with topological stress. One proposed mechanism is fork rotation, where the replisome rotates to transfer topological strain behind it, leading to the formation of additional precatenanes that are later resolved during DNA segregation. However, fork rotation has only been indirectly observed through the detection of excess precatenanes following reduced topoisomerase activity. In this thesis, we developed a modular DNA toolkit to mimic various topological scenarios in reconstituted DNA replication assays, utilizing single-molecule fluorescence microscopy. Our findings show that topological barriers ahead of the replication fork cause pausing and a reduction in replication rate. Despite these challenges, the replisome continues replication at nearly the same rate. We observed dynamic changes in the lagging strand product, particularly under topological stress, prompting further investigation. To better understand these dynamics, we developed a novel transverse flow imaging approach that allows spatial resolution of the parental, leading, and lagging strands during ongoing synthesis. Our observations suggest that the replisome can rotate at high speeds during replication, helping to manage topological tension. This rotation may lead to an increase in the number of polymerases near the replisome, indicating potential restart events or DNA damage sites. To facilitate the analysis of hundreds of single-molecule events, we also developed a software package called Mars (molecule archive suite). This open-source platform provides a flexible workflow for the reproducible analysis of image-derived properties and could serve as an universal solution for creating analysis pipelines to extract quantitative information from complex single-molecule datasets. In summary, we developed an assay to study topological barriers in single-molecule replication. Our transverse flow approach provided new insights

into fork rotation dynamics, where fork rotation was directly observed for the first time, bringing us closer to understanding the complex process of DNA replication.

Chapter 1 - Introduction and background

1.1. It is a DNA world

Deoxyribonucleic acid (DNA) and the process of DNA replication are fundamental to the survival of all forms of life, ranging from bacteria, to complex organisms such as vertebrates, as well as viruses. The faithful replication and transfer of the genetic code from the parent cell to its daughter cells is crucial for the organism's survival and proliferation. DNA, as the carrier of genetic information, must possess a high degree of structural durability. Recent advancements in DNA analysis have enabled the extraction and sequencing of DNA that is over one million years old, providing valuable insights into evolutionary processes (van der Valk et al., 2021). In addition to its remarkable durability, DNA's information density, which exceeds petabytes per gram, has made it a prime candidate for innovative data storage applications (Erlich & Zielinski, 2017; Gervasio et al., 2024; Yim et al., 2021). The unique properties of DNA allow the construction of programmable, self-assembling nanostructures with potential applications in drug delivery, bioimaging, and computational systems (Dey et al., 2021). Stretching the 4.6 megabases (Mb) chromosome of *Escherichia coli* (*E. coli*) from end to end shifts our perspective from the microscopic scale of cellular structures to the macroscopic scale, achieving a theoretical length of approximately 1.6 mm (Alberts et al., 2022). This massive polymer must be compacted within the confines of the cell while also remaining accessible to the enzymes that copy the genetic information and facilitate protein synthesis. In the following sections, we will establish the foundational principles for understanding both DNA structure and DNA topology, explore how this polymer can be described using physical equations, and examine how its organization within the cell, ensures accessibility to the molecular machinery involved in genetic processes.

1.1.1. DNA – a double helix

First reported in 1953 in a series of three back-to-back articles in *Nature*, DNA was revealed to adopt a double helical structure (Franklin & Gosling, 1953; Watson & Crick, 1953; Wilkins et al., 1953). This discovery marked a milestone in the history of science, ultimately leading to the awarding of the Nobel Prize to James Watson, Francis Crick, and Maurice Wilkins. However, it is important to acknowledge the substantial contributions of Rosalind Franklin, whose work was instrumental in this discovery

(Cobb & Comfort, 2023). Rosalind Franklin produced Photograph 51 using X-ray diffraction, a crucial piece of data that was later shown to Watson and Crick by Maurice Wilkins. Wilkins was also involved in the early study of DNA using X-ray crystallography. Watson and Crick used Franklin's data, along with other insights, to build the first accurate model of DNA. She deserves recognition as an equal contributor among the group of key scientists who achieved one of the most significant scientific breakthroughs to date (Cobb & Comfort, 2023).

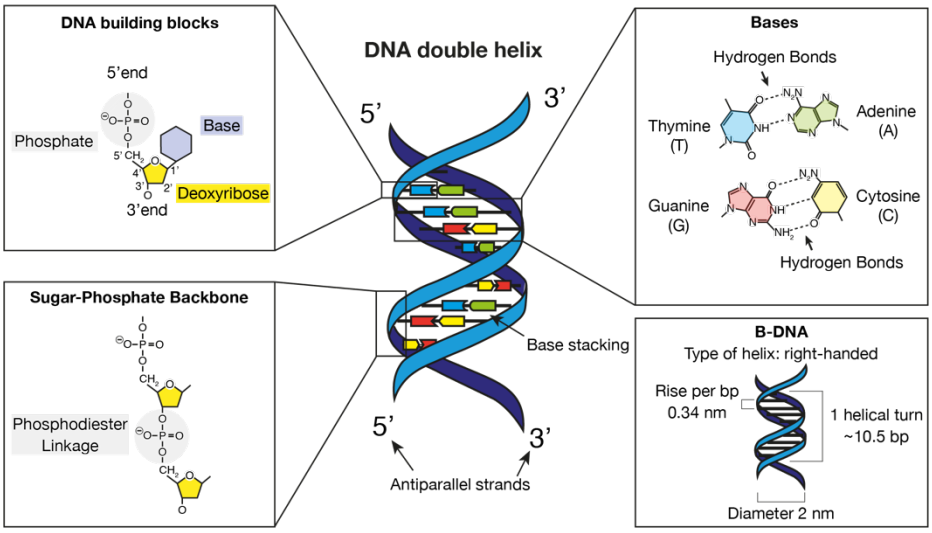


Figure 1-1 DNA double helix.
DNA building blocks consists of a phosphate, deoxyribose (sugar), and a base, together forming a nucleotide. Nucleotides are linked by phosphodiester bonds creating the sugar-phosphate backbone. The bases include thymine (T), adenine (A), guanine (G), and cytosine (C). T and A are complementary forming two hydrogen bonds, while G and C form three bonds. B-DNA, the most common form of DNA, is characterized by the features in the figure. (adapted from *Molecular Biology of the Cell*, Alberts et al. (2022))

DNA consists of two complementary polynucleotide chains running antiparallel to each other, with each chain representing one DNA strand (Figure 1-1). These two strands are held together by hydrogen bonds formed between the base portions of the nucleotides. A nucleotide is composed of a five-carbon sugar, to which a phosphate group and a base is attached. In the case of DNA, the sugar component is deoxyribose, which is attached to the phosphate group (Figure 1-1). The genetic information is encoded using four bases: adenine (A), thymine (T), cytosine (C), and guanine (G) where adenine pairs with thymine, and cytosine pairs with guanine (Figure 1-1). The backbone of each DNA strand is composed of a chain of alternating sugar and phosphate molecules, which are covalently bonded (Figure 1-1). Because of this

linking, each DNA strand has a polarity with the 5’ phosphate and the 3’ hydroxyl groups. They are referred to as the *3’ end* and the *5’ end* derived from the orientation of the deoxyribose sugar, which gives the DNA a directionality. (Alberts et al., 2022; Nelson & Cox, 2021)

The three-dimensional structure of DNA, the double helix, arises from the chemical and structural features of the two polynucleotide chains. The sugar-phosphate backbones are positioned on the outside of the helix, while all the bases are facing inward, forming hydrogen bonds that holding together the two individual strands. The purine bases (two-ring structure), A and G, pair with a pyrimidine base (single-ring structure) T and C, respectively. G and C form three hydrogen bonds while A and T only from two. This complementary base-paring ensures the most energetically favorable packing within the DNA. Base stacking supports equal spacing between each base pair (0.34 nm) as well as the two DNA strands (2 nm) for B-form DNA (Figure 1-1). The two strands twist into a right-handed helical structure, completing one full turn every 10.5 base pairs (bp). B-form DNA is the most stable variant under physiological conditions and will serve as the point of reference, when we talk about DNA in this thesis. In addition, DNA can exist in A-from and a Z-form configuration. In all three forms, the two strands are oriented in an antiparallel fashion, meaning that the DNA strands have opposite polarity. (Alberts et al., 2022; Nelson & Cox, 2021)

1.1.2. A polymer model for DNA

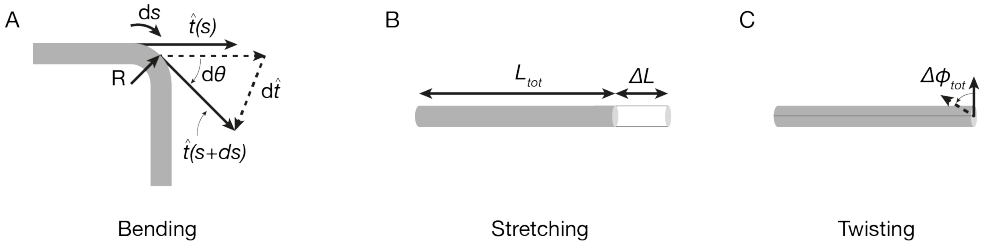


Figure 1-2 Deformation of a thin elastic rod.
(**A**) The bend vector is illustrated as a circular segment of the rod. Arc length or contour length is described by *s*. The tangent vector is moved by *ds* and compared to the position at *s*. The difference of these vectors can be described as $d\theta$. (**B**) Stretching of the thin rod. The total length L_{tot} is extended by ΔL ($u = \Delta L/L_{tot}$). (**C**) Twisting of the thin rod. The rod is twisted by $\Delta\Phi_{tot}$ ($\omega = \Delta\Phi_{tot}/L_{tot}$). (adapted from Biological Physics: Energy, Information, Life, Nelson (2014))

How can we describe the complex structure of DNA using a physical model? As discussed in the previous section, DNA is composed of building blocks that link

together to form two strands which twist around each other into a helical shape. When viewed on a larger scale, DNA can be modeled as a very long, thin rod (diameter ≪ length). In this section, we will explore the physical properties of DNA, focusing on its mechanical behavior. The following discussion on the derivation of the worm-like chain model is primarily based on Chapter 9.1.2 of "Biological Physics: Energy, Information, Life" by Nelson (2014).

DNA can be modeled as a long rod with a total length of L_{tot}. The geometry of the rod can be altered by applying force or torque. We consider small segments of length *ds* at a distance *s* along the rod. The segment is deformed by stretching *u(s)*, bending *b(s)*, and twisting *w(s)*, describing changes in length, tangent vector, and rotational orientation along the rotational axis in relation to the neighboring segments, respectively (Figure 1-2 A-C). In the context of DNA, when considering a molecule with *N* base pairs, we can ask the following questions. What is the difference between the contour length of natural DNA (0.34 nm x *N*) compared to the current extension (stretching)? How much is the base pair tilted relative to the next one along the helical axis (bending)? What is the number of base pairs required to complete a helical turn (twisting)? Twist is one important feature of DNA which we will discuss in more detail in a later section of the introduction (section 1.1.4). The twist density can be described as a measure of rotation $\Delta\Psi$ relative to the previous base pair along the DNA:

$$\omega = \frac{\Delta\psi}{0.34\ \mathrm{nm}} - \omega_0, \text{where } \omega_0 = \frac{2\pi}{10.5\ \mathrm{bp}} \frac{1\ bp}{0.34\ nm} \qquad 1.1$$

To describe the elastic energy cost *E* required to deform the thin, long elastic rod representing DNA, we divide the rod into many short segments at position *s* and length *ds*. The total energy cost can be obtained by summing the deformation energy *dE(s)* across all segments. Assuming each deformation follows Hooke's law (quadratic function: $U = ½\, k\, x^2$) for each type of deformation, we arrive at the following expression, where, k_BT represents the product of the Boltzmann constant and temperature:

$$\mathrm{dE} = \frac{1}{2} k_B T[Ab^2 + Bu^2 + C\omega^2 + 2Du\omega]ds \qquad 1.2$$

A is the bend persistence length and *C* is the twist persistence length. Bk_BT describes the stretch stiffness and Dk_BT accounts for twist-stretch coupling effects. We can simplify this equation by making assumptions about our long, thin rod. First, we assume a force regime where the rod is inextensible ($Bk_BT >> F_{applied}$), leading to a fixed

total length and omitting the stretching term. Furthermore, we assume that the rod is free to rotate at the attachment points leading to no contribution of the twist and twist-stretch coupling term. This leaves us with a simplified elastic rod model, also known as the Kratky-Porod model or more commonly as the worm-like chain model (WLC) (Figure 1-3 A):

$$E = \frac{1}{2} k_B T \int_0^{L_{tot}} ds\, Ab^2 \tag{1.3}$$

To test this model, we require a useful approximation, as the equation cannot be solved analytically. The WLC model can be approximated with the following equation (Bustamante et al., 1994; Marko & Siggia, 1995):

$$\frac{fA}{k_B T} = \frac{z}{L} + \frac{1}{4\left(1 - \frac{z}{L}\right)^2} - \frac{1}{4} \tag{1.4}$$

In this approximation, the product of the force f and the (bend) persistence length A is divided by k_BT. The equation scales with extension z over the contour length L. In this context, persistence length describes the rigidity of our polymer. As a comparison, at physiological salt concentrations, DNA has a persistence length of ~50 nm (Garcia et al., 2007; Porschke, 1991), while actin has a persistence length of around 17 µm (Khan et al., 2021).

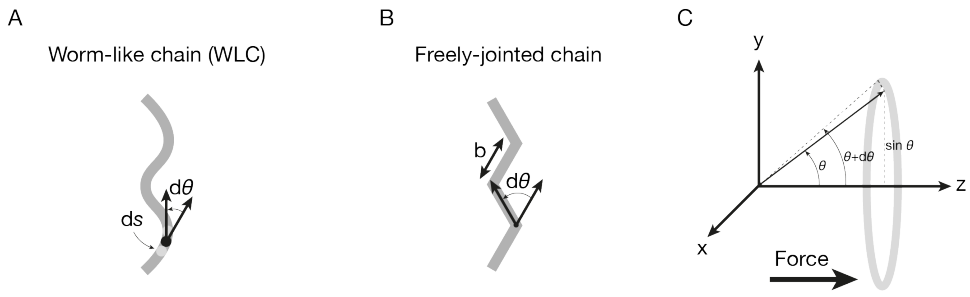


Figure 1-3 Comparison of polymer models for DNA.
(**A**) Worm-like chain (WLC). The curvature and rigidity of the chain is described by the average angle θ between adjacent monomers in the polymer chain. (**B**) Freely-jointed chain (FJC). The polymer chain is viewed as a chain of rigid segments connected by a flexible linker. The segments are described by the Kuhn length *b*. (**C**) The statistical weight of various conformations of a rigid segment in a uniform force field. The conformation change of the segment is limited by the sector between angle *θ* and *θ + dθ*. The sector is proportional to the area of a unit sphere with a radius of *sin(θ)*. (adapted from Sarkar and Rybenkov (2016))

An even simpler polymer model to describe the relation between force and extension is the freely-jointed chain (FJC) model first proposed by Werner Kuhn (Kuhn, 1934) (Figure 1-3 B). In this model, the polymer is described as a chain of rigid segments which are connected via completely flexible links allowing orientations in three dimensions. The only defining parameter is the segment length *b* also called Kuhn length. The rigidity of a polymer is determined by the length of its segments, with longer segments resulting in a more rigid polymer. The number of segments can be simply calculated by dividing the *L* (contour length) by *b* (*N* = *L*/*b*). We assume a random coil formation similar to diffusion or random walk (Sarkar & Rybenkov, 2016), where the average size of the coil scales with the square root of the polymer length $\langle R^2 \rangle = Lb$.

If we apply a force on the polymer chain to stretch it, it counteracts with an entropic, elastic force. In this constant force field, the segments can be described by a Boltzmann distribution:

$$W(\theta) = e^{\left(\frac{-\Delta G}{k_B T}\right)} = 2\pi \sin\theta \, e^{\left(\frac{Fb\cos\theta}{k_B T}\right)} \qquad 1.5$$

The free energy of a segment in the force field is described by ΔG. θ is the angle between the orientation of the force field and segment, where $b\cos\theta$ is the projection of the segment on the force. When multiplied with *F*, it results in the energy of the segment within the force field (Figure 1-3C). The statistical weight of the conformation is accounted for by $2\pi \sin(\theta)$. The average projection of the segments onto the force field, *z*, is described by the following expression: (Sarkar & Rybenkov, 2016)

$$z = \frac{\int_0^\pi b\cos\theta \, W(\theta) d\theta}{\int_0^\pi W(\theta) d\theta} = b\mathcal{L}\left(\frac{F\,b}{k_B T}\right) \qquad 1.6$$

Where $\mathcal{L}(x)$ is the Langevin function ($\mathcal{L}(x) = \coth(x) - 1/x$). Since we assume that all segments are equally influenced by the force, the relative extension *z*/*L* would be given by:

$$\frac{z}{L} = \mathcal{L}\left(\frac{F\,b}{k_B T}\right) \qquad 1.7$$

Now we assume a low extension of the polymer ($Fb \ll k_B T$). We can use a Taylor expansion of the Langevin function, which results in x/3 for the first term. The force at low extension can be calculated as follows:

$$F = \frac{3k_B T}{b\,L} z \tag{1.8}$$

We end up with an equation analogous to Hooke's law, which we also assumed for the worm-like chain earlier in this section. Interestingly, if we assume low force for the WLC expression (Equation 1.4) (z/L ≪ 1), we can make the following approximation using another Taylor expansion:

$$\frac{1}{\left(1 - \frac{z}{L}\right)^2} \approx 1 + 2\frac{z}{L} \tag{1.9}$$

We can express the force of the WLC as follows:

$$F = \frac{3k_B T}{2A\,L} z \tag{1.10}$$

We can see that equation 1.8 and equation 1.10 are the same if we use the relation *2A = b* for the relation between persistence length *A* and Kuhn length *b*. The freely-jointed chain model describes DNA extension at low forces quite well. However, at higher forces, the FJC model follows the qualitative trend well but fails to match quantitatively (Smith et al., 1992). When comparing experimental data and the WLC prediction from the continuous, isotropic polymer model using the persistence length A, the prediction is a quite good representation of the actual trend (Bustamante et al., 1994). The WLC works for forces less than 5-10 pN (Vilfan et al., 2009). Furthermore, new effects come into play when applying forces high enough to extend the length of the DNA beyond the contour length (Cluzel et al., 1996; Schakenraad et al., 2017). What is not described by the freely-jointed chain nor the worm-like chain model for DNA is the contribution of twist and how it influences the polymer structure. To summarize, the presented models are good approximations to estimate the force required to extend a DNA molecule that is free to swivel and release twist, while being extended to less than the contour length.

1.1.3. DNA topology

The double-stranded, helical structure of DNA provides a durable environment for the genetic code, promoting genome stability. However, when the two strands are accessed for the synthesis of DNA or RNA, they must be locally separated. Since free rotation is restricted in a cellular context due to topological barriers, this separation changes DNA topology. DNA is topologically constrained in both circular bacterial

chromosomes and eukaryotic chromatin. How DNA topology benefits chromosomal organization in bacteria (section 1.1.5) and the challenges it poses during cellular processes (sections 1.3 and 1.4.2) will be discussed in detail in later sections. (Koster et al., 2010; Nelson & Cox, 2021)

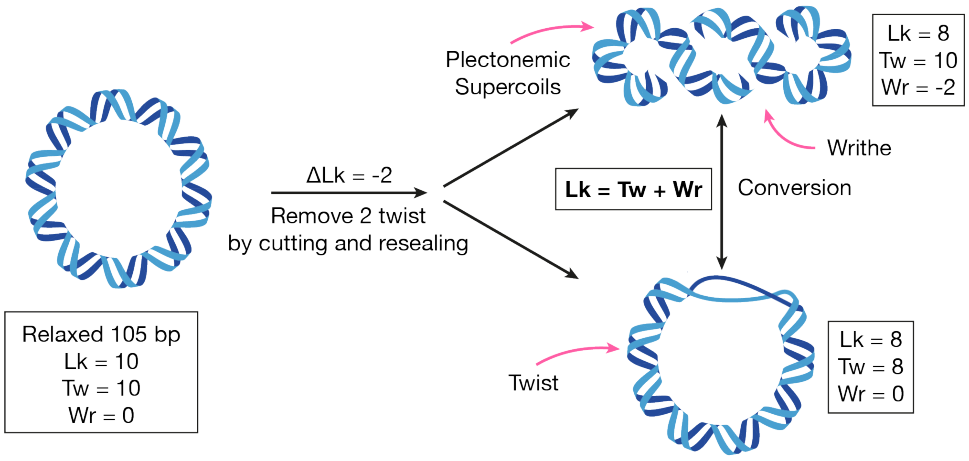


Figure 1-4 DNA Topology.
Linking number *Lk* describes the topological state of DNA. In this example the *Lk* is changed by -2 of a relaxed circular plasmid (*Lk* = 10). Lk change can be compensated by introduction of writhe (*Wr*) or at the expense of twist (*Tw*). (adapted from Koster et al. (2010))

To understand DNA topology and the associated benefits and challenges for cells, we first need to discuss the basic concept. Topology is only defined for DNA molecules with constrained ends. For example, closed-circular DNA such as plasmids (Koster et al., 2010) and linear DNA with ends that are fixed due to protein interaction have defined DNA topology. Topology is not defined for DNA with a single strand break (nicks or nicked DNA). To quantify the topological state of DNA, the linking number *Lk* is used which describes the number of helical windings around the central axis (twist) and the number of times the helix is crossing itself (writhe) (Koster et al., 2010) (Figure 1-4). The relationship between twist (Tw) and writhe (Wr) can be described with the following equation (Fuller, 1971; White, 1969):

$$Lk = Tw + Wr \tag{1.11}$$

Altering the linking number is only possible if a break is introduced and resealed. To describe the state of relaxed DNA, Lk_0 is used. For example, to calculate the linking number of a 105 bp long DNA molecule you divide the number by 10.5 bp which results

in Lk_0 = 10. The difference between the current linking number *Lk* and the relaxed state Lk_0 is described by linking number difference *ΔLk*. When overwinding or underwinding the DNA helix, *ΔLk* has positive or negative value, respectively. Twist can be transformed into writhe. Writhe describes DNA loops which are also called supercoils or plectonemes. When comparing two different sized DNA molecules under the same experimental conditions, the superhelical density σ is used, which is independent of DNA length (Equation 1.12). (Deweese et al., 2008; Koster et al., 2010; Nelson & Cox, 2021; Seol & Neuman, 2016)

$$\sigma = \frac{Lk - Lk_0}{Lk_0} \qquad 1.12$$

For example, consider a circular relaxed plasmid with 105 base pairs (bp) (Figure 1-4). It has an initial linking number of 10 and a superhelical density *σ* of 0. When the linking number *Lk* is reduced by 2 (*ΔLk* = -2), the number of twists is reduced by 2, resulting in *Tw* = 8. Alternatively, plectoneme formation can lead to a writhe change of -2. The superhelical density σ after the change of *Lk* is -0.2. In a plasmid ten times the size (1050 bp), a change of -20 in twist or -20 in writhe would need to be introduced in order to archive the same superhelical density.

1.1.4. Modeling the effects of twist on the DNA structure

In Section 1.1.2, we introduced a polymer models for DNA that assumed it was free to rotate, thereby neglecting the twist term in the description of force-extension behavior. However, altering the degree of twist can lead to significant structural changes, such as the aforementioned formation of plectonemes. Here, we will describe the behavior of DNA under varying levels of twist using physical equations. Experimental studies utilizing magnetic tweezers and angular optical traps have provided insights into the behavior of supercoiled DNA under applied forces and topological constraints validating these equations (Forth et al., 2008; Strick et al., 1996; Strick et al., 1998). In these experiments, one end of the DNA was attached to a surface and the other end was fixed to either a bead or quartz cylinder, resulting in both end of the DNA being torsionally constrained. Rotation of the bead or cylinder induced rotation in the DNA, thereby providing controlled manipulation of its topological state. These techniques have been instrumental in advancing our understanding of the behavior of DNA under the influence of twist.

The linking number (Equation 1.11), which describes the relationship between the twist and writhe components, remains constant unless altered by rotating the DNA strand. Initially, changes in the linking number ΔLk result in a change in twist (Strick et al., 1998). Retaining the term for the twist contribution from Equation 1.2, we can describe the energy stored in each twist as follows:

$$E_{twist} = \frac{1}{2}\frac{\mathrm{k_B T}\, C}{L}(\Delta\phi)^2, \text{where } \Delta\phi = 2\pi n \qquad 1.13$$

We model the DNA as a torsional spring and apply Hooke's law to describe its behavior. Let *n* represent the number of full turns applied by the magnet that controls the magnetic bead to which one end of the DNA is attached. As the length of the DNA molecule *L* increases, the twist energy stored per turn decreases (Equation 1.13) since they are inversely proportional. As long as the change in linking number is only going into changes in twist (and not writhe), the torque *Γ* increases linearly. The torque Γ can be expressed using the following equation (Strick et al., 2003; Vilfan et al., 2009):

$$\Gamma = \frac{\mathrm{k_B T}\, C}{L} 2\pi n \qquad 1.14$$

After applying additional rotations, the DNA undergoes the buckling transition, in which it forms plectonemes rather than accumulating additional twist (Strick et al., 1996). The buckling transition is force dependent and is defined by the buckling torque Γ_B. Once the buckling torque Γ_B is reached, changes in the linking number result in the formation of plectonemes with a radius R_p, reducing the extension by the circumference of the loop/plectoneme ($2\pi R_p$). To derive expressions for the buckling torque Γ_B and the plectoneme radius R_p, we can balance the torsional energy against the work required to from a loop and the bending energy involved in its formation (Strick et al., 2003; Vilfan et al., 2009):

$$E_{loop} = 2\pi R_p \mathrm{F} + \frac{\pi A k_B T}{\mathrm{R_p}} \qquad 1.15$$

The first term describes the work against a force *F*, while the second term accounts for the bending energy required to form a complete loop. By minimizing E_{loop} with respect to R_p, we can find expressions for the buckling torque Γ_B and the change in extension Δz as functions of the plectoneme radius R_p (Forth et al., 2008; Strick et al., 2003):

$$\Gamma_B = \sqrt{2k_B TAF} \quad 1.16$$

$$\Delta z = 2\pi \mathrm{R_p} = 2\pi \sqrt{\frac{k_B TA}{2F}} \quad 1.17$$

This simplified model correctly captures the trend qualitatively when compared to experimental data (Vilfan et al., 2009). The buckling torque Γ_B increases with force. To describe the buckling behavior, we use the bend persistence length *A*, as forming a loop requires bending the DNA to generate writhe. This is in contrast to twist deformation (Equation 1.14), which depends on the torsional persistence length *C*. As the external force increases, the loop size decreases. However, this model does not account for the handedness of DNA, which affects the buckling transition depending on whether the DNA is rotated clockwise or counterclockwise. At low force, the coiling behavior is symmetric for both negative (underwound) and positive (overwound) torque. As the force increases, the behavior becomes asymmetric: positive torque shifts the buckling point as described in equation 1.16, while for negative torque the DNA denatures locally (Strick et al., 1996) instead of forming loops. However, the model fails to accurately describe the behavior at lower forces. More complex models that better capture the trend at low forces are available, but they typically require numerical solutions (Vologodskii & Marko, 1997).

1.1.5. Chromosomal compaction and organization in bacteria

DNA must be bent and twisted to fit within the confines of a cell. If fully extended, the 4.6 MB chromosome in *E. coli* would measure approximately 1.6 mm in length, yet it occupies only 60% of the total cell volume (Wu et al., 2019). It has to be condensed more than a 1,000-fold in order to fit inside a cell with a 1 μm diameter (linear compaction = physical length/diameter) (Holmes & Cozzarelli, 2000). It is even more impressive for human cells, in which the genetic code has to be linearly compacted 200,000-fold (Holmes & Cozzarelli, 2000). The study of these compaction mechanisms presents a fascinating and dynamic field of research and have implications in many research disciplines. In this section, we will focus on the bacterial chromosome and the mechanisms involved in its compaction and organization.

The bacterial chromosome is organized into discrete bodies with soft and irregular outlines, known as nucleoids (Robinow & Kellenberger, 1994). In the early 1970s, a method was developed to gently lyse *E. coli* cells, allowing for the direct visualization

of nucleoids using electron microscopy (Kavenoff & Ryder, 1976). This led to the iconic image of the bacterial chromosome, revealing plectonemic loops emerging from a dense core structure. These observations inspired the development of the rosette model, in which plectonemes are proposed to organize the nucleoid. The resulting structure resembles a bottlebrush (Wang et al., 2013) (Figure 1-5). The complexity and density of this structure raise important questions about how it remains organized and compacted while still allowing access to the genetic material for essential cellular processes.

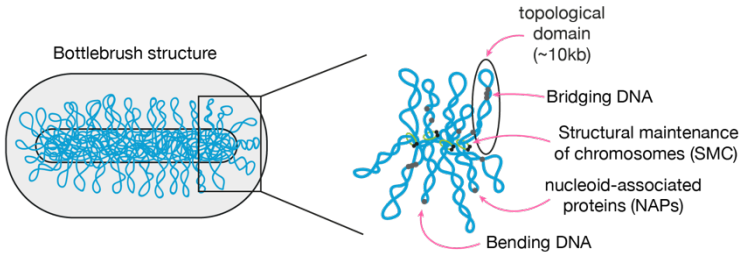


Figure 1-5 Bacterial nucleoid structure and organization.
The bacterial chromosome forms a bottlebrush-like structure, with plectonemic loops representing the bristles. A zoomed-in view reveals a complex network of DNA and nucleoid-associated proteins (NAPs), which organize the DNA through bending and bridging interactions. Structural maintenance of chromosomes (SMC) proteins play a crucial role in this organization. Topological domains within the nucleoid have an average size of 10 kb.

Topological domains, also known as plectonemic loops, represent a key mechanism for chromosome compaction (Figure 1-5). The DNA is underwound (negatively supercoiled) with the superhelical density σ maintained between -0.07 and -0.05 (Equation 1.12) (Fulcrand et al., 2016; Nelson & Cox, 2021). This supercoiling induces the formation of plectonemes, which bring distant regions of the DNA closer together. Supercoiling not only aids in organizing the chromosome but also plays a role in chromosome segregation (Joyeux, 2020; Jun & Wright, 2010; Sawitzke & Austin, 2000). While a single nick can relax a plasmid, relaxing an entire chromosome requires multiple nicks, suggesting that the chromosome is organized into domains with contained topology (Worcel & Burgi, 1972). These topological domains vary in size but average around 10 kb (Postow et al., 2004), leading to an estimated 460 such domains within the bacterial chromosome. These domains contribute to DNA condensation, DNA decatenation, DNA repair, and protection from relaxation, particularly by keeping double-strand break ends in close proximity (Postow et al., 2004).

Nucleoid-associated proteins (NAPs) are thought to act as boundary elements that restrict the free rotation of DNA, thereby creating and maintaining topological domains. A subgroup of these proteins is often compared to histones in eukaryotes, which is why they are called histone-like proteins. In *E. coli*, key NAPs include HU, IHF, Fis, and H-NS. These proteins compact and stabilize DNA loops by bending and bridging interactions, both at specific and nonspecific sites along the chromosome (Dillon & Dorman, 2010; Wang et al., 2013) (Figure 1-5). Cells lacking these factors exhibit defects in chromosome segregation (Wang et al., 2013). However, the overall condensation of the chromosome is not drastically affected, suggesting that another set of proteins plays a more critical role in constraining topological domains (Wang et al., 2013).

Structural maintenance of chromosomes (SMC) proteins have become a focal point of research investigating the mechanisms driving chromosome compaction (Brandao et al., 2021; Davidson et al., 2019; Ganji et al., 2018; Pradhan et al., 2023). The functions of SMC proteins significantly impact chromosomal organization and compaction and are highly conserved across all domains of life (Hirano, 2006). In eukaryotes, SMC proteins are involved in chromosome compaction, DNA repair, and sister chromatid cohesion, with different SMC variants such as condensin, SMC5/6, and cohesin fulfilling these roles (Nasmyth & Haering, 2005). In *E. coli*, the MukBEF complex serves as the functional and structural analog (Burmann et al., 2021; Rybenkov et al., 2014). *E. coli* cells can survive without the SMC complex at lower temperatures, but they exhibit severe defects in chromosome segregation and display decondensed nucleoids (Wang et al., 2013). SMC proteins may play a crucial role in constraining topological domains (Burmann & Lowe, 2023).

In contrast to depictions of the chromosome as a static structure composed solely of loops, it must actually remain dynamic in order to facilitate integral interactions with the genetic material (Guo et al., 2021; Janissen et al., 2024; Postow et al., 2004). The dynamic topological state of the chromosome is influenced by processes such as transcription and replication, which must be carefully regulated (Koster et al., 2010). Topoisomerases are the enzymes responsible for maintaining these changes in topology.

1.1.6. Topoisomerases: Mechanisms and cellular roles in DNA topology

Topoisomerases are a family of enzymes responsible for modulating DNA topology by introducing transient breaks in DNA strands, thereby maintaining chromosome structure (Bush et al., 2015; Vos et al., 2011). Processes such as transcription, replication, and repair require the unwinding and rewinding of double-stranded DNA, thereby altering its topological state by causing overwinding or underwinding. Additionally, DNA replication and repair can lead to the formation of entangled chromosomes, known as precatenanes, which can result in challenges such as segregation defects or DNA strand breaks (Bush et al., 2015; Helgesen et al., 2021; Vos et al., 2011; Wang et al., 2008; Zawadzki et al., 2015). Due to their critical role in maintaining genome integrity, topoisomerases are important targets for research, particularly for therapeutic applications (Bax et al., 2019; Pommier, 2013). Here we will discuss the categorization of topoisomerases, their general functions in cellular contexts, their working principles using bacterial DNA gyrase and Topoisomerase IV (Topo IV) as examples, and their broader significance in human cells.

The importance of topoisomerases is highlighted by their ubiquitous presence across all domains of life, where they play a critical role in resolving topological problems (Bush et al., 2015; Vos et al., 2011). Although the mechanisms by which various topoisomerases alter DNA topology differ, they share some key features (McKie et al., 2021). During their catalytic cycle, topoisomerases form a covalent DNA-topo intermediate through a phosphotyrosyl linkage between a nucleophilic tyrosine residue and the phosphate group of the DNA backbone. The nucleophilicity of the tyrosine is responsible for introducing single- or double-strand breaks in the DNA via nucleophilic attack (McKie et al., 2021; Vos et al., 2011). Topoisomerases are categorized into two classes. Type I cleaves only one strand of DNA, and Type II cleaves both strands and changes the topology. Enzymes with odd Roman numerals are classified as Type I, while those with even numerals are classified as Type II (Vos et al., 2011). Additionally, Type II topoisomerases are further subdivided into subtypes A, B, or C, based on structural, evolutionary, and mechanistic criteria (Bush et al., 2015; McKie et al., 2021). Given the complexity of topoisomerases, we will focus on two examples that are particularly relevant for later discussions, which are the bacterial topoisomerases DNA gyrase and Topoisomerase IV (Topo IV). However, since the working principles are conserved across prokaryotic and eukaryotic topoisomerases, insights gained from these examples can be applied more broadly to other organisms.

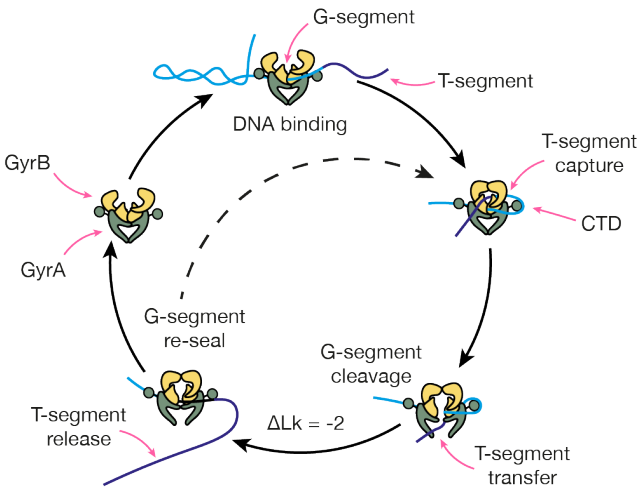


Figure 1-6 Mechanism cycle of DNA gyrase.
DNA gyrase binds to the G-segment (Gate segment) of the DNA. The T-segment (Transfer segment) is wrapped around the C-terminal domain (CTD) and captured. The G-segment is then cleaved, creating a double-strand break, through which the T-segment is passed. This process changes the DNA's linking number by -2. The G-segment is subsequently religated, and the T-segment is released. The cycle can either continue or the DNA gyrase may dissociate from the DNA. (adapted from Stracy et al. (2019))

DNA gyrase is a unique Type IIA topoisomerase primarily found in bacteria, though it is also present in plants and archaea (McKie et al., 2021). Gyrase has the ability to relax positive supercoils, introduce negative supercoils, and decatenate DNA in an ATP-dependent manner. Additionally, it can relax negative supercoils independently of nucleotides (Bush et al., 2015). The *E. coli* gyrase is a heterotetramer composed of two GyrA and two GyrB subunits (Figure 1-6). It is an ideal target for antibacterial drugs because it is essential in bacteria but absent in humans (Collin et al., 2011). The importance of DNA gyrase in bacteria lies in its ability to introduce negative supercoils into the chromosome, which is crucial for initiating cellular processes such as replication and transcription by facilitating DNA strand separation. Later in section 1.3.1, we will discuss its role in the elongation phase of DNA replication. Mutations in DNA gyrase lead to a significant decrease in replication and transcription efficiency (Kreuzer & Cozzarelli, 1979). The enzyme functions by binding to the G-segment of DNA, the region where a transient DNA break is introduced (Figure 1-6). The 'G' stands for gate. A second stretch of DNA, the T-segment, is wrapped around one of the C-terminal domains (CTD) of GyrA and passed through the G-segment, thereby introducing negative supercoils (Nollmann, Crisona, & Arimondo, 2007). Each catalytic cycle of gyrase results in a change in the topological state of DNA, with a *ΔLK* of -2

(Brown & Cozzarelli, 1979). The activity of gyrase depends on the force applied to the DNA and whether it is relaxing positive supercoils or introducing negative ones (Agarwal & Duderstadt, 2020; Nollmann, Stone, et al., 2007).

Topo IV is a bacterial Type IIA topoisomerase that plays a crucial role in decatenating replication products, relaxing both positive and negative supercoils, and knotting and unknotting DNA in an ATP-dependent manner (Bush et al., 2015). Topo IV is a heterotetramer composed of two ParC and two ParE subunits. While these subunits are homologous to GyrA and GyrB, Topo IV is not capable of introducing negative supercoils into DNA (Bush et al., 2015). However, Topo IV is approximately 100 times more active in decatenation in vivo compared to DNA gyrase (Bush et al., 2015). The discovery of Topo IV was linked to studies on partitioning defects, suggesting its involvement in chromosome segregation (McKie et al., 2021). Additionally, Topo IV interacts with MatP and SMC proteins from the *E. coli* MukBEF complex, which are involved in structuring and organizing the bacterial chromosome (Nolivos et al., 2016; Wang et al., 2013). Topo IV also supports replication and transcription, topics that will be discussed in more detail in section 1.3.

DNA gyrase is a key target for antibiotics in the fight against bacterial infections (Collin et al., 2011). It is also important to consider the implications of abnormal topoisomerase function in humans. A review by Pommier et al. (2022) provides a comprehensive summary of the roles human topoisomerases play in genome stability and organization. Human cells are highly complex, containing six specialized topoisomerases that interact with DNA, RNA, and mitochondrial DNA topology. The chromatin structure in eukaryotic cells presents a particularly challenging environment for maintaining DNA topology, due to the constant changes driven by replication, transcription, and the additional complexity of chromatin remodeling. However, the topological challenges encountered during replication and transcription in human cells are similar to those observed in bacterial replication, a topic that will be explored in more detail in section 1.3. Like their bacterial counterparts, eukaryotic topoisomerases are essential for supporting replication, transcription, and genome organization. For instance, the eukaryotic topoisomerase II (TOP2) appears to be recruited to support the SMC protein cohesin by eliminating topological barriers or DNA entanglements, allowing DNA to be translocated through the SMC ring. Topoisomerases also contribute to genome stability; for example, TOP3A is involved in DNA end resection and homology-directed repair. However, dysfunctional topoisomerase activity can

trigger the activation of repair pathways themselves. Insufficient activity can interfere with replication and transcription, leading to the accumulation of R-loops, G-quadruplexes, Z-DNA, catenanes, and knots. During their catalytic cycle, topoisomerases form topoisomerase cleavage complexes (TOPcc). If these complexes are not resolved by completing the catalytic cycle, they can become topoisomerase DNA-protein crosslinks (TOP-DPCs), which are associated with single-strand and double-strand breaks induced by environmental and metabolic factors. Understanding this mechanism is particularly beneficial, as the trapping of TOP1ccs is the mechanism of action for certain anticancer drugs. Notably, TOP1 is involved in removing topological challenges ahead of the replication fork.

In summary, topoisomerases play a crucial role across all domains of life, participating in the most essential cellular pathways. Before discussing how topoisomerases support DNA replication, it is necessary to first understand the mechanics of DNA replication itself.

1.2. DNA replication

1.2.1. Basic principle of a bacterial replisome

We have established that the chromosome is a complex and massive structure that must be accurately replicated and segregated to produce two new cells. This process must occur without errors to avoid altering the genetic sequence. It is essential to understand the working principles of the replisome before discussing the topological challenges that arise during replication that can threaten chromosome integrity. Replisomes are the molecular machines that carry out DNA replication. They consist of multiple proteins that dynamically assemble and disassemble (Alberts et al., 2022). In the following section, we will describe the working principles of a bacterial replisome, using the *E. coli* replication machinery as an example (Figure 1-7B). The overall process of replication is divided into three stages: initiation, elongation, and termination (Nelson & Cox, 2021; Xu & Dixon, 2018).

Initiation is the first stage of the replication process, during which the replisome assembles on the DNA and initiates the process of DNA replication. In circular double-stranded DNA (dsDNA), there is a single initiation site known as the origin of replication (oriC) (Xu & Dixon, 2018). This 250 bp sequence is rich in AT pairs and contains DnaA boxes, which are recognized by multiple copies of the replication initiation protein,

DnaA. The binding of DnaA triggers the melting of the oriC dsDNA through an ATP-dependent stretching mechanism (Duderstadt & Berger, 2013; Duderstadt et al., 2011). Once DnaA melts the oriC, creating the replication bubble, two copies of the helicase DnaB are recruited and loaded onto the exposed ssDNA. DnaC acts as the helicase loader, trapping DnaB in a right-handed lockwasher configuration (Arias-Palomo et al., 2013; Arias-Palomo et al., 2019). In solution, DnaB and DnaC form $DnaB_6$-$DnaC_6$ complexes, and a N-terminal domain of DnaA assists these complexes in locating the melted origin (Bleichert et al., 2017). The two $DnaB_6$-$DnaC_6$ complexes are loaded onto each of the two ssDNA strands in a 5'-3' orientation, corresponding to the lagging strand of the future replication fork, forming to replication forks. Due to the antiparallel nature of DNA, the complexes are loaded in opposite directions. DnaG, the primase, then interacts with $DnaB_6$, leading to the dissociation of DnaC (Makowska-Grzyska & Kaguni, 2010). DnaG is responsible for synthesizing the RNA primer required for leading-strand DNA elongation, as well as for repriming the Okazaki fragments on the lagging strand.

The elongation phase involves the synthesis of new DNA strands by the replisome (Figure 1-7A). This process can be divided into leading and lagging strand synthesis, which occurs simultaneously Due to the fixed directionality of polymerase synthesis and the antiparallel orientation of DNA strands, the two daughter strands are synthesized in different ways. The leading strand is synthesized continuously, while the lagging strand is synthesized in short segments known as Okazaki fragments (OFs) (Xu & Dixon, 2018). OF fragments are frequently reprimed by RNA primers, which are later replaced by DNA through the action of DNA polymerase I. The ssDNA is protected by single-stranded DNA-binding proteins (SSBs). It is important to note that the lagging strand contains nicks, which are subsequently sealed by DNA ligase (Xu & Dixon, 2018). The primary replicative polymerase in E. coli is the Pol III holoenzyme, which is composed of 10 different proteins. This complex can be divided into the polymerase core ($\alpha\varepsilon\theta$), the sliding clamp (β_2), and the clamp loader complex ($\delta\tau_n\gamma_{3-n}\delta'\psi\chi$) (Xu & Dixon, 2018). The β_2 clamp is loaded onto the RNA primer terminus by the clamp loader complex. Pol III is tethered to the clamp, which ensures high processivity and the rapid synthesis of DNA (Jergic et al., 2013).

Termination is the final phase of DNA replication, occurring when the two replication forks meet head-on at a point opposite the origin of replication (oriC) (Nelson & Cox, 2021). In this region, there are ten 23 bp repeats known as termination sites (*Ter*), which

serve as binding sites for the monomeric termination protein Tus (Berghuis et al., 2015; Elshenawy et al., 2015; Neylon et al., 2005). The Tus-*Ter* complex forms in a 1:1 ratio, with each complex having a permissive and a non-permissive face. The permissive face allows the replisome to pass, while the non-permissive face stalls it (Mulcair et al., 2006). The ten *Ter* sites are organized into two groups: the first five are permissive, followed by five non-permissive sites. This arrangement ensures that replication forks converge within the terminus region, facilitating subsequent chromosome segregation. Upon completion of termination, the cell is ready to divide into two new cells, initiating a new cycle of replication. The process of replication is very complex and involves many different proteins working together to create the new daughter strands. In the next section we will discuss a model replisome with less components resembling bacterial replication.

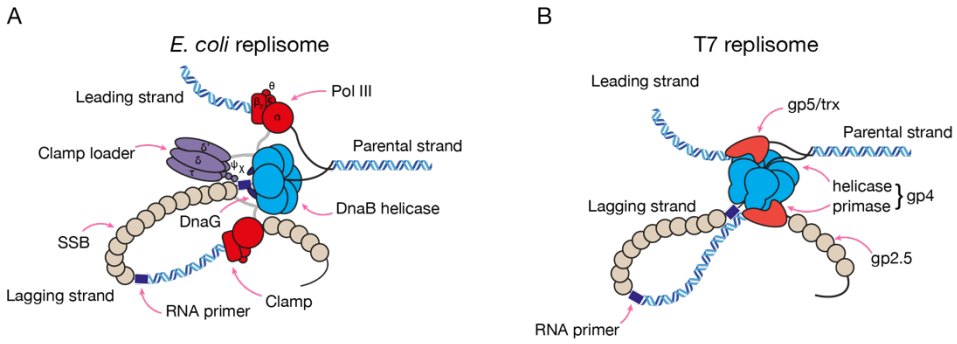


Figure 1-7 Replisome composition of *E. coli* and T7.
(**A**) *E. coli* replisome during elongation phase of replication. The parental DNA strand is separated into leading and lagging strands by DnaB (helicase). The leading strand is continuously synthesized by Pol III (αεθ). The β_2 clamp, located at the polymerase, enhances speed and processivity. Pol III is tethered to DnaB by interaction with the clamp loader complex ($\delta\tau_n\gamma_{3-n}\delta'\psi\chi$). The lagging strand passes through the helicase and is primed by DnaG (primase), leading to its discontinuous synthesis of Okazaki fragments (OF). Exposed ssDNA is protected by SSB (ssDNA-binding proteins). (**B**) T7 replisome during elongation phase of replication. The parental DNA strand is separated into leading and lagging strand by gp4, which functions as both the helicase and the primase. The leading strand is synthesized by polymerase gp5 together with the processivity factor thioredoxin (trx) from *E. coli*. Both the leading and lagging strand polymerases are directly attached to the helicase. The lagging strand is synthesized discontinuously as OF. Exposed ssDNA is protected by gp2.5, the ssDNA-binding protein that also ensures coordinated replication.

1.2.2. The T7 replisome

The T7 chromosome is a 40 kb long linear DNA molecule encoding 50 genes (Kulczyk & Richardson, 2016; Lee & Richardson, 2011). The T7 replisome can replicate DNA

using four proteins: the helicase-primase gene product 4 (gp4), the polymerase gene product 5 (gp5), the processivity factor thioredoxin (trx) from *E. coli*, and the ssDNA-binding gene product 2.5 (gp2.5) (Figure 1-7B). The helicase gp4 is a homohexamer that unwinds the dsDNA to form ssDNA, while also synthesizing the RNA primers used for the discontinuously synthesized OFs on the lagging strand. The leading strand is synthesized continuously by gp5 together with the cofactor thioredoxin (trx) from the host *E. coli*. To protect the ssDNA and for coordinated replication, gp2.5 is essential (Hernandez & Richardson, 2019). These four proteins form the minimal replication system from T7 (Kulczyk & Richardson, 2016).

In the case of *E. coli* and other replication systems, many more proteins must work together to synthesize DNA. This increased complexity poses challenges for reconstituting DNA replication in single-molecule assays or other biological studies when compared to the T7 system. However, the core principles of replisome architecture are remarkably similar across different organisms (O'Donnell et al., 2013; Yao & O'Donnell, 2010), allowing researchers to reduce complexity and focus on the physical behavior of the replisome. The T7 replication system is a simplified replisome that shares the fundamental mechanisms required for DNA duplication with the more complex systems. Recent structural studies have revealed striking similarities between the T7 replisome and those of bacterial and eukaryotic systems, making T7 an excellent model due to its simplicity (Gao et al., 2019). Complex single-molecule assays can first be conducted using the T7 system, providing a foundation for investigating the more intricate protein interactions in more complex systems, such as studying dynamic changes in replisome composition (Geertsema et al., 2014; Lewis, Spenkelink, Jergic, et al., 2017). Additionally, the human mitochondrial replication system shares similarities with T7, with individual proteins exhibiting sequence and structural homology (Shutt & Gray, 2006). These similarities have led to the hypothesis that some genes necessary for mitochondrial replication may have originated from an ancestor of T-odd phage (mitochondrial endosymbiosis). Developing methods with a simpler system like T7 can accelerate progress, as the reduced complexity allows for more straightforward experimentation. This makes T7 an ideal model for single-molecule studies. In summary, the simplicity of the T7 replisome has enabled comprehensive structural studies of the entire replisome (Gao et al., 2019; Kulczyk et al., 2017) and visualization of the replication process at the single-molecule level (Duderstadt et al., 2016; Geertsema et al., 2014; Hamdan et al., 2009; Lee et al., 2006).

1.2.3. The dynamics of a replisome

The relative simplicity of T7 replisome has allowed for structural studies of the entire replication machinery (Gao et al., 2019; Kulczyk et al., 2017). Modern cryo-electron microscopy (cyro-EM) has provided structural insights, including a helicase, primase, and polymerase, actively synthesizing DNA (Gao et al., 2019; Kulczyk et al., 2017). Such structures enable detailed insights into the replication process (Abid Ali et al., 2016; Fernandez-Leiro et al., 2015). However, structural studies often capture the protein complex in a static state, making it challenging to observe dynamic behavior. In contrast, single-molecule techniques have a lower resolution limit, but they provide the ability to study the compositional dynamics of these complexes, thereby offering critical insights into their function (Duderstadt et al., 2016; Geertsema et al., 2014; Lewis, Spenkelink, Jergic, et al., 2017; Lewis et al., 2020).

During DNA replication, the dsDNA must separate into a fork or Y-shaped junction, called the replication fork (Figure 1-8). The leading strand polymerase is positioned at the opening of the helicase gp4. Structural studies have identified a β-hairpin located on the polymerase, which aids in the separation of DNA between the leading strand polymerase and the helicase gp4 (Gao et al., 2019; Kulczyk et al., 2017). Within the helicase gp4, the ssDNA adopts a helical structure with 12 base pairs per turn, resembling the B-DNA conformation. This structure has a diameter of 2.3 nm, and the rise per base pair decreases from 0.34 nm in B-DNA to 0.29 nm. Notably, the helicases from *E. coli* (DnaB) and yeast (CMG) exhibit similar helical conformations to gp4 (Georgescu et al., 2017; Itsathitphaisarn et al., 2012). Additionally, the two lagging strand polymerases are directly attached to the primase, ready to synthesize the Okazaki fragments.

While the architecture of the T7 replisome appears static in most structural studies, single-molecule experiments have revealed a highly dynamic composition within the T7 replisome and across replisomes in general (Geertsema et al., 2014; Lewis, Spenkelink, Jergic, et al., 2017; Lewis et al., 2020; Spenkelink et al., 2019). The hexameric structure of the helicase gp4 provides binding sites for up to six DNA polymerase gp5 molecules. Single-molecule rolling circle experiments have demonstrated that up to six polymerases can be associated with the replisome at any given time (Geertsema et al., 2014). The number of polymerases is inversely correlated with the priming frequency on the lagging strand. Similarly, frequent polymerase

exchanges have been observed in *E. coli* during replication (Lewis, Spenkelink, Jergic, et al., 2017; Monachino et al., 2020). The number of Pol III* (Pol III holoenzyme lacking the β clamp) molecules present is concentration-dependent and can remain stably bound in the absence of excess components in solution. In summary, the replisome's composition is highly dynamic, allowing it to respond to environmental factors such as fluctuations in nucleotide concentrations or replisome components, thereby adapting to challenges encountered during replication.

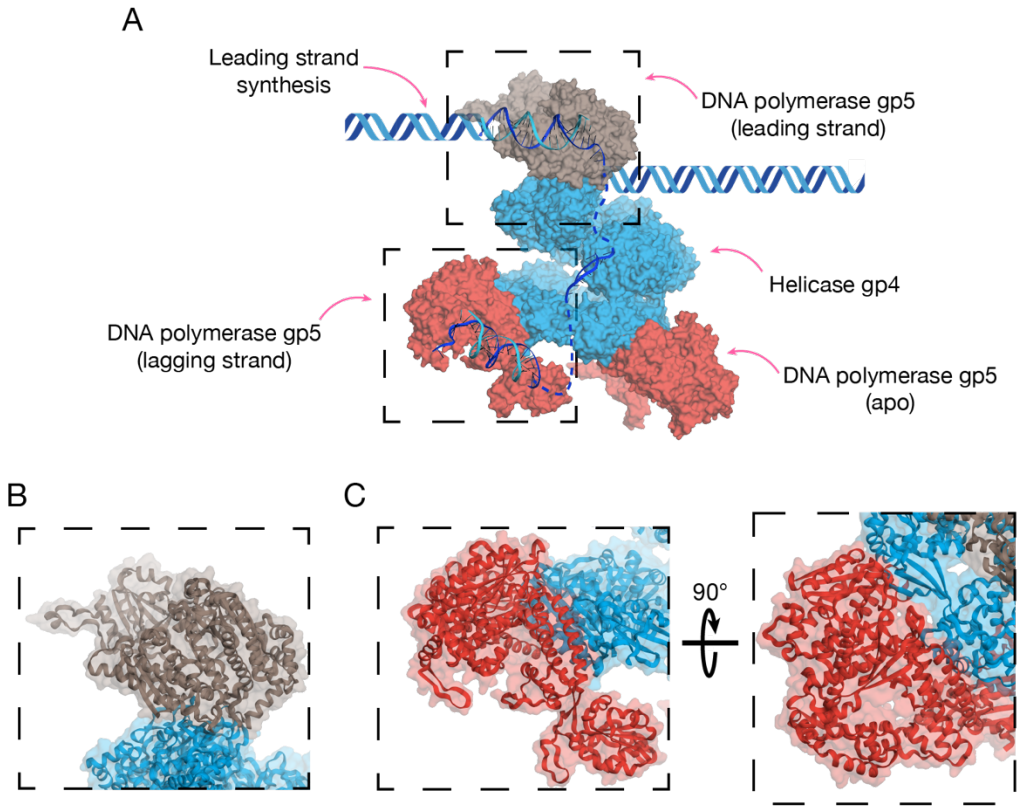


Figure 1-8 Structural rendering of T7 replisome.
(**A**) A replisome model composed of leading strand polymerase (gray), helicase (blue), lagging strand polymerase (red), and a second lagging strand polymerase (red, apo). Leading and parental strand DNA's are represented by DNA cartoon to illustrate potential orientation. (**B**) Zoom in on leading strand polymerase demonstrating extensive surface area at the protein-protein interface. (**C**) Zoom in on lagging strand polymerase and rotated by 90°, also demonstrating extensive surface area at the interface. To create the rendering, cryo-EM maps from Gao et al. (2019) (Appendix, Table 5-2) were used and fitted with crystal structures of the single components (Appendix, Table 5-1).

1.3. Topological challenges during DNA replication

The chromosome is composed of negatively supercoiled DNA organized in topological domains, which facilitates its compaction to fit within the cell. However, this compaction can pose significant challenges for processes that occur on DNA, particularly during replication. The topological challenges presented by constrained DNA are universal problems faced by replication machinery across all domains of life, given that these machines share a common architecture and similar mechanisms for DNA replication (Gao et al., 2019; Pommier et al., 2022; Pommier et al., 2016; Postow et al., 2001). Here, we will use the bacterial *E. coli* replisome as a representative example to explore these challenges.

The *E. coli* chromosome is generally negatively supercoiled, which facilitates DNA denaturation and the initiation of replication (Fulcrand et al., 2016; Nelson & Cox, 2021). However, as the replication complex progresses along the DNA template, the helicase unwinds the helical structure, resulting in a linking number change ΔLk of +50 to +100 ahead, depending on the replication rate considered for *E. coli* (Breier et al., 2005; Nelson & Cox, 2021; Reyes-Lamothe et al., 2008) (Figure 1-9A). This unwinding introduces an excess of positive supercoils ahead of the replication fork, which can cause fork stalling (Postow et al., 2001). If unresolved, this tension can lead to replication fork reversal and even fork collapse (Baxter, 2015; Bermejo et al., 2007). While the replisome generates positive supercoils ahead of the fork, it also forms precatenanes or interwound duplexes in the daughter strands behind the fork (Figure 1-9A). Persistent precatenanes can prevent chromosome segregation (McKie et al., 2021). Overcoming these topological challenges is critical for cell survival. In the following sections, we will discuss the mechanisms that cells use to resolve the strain caused by accumulated twists.

1.3.1. The role of topoisomerases in managing replication-induced topological stress

Topoisomerases are responsible for maintaining the topological state of the chromosome, ensuring its proper shape and function (section 1.1.6). They play an essential role in DNA replication by resolving the tension that builds up in the DNA. In *E. coli*, two key topoisomerases, DNA gyrase and Topoisomerase IV (Topo IV), are responsible for modulating topology ahead of and behind the replication fork (Figure

1-9A). DNA gyrase can relax positive supercoils and uniquely introduce negative supercoils (Koster et al., 2010; Postow et al., 2004; Vos et al., 2011). Topo IV primarily functions to remove precatenanes after replication, aiding in chromosome segregation (Helgesen et al., 2021; Wang et al., 2008; Zawadzki et al., 2015).

DNA gyrase and Topo IV, work together to overcome topological challenges during replication (Figure 1-9A). There are approximately 300 molecules of DNA gyrase per cell that are stably bound to DNA, whereas the number of Topo IV is 10-fold lower, with approximately 30 molecules bound to DNA per cell (Stracy et al., 2019). However, as discussed earlier, the chromosome is a complex structure composed of topological domains, with the topological state of DNA being constrained within these domains (Wang et al., 2013). Interestingly, Topo IV appears to be more closely associated with SeqA, a protein that binds to newly synthesized DNA, than with the replisome itself, thus resolving the precatenanes with a temporal delay (Helgesen et al., 2021). This observation suggests that DNA gyrase plays a more critical role than Topo IV in resolving topological strain during replication.

In the study by Stracy et al. (2019), an average of 12 DNA gyrase molecules were observed in close proximity to the replication fork, while the remaining DNA gyrase molecules modulate DNA topology outside of the topological domain. DNA gyrase enzyme performs approximately one cycle per second when resolving positive supercoiling, with the rate depending on the force applied to the DNA (Agarwal & Duderstadt, 2020; Nollmann, Stone, et al., 2007). Each cycle results in a linking number change (ΔLk) of -2. When multiplying this by the number of DNA gyrase molecules near the replication fork, the total linking number change achieved is -24 twists per second. However, the replisome introduces a positive linking difference (ΔLk) of +50 to +100 per second. This suggests that the number of DNA gyrase molecules near the replication fork is insufficient to resolve all the overwinding caused by the replisome, even when considering the activity of Topo IV (Stracy et al., 2019). This observation, also highlighted by Stracy et al. (2019), points to the possibility of an alternative pathway for relieving topological strain, potentially involving the replisome itself. This concept will be explored further in section 4.2.1, where we present a simple simulation to model the interaction between the replisome and DNA gyrase.

1.3.2. Fork Rotation as a mechanism to release topological strain

Another mechanism to remove the torsional strain ahead of the replication fork is fork rotation (Keszthelyi et al., 2016; Peter et al., 1998; Schalbetter et al., 2015; Wang et al., 2008). This process was first proposed by Champoux and Been (1980). When torsional strain builds up ahead of the replication fork, the replisome itself rotates, transferring the additional twist behind it, forming additional precatenanes (Figure 1-9A). You can also envision the replisome following the helical path of the DNA as it unwinds, effectively overcoming strain. This process still requires the action of topoisomerases, as unresolved precatenanes would interfere with chromosome segregation (Helgesen et al., 2021; Wang et al., 2008; Zawadzki et al., 2015). However, it solves the immediate problem of unfavorable, excess topological strain.

Fork rotation has been indirectly observed through the formation of precatenanes or catenated dimers in various replication gel-based assays (Hiasa & Marians, 1994; Schalbetter et al., 2015; Snapka et al., 1988; Sundin & Varshavsky, 1980, 1981). In these studies, the replication machinery was challenged by reduced or abolished topoisomerase function during the replication process. For instance, in studies involving Simian virus 40, greatly reduced topoisomerase II activity led to the formation of various forms of catenated dimer molecules (Sundin & Varshavsky, 1980, 1981). It was hypothesized that especially during the termination phase of replication, the formation of catenated dimers occurred due to the limited space available for topoisomerase action (Sundin & Varshavsky, 1980). Additionally, fork rotation has been observed in vitro during the elongation phase, where it supports replication by forming precatenanes behind the fork (Hiasa & Marians, 1994; Peter et al., 1998).

The study by Schalbetter et al. (2015) suggests that typically fork rotation is restricted to prevent the accumulation of DNA damage, but it is triggered under specific conditions, such as when topoisomerase activity is reduced, in hard-to-replicate regions, or during termination. This study was conducted in budding yeast and notably identified the fork protection complex Tof1/Csm3 as a regulator of fork rotation. Tof1/Csm3 are yeast homologs of the human fork protection complex Timeless/Tipin complex, suggesting similar implications in human replication.

In summary, fork rotation can support the replisome when topoisomerase activity is compromised, which can occur due to obstacles on the DNA that limit space for topoisomerases (Figure 1-9B). Such scenarios include reaching the end of a topological domain, encountering head-on conflicts between the replisome and RNA

polymerase, or during termination when two replication forks converge. Although fork rotation of the replisome has not yet been directly observed as physical rotation, it has been inferred from bulk assays. The next section will discuss methods for observing rotation in single-molecule studies and how topological stress influences transcription and replication at the single-molecule level.

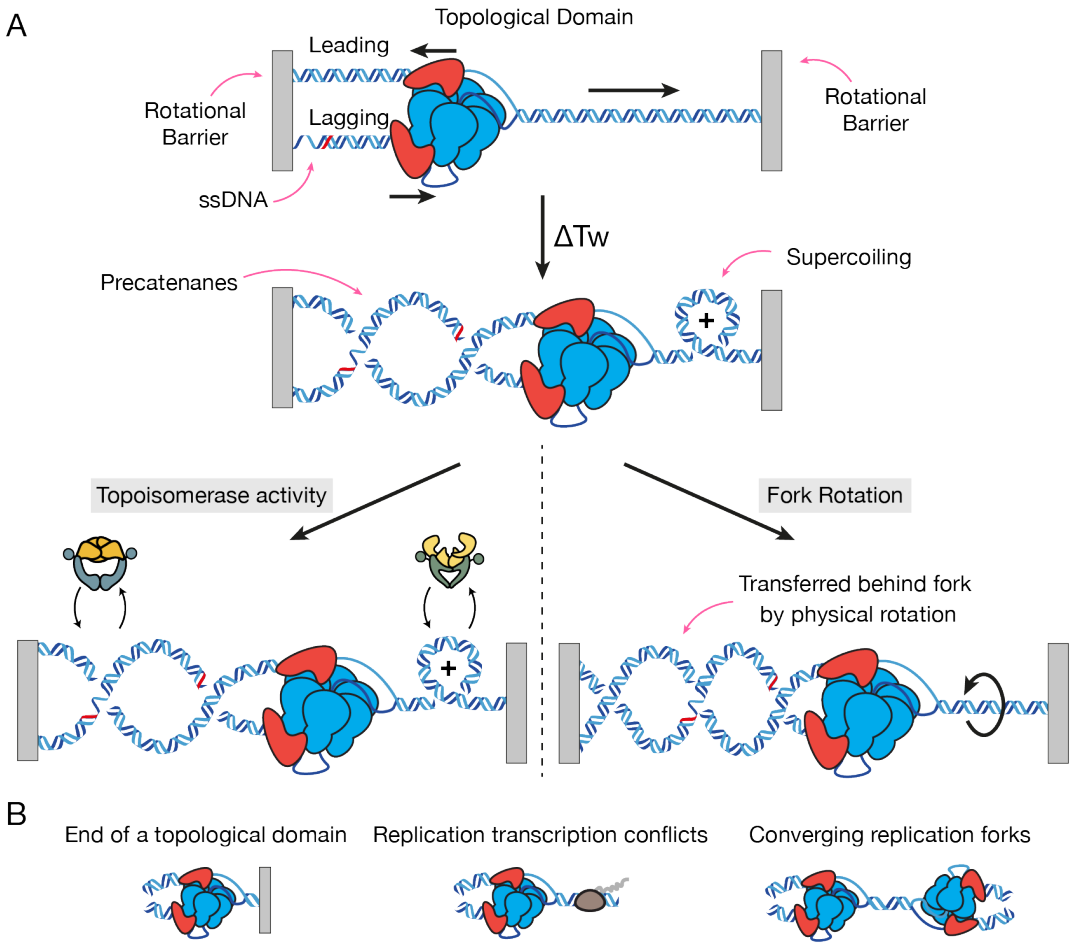


Figure 1-9 Topological challenges during replication.
(**A**) DNA replication occurs within a topological domain with constrained topology. The rotational barriers prevent the replisome from rotating to release the strain. When the helicase separates the DNA (ΔTw), this results in positive supercoiling ahead of the replisome as well as precatenanes behind. There are two possible pathways presented to resolve the strain. On the left, if the spacing allows, topoisomerases take care of topological strain ahead and behind the fork. On the right, due to space constraints, the fork itself rotates and transfers the strain behind the fork, which is later resolved. This increases the number of precatenanes. (**B**) Possible scenarios where fork rotation could release torsional strain. Fork rotation may release torsional strain when the replisome encounters obstacles like structural proteins, heads-on transcription replication conflicts, or during termination. In these scenarios space for topoisomerases on the DNA is limited.

1.4. Single-molecule techniques revealing hidden dynamics of molecular machines

The replisome, along with other molecular machines, is key to understanding the inner workings of life. However, due to their microscopic size, observing these molecular machines in action requires innovative techniques. Traditional methods, which observe millions of molecules simultaneously, often obscure the diversity of individual molecular states, providing only an averaged view of their behavior. In contrast, single-molecule techniques allow us to observe these processes at the level of individual molecules, revealing the full spectrum of their dynamic states (Leake, 2021; Zlatanova & van Holde, 2006). This is particularly important for biological processes such as DNA replication, where microstates may vary depending on environmental factors, contributing to the overall robustness of the process (Scherr et al., 2018). Without these techniques, alternative pathways and transient states would remain hidden within the ensemble average. Throughout the introduction, we have touched on these techniques and demonstrated their power, and now we will explore them in greater detail, with a specific focus on their application to transcription, replication, and the study of molecular rotation. The concept of observing rotation will be particularly important in Chapter 5, while the extensive research on RNA polymerase encountering topological stress during transcription offers a rich context for the discussion later on.

1.4.1. Observing rotation in single-molecule studies

First, we will examine how rotation has been studied at the single-molecule level by discussing four key methods that have successfully resolved rotation. We begin with the seminal work of Noji et al. (1997), where the rotation of the F_1-ATPase was directly observed (Figure 1-10A). The F_1-ATPase consists of a central rotor, the γ-subunit, with a radius of approximately 1 nm, and a stator barrel, the $\alpha_3\beta_3$-subunit, with a radius of around 5 nm. To visualize this rotation, the F_1-ATPase was immobilized on a slide surface, allowing the γ-subunit to rotate freely. A fluorescently labeled, biotinylated actin filament was attached to the γ-subunit through biotin-streptavidin interaction. Upon the introduction of ATP into the system, the actin filament rotated as a result of the γ-subunit's rotation. As discussed earlier, actin has a high persistence length, making it appear as a stiff rod. The observed rotation was unidirectional and ATP-dependent, occuring at a rate of up to four rotations per second. The torque generated was estimated to be as high as 45 pN nm, based on the length of the filament.

Subsequent studies, such as those by Adachi et al. (2007), improved imaging techniques, enabling the observation of substep rotations.

In another example, Sowa et al. (2005) observed steps in the rotational movement of the bacterial flagellar motor, a rotary machine that powers the swimming motion of bacteria by rotating its helical filaments like a propeller. The rotor, a ring with a diameter of approximately 45 nm, is embedded in the cytoplasmic membrane and is powered by an electrochemical ion gradient. To visualize this rotation, fluorescent beads were attached to the flagellar motor and imaged using back-focal-plane interferometry in combination with high-speed recording (Figure 1-10B). This method allowed the detection of 26 distinct steps per revolution.

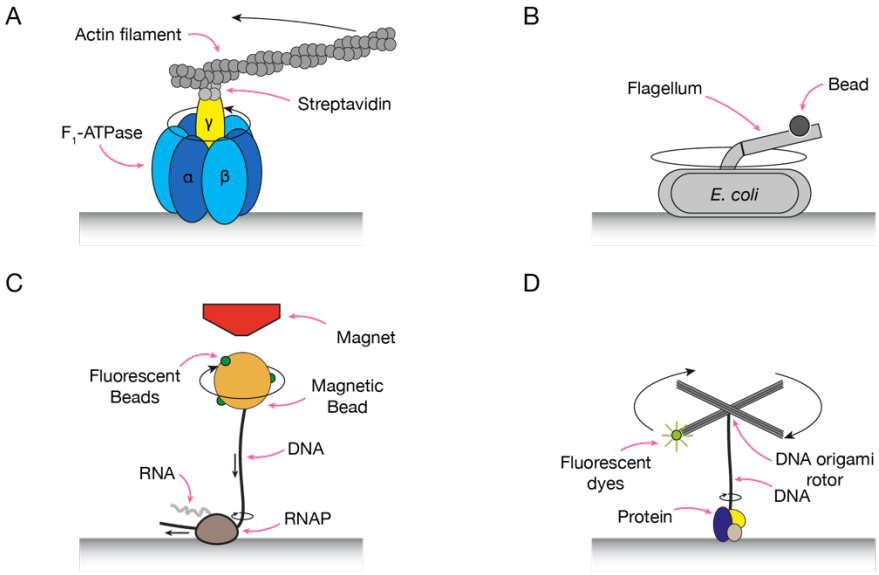


Figure 1-10 Different single-molecule assays to observe rotation by biomolecules.
(**A**) F_1-ATPase rotation observed by attaching an actin filament to the γ subunit. (**B**) Bacterial flagellar motor observed by attaching a (fluorescent) bead. (**C**) RNA polymerase fixed to a slide resulting in rotation in the DNA template observed by attaching a bead decorated with smaller fluorescent beads. (**D**) Observing rotation by protein (here RecBCD) by attaching a DNA template to a DNA origami rotor. (A), (B), (C), and (D) were adapted from Noji et al. (1997), Sowa et al. (2005), Harada et al. (2001), and Kosuri et al. (2019), respectively.

The F1-ATPase and the flagellar motor are prime examples of rotary machines functioning on a cellular level. However, as discussed earlier, other proteins and protein complexes also exhibit rotational behavior, even if they are not classical rotary motor proteins. For instance, Harada et al. (2001) demonstrated that when the RNA

polymerase (RNAP) from *E. coli* was immobilized on a microscope slide, the DNA template rotated as RNAP synthesized RNA (Figure 1-10C). This rotation was observed by attaching a bead decorated with fluorescent dyes to the free end of the DNA template, with the bead held in place by a conical magnet. The beads rotated clockwise, corresponding to the right-handed helical structure of B-DNA. The rotation rate was dependent on the NTP concentration, and the torque generated by RNAP was estimated to be greater than 5 pN nm. However, since RNAP was immobilized on the surface, it remains unclear whether DNA or RNAP itself rotates within cells. The movement of RNAP on DNA will be further explored in the next section, which focuses on topological challenges during transcription.

In the final example, Kosuri et al. (2019) developed an innovative approach called Origami-Rotor-Based Imaging and Tracking (ORBIT) to visualize rotation caused by the RecBCD helicase, which is involved in DNA repair, as well as RNAP. The method utilizes a DNA origami structure shaped like a cross, with one arm labeled with fluorescent dyes to track its movement. Each arm of the DNA origami rotor is 80 nm long, providing an amplified signal of rotational movement while maintaining high torsional stiffness and low hydrodynamic drag (Figure 1-10D). In the study, RecBCD was immobilized on a slide surface, followed by introduction of a DNA template along with the origami rotor attached. This method offered high spatial resolution, allowing researchers to resolve the initiation, pausing, and backtracking of the RecBCD enzyme. When applied to RNAP, the technique achieved a resolution of single base-pair steps, with each step corresponding to a 35° rotation. Compared to previous studies where a bead was attached to DNA, ORBIT overcame the obscuring effects of Brownian fluctuations, thus enhancing the readout quality. All the techniques discussed share a common feature, which is the requirement of attaching an additional object, such as another protein (like actin), a fluorescent bead, or a DNA origami structure, to resolve the rotational movement.

1.4.2. Single-molecule studies of topological barriers during transcription

Previously, we have previously discussed the topological challenges encountered during DNA replication. Before we dive into methods for observing how topological strain affects the transcription process, it is essential to clarify the specific challenges involved. Similar to DNA replication, RNAP alters the topological state of DNA during transcription. To synthesize RNA, RNAP must unwind the DNA, creating positive

supercoils ahead of itself while compensating by introducing negative supercoils behind (Figure 1-11). This phenomenon, known as the twin-supercoiled-domain model, was first proposed by Liu and Wang (1987). The induction of supercoils has been indirectly confirmed using topoisomerases in bulk plasmid studies (Leng et al., 2004; Liu & Wang, 1987) and has also been examined at the single-molecule level (Chong et al., 2014; Janissen et al., 2024; Ma et al., 2013). Recently, direct experimental validation of the twin-supercoiled-domain model was achieved through a single-molecule fluorescence study (Janissen et al., 2024). This study provided direct visualization of supercoiling as it occurred both ahead of and behind the RNAP. In this section, we will explore the single-molecule approaches used by Chong et al. (2014) and Janissen et al. (2024) in greater detail to understand how topological barriers can be studied at the single-molecule level.

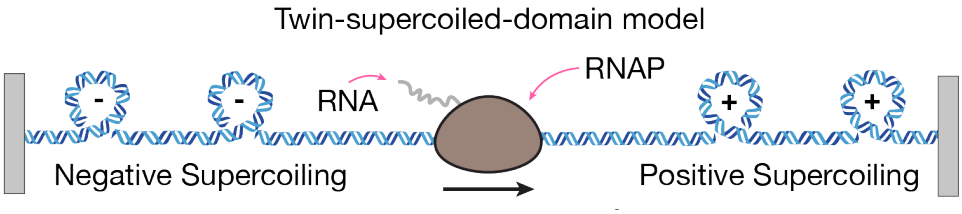


Figure 1-11 Twin-supercoiled-domain model.
RNA polymerase (RNAP) simultaneously introduces positive supercoiling ahead and behind itself, respectively, when transcribing RNA.

Let's begin with the study by Janissen et al. (2024). Before delving into the technical aspects, we'll first summarize the key findings. The researchers developed a single-molecule assay to directly visualize the twin-supercoiled domains generated by RNAP during transcription. They demonstrated that supercoils were introduced equally both ahead and behind the RNAP. By cutting the RNA transcript with RNAase, they confirmed that the viscous drag of the transcript was not necessary for supercoil formation, challenging the existing explanation for the twin-supercoiled domain model, which included transcript drag as a contributing factor. To conduct this study, they engineered a torsionally constrained 31 kb linear DNA construct and equipped it with a T7A1 promoter for transcription initiation. The DNA ends were ligated with multiple biotin handles, which interacted with the microscopy slide surface to create multiple attachment sites, thereby mimicking a topological domain within a chromosome. This study used the *E. coli* RNAP. Notably, they did not control the DNA's topological state. Supercoiling was solely introduced by the RNAP. They tracked both the position of

RNAP and the size of the formed plectonemes, and they were able to observe and quantify the supercoils generated ahead and behind the RNAP. The use of an intercalating dye, SYTOX Orange, allowed them to detect the handedness of the supercoiling, as the dye signal varies with handedness and the degree of supercoiling (Kolbeck et al., 2024). Additionally, by introducing a nick in the DNA via photo-induced damage, they observed that the relaxation time of the supercoils increased when RNAP was bound, indicating higher viscous drag. To further investigate the role of the RNA transcript, they introduced RNase to cut the transcription product, which confirmed that RNAP alone could generate supercoils. Janissen et al. (2024) elegantly validated the twin-supercoiled domain model and demonstrated how a single-molecule approach can yield information rich data by adapting the basic assay for various types of measurements.

Chong et al. (2014) conducted an in-depth study exploring how the stochastic buildup of supercoiling could lead to transcriptional bursting in bacteria. The researchers observed that transcription of highly expressed genes often occurs in stochastic bursts, and their study links this phenomenon to supercoiling dynamics. They developed an assay to monitor transcription of individual molecules in real-time, revealing that the buildup of positive supercoiling during elongation can slow down RNAP and even prevent initiation. The addition of DNA gyrase helps to resume transcription by relieving this supercoiling-induced stress. Using single-cell mRNA counting via fluorescence in situ hybridization (FISH), they demonstrated that transcriptional bursting depends on DNA gyrase concentration. They propose that the stochastic binding and rebinding of DNA gyrase at highly expressed genes underlies the observed transcriptional bursts. For this single-molecule study, they created linear DNA substrates with one or multiple biotin molecules at the ends, enabling the DNA to attach to a slide surface through single or multiple streptavidin interactions. They examined the behavior of both T7 and E. coli RNAP, initiating transcription at their corresponding promoters. They monitored transcription by staining the RNA product with an RNA dye, where the dye signal increases as the mRNA lengthens. During transcription on a 12 kb linear substrate, they observed a reduction in transcription rate when multiple biotins anchored the DNA to the surface, indicating that the RNAP was slowed due to the buildup of positive supercoils ahead of it. Adding DNA gyrase restored the elongation rate to that of transcription on singly-tethered DNA. To simulate a topological domain, they created a more complex assay using circular plasmids attached to the slide surface via multiple biotin linkages, effectively creating

a diffusion barrier for torsional strain. They initiated transcription from a promoter and included a termination site on the plasmid. Even without topoisomerase, initiation was not disturbed. When both Topo I and DNA gyrase were present, they found no impact on initiation rates. However, when only Topo I was added, leading to the buildup of positive supercoiling, initiation rates declined over time. This reduction could be reversed by adding DNA gyrase, which removed the positive supercoils. This study demonstrated at the single-molecule level that varying DNA gyrase concentrations around RNAP could lead to different initiation rates, potentially explaining the stochastic nature of transcription in highly expressed genes.

In summary, we discussed two examples of how single-molecule approaches can be used to study the topological challenges during transcription. Creating these challenges involved multiple attachments between streptavidin and biotin on a slide surface, which allowed RNAP to generate supercoils without rotating, likely due to a balance of positive and negative twists. The final section will explore how similar topological challenges have been studied in DNA replication at the single-molecule level.

1.4.3. Single-molecule studies of topological barriers during replication

The topological challenges during replication are more intricate than those encountered during transcription due to the production of two daughter strands and the involvement of the multi-protein replisome complex (section 1.3) (Figure 1-12A). Replication involves duplicating the entire genome, a massive structure, which adds to the difficulty of studying it at the single-molecule level. To date, only one study has investigated the influence of topological strain on the replisome using single-molecule techniques. In this final section, we will review that study and introduce the specific aims of this thesis.

The study by Kurth et al. (2013), titled "A solution to release twisted DNA during chromosome replication by coupled DNA polymerases", explored a potential mechanism for releasing topological strain during replication. We will first summarize the key findings and then explain the experimental procedures. As discussed in earlier sections, the replisome alters the topological state of the chromosome due to the connections among its various components. Kurth et al. (2013) studied the *E. coli* replisome using both single-molecule and ensemble assays and proposed that the replisome can resolve topological issues independently of topoisomerases through a

mechanism named "signal release." They observed that the lagging strand polymerase frequently detaches (uncouples) from the template strand, leaving incomplete Okazaki fragments and resulting in single-stranded regions. Notably, this mechanism operates independently of primase activity, indicating that it is not triggered by protein interactions. In addition, when the lagging strand is uncoupled using a separately primed plasmid, the lagging strand polymerase exhibits high processivity. In summary, the study suggests that polymerase coupling reduces processivity and leads to the release of the lagging strand polymerase from the DNA.

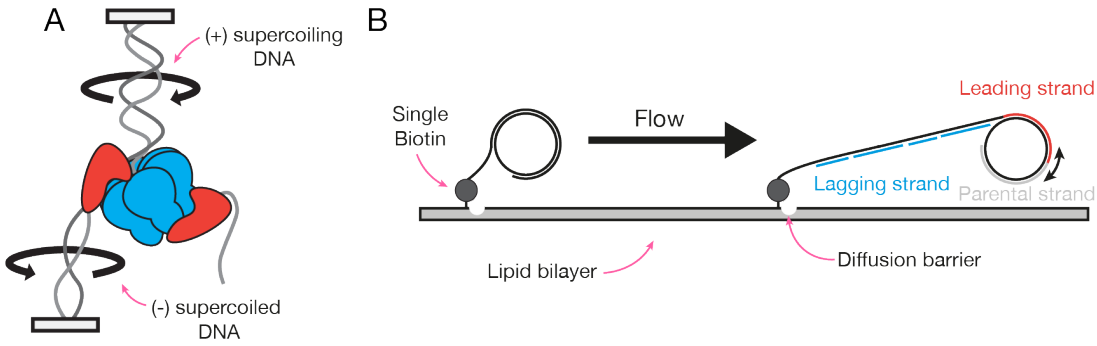


Figure 1-12 Rolling circle replication assay.
(**A**) Topological challenge during replication. Arrows indicate rotational direction of at the parental and leading strands. Lagging strand is free to rotate because of ssDNA. (**B**) Rolling circle approach to study DNA replication used in Kurth et al. (2013). Lagging strand is attached using single biotin attachment. Rolling circle substrate can be replicated indefinitely and is stretched out by flow. Rotation shown in (A) of parental and leading strand are canceled out due to being connected (indicated by black arrow).

The methodology of the single-molecule assay in this study utilized a TIRF (Total Internal Reflection Fluorescence) microscopy setup. The DNA template for replication was a rolling circle substrate, which was attached to the slide surface via a single biotin link on the lagging strand (Figure 1-12B). The rolling circle substrate is particularly suited for this type of study as it can undergo multiple replication cycles, presenting a preformed replication fork for the replisome to load. The biotin attachment was anchored to a lipid bilayer and moved along the surface until it encountered a diffusion barrier. When extended, these molecules resemble a curtain, hence the name "DNA curtain" for this assay. The DNA was stained with the fluorescent dye Yo-Pro1, which selectively binds to double-stranded DNA (dsDNA), allowing visualization of gaps in the lagging strand when extended by flow. As the flow increased, more single-stranded

DNA (ssDNA) gaps became visible, revealing the gaps in the lagging strand. Higher flow rates exerted greater force on the replisome.

In contrast to the two studies discussed in section 1.4.2, this assay utilized a single biotin attachment, with the rolling circle substrate connected only via the lagging strand. The lagging strand had areas with ssDNA, which could potentially relieve topological strain. The rolling circle template's configuration, where the leading and parental strands are connected, offers a mechanism to cancel out the twist introduced into these strands due to their opposite signs. In comparison, Chong et al. (2014) required multiple biotin attachments to introduce a topological challenge in their substrate. Creating a topologically constrained rolling circle substrate is more complex and would involve many additional steps, and thus it would not be feasible for use in the same single-molecule study (Kuhn et al., 2002). However, the mechanism proposed by this study, involving the detachment of the lagging strand polymerase, could indeed be a valid method for alleviating topological stress. This prompts the question of how a single-molecule study can effectively explore the topological challenges faced by the replisome.

1.5. Aims of this thesis

As discussed in this chapter, DNA is a complex polymer capable of bending and twisting into intricate structures, which is crucial for storing genetic information in chromosomes. However, chromosomes are dynamic and densely populated with various proteins that support their structure, along with molecular machines that use DNA as a template for information transfer. When stripped of structural proteins, these molecular machines must contend with the helical structure of DNA, which can twist and impose physical strain. While many studies have focused on transcription and its response to topological stress (section 1.4.2), the replisome, one of the cell's most important machines, has not been as extensively studied, particularly in single-molecule experiments that reveal dynamic interactions (section 1.4.3). Recent research has shown that topoisomerases, which resolve topological strain, may be insufficient in keeping up with the replisome's speed, raising the possibility of alternative strain relief pathways (section 1.3.1). Additionally, in some situations, limited space on the chromosome forces the replisome to manage topological challenges independently. The concept of fork rotation, where strain is transferred from ahead of the replisome to behind it, offers a temporary solution, though it has only been observed indirectly.

The primary goal of this thesis is to create and observe topological challenges encountered by the replisome in order to investigate changes in replication dynamics. Topoisomerases were excluded from the experiment to observe how the replisome independently manages topological strain and to reveal its intrinsic coping mechanisms.

Study the replisome encountering a topological challenge is a challenging goal that can be subdivided into three distinct aims. The first aim was to the development of an open-source analysis platform for single-molecule experiments (Mars). Using single-molecule approaches presents unique challenges due to the complexity of datasets. Tracking and analyzing data from numerous single molecules can be difficult without standardized methods. Furthermore, new experimental procedures often require flexible workflows to accommodate additional analysis steps. To address these challenges, we developed Mars, an open-source platform compatible with Fiji and ImageJ2, as detailed in chapter 3. Mars was designed to facilitate the tracking of single-molecule data from raw images to final plots, providing a flexible and reproducible solution for analyzing image-derived data and promoting the discovery of new biological phenomena.

With appropriate analysis tools in hand, the second aim was to develop an imaging platform to study topological barriers at the single-molecule level. Chapter 4 presents the assay concept, where we reconstituted replication on DNA while tracking positional changes over time. Initial estimates suggested that topological challenges could affect replication dynamics, as reflected in changes in rate and processivity. We developed a modular DNA toolkit to create substrates that allow the replisome to resolve these spatial challenges. By mixing different modules, we were able to mimic various topological challenges and observe the replisome's response. We verified that the constructed DNA was topologically constrained and imaged it under calibrated conditions to avoid measurement artifacts. Finally, we validated the replication system and explored various readout methods.

After confirming the viability of our imaging concept, the third aim was to investigate the replisome's real-time response to a topological challenge. In chapter 5, we will explore the biological implications of simulating a topological challenge ahead of the replisome, similar to scenarios that trigger fork rotation. This investigation sought to answer how the replisome behaves when topoisomerase activity is constrained due to

spatial limitations. Although we did not observe a significant impact on replication dynamics, visual differences in replication reactions and limitations in resolution led to the development of a new imaging approach, the transverse flow method. This advancement supported our aim by enabling us to observe replisome rotation without additional components. These experiments investigated replisome dynamics when challenged by a topological barrier, with a primary focus on developing a single-molecule assay.

Chapter 2 - Materials and methods

2.1. Materials

2.1.1. Chemicals and Recombinant Proteins

Chemical / Recombinant Protein	Source	Identifier
1 Kb Plus DNA Ladder	Thermo Fisher Scientific	Cat# 11578636
3-Aminopropyltriethoxysilane (APTES)	Carl Roth	Cat# 2328.1
Capto HiRes Q 5/50	Cytiva	Cat# 29275878
LD655-T7 gp5	This study	N/A
Amicon Ultra-15 Centrifugal Filter units	Merck	Cat# UFC901024
Biotin-PEG-Succinimidyl Carbonate (MW 5000)	Laysan Bio	Cat# Biotin-PEG-SC-5000-1g
dNTP Bundle	Jena Bioscience	Cat# NU-1005S
HisTrap HP 5 ml	Cytiva	Cat# 17524802
HiPrep 26/10 Desalting	Cytiva	Cat# 17508701
HiTrap Heparin HP 5 ml	Cytiva	Cat# 17040703
HiTrap Q HP 5 ml	Cytiva	Cat# DOWC634165S
Lambda DNA	New England Biolabs	Cat# N3011S
LD655-CoA	Lumidyne Technologies	Custom synthesis
mPEG-Succinimidyl Carbonate (MW 5000)	Laysan Bio	Cat# MPEG-SC-5000-5g
NTP Bundle	Jena Bioscience	Cat# NU-1014L
Phusion High-Fidelity DNA polymerase	New England Biolabs	Cat# M0530S
PCA	Sigma-Aldrich	Cat# 03930590
PCD	Produced in our lab	
QIAquick Gel Extraction Kit	Qiagen	Cat# 28706
QIAquick PCR Purification Kit	Qiagen	Cat# 28106
Qiagen Plasmid Maxi Kit	Qiagen	Cat# 12165
Restriction endonucleases	New England Biolabs	N/A
Sephacryl S-1000 SF Tricorn 10/300 GL	GE Healthcare	N/A
Streptavidin	Sigma-Aldrich	Cat# S4762
Superdex 200 Increase 10/300 GL	Cytiva	Cat# 28990944
SFP synthase	Produced in our lab	N/A
SYLGARD 184 Silicone Elastomer Kit	VWR	Cat# DOWC634165S
SYTOX Orange Nucleic Acid Stain	Thermo Fisher Scientific	Cat# S11368
T4 DNA Ligase	New England Biolabs	Cat# M0202S
Taq Polymerase	New England Biolabs	Cat# M0267S
Trolox	Sigma-Aldrich	Cat# 648471

Table 2-1 Chemicals and recombinant proteins.

The chemicals, that are not listed, were purchased from Sigma-Aldrich or Carl Roth. PCD was prepared according to Senavirathne et al. (2018). SFP synthase was prepared according to Yin et al. (2006). Plasmids and initial proteins of gp4, gp5, and gp2.5 were kindly provided by Prof. Dr. Samir Hamdan and Dr. Masateru Takahashi from the KAUST institute, Saudi Arabia.

2.1.2. Essential micromirror TIRF microscope parts

Microscope part	Source	Lot Number
808 nm 250 mW single mode laser module	Lumics	Cat# LU0808M250
ET520/40m emission filter	Chroma	Cat# ET520/40m
iXon Life 888 EMCCD camera	Andor	Cat# iXon Life 888
Laser Box: OBIS	Coherent	Cat# 1228877
Micromirror TIRF system	Mad City Labs	Cat# MicroMirror TIRF System
OBIS 488nm LS 120mW laser	Coherent	Cat# 1277614
OBIS 532nm LS 120mW laser	Coherent	Cat# 1280765
OBIS 637nm LX 100mW laser	Coherent	Cat# 1193838
OptoSpin25 filter wheel	Cairn Research	Cat# Optospin25
OptoSplit II dual emission image splitter	Cairn Research	Cat# OptoSplit II
Prime BSI (sCMOS)	Photometrics	Cat# Prime BSI
T635lpxr dichroic mirror	Chroma	Cat# T635lpxr
TIRF Lock	Mad City Labs	Cat# TIRF Lock
ZET532/640m emission filter	Chroma	Cat# ZET532/640m
ZT488rdc dichroic mirror	Chroma	Cat# ZT488rdc
ZT532rdc dichroic mirror	Chroma	Cat# ZT532rdc
ZT647rdc dichroic mirror	Chroma	Cat# ZT647rdc
ZET488/10x excitation filter	Chroma	Cat# ZET488/10x
ZET532/10x excitation filter	Chroma	Cat# ZET532/10x
ZET640/10x excitation filter	Chroma	Cat# ZET640/10x
Apo N TIRF 60 x oil-immersion TIRF objective (NA 1.49)	Evident	Cat# UPLAPO60XOHR

Table 2-2 Essential micromirror TIRF microscope parts.

2.1.3. Oligonucleotides

All oligonucleotides were synthesized by Eurofins Genomics located in Ebersberg, Germany.

Name	Sequence
ORT-1	[BIO] AGGTCGCCGCCC
ORT-2	[PHO] TCGAGGGCGGCGACCT
ORT-3	[PHO] CTAGAGACAGCAAGTTGGACAATCCATCTCGTTCTATCACTAATGCAG GGAGGATTTCAGATATGGCAACTAGTATGCCG
ORT-4	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATGGATTGTCCAACTTGCTGTCT
ORT-5	[BIOTEG] TTTTTTTTTCGGCATACTAGTTGCCATATCTGAAATCCTCCCTGC
ORT-8	[PHO] GGCCCGGCATACTAGTTGCCATATCTGAAATCCTCCCTGC
ORT-9	TATTAGCGGCCGCGATTGTTCTTTATTCATTTT
ORT-10	CCCATCGGAAAACTCCTGCTTTAGC
ORT-11	TTACCGCATACCAATAACGCTTCAC
ORT-12	GAAACTCAACATCGTCATCAAACGC
ORT-13	[PHO] CTAGAGACAGCAAGTTGGACAATCCATCTCGTTCTATCACTAATGCAG GGAGGATTTCAGATATGGCAACTAGTATGCCG

Table 2-3 Sequences of oligonucleotides

2.1.4. Additional equipment

Equipment	Source	Identifier
Avanti J-26S XP Centrifuge	Beckman Coulter	Cat# B22987
NanoPhotometer NP80	IMPLEN	Cat# NP80
NGC	BioRad	N/A
Polyethylene tubing (0.58 x 0.96 mm)	A. Hartenstein	Cat# SX05
Precision coverslips #1.5H, 22x22 mm	Marienfeld	Cat# 0107052
Syringe Pump	World Precision Instruments	Cat# AL1000-220
Whatman Unicore Punch Kit - 1mm	Merck	Cat# WB100073
Zepto plasma cleaner	Diener Electronic	Cat# Model 2 base unit type A

Table 2-4 Additional equipment

2.2. Python-based simulation of DNA replication and Supercoiling

To implement the simulation of replication behavior described in section 4.2.1, the following steps of the code were outlined (Figure 2-1). The assumptions included a topological domain size of 10 kb, a gyrase activity rate of 1 cycle per second with a DNA footprint of 140 bp, and a replication rate of 500 bp per second. The initial superhelical density ('σ_{start}') was set to -0.07, while the stalling superhelical density ('σ_{max}') was set to 0.1. The simulation was carried out using a nested loop with conditional branching in Python (Figure 2-1).

The loop continued while there was still DNA left to replicate ('Basepairs' > 0). The available space for gyrase molecules ('Space') was calculated by dividing the remaining DNA ('Basepairs') by the space required for gyrase interaction ('GyrSp'). The variable 'Space' represented the number of binding sites available for gyrase. The number of gyrase molecules ('NumGyr') was randomly generated using a gamma function and compared to the available space. If there were more binding sites than gyrase molecules, 'NumGyr' remained the same. Otherwise, the loop checked if 'Space' was greater than 0, and 'NumGyr' was updated to be equal to 'Space'.

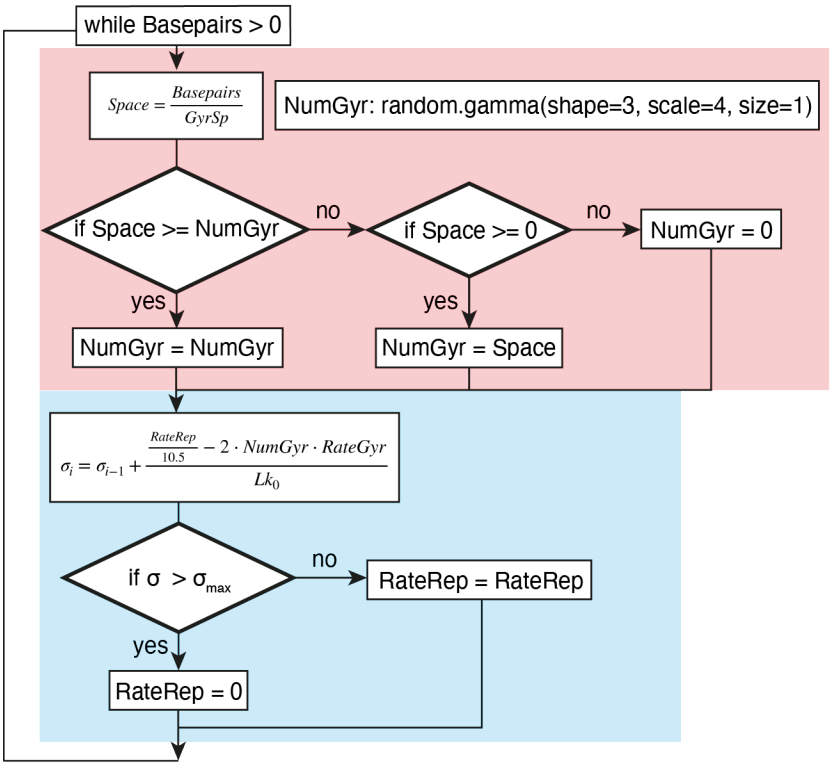


Figure 2-1 Schematic workflow of the DNA replication simulation.
The flowchart illustrates the steps and conditional logic used in the Python-based simulation of DNA replication dynamics in a topologically constrained environment. The assumptions for the simulation include a topological domain size of 10 kb, a gyrase activity rate of 1 cycle per second with a DNA footprint of 140 bp, and a replication rate of 500 bp per second. The simulation starts by initializing the superhelical density (σ_{start} = -0.07) and a stalling superhelical density threshold (σ_{max} = 0.1). A nested loop with conditional branching calculates the available space for gyrase molecules ('Space') and generates a random number of gyrase molecules ('NumGyr') using a gamma function. Depending on the available space, the simulation updates the superhelical density ('σ_i') by accounting for the activities of both the replisome and gyrase. When the current superhelical density reaches the stalling threshold ('σ_{max}'), the replication rate ('RateRep') is set to zero, indicating a stall in replication. The loop continues until no DNA remains to be replicated.

Next, the superhelical density ('σ_i') was updated by adding the value corresponding to the activity of the replisome and subtracting the effect of the gyrase activity. In the final 'if' statement, the current superhelical density ('σ_i') was compared to the stalling superhelical density ('σ_{max}'). If the stalling threshold was reached, the replication rate ('RateRep') was reduced to 0. Otherwise, 'RateRep' remained unchanged. The loop continued until no DNA was left to replicate.

2.3. Protein purification and labeling

2.3.1. Expression and purification of T7 gp2.5 protein

The expression and purification of the T7 bacteriophage gp 2.5 protein were based on a previously described protocol (Rezende et al., 2002). Unlabeled gp 2.5 was expressed using a pET-17b vector in *E. coli* BL21 star cells. The cells were cultured in 1 liter of Luria-Bertani (LB) medium (Ampicillin resistance) at 37°C until reaching an optical density (OD) of 1.0. Before induction, a 10 ml aliquot of the culture was collected for a mini-prep to check for mutations. Protein expression was induced with 1 mM IPTG, and the culture was incubated at 37°C for an additional 4 hours. Cells were harvested by centrifugation at 4,000 rpm for 15 minutes, washed with 35 ml of PBS, and pelleted again at 4,000 rpm for 15 minutes. The supernatant was discarded, and the cell pellets were stored at -80°C until sequencing confirmed no mutations. The pellets were then resuspended in 30 ml of lysis buffer (Buffer A, low salt: 50 mM Tris-Cl, pH 7.5, 0.1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, 50 mM NaCl). The lysis mixture was supplemented with 1x protease inhibitor cocktail (100x stock: 200 µM Pepstatin, 200 µM Leupeptin, 100 mM PMSF, 106 mM Benzamidine, 100 µg/ml Aprotinin), 1x lysozyme (100x stock: 10 mg/ml lysozyme), and 1x DNase I (100x stock: 10 mg/ml). The cells were lysed by sonication in three rounds (5 minutes each, 4 cycles per round, at ~30% power). The cell lysate was cleared by centrifugation at 16,000 rpm for 35 minutes at 4°C using a JA 25.5 rotor.

The cell lysate was first applied to a HiTrap Q HP anion exchange column (5 ml) equilibrated with Buffer A (low salt). The protein was eluted using a gradient of 50 mM to 1 M NaCl. The peak fractions containing the protein of interest were identified, pooled, and concentrated using a 10,000 MWCO Amicon Ultra Centrifugal Filter unit. The concentrated protein was then applied to a HiPrep 26/10 Desalting column

equilibrated with Buffer A (low salt). The peak fractions were pooled and further purified on a HiTrap Heparin column (5 ml). The protein was eluted using a 2 M NaCl gradient. Initially with a linear gradient up to 60%, followed by a step elution to 100% of the elution buffer. The peak fractions were pooled, concentrated again with a 10,000 MWCO Amicon Ultra Centrifugal Filter unit, and loaded onto a Superdex 200 Increase 10/300 gel filtration column equilibrated with 2x Buffer A. For the final purification step, the pooled fractions were applied to a Capto HiRes Q 5/50 column equilibrated with 2x Buffer A. The protein was eluted using a 2 M NaCl gradient, with a linear gradient up to 60%, followed by a step elution to 100% of 2 M NaCl. The peak fractions were analyzed by SDS-PAGE, and those with sufficient purity were kept separately. A nuclease test was performed on each fraction. For storage, the fractions were mixed 1:1 with 100% glycerol and stored at -80°C for long-term storage or at -20°C for immediate use. The final protein concentration was determined using a spectrophotometer (NP80) by measuring absorbance at 280 nm, and protein purity was assessed by the 260/280 ratio. SDS-PAGE gel analysis was performed between each chromatography step to monitor the protein purity.

2.3.2. Expression, purification, and labeling of LD655-Gp5/trx

The expression and purification of T7 bacteriophage gp5/trx were based on the previously described protocol by Johnson and Richardson (2003). Internal YBBR-labeled gp5 was cloned into a pRSFDuet vector, and thioredoxin (trx) was cloned into a pET Duet vector. Both plasmids were co-transformed into *E. coli* BL21 star cells and plated on LB agar plates containing kanamycin (Kan) and ampicillin (Amp) for selection. A preculture was grown overnight in LB medium with Amp and Kan and then diluted 1:100 in 2 liters of TB medium containing Amp, Kan, and K salts. The cells were grown at 37°C until reaching an OD of 1.0, after which protein production was induced with 1 mM IPTG. The culture was further incubated at 37°C for an additional 4 hours. The cells were harvested by centrifugation at 4000 rpm for 15 minutes, washed with 35 ml of PBS, and centrifuged again at 4000 rpm for 15 minutes. The resulting cell pellet was stored at -80°C. To enhance lysis, the cells underwent a freeze-thaw cycle by thawing in a water bath, refreezing in liquid nitrogen, and thawing again. The pellet was resuspended in 50 ml of lysis buffer (25 mM HEPES-KOH, pH 8.0, 1 mM DTT, 5% (v/v) glycerol, 500 mM KCl, 20 mM imidazole). The lysis mix was supplemented with 1x protease inhibitor cocktail, 1x lysozyme, and 1x DNAase I. The cells were lysed by

three rounds of sonication (5 minutes, 4 cycles, ~30% power). The cell lysate was cleared by centrifugation at 15,000 rpm for 30 minutes at 4°C using a JA 25.5 rotor.

In the first step of chromatography, the cell lysate was applied to a HisTrap (5 ml) column equilibrated in lysis buffer. The column was washed sequentially with lysis buffer and wash/desalting buffer (25 mM HEPES-KOH, pH 8.0, 1 mM DTT, 5% (v/v) glycerol, 200 mM KCl). The protein was then eluted using a gradient of 20 mM to 250 mM imidazole. The elution buffer contained less salt than the lysis buffer (25 mM HEPES-KOH, pH 8.0, 1 mM DTT, 5% (v/v) glycerol, 200 mM KCl, 250 mM imidazole). Peak fractions were pooled and concentrated using a MWCO 10,000 Amicon Ultra Centrifugal Filter unit. After concentrating, the sample was applied to a HiPrep 26/10 Desalting column equilibrated in wash/desalting buffer to remove imidazole. The protein was then incubated with TEV protease at a 1:50 ratio and 10 µl DNAse overnight at 4°C to remove the His-tag from trx and DNA. The sample was subsequently applied to a HisTrap (5 ml) column equilibrated in wash/desalting buffer, and the flow-through was collected. Next, the sample was loaded onto a HiTrap Heparin (5 ml) column equilibrated in wash/desalting buffer. The protein was eluted using a salt gradient from 200 mM to 1 M KCl. Peak fractions were pooled and concentrated using a MWCO 10,000 Amicon Ultra Centrifugal Filter unit. The sample was then applied to a Superdex 200 Increase 10/300 gel filtration column equilibrated in gel filtration (GF) buffer (25 mM HEPES-KOH, pH 8.0, 1 mM DTT, 10% (v/v) glycerol, 150 mM KCl, 0.1 mM EDTA). The protein eluted as a single symmetric peak, corresponding to an approximate molecular weight of 95 kDa (a 1:1 complex of gp5 and trx). Peak fractions were pooled and concentrated. A portion of the protein samples was taken for labeling reactions, while the remaining samples were snap-frozen and stored at -80°C.

To produce LD655-gp5 in complex with Trx, ybbR-gp5/Trx was mixed with SFP synthase and LD655 CoA at a molar ratio of 1:1:1.5 in GF buffer supplemented with 10 mM $MgCl_2$ and incubated at 30°C for 2 hours. The sample was then applied to a Superdex 200 Increase 10/300 gel filtration column equilibrated in GF buffer. Peak fractions were pooled and concentrated using an Amicon Ultra Centrifugal Filter unit. Aliquots were snap-frozen and stored at -80°C. Labeling efficiency was estimated to be approximately 94%, based on the extinction coefficients of the gp5/trx complex and LD655. The final protein concentration was determined using a spectrophotometer (NP80) by measuring absorbance at 280 nm, and protein purity

was assessed by the 260/280 absorbance ratio. SDS-PAGE gel analysis was performed between each chromatography step to monitor protein purity.

2.4. DNA substrate preparation

2.4.1. DNA Handle preparation with multiple biotins for topology studies

To torsionally constrain the DNA substrates, DNA handles with multiple biotins incorporated into the DNA sequence were prepared using PCR. Two different handles were created, each complementary to a specific restriction binding site: one for XhoI and the other for NotI. The preparation steps were identical for both handles, except for the forward and reverse primers used, the input template in the second PCR reaction, and the restriction digest step. The following process applies to both versions, with any differences indicated as necessary.

In the first PCR step, a template was created using lambda DNA. Each 50 µl reaction mixture contained 50 ng of lambda DNA, 6 units of Phusion High-Fidelity DNA polymerase, PCR reaction buffer (1x HF Buffer), 200 µM of each dATP, dCTP, dGTP, and dTTP (from the dNTP bundle), 3% DMSO, and the appropriate forward and reverse primers (NotI: ORT-9, ORT-10; XhoI: ORT-11, ORT-12). The steps of the PCR program are presented in Table 2-5. After the PCR, 5 µl of each reaction was loaded onto a 1% TBE agarose gel and electrophoresed at 90 V for 1.5 hours.

Steps	Temperature	Duration
Initial Denaturation	98°C	3 min
30 cycles	98°C	30 s
	68°C	30 s
	72°C	3 min
Final Extension	72°C	10 min
Hold	4°C	-

Table 2-5 PCR Program to create DNA template

The PCR product was purified using the QIAGEN QIAquick PCR Purification Kit, following the manufacturer's instructions, and eluted in 30 µl of the provided elution buffer. To incorporate biotin molecules into the DNA handle, the concentration of dTTP in the PCR reaction was reduced and supplemented with Biotin-11-dUTP. Each 50 µl reaction contained 50 ng of the purified template from the previous PCR, 15 units of

Taq DNA polymerase, 1x ThermoPol Reaction Buffer, 200 µM each of dATP, dCTP, and dGTP, 130 µM dTTP, 70 µM Biotin-11-dUTP, 3% DMSO, and the same forward and reverse primers (NotI: ORT-9, ORT-10; XhoI: ORT-11, ORT-12) as in the previous reaction. The reaction steps are detailed in Table 2-6.

Steps	Temperature	Duration
Initial Denaturation	95°C	3 min
35 cycles	95°C	30 s
	60°C	30 s
	68°C	3 min
Final Extension	68°C	10 min
Hold	4°C	-

Table 2-6 PCR Program to integrate Biotion-11-dUTP

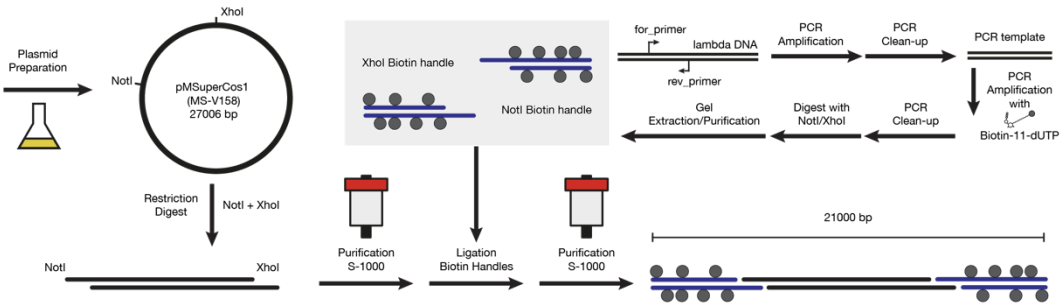


Figure 2-2 Preparation of biotin handles and linear, biotinylated DNA.
Steps to prepare 21 kb linear substrates with multiple biotins handles. In the upper right corner, the steps for creating the biotin handles are displayed.

The PCR product was purified using the QIAquick PCR Purification Kit, following the manufacturer's instructions, and eluted in 30 µl of the elution buffer. To create sticky ends for binding to the DNA, the handles were digested with either 30 units of NotI-HF or XhoI in a digest buffer (1x rCut Smart Buffer) containing 2 µg of DNA in 50 µl reactions. The reaction mix was incubated at 37°C for 3 hours, followed by heat inactivation at 65°C for 20 minutes. For the final purification of the handles, the reactions were loaded onto a 0.75% TBE agarose gel and run at 90 V for 1 hour. The bands were cut out with a scalpel, and the handles were extracted using the QIAquick Gel Extraction Kit, with each wash step performed twice. The final product was eluted in 30 µl, and its concentration was measured. The handles were freshly prepared for each DNA preparation requiring biotin handles and were stored on ice in a cold room or fridge until the ligation step.

2.4.2. Preparation of linear, biotinylated DNA for single-molecule nicking frequency quantification

To ensure the DNA remained torsionally constrained and to create a substrate for quantifying nicking frequency, a linear DNA with the NotI and XhoI handles ligated to opposite ends was prepared (Figure 2-2). The DNA backbone used for this substrate was pMSuperCos1 (MS-V158), created by Dr. Matthias Scherr, a former member of the Duderstadt Lab.

A 5 ml preculture of LB (Amp) was inoculated using a single colony from a freshly streaked *E. coli* DH5α LB-Amp plate. The preculture was grown for 8 hours, then diluted 1:1000 into a 100 ml LB-Amp culture and grown overnight at 37°C. The next day, cells were harvested by centrifugation (4000 rpm, 15 min), and the plasmid was extracted using the QIAGEN Plasmid Maxi Kit following the manufacturer's instructions. The application of the sample, washing, and elution steps were performed in the cold room, with the wash step repeated three times instead of the standard two. The plasmid was resuspended in 10 mM Tris-HCl, pH 8.0, and 0.1 mM EDTA. Samples were taken from the lysate, flow-through, wash, and elution steps and checked on a 1% TBE agarose gel (90 V, 1.5 hours).

To create sticky ends, 125 µg of plasmid was digested overnight at 37°C with 125 units of XhoI and NotI-HF in a 500 µl reaction volume using 1x rCutSmart buffer. The digestion was performed inside an incubator to prevent precipitation at the lid. To prepare the purification system for separating the larger DNA fragments (21 kb) from the smaller DNA (6 kb), all components, including the sample loop, syringe port, fraction collector, and tubing, were soaked in 0.5 M NaOH for 30 minutes to eliminate any remaining nucleases from previous purifications. Before loading the sample onto a Sephacryl S-1000 SF Tricorn 10/300 gel filtration column for size exclusion chromatography, 0.1% SDS was added to quench the digestion reaction. The column was equilibrated with 10 mM Tris-HCl (pH 8.0), 300 mM NaCl, and 1 mM EDTA (S-1000 Buffer). Fractions were collected, and the peak fractions were analyzed by loading them onto a 1% TBE agarose gel. Desired fractions were pooled, and three times the volume of 100% ethanol (stored at -20°C) was added to the pooled fractions. The sample was stored overnight at -20°C. The DNA was precipitated by centrifugation at 15,000 x g for 1 hour. The supernatant was discarded, and the DNA pellet was washed with 5 ml of 70% ethanol at room temperature. After repeating the

centrifugation for 20 minutes, the supernatant was discarded, and the pellet was dried for 10 minutes at room temperature. The linearized plasmid was reconstituted in 300 µl of 10 mM Tris-HCl (pH 8.0) and stored in the cold room until the ligation step.

To construct a torsionally constrained linear DNA substrate, handles with multiple Biotin-11-dUTP were ligated onto the linearized DNA. The handles were prepared as described in section 2.4.1. The reaction mixture contained 18 µg of linearized DNA, 2000 units of T4 DNA Ligase, 1x T4 DNA Ligase Buffer, 2 mM ATP, 2 µg of NotI handle, and 2 µg of XhoI handle, in a total volume of approximately 110 µl. The reaction was incubated overnight at 16°C. To separate excess handles from the final DNA product, the reaction mixture was injected into a Sephacryl S-1000 SF Tricorn 10/300 gel filtration column equilibrated with S-1000 buffer. The fractions were analyzed on a 1% TBE agarose gel, and the desired fractions were pooled. These were then mixed with three times the volume of 100% ethanol at -20°C, and the ethanol precipitation was carried out as described previously. The DNA pellet was reconstituted in 100 µl of 10 mM Tris-HCl, pH 8.0, and 0.1 mM EDTA. Small aliquots (10 µl) of the final DNA substrates were snap-frozen in liquid nitrogen and stored at -80°C. The DNA substrate was tested on the microscope for binding and the creation of supercoils.

2.4.3. Preparation of forked, biotinylated DNA for single-molecule replication assays

To study DNA replication, a pre-primed DNA forked substrate was created to bind to the surface through streptavidin-biotin interactions. Biotins were integrated into the DNA substrate, allowing it to either rotate freely or be torsionally constrained, depending on the number of biotin molecules present in the DNA sequence. The preparation of these substrates followed a common workflow (Figure 2-3), with variations only in the specific DNA modules attached to the sticky ends of the 27 kb plasmid backbone. While a similar approach has been employed previously to investigate DNA replication (Lewis et al., 2020; Mueller et al., 2020), these earlier studies did not include the integration of topological barriers within the substrates.

The plasmid was amplified, harvested, and purified according to the method described in section 2.4.2. The purified plasmid was then digested with XbaI and XhoI to create sticky ends (XhoI-XbaI lin. V158). The digestion reaction, which contained 125 µg of plasmid DNA, 125 units each of XhoI and XbaI, and 1x rCutSmart buffer, was incubated at 37°C overnight. To stop the reaction, 0.1% SDS was added, and the mixture was

subsequently injected into a Sephacryl S-1000 SF Tricorn 10/300 gel filtration column equilibrated with S-1000 buffer. The collected fractions were analyzed on a 1% TBE agarose gel and the desired fractions were pooled. Ethanol precipitation was then performed by adding three volumes of 100% ethanol at -20°C, following the protocol described in section 2.4.2. The resulting DNA pellet was resuspended in 300 µl of Tris-HCl, pH 8.0 and stored on ice until the ligation step.

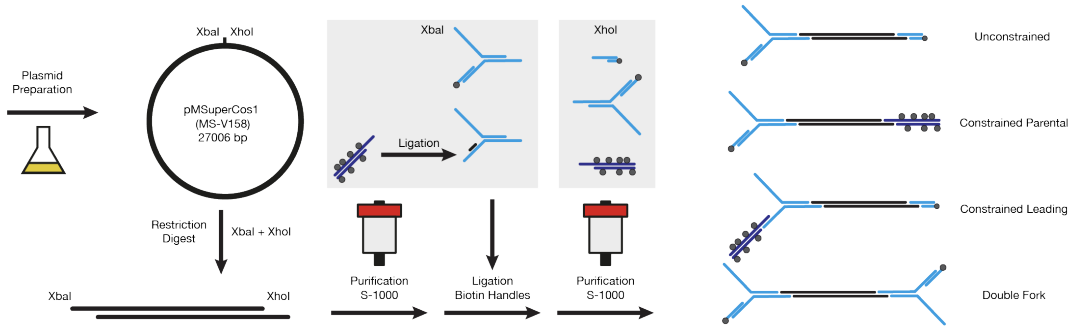


Figure 2-3 Preparation of different forked substrates for replication studies. Steps to prepare the different forked substrates for replication studies are laid out. Plasmid was prepared and digested with XbaI and XhoI. Linearized DNA with sticky ends was purified by a gel filtration column followed by a ligation step of different functional handles. The ligation mix was cleaned up by size exclusion and stored until further use.

Depending on the substrate design, different fork and end modules were prepared prior to the ligation step. To create the fork and end pieces containing a single biotin, oligos were annealed in Duplex Buffer (30 mM HEPES-KCl, pH 7.5, 100 mM KOAc). For the end piece with a XhoI matching sequence and a single biotin, oligos ORT-1 and ORT-2 were mixed in a 6:1 ratio, heated to 95°C for 5 minutes, and then cooled to 4°C at a rate of 0.5°C per minute (Table 2-7). The final concentration was adjusted to 10 µM. To create the fork with a XbaI matching sequence and a single biotin, oligos ORT-3, ORT-4, and ORT-5 were mixed in a ratio of 1:6:60 and annealed following the same protocol (Table 2-7). Alternatively, to create a fork with a XhoI matching sequence containing one biotin, oligos ORT-13, ORT-4, and ORT-5 were mixed and annealed in the same manner. The final concentration for this preparation was set at 1 µM.

The steps to create an end module with multiple biotins incorporated are described in section 2.4.1. The fork module containing multiple biotins was prepared by ligating the NotI biotin handle (Section 2.4.1) to a pre-annealed fork with a four-base pair overhang

matching NotI. This fork, made up of oligos ORT-3, ORT-4, and ORT-8, was prepared as previously described, with the exception of a different mixing ratio of 1:6:12 (final concentration of 1 µM). The annealed fork and the NotI biotin handle were then combined in a 1:3 ratio in a reaction volume of 65 µl containing 2000 units of T4 DNA Ligase and 1x T4 DNA Ligase Buffer. The ligation reaction was carried out overnight at 16°C.

Steps	Temperature	Duration
Initial Heating	95°C	5 min
Cool Down	ΔT = -0.5°C	1 min
Hold	4°C	-

Table 2-7 PCR program to anneal oligos for forked substrate

To create the final configurations of the forked substrates, a ligation was performed in 1x T4 DNA Ligase Buffer with 2000 units of T4 DNA Ligase, 18 µg of linearized plasmid (XhoI-XbaI lin. V158), and 2 mM ATP at 16°C overnight, using different DNA modules. To produce the substrate that is free to rotate at both the fork and the end, the end module with a single biotin (ORT-1, ORT-2) and the fork module with a single biotin (ORT-3, ORT-4, ORT-5) were added to the ligation reaction at a final concentration of 140 nM each. For the substrate with a torsionally constrained end and a rotatable fork, the XhoI biotin handle (2 µg in total) and the single biotin fork (ORT-3, ORT-4, ORT-5, final concentration 140 nM) were added. To create the substrate with a rotatable end and a torsionally constrained fork, the end module with a single biotin (ORT-1, ORT-2, final concentration 140 nM) and the complete ligation reaction from the fork with the NotI biotin handle (ORT-3, ORT-4, ORT-8, NotI biotin handle) were used. For the substrate with two replication forks, one on each end, the two fork modules matching XhoI (ORT-3, ORT-4, ORT-5) and XbaI (ORT-13, ORT-4, ORT-5) were added at a final concentration of 140 nM.

The following steps were performed identically for each substrate. To remove any excess DNA modules that were not ligated to the final substrate, the reaction mix was injected into a Sephacryl S-1000 SF Tricorn 10/300 gel filtration column equilibrated with the S-1000 buffer. The fractions were analyzed on a 1% TBE agarose gel, and the desired fractions were pooled. The pooled fractions were combined with three times their volume of 100% ethanol and stored at -20°C. Ethanol precipitation was conducted as described in section 2.4.2. The DNA pellet was reconstituted in 100 µl

of Tris-HCl, pH 8.0, and 0.1 mM EDTA. Small aliquots (10 µl) of the final DNA substrates were snap-frozen in liquid nitrogen and stored at -80°C. The substrates were then tested on the microscope for binding and length measurements.

2.5. Single-molecule assays

2.5.1. Total internal reflection fluorescence microscopy (TIRFM)

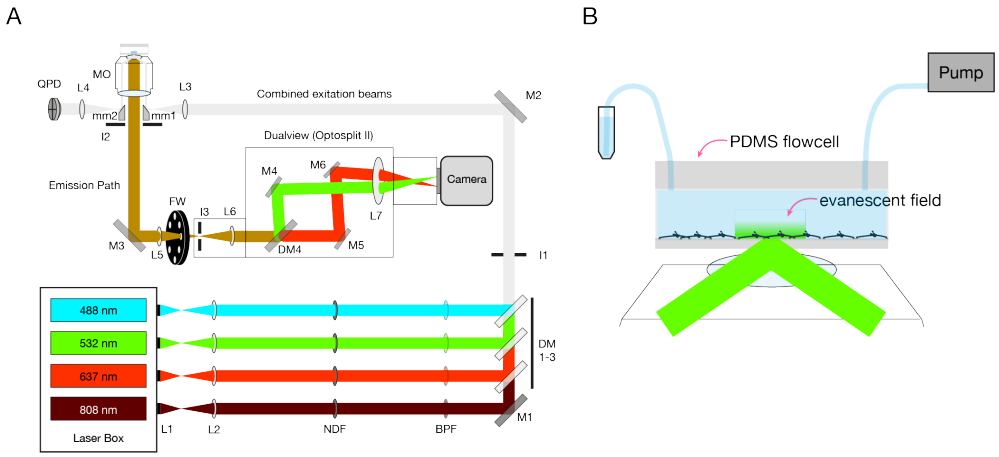


Figure 2-4 Schematic layout of the custom mmTIRF microscope.
(**A**) Schematic representation of the optical setup of the custom-modified mmTIRF microscope. Four individual laser sources (488, 532, 637, and 808 nm) are expanded and collimated by a Keplerian beam expander consisting of lenses (L1, L2). Neutral-density filters (NDF) adjust the laser intensity, and band-pass filters (BPF) ensure correct wavelength selection. The beams are combined using a broadband mirror (M1) and long-pass dichroic mirrors (DM1-3), and their diameter is refined with an iris (I1). The beam is redirected by mirrors (M2) to the entry micromirror (mm1) and focused through a lens (L3) near the back aperture of the microscope objective (MO). The sample, placed in a flow cell, is positioned with nanometer precision using a piezo-controlled stage. The exit beam is collected by the exit micromirror (mm2) and directed to a quadrant photodiode (QPD) detector, enabling autofocus via a feedback loop using an infrared laser (808 nm). The emission pathway configuration showing scattered light reduction with iris (I2) and broadband mirrors (M3). The emission light passes through the filter wheel (FW) to remove residual scattered light and is split into two channels (<635 nm and >635 nm) by a long-pass dichroic mirror (DM4) in the OptoSplit II system, with the channels aligned using mirrors (M4, M5, M6) and focused onto separate areas of the sCMOS/EMCCD camera for dual-emission imaging. (**B**) Schematic representation of the flow cell placed on top of the microscope objective. Total internal reflection fluorescence (TIRF) creates an evanescent field that selectively illuminates DNA molecules close to the slide surface. The flow cell is connected to a buffer reservoir and a syringe pump or other microfluidic device to control the flow of reagents and maintain the desired conditions during imaging.

Total internal reflection fluorescence microscopy (TIRFM) was used throughout this study for all single-molecule assays. A co-localization single-molecule spectroscopy (CoSMoS) technique was employed on a custom-modified mmTIRF microscope ("Winky") based on a system from Mad City Labs, as previously described (Larson et al., 2014) (Figure 2-4A). All essential microscopy components are listed in Table 2-2. The mmTIRF was equipped with four individual laser sources at wavelengths of 488, 532, 637, and 808 nm (OBIS 488nm LS 120 mW, OBIS 532 nm LS 120 mW, OBIS 637 nm LX 100 mW, and 808 nm 250 mW single mode laser). A Keplerian beam expander, comprising two lenses (L1 and L2), expanded and collimated all individual laser beams. Multiple neutral-density filters (NDF) were then used to attenuate the laser intensity, with one filter reducing the beam intensity by a factor of 10. A band-pass filter (BPF) appropriate for the laser wavelength ensured the correct wavelength in each beam line. Individual laser beams were combined using an aligned broadband mirror (M1) and several long-pass dichroic mirrors (ZT488rdc, ZT532rdc, and ZT647rdc) (DM1-3), selected based on their spectral properties to reflect and transmit the required wavelengths. An iris (I1) was placed to adjust the beam diameter, followed by another periscope/broadband mirror (M2) to redirect the beam.

The beam was aligned to the entry micromirror (mm1) and focused through another lens (L3) near the back aperture of the microscope objective (MO, Apo N TIRF 60x oil-immersion TIRF objective, NA 1.49). The position of mm1 was adjustable to change the TIRF angle. The sample, placed on top of the MO in a flow cell connected to a microfluidic system (Figure 2-4B), could be adjusted in the xyz directions with low nanometer precision using a piezo-controlled microscope stage (Mad City Labs' mmTIRF system). The exit beam was collected by the exit micromirror (mm2) and directed to a quadrant photodiode (QPD) detector (Mad City Labs: TIRF Lock system). A feedback loop utilized the 808 nm infrared laser to autofocus during acquisition, with the main microscope control software (Micromanager) adjusting the z-position of the microscope stage based on the QPD readout. Scattered light beneath the MO in the emission path was reduced by an iris (I2) right below the two micromirrors (mm1 and mm2), increasing the signal-to-noise ratio of the emission fluorescence. A collimating lens (L5) was placed next and the emission light was redirected by a broadband mirror at a 45° mount (M3) toward the filter wheel (FW, Optospin). The emission filter in the FW removed residual scattered light from the emission path (ZET532/640nm emission filter for 532/637 nm). The emission light was collimated inside the OptoSplit II dual emission image splitter. The emission light was split into >635 nm and <635 nm

channels using a long-pass dichroic mirror (DM4, T365lpxr), creating two channels aligned with internal broadband mirrors (M4, M5, M6). The two channels were focused on different areas of the imaging camera, creating separate images for >635 nm and <635 nm.

The microscope was custom-built, and parts were replaced over time. However, the working principle remained unchanged. Two different cameras were used during the project: initially, a Photometrics PrimeBSI sCMOS camera and later, an Andor iXon Life 888 EMCCD camera with comparable specifications. The CMOS camera was used in 2x binning mode to match the pixel size of the EMCCD in 1x binning mode. All single-molecule experiments were performed on the described mmTIRF microscope in a temperature-controlled room at 22.5 ± 0.5 °C. To visualize the entire DNA molecule, mm1 was adjusted slightly out of TIRF in the highly inclined and laminated optical sheet (HILO) mode. LD655 dyes were excited with the 637 nm laser, while SYTOX Orange dyes were excited with the 532 nm laser. Dyes were excited sequentially in a multicolor experiment every 5 seconds for up to 200 frames, with exceptions noted in the method description of individual assays. All microscope parts were controlled by Micromanager v2.0.0 (Edelstein et al., 2010; Schneider et al., 2012) and custom BeanShell scripts.

2.5.2. PDMS flow cell preparation

Two different flow cell configurations were utilized in this study. The first design, referred to as the 'linear flow cell,' featured three linear flow lanes, each with an inlet and outlet on opposing sides, allowing for three separate experiments (Figure 2-5A). The second design, the 'transverse flow cell,' included two perpendicular flow lanes intersecting in the center, equipped with two inlets and two outlets, facilitating the application of transverse flow (Figure 2-5B).

To make either type of flow cell, a silicon wafer with the negative structure of the flow lanes was produced by AMO using plasma etching (Figure 2-5C). The wafer was placed in a plastic box to serve as a mold. The flow cell was constructed from polydimethylsiloxane (PDMS) by thoroughly mixing two components of the Sylgard 184 Silicone Elastomer kit according to the manufacturer’s instructions, in a ratio of 1:10. This mixture was poured into the mold containing the silicon wafer and degassed inside a desiccator until no air bubbles were visible on the surface, then kept under reduced pressure overnight. The mold, along with the PDMS, was baked at 60°C

overnight to harden the material. The following day, the flow cells were carefully cut out of the mold using a razor blade to match the dimensions of the coverslips and inlets were punched using the Whatman Unicore Punch Kit (1.00 mm).

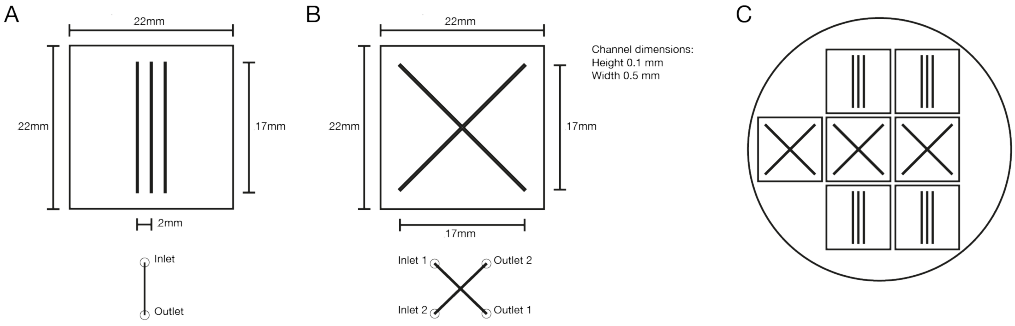


Figure 2-5 Patterns for PDMS flow cell.
(**A**) Schematic representation of the linear flow cell configuration, featuring three parallel flow lanes, each with a single inlet and outlet. (**B**) Schematic of the transverse (x-shaped) flow cell configuration, showing two perpendicular lanes intersecting at a right angle, with each lane having its own inlet and outlet. (**C**) Illustration of the silicon wafer design created by plasma etching, used as a mold for the flow cell fabrication.

2.5.3. Functionalization of PEG-Biotin slides for microscopy

Glass coverslips (22 x 22 mm, Marienfeld) were first cleaned using a Zepto plasma cleaner for 15 minutes, then transferred to a glass container filled with acetone and incubated for 5 minutes. To silanize the coverslips, they were incubated in acetone containing 2% (v/v) 3-aminopropyltriethoxysilane for 1 minute. The reaction was quenched by adding an excess of deionized water (ddH_2O). The coverslips were subsequently rinsed with ddH_2O, dried with compressed air, and baked at 110°C for 1 hour. For functionalization, the coverslips were incubated overnight at room temperature with a fresh solution of 0.6% (w/v) Biotin-PEG-Succinimidyl Carbonate (MW 5000) and 15% (w/v) mPEG-Succinimidyl Carbonate (MW 5000) in 0.1 M $NaHCO_3$. After incubation, the coverslips were rinsed with ddH_2O, dried with compressed air and incubated again with a fresh Biotin-PEG/mPEG solution as described above. Finally, the coverslips were rinsed with ddH_2O, dried with compressed air, and stored in a desiccator under vacuum.

2.5.4. Flow cell preparation

A functionalized PEG-Biotin microscopy slide was incubated with 0.2 mg/ml streptavidin dissolved in blocking buffer (20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 2 mM

EDTA, 0.2 mg/ml BSA, 0.005% (v/v) Tween-20) for 30 minutes. The slide was then rinsed with deionized water (ddH_2O) and dried with compressed air. To create the flow channels, a PDMS block with either a linear or transverse pattern was placed on top of the slide. Polyethylene tubes (inner diameter 0.58 mm) were inserted as inlets and outlets. The inlet tube was placed in a tube containing blocking buffer, while the outlet tube was connected to a syringe pump to establish directed buffer flow. The flow channel was flushed with blocking buffer, incubated for 15 minutes, and then washed at a high flow rate of 200 µl/min for 1 minute.

The DNA tethering process varied depending on the flow cell configuration. For the linear flow cell configuration, 5 pM of forked, biotinylated DNA was introduced at a flow rate of 17 µl/min for 24 minutes in reaction buffer (40 mM Tris-HCl, pH 7.4, 50 mM Potassium Glutamate, 10 mM Magnesium chloride, 0.1 mg/ml BSA, 10 mM DTT) supplemented with 200 µM chloroquine. Unbound DNA was then washed out using reaction buffer containing 300 µM ATP/CTP, 600 µM dNTPs, and 150 nM SYTOX Orange at a flow rate of 20 µl/min for 10 minutes, and the DNA was imaged to check the quality of the DNA tethers.

For the transverse flow cell configuration, inlet 2 and outlet 2 were clipped off with metal clips to prevent side flow. Then, 5 pM of forked, biotinylated DNA was introduced at a flow rate of 20 µl/min for 24 minutes in reaction buffer. Unbound DNA was washed out with reaction buffer containing 300 µM ATP/CTP, 600 µM dNTPs, and 150 nM SYTOX Orange at a flow rate of 20 µl/min for 10 minutes, while inlet 2 and outlet 2 remained clipped. The clips were then removed from inlet 2 and outlet 2, and inlet 1 and outlet 1 were clipped. Washing buffer was introduced at a flow rate of 20 µl/min for 5 minutes to ensure an even distribution of SYTOX Orange throughout the flow cell. The DNA field was checked to confirm the proper formation of arched molecules.

2.5.5. Single-molecule photoinduced nicking assay

To introduce negative supercoils in the linear, biotinylated DNA, the DNA application process in the transverse flow configuration was slightly modified. The workflow for inducing negative supercoils in DNA was adapted from the protocol by Ganji et al. (2016). During the DNA tethering step, the linear, biotinylated DNA was mixed in reaction buffer containing 500 nM SYTOX Orange. The tube was covered with aluminum foil to protect the DNA from light exposure and to avoid photo damage. The DNA was applied following the same procedure described for the transverse flow cell

configuration, with the SYTOX Orange concentration reduced to 150 nM during the subsequent wash steps, which facilitated the formation of plectonemes.

The specific imaging conditions tested are detailed in the results section. For conditions involving an oxygen scavenging system, the reaction buffer was supplemented with 1 mM Trolox (pre-aged under UV light for 6 minutes), 2.5 mM PCA, and 0.21 units/ml of PCD. The collection scripts were adjusted to be the same as the actual imaging conditions, with only the number of timepoints increased to accommodate the desired data acquisition.

2.5.6. Single-molecule linear replication assay

To initiate the T7 replication reaction in the linear configuration, a solution containing 2.5 nM gp4 (hexamer), 20 nM gp5-trx (interchangeably referred to as LD650-gp5 and gp5), and 750 nM gp2.5 in reaction buffer supplemented with 300 μM ATP/CTP, 600 μM dNTPs, and 150 nM SYTOX Orange was introduced into the prepared flow cell at a flow rate of 150 μl/min for 1.5 minutes. The inlet and outlet tubes were then clipped to prevent any flow fluctuations or external force. The proteins were allowed to incubate for 1.5 minutes before starting image collection, with frames captured at 5 second intervals.

2.5.7. Single-molecule transverse flow replication assay

To visualize the T7 replication reaction in the transverse configuration, 2.5 nM gp4 (hexamer), 20 nM gp5-trx (LD650-gp5 and gp5 used interchangeably), and 750 nM gp2.5 in reaction buffer supplemented with 300 μM ATP/CTP, 600 μM dNTPs, and 150 nM SYTOX Orange were introduced at a flow rate of 20 μl/min from inlet 2 to outlet 2 of the prepared flow cell, while inlet 1 and outlet 2 were clipped to prevent unwanted buffer flow. After 3.5 minutes, image collection began, maintaining a steady buffer flow of 20 μl/min to stretch the lagging strand product and visualize replication. For slow collection, images were taken every 5 seconds. In continuous mode, image collection started 4 minutes and 50 seconds after initiating flow, with the collection script cycling between the green and red lasers for 100 frames, resulting in 200 ms timesteps between images.

2.6. Single-molecule data analysis

All raw data from the single-molecule experiments were processed in Fiji (Schindelin et al., 2012) using the Mars software suite (Huisjes et al., 2022) (see also Chapter 3). All plots and visualizations were generated using Python.

2.6.1. Nick event analysis, quantification, and fitting

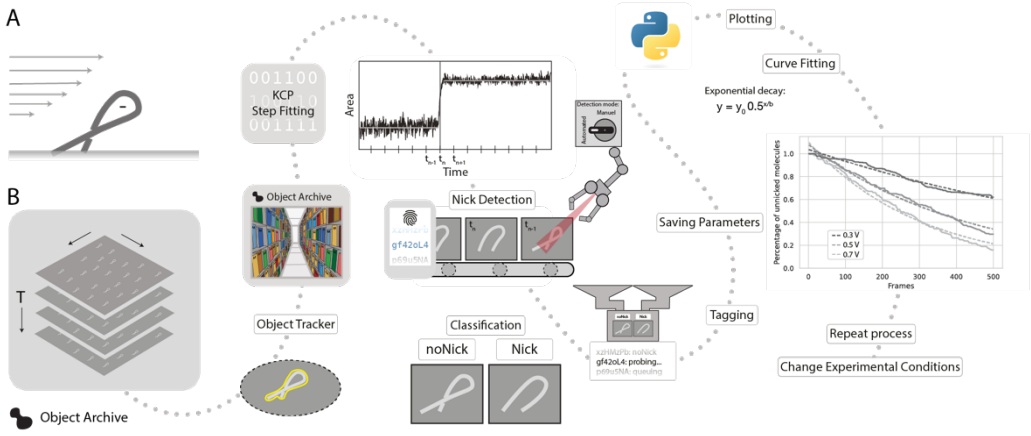


Figure 2-6 Mars workflow to analyze nick events.
(**A**) Schematic of the nicking assay. Negative supercoiling was induced in a constrained, linear DNA substrate, which was then imaged under flow conditions. (**B**) Overview of the analysis pathway, beginning with the raw image stack on the left and ending with the final plot on the right. DNA shapes were tracked using the 'Object Tracker' tool, and the data was saved in an Object Archive. Kinetic change point analysis was performed to detect step transitions in the trace of area over time, corresponding to nick introduction. Nicks were identified either manually or through the detected step transitions. Molecules were tagged, and the archive was imported into Python for final analysis. Curves were fitted using an exponential decay model to determine the half-life. The process was repeated for different conditions.

To analyze nicking events, the following workflow was applied. Supercoiled DNA molecules were marked in the video using the ROI tool in Fiji/ImageJ2. The 'Object Tracker' tool from Mars was then used to fit the shapes based on the ROI windows, creating an Object Archive. When a single nick occurred, the area of the DNA molecule increased, which could be analyzed using the 'Kinetic Change Point Finder' for step fitting (Hill et al., 2018) (Figure 2-6B). Each molecule was also inspected manually. Those without a nick were tagged as 'noNick,' and the timepoint of the nicking event, if any, was added to the molecule's parameters.

$$\%\ of coilable\ molecules = \frac{Number\ of\ constrained\ molecules}{Total\ number\ of\ double\ tethered\ molecules} \cdot 100 \qquad 2.1$$

To assess the quality of the DNA, the number of constrained molecules at the first timepoint was compared to the total number of molecules forming a double tether. For each experiment, the percentage of constrained molecules was calculated by dividing the number of constrained molecules by the total number of double-tethered molecules and multiplying the result by 100 (Equation 2.1).

The archives were imported into Python for further analysis. To create survival curves, the percentage of unnicked molecules at each time point was calculated by dividing the number of unnicked molecules at a specific time point by the total number of constrained molecules present at the start of the measurement. This percentage was then plotted against time (in frames). The half-life was determined by fitting the curves with an exponential decay model, where *b* represents the half-life of the curve and y_0 denotes the initial quantity (here equal to 1):

$$y = y_0 0.5^{x/b} \qquad 2.2$$

2.6.2. DNA substrate length measurement

To calculate the length of individual molecules, DNA molecules were identified in the video using the 'DNA Finder' tool from Mars, which provided the coordinates for the top and bottom ends of each DNA molecule (DNA top: (X_1, Y_1), DNA bottom: (X_2, Y_2)). These coordinates were imported into Python. The length of each DNA molecule L_{DNA} was then calculated in pixels using the Pythagorean theorem (Equation 2.3). The mean length was determined from all molecules, and the L_{DNA} values for each molecule were plotted in a histogram.

$$L_{DNA} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \qquad 2.3$$

The length of the molecules L_{DNA} was multiplied by the pixel size after magnification (156 nm) to convert the measurement from pixels to nanometers. This value was then compared to the crystallographic length, which was calculated by multiplying the number of base pairs for each substrate by 0.34 nm, representing the rise per base pair in B-DNA.

2.6.3. Replication rate, processivity, pause analysis in the linear flow cell

To analyze replication events in the linear flow cell configuration (Figure 2-7A), a dedicated analysis workflow was developed in Mars (Figure 2-7B). Active molecules that showed blob a formation of the lagging strand were marked using the ROI tool. The 'Peak Tracker' tool from Mars was employed to track the replication reaction by following the DNA blob, creating a Single Molecule Archive (DoG filter radius 2, detection threshold 3). Identified positions were corrected for sample drift occurring during the measurement using the 'Drift Corrector.' Immobile peaks were identified by calculating the variance in the y-coordinate and were used as reference points to calculate drift.

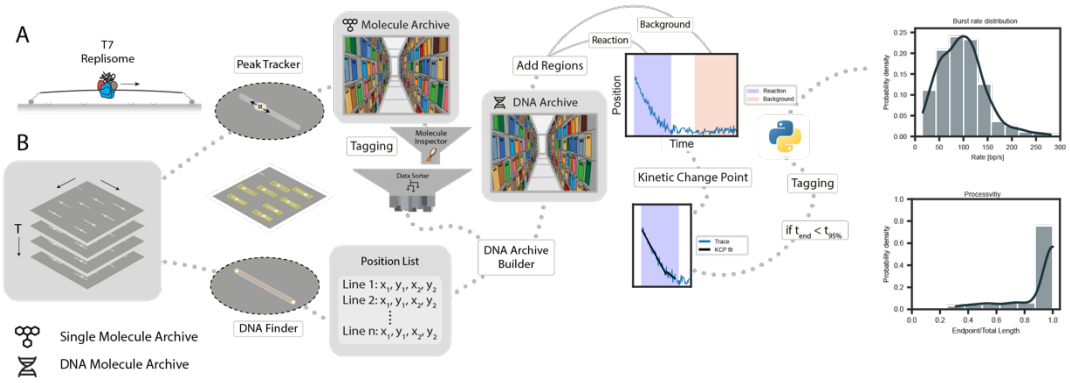


Figure 2-7 Mars workflow to analyze rate and processivity in linear flow cell. (**A**) Schematic of the replication reaction in the linear flow cell. Replication was visualized by observing the accumulation of the lagging strand product moving along the DNA. (**B**) Overview of the analysis workflow, starting with the raw image stack on the left and ending with the final plot on the right. The DNA blob for individual molecules was tracked using the 'Peak Tracker' tool. Simultaneously, the 'DNA Finder' identified the corresponding linear DNA molecules and saved their coordinates. A Molecule Archive was created to store the tracked DNA blobs and was subsequently used to generate a DNA Archive by matching the coordinates. Regions were defined for each trace to analyze replication rates ('Reaction') and, when possible, a background region ('Background'). Kinetic change point (KCP) analysis was performed on each trace to determine the replication rate. Molecules were filtered if the reaction exceeded the 95% confidence interval, based on calibration from the nicking assay. The archives were imported into Python to generate final histograms of the rate and processivity distributions.

Simultaneously, DNA molecules were fitted using the 'DNA Finder' tool in the final frame of the video. This data was correlated with the positional information of the DNA blob, creating a DNA Molecule Archive. The DNA blob was placed within the reference frame of the linear DNA molecule, allowing for the description of kinetics in terms of base pairs along the DNA molecule. An activity region was defined for each molecule, marking where the replication reaction occurred. Molecule kinetics were fitted using

the 'Kinetic Change Point Finder' (KCP) tool, with either a background region input to calculate noise in the trace or a manually set value for σ. Molecules were manually sorted using tags, excluding those not properly tracked or with a reaction window exceeding the set timepoint from the 95% confidence interval. The segments table generated by the KCP command was imported into Python for further analysis.

Rates lower than 15 bp/s (representing a 6-fold reduction compared to the median rate) were considered pausing events, while rates resulting from incorrect fits were excluded. The rates were plotted in a histogram, with the bin size determined by the square root of the number of observations, and a kernel density estimation was overlaid on the distribution. The segments table included the endpoint of the reaction, and molecules were corrected for the direction of replication. Molecules with completed replication reactions were manually tagged as 'Complete.' Endpoints were plotted as the percentage of DNA replicated and a kernel density estimation was plotted over the data. The pausing behavior was manually marked using the 'Region' tool in Mars, and the number of pauses was plotted as a feature.

2.6.4. DNA blob shape analysis in the linear flow cell

DNA product shape analysis was carried out as follows. Tables containing pixel values were created for each molecule using the ‘DnaArchiveKymographBuilder’ tool from mars-kymograph. This tool generated a single array of pixel values for each timepoint by averaging 5 pixels on either side of the marked DNA molecule in each DNA archive (Figure 2-8A). The resulting tables were imported into Python for further analysis. Each line of pixel values was smoothed using a moving average of 2 pixels.

A global threshold for each table was calculated using the ‘threshold_otsu’ function from the scikit-image package (van der Walt et al., 2014) (Figure 2-8B). Pixels with values above this threshold were considered DNA signal (white), while those below were considered background (black). The number of white pixels per timepoint (one row of pixels presented a time point), representing the size of the DNA blob, was counted. The average size was calculated over five timepoints near the end of the reaction window, specifically 5 timepoints before the last one, to avoid measuring the blob size at the tip of the DNA molecule. The mean size for all molecules was computed and displayed as a violin plot. To create line profiles, the pixel values after applying the moving average were plotted for each timepoint.

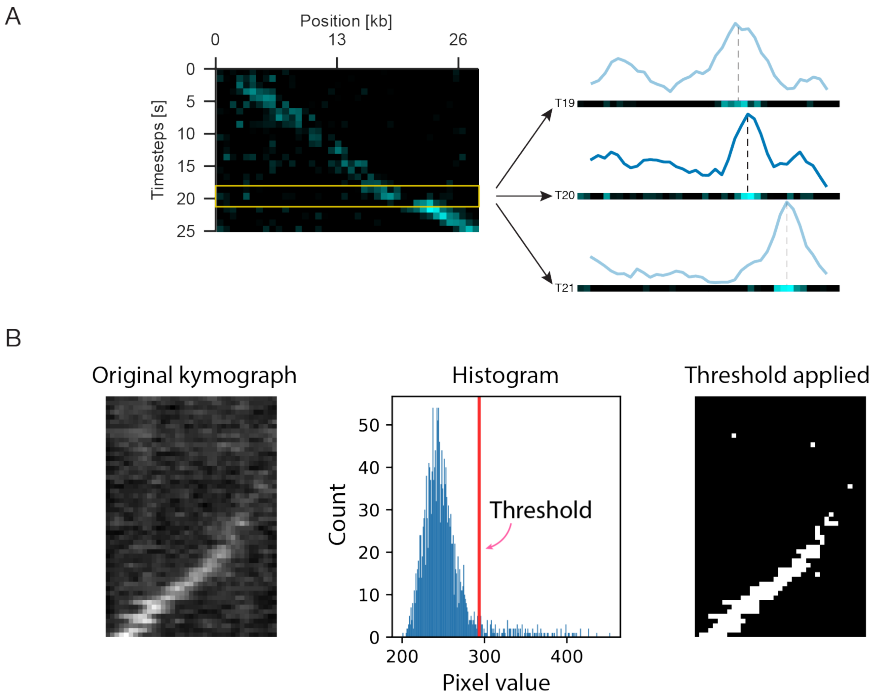


Figure 2-8 DNA blob shape analysis.
(**A**) Illustration of the process for converting kymograph time points (one row of pixels) into line profiles. (**B**) Original kymograph transformed into black and white version using a calculated threshold ('threshold_otsu').

2.6.5. Quantification of labeled gp5 protein and polymerase estimation

The average intensity of the labeled protein gp5 was quantified by analyzing bleaching steps in surface-immobilized gp5 molecules. Bleach step analysis was performed separately for each experiment. Immobile spots in the protein channel were identified, and the 'Peak Finder' command in Mars was used to create regions of interest (ROI). The 'Molecule Integrator' command then integrated the peaks over time, generating a time trace of the background-corrected intensity for each spot. Kinetic changepoint analysis was applied to fit each bleaching step, with a confidence threshold of 0.99 and an experimentally determined σ value for noise, calculated from the last 10 time points of each trace (Duderstadt et al., 2016; Hill et al., 2018). The final bleaching step for each trace was marked, and a Python script was used to calculate the intensity difference before and after bleaching. The mean value from all accepted molecules was used to normalize the intensity of each replisome, estimating the number of polymerases present.

To correct for the beam profile in the protein channel, the final frame was used, and a Gaussian blur with a radius of 40 pixels was applied using the 'Beam Profile Corrector' tool in Mars. Moving protein blobs were tracked with the 'Peak Tracker,' which integrated the intensity over time using an inner radius of 3 pixels and an outer radius of 12 pixels to subtract the background signal. The resulting intensity was then divided by the estimated step size for each experiment to calculate the number of polymerases.

2.6.6. Rate and processivity analysis for transverse flow

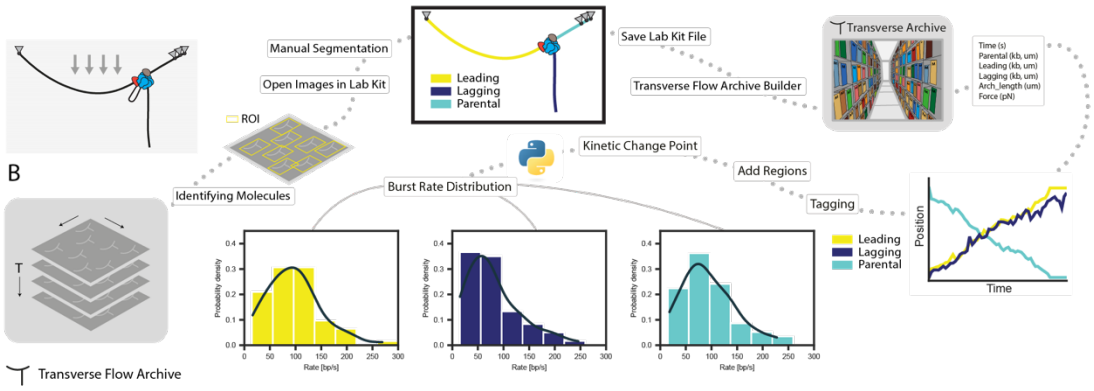


Figure 2-9 Mars workflow to analyze transverse flow replication events. (**A**) Schematic of the replication reaction in the transverse flow assay. The replication was visualized by stretching the lagging strand product using a transverse flow. (**B**) Overview of the analysis pathway, illustrating the steps from the raw image stack (left) to the final plot (right). The video was processed in LabKit, where the strands were manually segmented into 'Leading', 'Lagging', and 'Parental'. The segmented files were saved, and a custom importer ('Transverse Flow Archive Builder') was used to build an archive containing data on the lengths of different DNA arms, arch length, and applied force over time. Molecules were reviewed and tagged in the archive, with regions defined for kinetic change point analysis. The results were then imported into Python to generate the final plots of burst rate and processivity distributions.

Replication events in the transverse flow configuration (Figure 2-9A) were analyzed using the following workflow (Figure 2-9B). Transverse flow replication events were identified and marked using the ROI tool in ImageJ. The image file was then opened in Labkit (Arzt et al., 2022), where the replication events were manually segmented. The 'Leading', 'Lagging', and 'Parental' strands were traced for each frame. The segmentation file was saved and imported into Mars using a custom importer named 'Transverse Flow Archive Builder'. This process created a TransverseFlowArchive that contained the lengths of each segment at different time points, measured in pixels and

then converted to base pairs. Each molecule had their own UID and position information.

Molecules were sorted using the tag functionality, and reaction regions were defined for each trace. Kinetic changepoint analysis was applied to extract replication rates for the ‘Leading’, ‘Lagging’, and ‘Parental’ strands. Replication rate and processivity were determined in a manner similar to that used for the linear flow cell configuration.

2.6.7. Analysis of lagging strand arm length dynamics

Lagging strand arm length from each accepted trace was imported in Python by extracting it from the archive. The arm length was only considered within the reaction window. An Augmented Dickey-Fuller unit root test was performed using ‘adfuller’ function from ‘statsmodels’ to confirm that the traces are non-stationary. In the next step, the difference between x(t) and x(t-1) was calculated to make the trace stationary. This was confirmed by an augmented Dickey-Fuller unit root test. The variance of each trace was calculated and plotted as a violin plot.

2.6.8. Estimating forces using the worm-like chain model

Force was estimated using the worm-like chain (WLC) model using equation 1.4, assuming a persistence length A of 50 nm and a temperature T of 296.15 K. The contour length was set equal to the leading strand extension at each timepoint, while the current extension z was determined by the lagging strand length at each corresponding timepoint. A new column containing these force values was added to the TransverseFlowArchive. The archive was then imported into Python for further analysis. Forces were grouped into bins with a size of 3 kb, and a moving average with a step size of 1.5 kb was applied. Values below 5 kb were excluded from the analysis. The median force for each bin was calculated and plotted. The resulting data were fitted with a line plot to illustrate the trend in forces.

Chapter 3 - Mars: An adaptable solution for data management and reproducible analysis of biological datasets

The rapid evolution of cutting-edge imaging techniques is yielding increasingly intricate datasets, unveiling the temporal dynamics of individual cells and biomolecules. Single-molecule methods in particular grant access to elusive intermediates within intricate molecular pathways. Despite this, a scarcity of standards for processing these data-rich datasets poses challenges for broad dissemination. In this context, we developed Mars—an open-source platform created for the storage and processing of image-derived biomolecule properties. Mars presents Fiji/ImageJ2 commands scripted in Java for commonplace single-molecule analysis tasks. Employing a Molecule Archive architecture adaptable to complex analysis workflows.

To enrich biomolecule feature exploration, Mars boasts a comprehensive graphical user interface developed in JavaFX. Equipped with charting, tagging, region highlighting, scriptable dashboards, and interactive image views, this interface enhances user interaction. The interoperability of ImageJ2 ensures Mars can seamlessly integrate with various environments, including Python-coded ones like PyImageJ, facilitating interactive scripting and visualization. Mars stands as a flexible solution for reproducible analysis of image-derived properties. Its capabilities support the discovery and quantitative classification of novel biological phenomena, utilizing an open data format accessible to everyone.

This chapter has been adapted from parts of my earlier work titled ‘Mars, a molecule archive suite for reproducible analysis and reporting of single-molecule properties from bioimages,’ which was published in eLife in 2022. The original study's findings are essential for understanding the concept used for analyzing the single-molecule experiment presented in this thesis and are therefore revisited in this context. Furthermore, the analysis workflows developed to analyze the data from this thesis are all based on Mars and serve as an example for the capability of adaptability to various types of single-molecule data and workflow challenges.

Nadia M Huisjes*, **Thomas M Retzer***, Matthias J Scherr, Rohit Agarwal, Lional Rajappa, Barbara Safaric, Anita Minnen, Karl E Duderstadt

*These authors contributed equally

PMID: 36098381

DOI: https://doi.org/10.7554/eLife.75899

3.1. Challenges of scientific discoveries in a fast-passed environment

A major part of scientific discovery is the data analysis which can, if not documented in a reproducible fashion, slow down scientific advancement. New methods come with unique and complex datasets which call for efficient workflows to extract properties and the possibility to share and retrace the steps of extracting key features. It is a known issue to date (Carpenter et al., 2012; Eliceiri et al., 2012; Lerner et al., 2021; Meijering et al., 2016; Ouyang & Zimmer, 2017). In order to overcome these challenges software frameworks to curate images and workflows have been created (Allan et al., 2012; Kvilekval et al., 2010). Standards for reporting metadata have been developed and public archives were created to foster data availability (Goldberg et al., 2005; Williams et al., 2017). However, there is a lack of standard workflows for reproducibly tracking steps in an analysis process and reuse of image-derived properties, and the existing ones are limited in their adaptability to other experimental configurations (Greenfeld et al., 2015; Ingargiola et al., 2016). Meanwhile, the datasets in question are diverse and information-rich containing information on position, shape, composition and conformation over time of single cells and even protein complexes. The non-uniform format of these datasets obscures key features by hindering transparent and easy access. As a result, scientific discoveries are slowed down since biological features stay hidden, due to complicated and narrow focused workflows.

Single-molecule techniques are great examples for information rich, heterogenous dataset types where molecules can represent a rare intermediate state in a molecular process. Studies on conformation dynamics of membrane transport (Akyuz et al., 2013; Erkens et al., 2013), assembly and transcription by RNA polymerase (Baek et al., 2021; Chong et al., 2014; Duchi et al., 2016; Janissen et al., 2024), and DNA replication are some examples (Duderstadt et al., 2016; Geertsema et al., 2014; Scherr et al.,

2022; Tanner et al., 2009; Ticau et al., 2015). Complexes such as replication, transcription, DNA repair machineries as well as motor proteins can be spatially tracked on biological structures aiding to discover exchange events and conformational changes which influences the mentioned processes (Crickard et al., 2020; Lewis et al., 2020; Niekamp et al., 2021; Scherr et al., 2022). The influence of forces and torques on biological interactions are studied by attaching microspheres on macromolecules (Agarwal & Duderstadt, 2020; Dulin et al., 2013; Lebel et al., 2014; Neuman & Nagy, 2008; Revyakin et al., 2006; Strick et al., 1998). Technological improvements in the fluorescent signal of fluorophores, camera sensors and new imaging strategies make time-resolved live cell imaging possible (Gao et al., 2012; Tokunaga et al., 2008). These improvements have led to the discovery of frequent exchange of components in the replisome complex (Beattie et al., 2017; Kapadia et al., 2020) and the dynamics of target site search of the transcription complex (Chen et al., 2014).

In order to extract biological findings from these multidimensional datasets, a complex and long multi-step process needs to be applied. First the images are processed to create time traces for each molecule, followed by a clustering step according to its features. Tools are used to extract and monitor changes over time resulting in tabular data presenting the feature changes on a time trajectory. The time traces are evaluated manually or filtered automatically based on unique features of each trace. Most of the existing workflows are based on unidirectional transformation. Re-evaluating in the context of the original images at this step is almost impossible because of no traceable connection. Transferring the data to a separate data visualization platform to create figures for publication just adds to the issue. In order to work with large datasets, the size is reduced to deal with limited storage capacity by removing excluded observations. When realizing that the applied criteria need readjustments the whole process needs to be repeated which is very time consuming. Furthermore, including additional experiments is not straight forward in this workflow.

To overcome the stated issues presented, we developed Mars (Molecule Archive Suite) which consists of a collection of Fiji/ImageJ2 (Pietzsch et al., 2012; Rueden et al., 2017; Schindelin et al., 2012) commands and is integrated in the well-established Fiji environment. Mars provides an image processing workflow and facilitating image-derived properties in an adaptable Molecule Archive architecture (Figure 3-1). The framework enables a full analysis pipeline from image processing until creating a final

figure for publishing. Very large datasets can be merged, processed in parallel while being stored virtually. Traces can be explored in a clear and playful user interface in a fast manner, while creating sub groups on the fly using tags. Non-destructive tag-based filtering and grouping also makes sure that no observation is removed from the dataset and assigning universally unique ids (uids) to each record, which ensures a fully traceable workflow. The history of each molecule and dataset remains traceable throughout a complex analysis pipeline including multiple data merging steps. By design after evaluating, filtering criteria can be adjusted and the molecules can be retagged accordingly while tracking the used criteria in the log. The final archive, which can be shared, contains all the information needed to reproduce the workflow including the original image data.

The working principle of Mars is modular which makes it ideal to adapt the workflow according to the data specification while the working principle of reproducibility stays true for each adaptation. This chapter will focus in detail on the working principles of Mars. To illustrate the capability of Mars, one workflow will be presented to highlight the flexibility and adaptability of the data structure. In addition, Mars was utilized for all analysis tasks required for the preparation of this thesis.

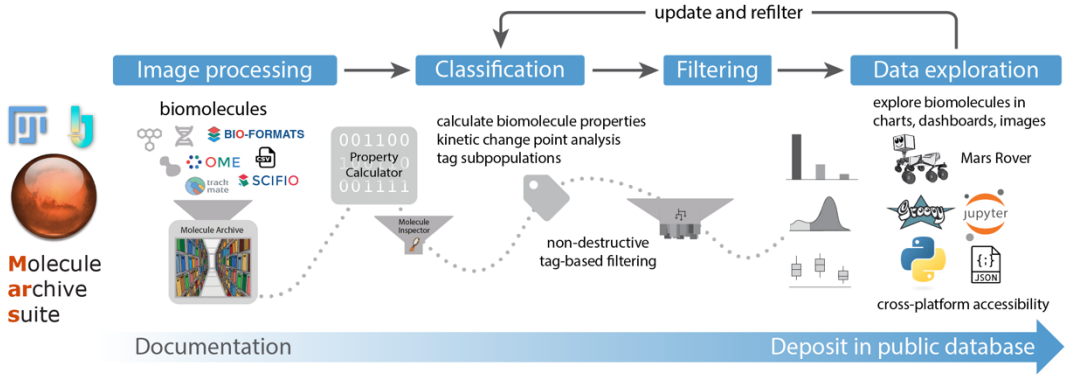


Figure 3-1 An overview of the Mars workflow.
Each process of reproducible data analysis with Mars begins with image processing, followed by iterative rounds of classification and filtering based on features in the data, and concludes with the final stage of data exploration and deposition into a public database. (Reprinted from Huisjes et al. (2022))

3.2. Results

3.2.1. Widespread drawbacks of conventional single-molecule image-processing analysis pipelines

With increasing complexity of the studied biological systems, the analysis platforms also need to be more vigorous and robust. This comes with the caveat of having multiple stages, to extract the desired quantitative information, where pitfalls can arise at different places. Mars was developed with these struggles in mind to overcome and create a platform eliminating these issues which will be explained in the following. These issues will be described using a general example of a workflow. This helps readers with no prior knowledge of single-molecule workflows to understand the subject. Furthermore, it will be highlighted how Mars resolves the issues of data readability and improves reproducibility of analysis steps.

Many single-molecule experiments involve capturing images that track the position and intensity of fluorescent labels attached to biomolecules over time. For instance, consider a basic experiment where fluorescently labeled polymerases move along DNA molecules during synthesis (discussed in more detail in the presented workflow). Researchers typically aim to monitor the polymerase position and fluorescence intensity across a series of images. From these raw images, they seek to determine the polymerase synthesis rate, the sizes of the synthesized products, and the number of polymerases associated with each DNA molecule. This analysis involves identifying the locations of fluorescent spots in each image and then linking spots from the same molecules across consecutive frames. In a typical single-molecule workflow, tracking data are compiled into a table with columns for image number, molecule position, and an index number for each track. Several filters are applied to select polymerases that move a minimum distance and exhibit fluorescence intensity consistent with a single dye. After this automated filtering, tracks are manually reviewed to eliminate those with tracking errors. In a standard workflow, these filtering steps result in the removal of rejected tracks from the table. Finally, a conversion factor relating DNA length to polymerase coordinates is used to calculate the synthesis rate and product length. This process is repeated across multiple videos, generating a final merged table with reindexed global track numbers.

This workflow includes multiple drawbacks that lead to increased analysis time, impede novel scientific discoveries, result in lack of reproducibility, and potentially lead to errors. The biggest design flaw of conventional analysis pipelines is the unidirectionality of each step. Rejected tracks are actually deleted in manual or automated filtering steps. This makes it impossible to adjust filtering parameters which means the analysis needs to be repeated fully with updated criteria. Furthermore, no robust and easy to use method for documenting filtering parameters, used scripts, manual rejection criteria, and the order of steps is in place. Evaluating the data in context to the original video including corresponding metadata is almost impossible since multiple videos/datasets were merged and reindexed losing connecting information. If one subset needs to be removed due to mistakes in sample preparation, wrong experimental conditions, or human error, it is not possible to identify and link tracks back to the original video in the final merged analysis file. To remove the rejected dataset, the whole analysis needs to be repeated from the beginning. Moreover, a quick re-evaluation to test hypotheses or new models is made difficult by the structure of workflows. For example, what if induced DNA damage is obscuring the replication rate distribution and the reaction window which is considered for analysis should be changed? What if the threshold, which is considered a pause, was set too high after examine the reaction in more detail? All the analysis steps would need to be repeated again.

Mars was developed with these issues in mind providing image processing commands designed for single-molecule applications. An adaptable Molecule Archive structure is in the center, which keeps a record of analysis steps, molecules, and metadata of experiments. The build-in graphical interface offers an intuitive and simple surface for on the fly data exploration. The molecules are filtered employing a non-destructive tag-based approach where rejected molecules are marked, not deleted. Molecules can simply be re-tagged using updated criteria supporting an unidirectional workflow without redoing the initial analysis. This approach also enables a simple and fast way to evaluate subsets of molecules to examine data on a deeper level. Each record gets a uid assigned, as well as a metadata id, which makes it possible to trace a molecules history from the raw video to the final dataset. The metadata ids can be used to filter out subsets of experiments, if the datasets are discovered to be compromised. In summary, once created, the Molecular Archive structure provided by Mars enables an in depth, reversible, documented framework for single-molecule data exploration.

3.2.2. Molecule Archive architecture

Molecule Archives offer a flexible, standardized framework designed for the storage and handling of image-derived properties. The whole process is adaptable to different experimental configurations. The system encompasses three types of records: *Properties*, *Metadata*, and *Molecule* (Figure 3-2). The types are defined independently of implementation details, which ensures abstraction. This design choice is important to enable the Molecule Archives to accommodate a diverse range of biomolecules, metadata, and property implementations. Additionally, it enables the development of new implementations that integrate effortlessly with the existing framework, algorithms, and user interfaces. To streamline the retrieval of records and the merging of datasets, each *Molecule* and *Metadata* record is assigned a human-readable, base58-encoded UID. Using these unique identifiers simplifies indexing, enhances scalable processing via UID-to-record maps supporting multithreaded tasks, and maintains consistent record traceability throughout analysis workflows.

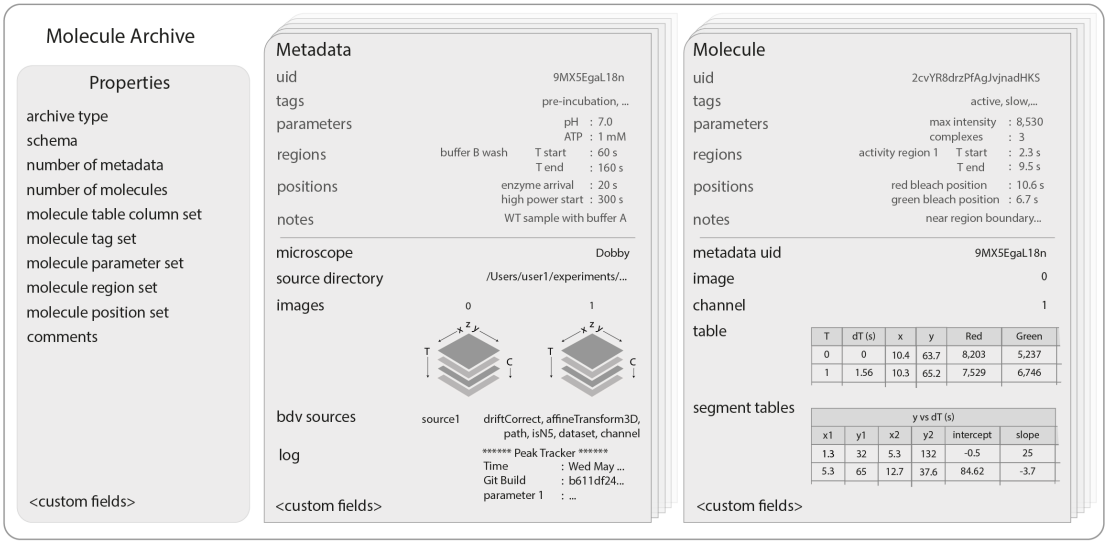


Figure 3-2 Molecule Archive structure.
Schematic representation of Molecule archive structure, which consists of three record types: *Properties*, *Metadata*, and *Molecule*. Global information about the Molecule Archive contents are stored in the single *Properties* record. The *Metadata* records store information about the recorded images from the experiment such as image dimensions and the analysis log. Molecule-specific information is stored in the *Molecule* records such as position over time and intensity. (Reprinted from Huisjes et al. (2022))

Molecule Archives feature a single *Properties* record that details the type and global contents of the Molecule Archive. This record includes comprehensive information

such as the count of *Metadata* and *Molecule* records and the unique names assigned to tags, parameters, regions, positions, and table columns. It houses global comments about the Molecule Archive, typically containing crucial details on the analysis strategies and naming conventions for tags, parameters, regions, positions, and other relevant fields. These comments serve as a guide to orient researchers who may not have been involved in the initial analysis. A user-friendly Markdown editor is included to enhance organization and clarity of these comments.

Metadata records in Molecule Archives capture essential experimental information related to the images used in biomolecule analysis. To maximize compatibility and reusability across a wide array of image formats, they include details such as image dimensionality, timepoints, channels, filters, camera settings, and other microscopy specifics, all stored in the universally recognized OME format (Goldberg et al., 2005).

Molecule Archives do not store the raw images, solely the properties derived from the images and the source directory containing the raw images. This linkage enables fluent data exploration in context of the original files. This functionality is enabled through integration with big data viewer (BDV) technology (Pietzsch et al., 2015) and is compatible with HD5 and N5 formatted images (Stephan Saalfeld, 2022). The *Metadata* records also store image file locations, coordinate transformations, and various settings. Universal time events of an experiment influencing all the molecules can be saved as Metadata positions and regions which are accessible for kinetic scripting and kinetic analysis. A key-value format is used to store global experiential conditions, for example buffer composition and the reaction temperature. To organize and categorize experiments, *Metadata* records can be tagged and filtered by entire experiments using the assigned tags. The whole analysis history, including used commands and settings for processing the data, is stored in a *Metadata* log and also updated if the user continues processing the data.

Molecule records in Molecule Archives are structured to efficiently store common image-derived properties of biomolecules. Each record typically includes a table documenting the position and intensity of biomolecules over time, the uid linking to the *Metadata* record with primary experimental details, and the image index containing the molecule. Regions and timepoints of interest (e.g. burst activity, bleaching step) can be marked for each molecule separately enabling kinetic studies and event-specific calculations. Additionally, the records store global properties of biomolecules,

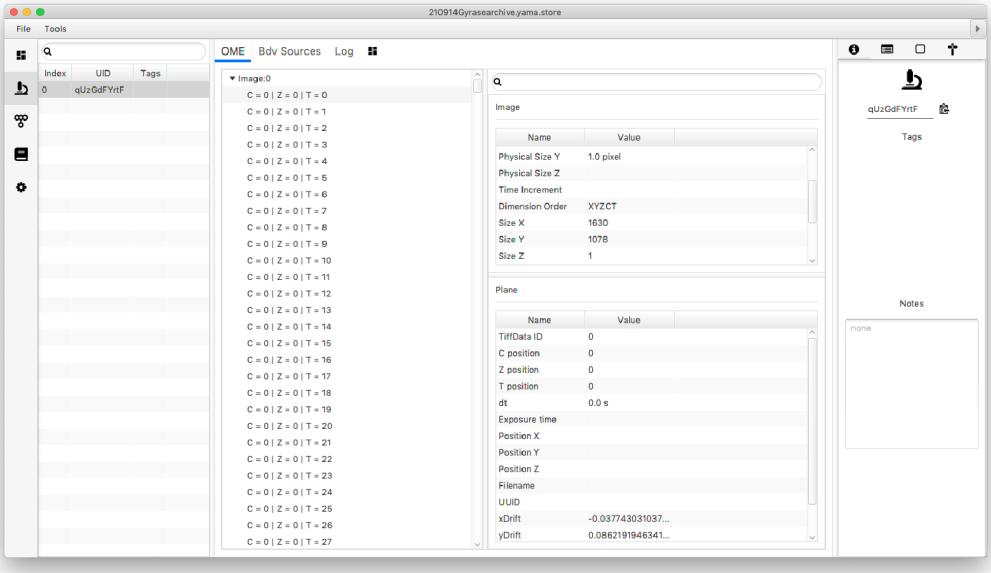
formatted as key-value pairs such as number of pauses, mean intensity, and standard deviation of a coordinate. Mars comes with the kinetic change point (KCP) command which is a tool to identify unbiased linear changes and steps (Hill et al., 2018). This can be used to characterize replication rates, changes in a FRET state, or dye bleaching events. The result of a KCP analysis is stored in separate segment tables. Custom tags and notes are used to categorize each *Molecule* record enabling analysis of only certain molecules.

Molecule Archives are highly versatile and can be created, stored, and accessed in various formats across multiple computing environments, ranging from traditional desktop applications to command-line interfaces suitable for parallel processing on high-performance clusters. Molecule Archives are encoded in JSON format, employing the field schema which is described in figure 3-2. Molecule Archives are compacted into single files with a yama extension. Applying smile encoding and compression of the molecule tables the file size is minimal. However, files can be stored and reopened as plaintext JSON for simplicity and accessibility. To overcome physical memory limitations, Mars can utilize the virtual storage mode which enables processing of very large datasets. In this mode molecule entries are loaded only if demanded by the user, supported by an efficient filesystem-backed record hierarchy. This setup not only conserves memory but also facilitates a robust and adaptable framework for conducting multi-step analysis workflows on large datasets, ensuring high performance and scalability.

3.2.3. Mars Rover: Interactive exploration of molecular features and images

We employ single-molecule techniques to explore new biological phenomena, a process that typically necessitates manual examination of individual biomolecules in raw images and their corresponding image-derived measurements. To streamline this methodology, we developed a JavaFX-based user interface, named Mars Rover. This interface facilitates comprehensive access to data stored within Molecule Archives. Mars Rover has been seamlessly integrated into the Fiji platform, providing accessible windows for all active Molecule Archives. Furthermore, these archives are readily available as both inputs and outputs for commands in Fiji/ImageJ2 and in the scripting languages supported by the platform.

A



B

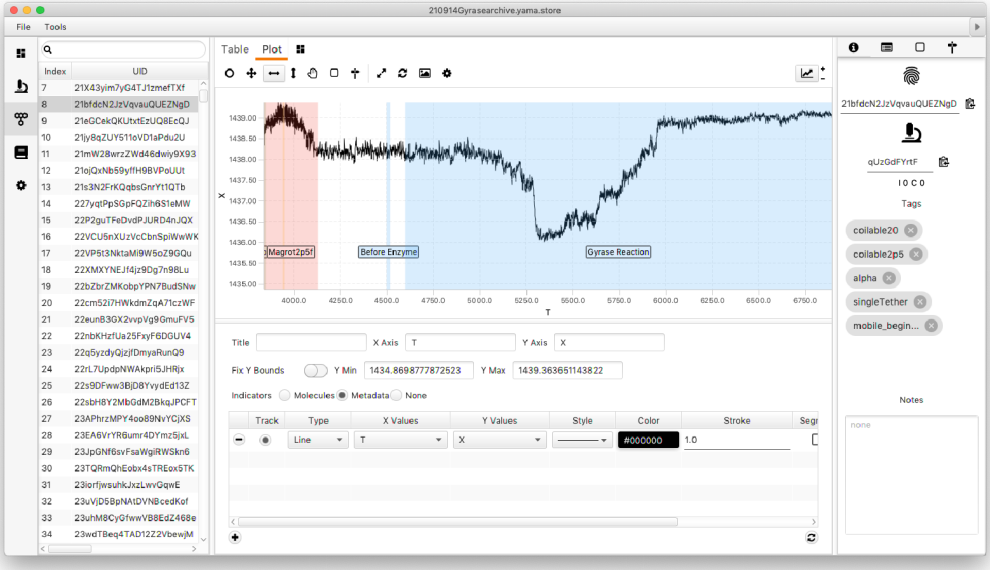


Figure 3-3 Mars rover window.
(**A**) Open microscopy environment (OME) image metadata information is shown in the metadata tab for each analyzed image. Metadata-specific tags are displayed here as well as other record options. The 'BDV Sources' tab is for facilitating Big Data Viewer file connections. Each experiment has its own uid for traceability. (**B**) The Mars Rover molecule tab contains each Molecule stored and organized using uids. User-friendly data exploration tools such as a trace plotter displaying region highlighting, and classification tools such as tags and notes, are highlighted. Each molecule has its own uid connected to the metadata uid to trace molecule history. (Reprinted from Huisjes et al. (2022))

In the Mars Rover interface, windows displaying Molecule Archives feature tabs and subpanels, offering comprehensive access to all aspects of *Molecule* and *Metadata* records, along with global properties, comments, and interface settings (Figure 3-3).

Scriptable chart widgets and detailed information on the Molecule Archive can be created and displayed on a customizable global dashboard tailored to your specific workflow. The *Molecule* tab is equipped with an interactive chart panel that illustrates the time evolution of biomolecule properties, complemented by tools for highlighting specific regions and positions for each trace. Each measurement can be evaluated in context of the original measurement and raw images using BDV integration (Pietzsch et al., 2015).

The Mars Rover was designed to offer broad customization options. All tabs and panels are structured within interfaces, separate from implementation specifics, enabling the extension of Mars Rover to incorporate custom icons and display elements according to the type of Molecule Archive. This design allows for ongoing improvements and addition of new workflows and by building upon the existing core architecture in future updates, while keeping the same comprehensive user experience.

3.2.4. Collection of commands for single-molecule image processing and analysis in Fiji/imageJ2

Mars is equipped with an extensive collection of Fiji/ImageJ2 commands tailored for a variety of common tasks in single-molecule image processing and analysis (Table 5-3). These tools enable users to locate, fit, integrate, and temporally track intensity peaks and objects within images. Additional commands are available for adjusting non-uniform excitation beam profiles and for applying transformations and colocalization filters to regions of interest. Beyond image processing, Mars offers Molecule Archive commands that facilitate opening, merging, and transforming datasets, as well as performing kinetic change point analysis. Mars also supports interoperability with several widely used formats, including importers for TrackMate, single molecule datasets (SMD) (Greenfeld et al., 2015), and LUMICKS h5 files from optical tweezers experiments.

The Mars commands can simply be accessed by installing via the Fiji updated site becoming available in a submenu within the plugins menu. Settings for commands are available in a dialog window. The image processing commands use the active image in Fiji as an input and some even offer a preview option before processing the whole set of images. The tracking and fluorescence integration commands create Molecule Archives which open in a new window upon completion of the command. Using the

previous mentioned Mars Rover, the user can explore all biomolecules in the Molecule Archive window. Progress is saved in two files. The yama extension contains all the primary Molecule Archive data while the yama.rover extension stores the Mars Rover settings allowing for restoring window states and dashboard widgets.

The typical workflow in Mars begins with the creation of a Molecule Archive by processing an image sequence, followed by the analysis and tagging of biomolecule features. Most of the time custom scripts are needed to manipulate the Molecule Archives in order to prepare the data for final publication-quality figures. The Mars workflow is well integrated with Python and Jupyter notebooks. Archives can be opened and analyzed using Python libraries. We created a step-by-step guide showcasing common workflows involving this strategy. They are freely available on our documentation site.

The design of ImageJ2 commands in a UI-agnostic format ensures their versatility across various contexts, regardless of whether the graphical user interface (GUI) is used or not. Commands can be merged into large scripts with multiple stages suitable for creating complex workflows which are executable on high-performance computing clusters. This enables the user to analyze multiple experiments in parallel efficiently, while not using a user interface. This results in the user not having to input every single step via a dialog window.

3.2.5. Mars interoperability using TrackMate as an example

Fiji, a robust open-source platform for scientific image analysis, offers well-established software tools suitable for image processing tasks. The tools can be combined in numerous different ways due to the modular structure of sorting them into plugins. Mars is well integrated with the Fiji workflow, enhancing interoperability on the platform. To showcase the full integration, an example is presented. Result from TrackMate (Tinevez et al., 2017), a Fiji plugin for single-particle tracking consisting of more complex algorithms and a user interface, can be exported to Mars using an action called ‘go to Mars’ (source code available at https://github.com/duderstadt-lab/mars-trackmate). This was added to the TrackMate panel to directly convert it to a Molecule Archive. It requires no additional steps to set it up. TrackMate complements the Mars workflow with a more complex tracking strategy, while the build-in Mars Peak Tracker, with its straightforward interface, offers a simple single-molecule tracking option for routinely tracking tasks. For example, offering the possibility to track object

shapes and positions using machine learning algorithms (Ershov et al., 2022) and facilitating the exported results in a Mars Object Archive.

3.2.6. Application of Mars

To analyze the data for this thesis, three distinct workflows were developed using Mars, demonstrating its potential use cases. These workflows were created to fill the gap in analysis tools for complex datasets. First, Mars's shape-tracking capability was utilized to detect nicks in supercoiled DNA, with results used to calibrate the imaging conditions of the mmTIRF microscope. Second, Mars was employed to track DNA replication over time, which was crucial for extracting replication burst rates and processivity. Third, the interoperability and flexibility of Mars enabled the development of a tool to import segmentation results from LabKit into the Molecule Archive framework. Detailed descriptions of these workflows are provided in the methods chapter of this thesis.

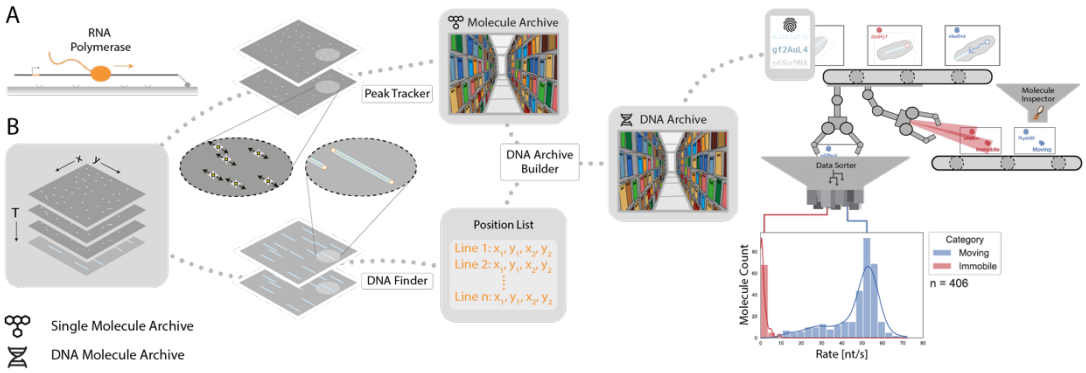


Figure 3-4 Workflow example: Tracking of RNA polymerase during transcription.
(**A**) Schematic representation of the RNA polymerase assay. Surface-immobilized 21 kb DNA containing a promoter sequence was incubated with fluorescently-labeled RNA polymerase. The transcription process was tracked over time using fluorescent signal. (**B**) Representation of the analysis pathway using Mars tools. Starting from the left, the 'Peak Tracker' found positions of fluorescent spots and tracked them over time creating a Molecule Archive. In parallel, the DNA position was extracted using the 'DNA Finder' appearing as a long, line-shaped object. The coordinates of DNA were used to create a DNA Archive transforming the polymerase positions on the reference frame of the DNA. The traces were sorted and classified to create the final plot showing various transcription rates (nt/s). A gaussian fit revealed a population average of 53±3.6 nt/s with n being the number of molecules. (Reprinted from Huisjes et al. (2022))

The experimental work was carried out on a mmTIRF microscope, a popular technique playing an important role for scientific discovery. It is utilized for studies of bio-macromolecular structures and their functionality. For example, it was used to study DNA replication (Dequeker et al., 2022; Ha et al., 2002; Lewis et al., 2020), rotational movement of the flagellum motor (Sowa et al., 2005), polymerization of actin filaments (Amann & Pollard, 2001), and processes in vivo (Vizcay-Barrena et al., 2011).

One workflow showing how to analyze such datasets using Mars is illustrated in the following mmTIRF study of the kinetics of a fluorescently-labeled RNA polymerase transcribing on an immobilized, promoter-containing, 21 kb DNA molecule (Scherr et al., 2022) (Figure 3-4A). In the presence of all four nucleotides, RNA polymerase successfully initiated transcription from the promoter and progressed along the DNA, which was visualized in real-time by measuring fluorescent emission upon excitation. Upon completion of transcription, the DNA was post-stained with SYTOX Orange to reveal the positions of the DNA molecules in the final frames of the video. By correlating the movement of RNA polymerase with the position of the DNA molecule, valuable insights into the processivity and progression rates of the polymerases were obtained.

To quantitatively analyze the data, a beam profile correction was initially applied to eliminate the non-uniform laser excitation in the field of view, caused by the Gaussian beam profile of the lasers used in TIRF microscopy. The ‘Peak Tracker’ (Figure 3-4B) then identified and recorded the location of each fluorescent spot across the progressing frames. The tracking results were stored in a Single Molecule Archive. Subsequently, the identified positions had to be corrected for sample drift occurring during the measurement using the ‘Drift Corrector’. The final frames of the video, showing the DNA molecules, were processed with the ‘DNA Finder’ to detect and measure the coordinates of all DNA molecule positions. The polymerases were transformed in the reference frame of the DNA to create a DNA Molecule Archive. The resulting DNA Molecule Archive contained the positional data of all polymerase molecules associated with the DNA. This information is the basis of the kinetic studies. To create sub-populations containing groups of molecules sharing kinetic features, molecule tags can be assigned for filtering and sorting. Utilizing the described feature, the analysis revealed a population-specific rate distribution (Figure 3-4B) with an observed transcription rate of 53 ± 3.6 nt/s. The result aligns well with previous studies,

where the reported transcription rates were between 40 and 80 nt/s (Thomen et al., 2008).

Additional examples with step-by-step instruction can be found on our website which were created to highlight other features of Mars such as handling very large datasets or single-molecule FRET.

3.3. Discussion

Day by day the bioimaging technologies are rapidly advancing catalyzed by innovative methods. Extensive, information-rich datasets are generated by these methods monitoring the time evolution of complex biological systems with deep spatial and temporal resolution. In order to fully leverage these technological breakthroughs and to ensure accurate quantification, it is crucial that these observations are efficiently, reproducibly analyzed, classified, and shared. Because of the information-rich and powerful nature of single-molecule imaging techniques, they managed to transition from niche applications to become central building block in molecular biology research, enabling to see phenomenon hidden in other applications. Unfortunately, reporting standards and common formats for storing are lacking slowing down exchange of information about the data and the analysis. Common file formats, for example for single-molecule FRET (Greenfeld et al., 2015), capable of storing raw photon data from point detectors or time-binned trajectories from images (Ingargiola et al., 2016), are often restrictive and not well-suited for adapting to other experimental approaches. Mars addresses this limitation by offering an extensive list of single-molecule image processing commands, an user-friendly graphical interface for data explorations, and a flexible and versatile Molecule Archive file format. The format ensures flexible storage and reuse of image-derived data across a wide array of experimental setups and assay designs. Mars is available as a collection of SciJava commands via the Fiji update site, which makes it available to a large community. Furthermore, Mars is developed as an open-source project with the source code being available on GitHub. In addition, a documentation website was created illustrating the basic functions with tutorials and example workflows. As a part of the Scientific Community Image Forum (Rueden et al., 2019), Mars benefits from a dynamic community platform that supports new users and provides advanced users with resources to tackle complex image analysis challenges.

The workflows presented demonstrate the versatility of basic collection of modular commands and a Molecule Archive structure, which can be applied across different experimental set-ups. However, by creating a general applicable and flexible workflow, the commands have limitations and may not address every challenge to a full extend. That is the reason why interoperability between Mars and other platforms available in Fiji is enabled. The TrackMate integration gives the option for a broad spectrum of particle tracking capabilities within the Molecular Archive format by exporting the tracking results seamlessly. Furthermore, Mars also integrates the SCIFIO framework (Hiner et al., 2016) to convert various image formats into OME format and includes a dedicated image reader supporting Micromanager (Edelstein et al., 2010), which is very commonly used in the single-molecule imaging community. This integration ensures that Mars remains adaptable to new image formats as they are developed and as corresponding SCIFIO or Bio-Formats readers become available (Linkert et al., 2010). The way Mars is structured, it has several built-in extension mechanisms to create new solutions if the integrations fall short. As shown for the transverse flow datasets (Chapter 5), it is possible to create custom Molecule Archive types and user interface elements to facilitate data exploration and tailor-made solutions. The advantage is not being limited to a fixed set of data points for each molecule but being able to adapt accordingly. Designed with scripting and command development in mind, Mars facilitates the analysis of data from emerging techniques beyond traditional single-molecule applications by expanding its existing framework. Furthermore, as illustrated in the example, Molecule Archives can be directly accessed in Python environments. Python enables the user to create finalized plots for scientific publishing, offering data manipulation and visualization libraries. The Molecule Archives can be loaded as JSON files or by using PyImageJ (Curtis Rueden, 2021). The direct interaction between the Python environment and the Molecule Archives gives the option of accessing single molecules using the UIDs or the entirety of the archive.

Single-molecule imaging techniques have become widely adopted and are now essential for uncovering new biological phenomena. Despite their popularity, there is a notable gap in data reporting standards. Detailed records of processing histories and the associated raw images, which are crucial for reproducing the analysis process, are sporadically provided. The Molecule Archive format presented in this chapter aims to simplify the process for researchers to accurately report their findings and enhance reproducibility. Information-rich datasets are more accessible, enabling reuse and

reevaluation. Moreover, this development is expected to boost confidence and the quantitative accuracy of results. However, using the Mars framework supports reproducibility but it is not guaranteed. The responsibility for reporting findings in a reproducible manner, adhering good scientific practice recommended for single-molecule imaging experiments, lays with the individual researcher (Lerner et al., 2021). For example, scripts used during the analysis process in Mars should be version controlled, published publicly, and all necessary parameters need to be reported in the log. To guarantee reproducibility, external software used in the workflow needs to be reported including the version number and the used settings.

Most discoveries of new biological phenomena through single-molecule observations involve time-consuming manual classification of individual molecules and characteristics. In addition, when manually classifying molecules, user bias could be added unintentionally. Recent advancements in the field of machine learning could help to automatically cluster and classify datasets (Kapadia et al., 2021; Thomsen et al., 2020). However, robust training datasets are needed to create accurate algorithms. Mars offers a great starting point for creating such datasets. The tagging system of Mars connects the feature with a time series which is ideal for machine learning based classifiers. Since very large datasets can be stored in virtual storage, preparation of training datasets and handling them is simplified. The direct interactions between Python and Molecule Archives ensure access to complex and extensive machine learning libraries such as Tensorflow and PyTorch (Ansel et al., 2024; Martín Abadi et al., 2015).

Mars offers a complete workflow solution to analyze complex datasets while being able to adapt to new challenges. Furthermore, working in conjunction with other platforms offering specialized solutions. Future work will build on the framework supporting scientific progress by aiding reproducibility of analysis steps connected to the original image content. Possessing the ability to speed up the process by creating unbiased pipelines using machine learning.

Chapter 4 – Development of a single-molecule platform for topology studies

To conduct the topology studies on the mmTIRF setup presented in chapter 5, extensive method development was necessary. A suitable experimental design is as important as the biological findings, since poor design can lead to artifacts and inaccurate interpretations. To ensure a robust experimental framework, control experiments must be performed with appropriate DNA substrates. The substrates for microscopy must be carefully designed because making and purifying each one involves multiple steps of complex biochemistry, which takes many days up to a week. For example, preparing the forked DNA substrates presented in this thesis required seven days. Furthermore, information from the literature can be used to predict possible outcomes of the experiments, which helps in understanding the timescales on which the proteins act, as well as potential limitations in observations. This process aids in formulating alternative strategies to overcome these limitations and lays the groundwork for further expanding the assay.

This chapter provides a detailed discussion of the key conceptual features of the modular DNA toolkit used in this study. The primary objective of creating a substrate for topology studies is explained, with outcomes for replication contextualized against existing knowledge, supported by calculations and simple simulations that reinforce the claim that an additional pathway is necessary to resolve topological strain. Basic concepts of the single-molecule assay utilized throughout Chapter 5 are also explained. Additionally, an assay specifically developed to measure the extent of DNA nicking under imaging conditions is introduced and discussed. The findings from this assay were used to establish a threshold to reduce the number of false-positive molecules and prevent artifacts, as nicks would lead to relaxed DNA molecules.

4.1. Introduction

Single-molecule techniques offer the possibility to study biomolecular behavior that would otherwise be obscured by the ensemble average of a large number of molecules. In many experiments, the key element to study interactions and activities, the key element is the DNA molecule, which can serve various functions. For instance, DNA can act as a template in replication (Duderstadt et al., 2016; Lee et al., 2006;

Lewis, Spenkelink, Schauer, et al., 2017; Lewis et al., 2020; Loparo et al., 2011) and transcription (Janissen et al., 2024; Kosuri et al., 2019; Scherr et al., 2022), a track for demonstrating loop extrusion by SMC proteins (Ganji et al., 2018; Pradhan et al., 2023), or as a single polymer chain to probe physical properties (Ganji et al., 2016; Perkins et al., 1994; Strick et al., 1996; Strick et al., 1998). A longer DNA molecule is particularly advantageous for studying these interactions, as it provides a flexible, micrometer-long tether, allowing for increased resolution of interactions when observed under a microscope.

In many cases, λ phage DNA has been used as the basis for creating substrates due to its 48.5 kb length and commercial availability. However, the standard λ phage DNA available is not always suitable for single-molecule experiments (Mueller et al., 2020), especially when studying DNA topology, because a single nick in the DNA can obscure the results. Unlike in magnetic tweezers or optical traps experiments where the topological integrity of the substrate can be verified by performing a coiling check to observe the buckling transition (Agarwal & Duderstadt, 2020; Forth et al., 2008; Strick et al., 1996; Strick et al., 1998), such checks are not possible with a standard fluorescent microscope, which lacks direct topological control. Plasmid DNA offers a better starting point for creating such substrates. Plasmids can be easily produced, stored, and modified using commercial enzymes and biochemical techniques.

The rolling circle approach, which has yielded a vast collection of replication studies at the single-molecule level, uses a plasmid-based substrate (Geertsema et al., 2014; Kurth et al., 2013; Lewis, Spenkelink, Jergic, et al., 2017; Spenkelink et al., 2019; Tanner et al., 2009). The basis of the rolling circle amplification is a circular plasmid containing a fork structure, where the replication machinery can load. The process can potentially continue indefinitely depending on the processivity and the experimental conditions (Monachino et al., 2018). To adapt this approach to single-molecule studies, one biotin is introduced on the lagging strand arm and the lagging strand product of the replication process can then be stretched out by flow (Tanner et al., 2009). Unfortunately, introducing a topological challenge into the rolling circle substrate is challenging (Kuhn et al., 2002). By design, the leading and parental strands are connected on the circular plasmid, thereby canceling out the linking number changes ahead and behind the replication fork due to the opposite signs of the change. This limitation makes the rolling circle strategy unsuitable for this research aim.

To address the challenges associated with using single-molecule microscopy to understand how the replisome deals with topological challenges, we developed a modular DNA toolkit with interchangeable end pieces to create different topological challenges for the replisome. While other modular substrate designs have previously been used in past single-molecule studies of DNA replication, they lacked topological challenges for the replisome (Lewis et al., 2020; Mueller et al., 2020). To date, no published strategy exists for creating a DNA substrate for replication studies that introduces topological challenges on a fluorescent microscope. Therefore, we established an adaptable workflow to create simple, modular long DNA substrates for TIRF studies, allowing us to introduce topological challenges to the replisome at different parts of the replication fork using a plasmid-based system.

4.2. Results

4.2.1. Replisome progression and gyrase interaction: Simulating topological challenges

Before describing how we created substrates to study topological challenges, it is important to consider whether mechanisms beyond topoisomerase interactions are required to resolve topological strain. As discussed in the introduction, the process of introducing and resolving positive DNA supercoiling is a tug of war between the replisome and topoisomerases. To better understand the balance between these forces, we developed a simple simulation that incorporates experimental findings. This section will first outline the main assumptions of the simulation, followed by an explanation of its mechanics, and finally, the presentation of the results

We made the following assumptions (Figure 4.2A): (1) The replisome has a rate of 500 base pairs per second, based on the accepted rate of the *E. coli* replisome of 500 to 1,000 base pairs per second (Breier et al., 2005; Nelson & Cox, 2021; Reyes-Lamothe et al., 2008). These rates are averages across the entire replication cycle. (2) The *E. coli* chromosome, approximately 4.6 Mb in size, is organized into topological domains that vary in size and positioning. For our calculations, we assumed a topological domain size of 10 kb, which is reported as the average size. Within each of these domains, the topological state of the DNA is contained (Postow et al., 2004). (3) As discussed in the introduction (section 1.1.3), DNA typically has a superhelical density σ between -0.05 and -0.07, which aids in initiating processes on DNA and in

compacting the large chromosome within the cell (Fulcrand et al., 2016; Nelson & Cox, 2021). Therefore, we assumed an initial superhelical density, σ_{start}, of -0.07. (4) We also needed to define a stalling superhelical density σ_{stall}. We will calculate this in detail in section 4.2.3. We assume the stalling superhelical density σ_{stall} to be 0.10. (5) While the replisome introduces positive twist, the topoisomerase DNA gyrase relaxes positive twist and introduces negative twist at rates of approximately 1 cycle per second and 0.5 cycles per second, respectively (Agarwal & Duderstadt, 2020; Nollmann, Stone, et al., 2007). One cycle of gyrase changes the linking number ΔLk by 2. For simplicity, we assumed the rate for positive relaxation to be 2 twist per second per gyrase molecule. (6) Gyrase wraps around 128 to 140 base pairs of DNA (Baker et al., 2011; Gore et al., 2006). We assume a DNA footprint of 140 bp. (7) Single-molecule fluorescent imaging revealed that an average of 12 gyrase molecules are closely associated with the replisome (Stracy et al., 2019).

The goal of the simulation was to calculate the current superhelical density σ (Equation 1.12) within a single topological domain, which defines the spatial constraints (Figure 4.2B). At each timestep, σ was updated and compared to the stalling superhelical density, σ_{stall}. The replisome introduced positive twist based on the replication rate divided by 10.5 (+), while gyrase relaxed positive twist (-) according to the number of gyrase molecules bound to the DNA multiplied by the relaxation rate. The number of gyrase molecules was modeled using a gamma distribution based on the findings of Stracy et al. (2019). This number was adjusted if there were spatial limitations as the replisome approached the end of the topological domain, as reduced space would limit the binding options for gyrase molecules.

We ran the first simulation with a distribution of gyrase molecules centered around a mean of 12. In the initial phase (up to 10 seconds), the σ increased (Figure 4.2C). After 10 seconds, the σ reached its maximum threshold of σ_{stall}, leading to the first pausing event of the replisome, indicated by a lack of change in the 'Position on DNA' and a replication rate of zero. Once gyrase resolved the twist, the replisome resumed replication. However, additional pausing events occurred before the replisome reached the end of the topological domain. In Figure 4.2D, the number of gyrase molecules is plotted alongside the mean number of gyrase molecules and the maximum binding sites available at any given time. Due to the gamma function's randomization of gyrase molecules, there were moments when more gyrase molecules could act on the DNA than the average number of the distribution. Beginning around

20 seconds, fewer binding sites were available than the average needed, leading to a smaller reduction in superhelical density σ between 20 and 30 seconds as twist accumulated.

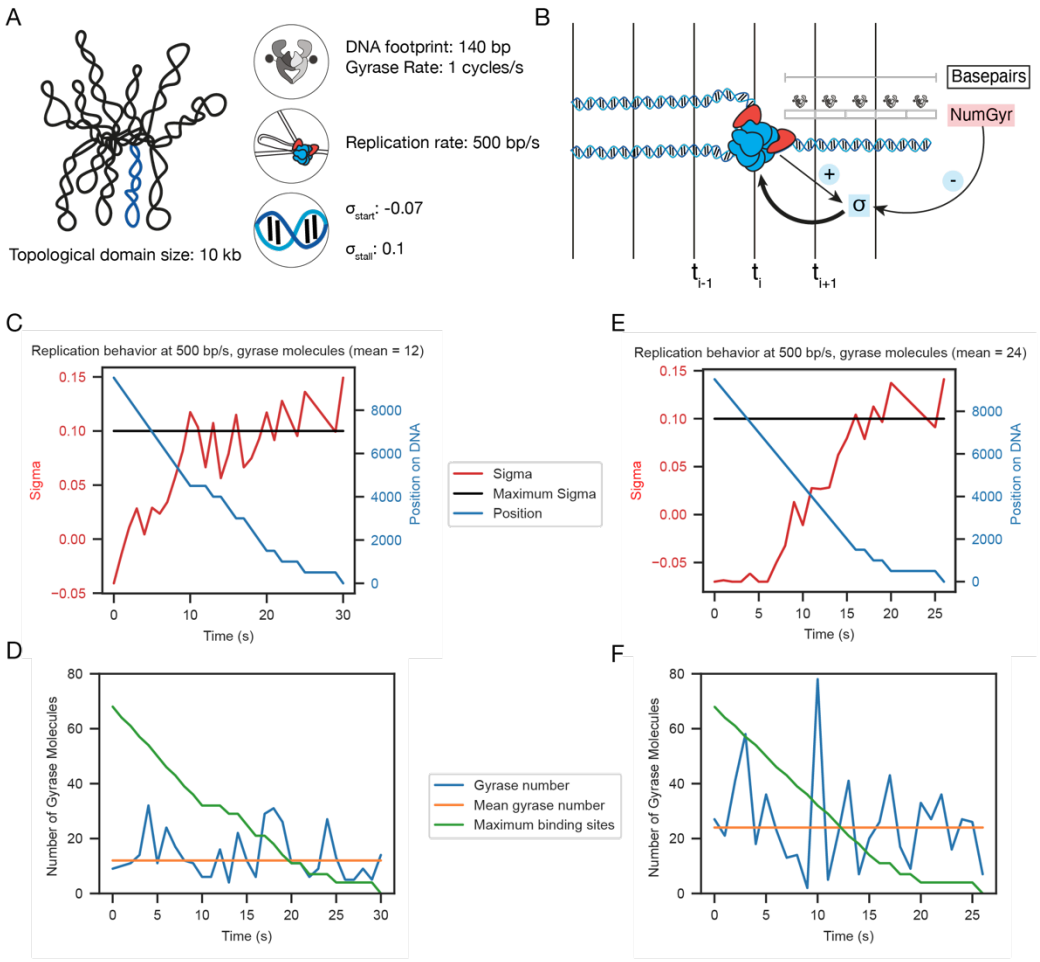


Figure 4-1 Replisome progression and gyrase interaction simulations.
(**A**) Assumptions made for the simulations. The displayed values were used to run the simulation. (**B**) Visual depiction of the simulation. The amount of DNA ahead of the replisome is checked for binding sites for DNA gyrase. The number of gyrase molecules (NumGyr) reduce the current superhelical density (σ) by the product of their count times their rate. The replisome adds to σ with each time step (t). The current σ is then compared to the maximum superhelical density (σ_{stall}), which can lead to pausing of the replication process. (**C**) Result of the simulation using a distribution with an average of 12 gyrase molecules (represents the situation inside the cell). The plot shows the σ value (red), the stalling σ_{stall} (black), and the position of the replisome (blue). Flat lines in the position indicate a pause event. (**D**) The number of gyrase molecules over time for the simulation in (C). The actual number of gyrase molecules at a given time point (blue), the mean number of gyrase molecules according to the input (orange), and the maximum binding sites/maximum number of gyrase molecules (green) are plotted. (**E**) Result of the simulation using a distribution with a mean of 24 gyrase molecules. (**F**) The number of gyrase molecules over time for the simulation in (E).

Could the issue of accumulating twist and frequent pausing be mitigated by increasing the number of gyrase molecules? To test this, we doubled the mean value of the distribution to 24 gyrase molecules and ran the same simulation scenario (Figure 4.2E). The maximum superhelical density that gyrase could maintain was set to the initial value, reflecting the in vivo condition where gyrase keeps supercoiling at this level. At the start of the replication process, σ was maintained at a high level close to the starting value σ_{start}. However, as replication progressed, σ shifted to positive values. Around 16 seconds, the first pause occurred, followed by two more pause events before reaching the end of the topological domain. When we analyzed the number of gyrase molecules over time, we observed that after 12 seconds, the number of available binding sites was fewer than the mean number of gyrase molecules (Figure 4.2F). The loss of the advantage conferred by additional gyrase molecules resulted in a situation similar to having fewer molecules overall, leading to further pausing events. This simple thought experiment demonstrates that even with topoisomerases present, overcoming topological strain is more complex than it might initially appear and highlights the dynamic nature of the process. Furthermore, the role of the replisome and its mechanisms for tolerating strain were not fully accounted for. The remaining part of this chapter will outline the basic concept of the assay developed to explore these mechanisms further.

4.2.2. Design of a modular DNA toolkit for topology studies

All substrate types were constructed using the same plasmid backbone (~27 kb), derived from the SuperCos1 cosmid vector, which was modified to include multiple cloning sites for unique restriction enzymes (Scherr et al., 2022). This design allowed for easy production of the plasmid in *E. coli* DH5α cells and subsequent isolation for further use. The plasmid was digested with unique restriction enzymes to create sticky ends (XbaI and XhoI), facilitating the ligation of functional elements. These functional elements included an end pieces with a single biotin, an end piece multiple biotin, a primed synthetic replication fork and a constrained synthetic fork (Figure 4-1A).

A single biotin end was created by annealing two DNA oligos, one of which was modified with a biotin. The surface of the microscope slide was functionalized with streptavidin, allowing it to interact with the biotin on the DNA molecule. Annealing the oligos produced a sticky end that was designed to match the overhang on the digested plasmid backbone. To create topologically constrained DNA, DNA handles were

generated in a separate PCR reaction using nucleotides modified with a biotin. Multiple biotin interactions with the surface prevents the DNA molecule from swiveling around the bond, effectively locking it into a fixed configuration. Similar DNA handles have been employed in magnetic tweezers and optical trap experiments (Agarwal & Duderstadt, 2020; Bryant et al., 2003; Strick et al., 1998). The created handles had sticky ends with the compatible with NotI and XhoI restriction sites.

The synthetic replication fork with a single biotin was created by annealing oligos matching either the XbaI or XhoI sites. The primer used for this process was biotin-modified, allowing it to attach to the streptavidin-functionalized surface of the slide. By incorporating the biotin modification on the primer, the DNA was double-tethered on the surface of the slide, with the primer positioned to initiate replication, indicating a functional assembly of the substrate. The other arm of the fork, representing the future lagging strand, was where the T7 helicase would load. To create the synthetic fork with multiple biotin handles, the fork was assembled by annealing oligos, followed by an additional ligation step to attach a multi-biotin handle matching the NotI site.

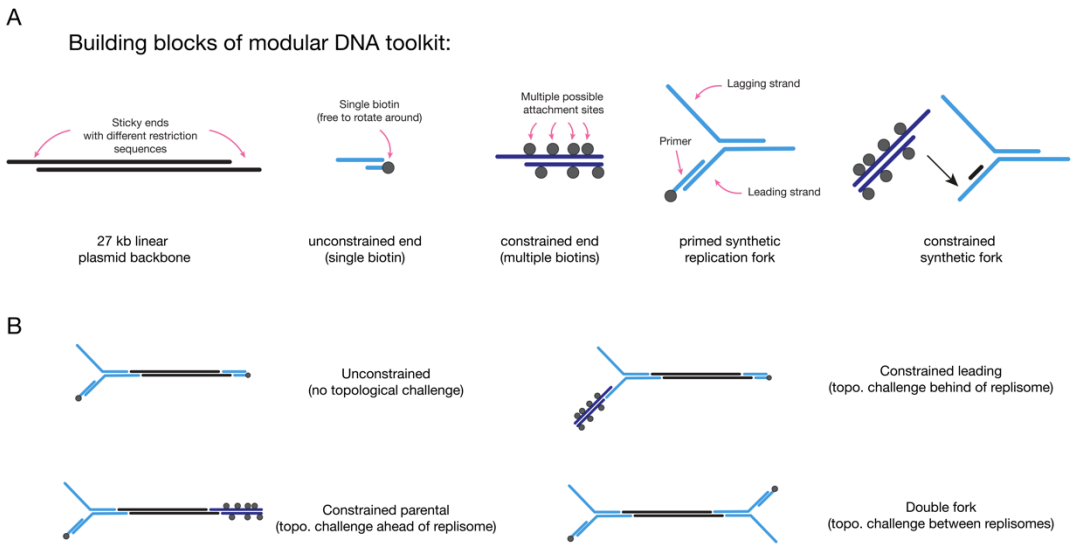


Figure 4-2 DNA modular toolkit.
(**A**) Schematic representation of the different building blocks in the modular DNA toolkit used to create substrates for single-molecule replication assays that address topological challenges. Key features of the building blocks are highlighted. (**B**) Potential combinations of building blocks to mimic various topological challenges. Depending on the number and placement of biotin molecules, the substrates were designed to be either topologically unconstrained, constrained behind the replication fork, constrained ahead of the replication fork, or to create topological stress between two converging forks.

Utilizing matching sticky ends allowed for various configurations of the DNA substrates (Figure 4-1B). When combining a fork with a single biotin and an end piece with a single biotin, an unconstrained substrate was created, which was later used to study replication without the influence of topological stress. To mimic a replication scenario where topoisomerase interaction is abolished in the parental part of the fork, a multi-biotin handle was introduced ahead of the replication fork, constraining the parental strand and creating topological stress ahead of the replisome. Introducing a multi-biotin handle at the leading strand part of the replication fork induced topological stress behind the replisome, simulating a scenario where topoisomerase activity is abolished behind the replisome. Lastly, by combining two forks, the interaction of two approaching replisomes could be studied, allowing for the investigation of whether topological stress accumulates between the forks, mimicking the termination phase of the replication cycle.

4.2.3. Estimation of replisome progression on DNA substrates

Considering that the replisome can tolerate a certain extent of torsional strain, it is important to estimate how far it would replicate on the DNA substrate before stalling, assuming the DNA behaves like a closed circular plasmid. If the movements cannot be resolved with the proposed readout methods, the assay design may need adjustments. The superhelical density σ serves as a useful measure to compare the amount of supercoiling between different systems (Equation 1.12). To calculate the superhelical density σ, the linking number of the relaxed state (Lk_0) is subtracted from the current linking number Lk and then divided by Lk_0. To determine Lk_0, the number of base pairs in the DNA is divided by 10.5, assuming it is in the B-DNA form (Nelson & Cox, 2021).

Previously published DNA data provided an estimate for σ_{stall}. McGlynn, Lloyd, and Marians demonstrated that replication could be stalled on a circular plasmid in a bulk assay by allowing torsional strain to build up, which was then used to synchronize the replisomes upon release (McGlynn et al., 2001). The plasmid used in these reactions was initially negatively supercoiled, allowing replication to initiate without the need for topoisomerases. The replisome stalled after synthesizing 1 kb of product on a plasmid with a total size of 6 kb. The Lk_0 of the plasmid is 571. The initial superhelical density σ was between -0.07 and -0.05 based on literature values (Fulcrand et al., 2016; Nelson & Cox, 2021). Table 4-1 presents the superhelical density and corresponding linking

number for the plasmid. During replication of the 1 kb, the linking number increases by +95, which is calculated by dividing the number of replicated base pairs by 10.5 bp. This change in linking number is added to the starting linking number Lk_{start} to determine the linking number when the replisome stalls, and the superhelical density is then calculated.

Superhelical Density (Start)	Linking Number (Start)	Linking Number (Stalling)	Superhelical Density (Stalling)
-0.07	531	626	0.10
-0.05	542	637	0.12

Table 4-1 Comparison of start and stalling superhelical density

Using the information on the superhelical density σ_{stall} at which the replisome stalls, we estimated how far the replisome might replicate in the single-molecule assay. The substrate used is approximately 27 kb in length, with a superhelical density of 0, as it is assumed to be fully relaxed before the replication reaction. The Lk_0 for this substrate is 2,571. Based on this, we can estimate the number of base pairs replicated before stalling. If the stalling density is assumed to be 0.10, the replisome would stall after replicating approximately 2,880 bp. If the density is 0.12, stalling would occur after 2,828 bp. Assuming fully stretched B-DNA, where one base pair corresponds to a length change of 0.34 nm, and given that our setup has a pixel size of 156 nm (representing 459 bp), the replisome could move about 6 pixels before stalling, covering approximately 10% of the substrate. To conclude, if topological strain merely led to complete stalling of the replisome, it would still be detectable with the current microscope setup.

4.2.4. Length distribution of DNA constructs

To validate the homogeneity and quality of the substrates, they were visualized at the single-molecule level using mmTIRF (Figure 4-3A). The biotin-labeled DNA molecules were attached to a streptavidin-coated microscopy slide through the flow channel created by the PDMS. Chloroquine was present during application of DNA to the slide to ensure that the DNA remained stretched. The individual DNA molecules were visualized by introducing SYTOX Orange, a fluorescent dye specific to dsDNA. The flow channel was washed with a buffer containing the dye, and the buffer flow was then stopped to differentiate between singly attached and doubly attached molecules.

Singly attached molecules appeared as DNA blobs at the attachment site, while doubly attached molecules formed linear DNA strands stretched between two attachment sites (Figure 4-3B).

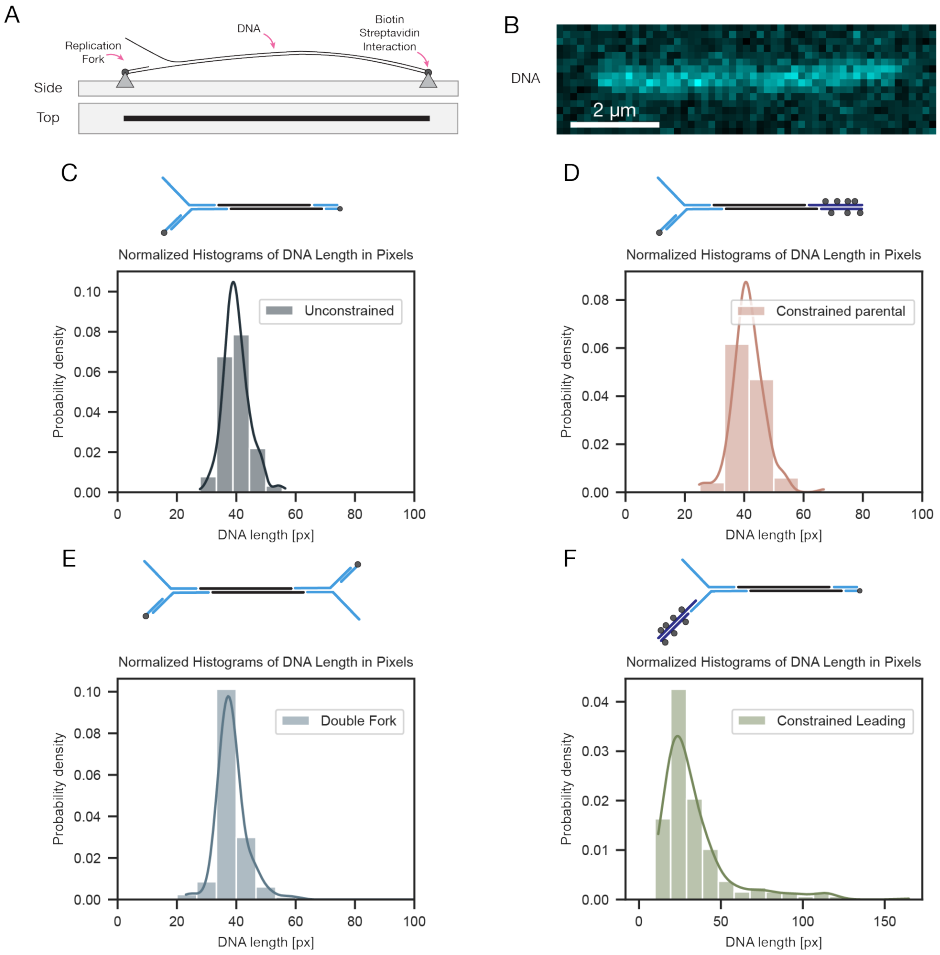


Figure 4-3 Length distribution of DNA substrates.
(**A**) Schematic of the linear attachment of replication substrates. (**B**) Example molecule on the microscope. (**C**) Length distribution for the unconstrained substrate (n = 346). (**D**) Length distribution for the constrained parental substrate (n = 179). (**E**) Length distribution for the double fork substrate (n = 247). (**F**) Length distribution for the constrained leading strand substrate (n = 335). Kernel density estimation was plotted on top for all four distributions. The value n describes the number of individual DNA molecules.

The lengths of individual DNA molecules were measured for the different substrates to assess uniformity. On average, the length of the unconstrained substrate was 6.3 ± 0.7 µm (Mean±SD) (Figure 4-3C), corresponding to 68 ± 7% of the crystallographic length (9.2 µm). Adding the biotin handle to the constrained parental substrate was expected to increase the length by 0.2 µm, which was reflected in the measured mean

length of 6.5 ± 0.8 µm (Figure 4-3D), or 69 ± 9% of the crystallographic length (9.4 µm). The double fork substrate had an average length of 6.0 ± 1.1 µm (Figure 4-3E), or 65 ± 12% of the crystallographic length. In a similar study, reported linear DNA molecules at 98% of the crystallographic length under comparable conditions (Mueller et al., 2020). The DNA substrate used in that study was approximately 18 kb, 9 kb shorter than the DNA used here, which might explain the observed differences in stretching. Since all three substrates exhibited similar stretching with minimal variation around the mean length, it can be assumed that they were assembled correctly.

In contrast, the constrained leading substrate exhibited a shorter average length (5.2 ± 3.4 µm) with significant variability (Figure 4-3F). Some DNA molecules were as long as 2.7 times the crystallographic length, which suggests that multiple DNA molecules may have combined to form longer strands. This indicates that the current strategy for creating the constrained leading substrate is insufficient and requires improvement.

Substrate	Crystallographic length [µm]	Length: Mean±SD [µm]
Unconstrained (n = 346)	9.2	6.3±0.7
Constrained Parental (n = 179)	9.4	6.5±0.8
Double Fork (n = 247)	9.2	6.0±1.1
Constrained Leading (n = 335)	9.4	5.2±3.4

Table 4-2 Crystallographic and mean length of DNA substrates

4.2.5. Validation of creation of topologically constrained DNA

To validate the proposed strategy of using multiple biotin interactions on the microscope slide to produce topologically constrained DNA and create a topological challenge, we prepared a linear substrate (21 kb) with the same handles used for the forked substrates (matching overhangs for NotI and XhoI). The substrate preparation followed the same steps as those used for the forked substrates, utilizing the same microscope setup, including slides, flow cells, and buffers. The transverse flow cell was employed for the experiments. The protocol for preparing supercoiled DNA was adapted from Ganji et al. (2016).

To induce negative plectonemes, linear DNA with multiple biotin handles on both ends was introduced into the transverse flow cell while using a higher concentration of SYTOX Orange (500 nM) (Figure 4-4A). SYTOX Orange is a monomeric intercalating dsDNA stain commonly used to stain DNA and modify its topological state (Ganji et

al., 2016; Ganji et al., 2018; Kim et al., 2018; Kolbeck et al., 2024; Lewis et al., 2020). The DNA molecules attached linearly to the slide. After immobilization, unbound molecules were washed out, and the SYTOX Orange concentration was reduced to 150 nM. The intercalator initially underwound the DNA, making it stiff, and reducing the dye concentration transformed the twist into writhe (Gudnason et al., 2007) since the substrate was torsionally constrained by multiple linkages between the biotins on the DNA and streptavidin on the slide surface (Figure 4-4A,B). Following the washing step, the flow direction was switched to transverse flow, creating arched molecules with more exposed plectonemes (Figure 4-4B). Over time, photo-induced damage introduced nicks, indicated by a shape change and the disappearance of the plectoneme (Figure 4-4B,C). This feature could be used to identify properly formed molecules. To assess DNA quality and purification efficiency, the number of constrained molecules was counted and divided by the total number of double-tethered molecules, yielding a percentage of 68.2 ± 10.6% (Mean ± SD) unnicked molecules across eight different experiments. Compared to other purification protocols for creating substrates for force spectroscopy, the percentage of coilable molecules ranges from 35% to 80% (Bell & Molloy, 2022; Paik et al., 2013; Papini et al., 2019; Seol & Neuman, 2011) (Table 4-3). The substrate created in this study was longer than those in other studies, increasing the likelihood of DNA breaks. In summary, we demonstrated the effectiveness of our preparation process in creating nick-free DNA. The resulting number of constrained molecules was in the upper range compared to other studies, even though the longer DNA length would typically increase the likelihood of accumulating nicks.

Study	Papini et al.	Bell & Molloy	Seol & Neuman	This study	Paik et al.
Substrate length	1.4 kb	10.1 kb	6 kb	~21 kb	~7.4 kb
Coilable molecules	35%	40%	60%	68%	80%

Table 4-3 Comparison of constrained substrate preparation

4.2.6. Single nick detection assay for assessing DNA integrity during imaging

As reported in section 4.2.5, DNA molecules change shape once a nick is introduced. Building on this observation, we developed an assay with single-nick resolution to estimate the amount of nicking over time. We used the results from the experiments to establish a reaction window, where molecules considered for analysis were within

the 95% confidence interval. Molecules within this interval were deemed undamaged by the imaging conditions, meaning no nick was introduced. Nicks in the DNA can release topological tension, potentially leading to artifacts in the measurements. Direct monitoring of nicks introduced during the main replication was not feasible since a nick did not alter the appearance of the forked substrate. Only double-strand breaks could be observed, as they caused the DNA to split into two pieces, visible as two blobs in the linear flow cell and two shorter DNA strands in the transverse flow cell. Additionally, this assay was instrumental in optimizing imaging conditions, balancing the need for higher laser power to improve contrast with the risk of photoinduced damage to the substrate.

DNA was prepared like described in section 4.2.5 and the DNA was monitored for 500 frames (Figure 4-4A-C). Once the plectonemes disappeared, the timepoint was marked. This was performed for different conditions. The image conditions were matched to the imaging conditions of the replication assay. For the first comparison, the laser power used to excite SYTOX Orange was varied from 0.3 V to 0.7 V. Increasing the voltage resulted in increase of photoinduced damage indicated by the lower percentage of unnicked molecules over time (Figure 4-4E). The experimental data was fitted with an exponential decay to estimate the half-life of the molecules. The half-life increases from 211 ± 1 (± SD) frames for 0.7 V to 657 ± 5 (± SD) frames for 0.3 V. In single-molecule experiments, oxygen scavenging systems (OSS) such as PCD, PCA, Trolox are frequently added to the reaction to increase lifetime of dyes by reducing oxidative damage (Aitken et al., 2008). We were interested to see the effect on our system and if it could increase the half-life by reducing photoinduced damage. We performed the same assay with the addition of PCD, PCA and Trolox. The condition with the OSS strongly reduced the amount of nicking (Figure 4-4F). The half-life almost triples from 211 ± 1 (± SD) frames to 602 ± 4 (± SD). We also tested whether reducing exposure time while increasing laser power could maintain image contrast without increasing DNA damage. To this end, we doubled the laser power from 1 V to 2 V and halved the exposure time from 200 ms to 100 ms. The resulting survival curves for these two conditions almost completely overlapped, with nearly identical half-life values (1 V, 200 ms: 340 ± 2 frames (± SD); 2 V, 100 ms: 325 ± 2 frames (± SD)) (Figure 4-4G). This indicated that the dosage was effectively the same. Now we could start imaging replication reaction.

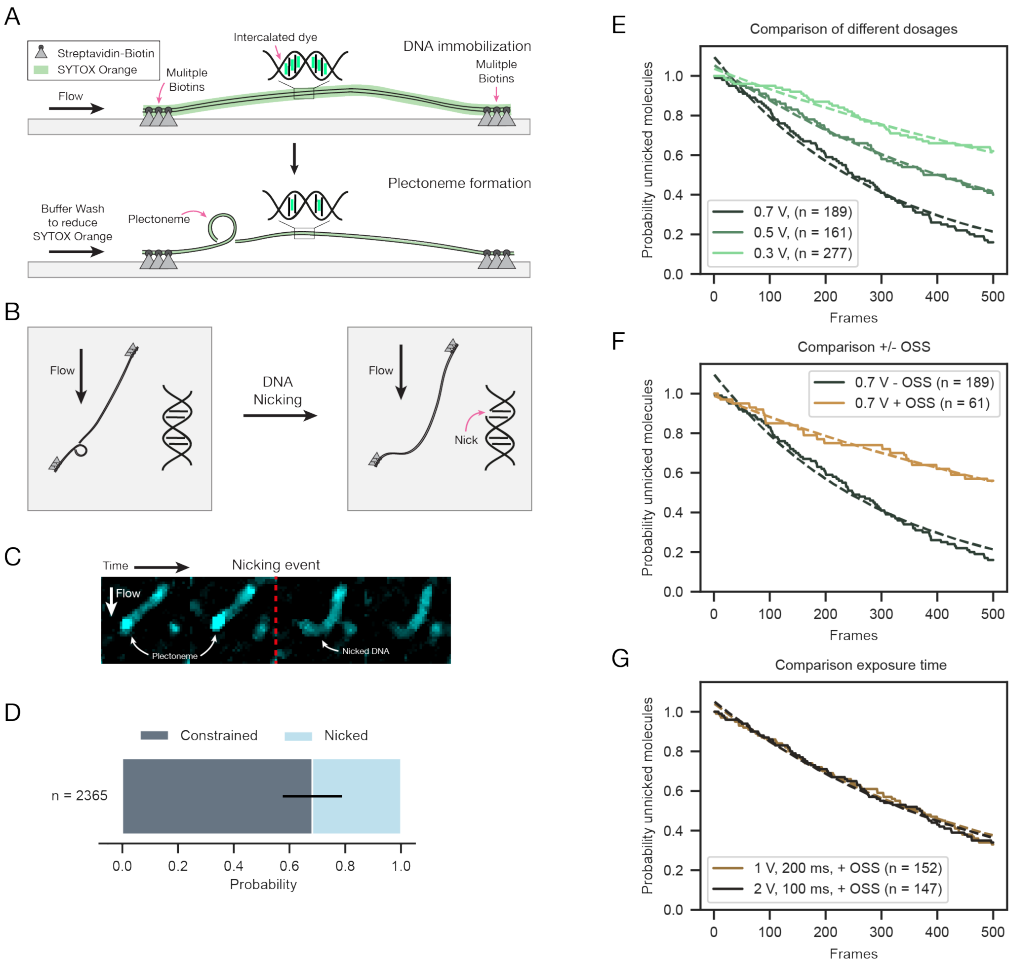


Figure 4-4 Topological constrained DNA and nicking frequency studies. (**A**) Schematic of plectoneme formation in constrained DNA by varying the concentration of intercalating dye. (**B**) Schematic of the shape change induced by a single nick. (**C**) Example molecule over time, where nick was introduced leading to loss of plectonemes. (**D**) Quantification of constrained DNA and nicked DNA observed on the microscope, with displayed percentages representing the combined probability of both conditions. (**E**) Survival curve of unnicked molecules under different laser powers. (**F**) Comparison of survival curves of unnicked molecules with and without oxygen scavenging system (OSS). (**G**) Comparison of survival curve for unnicked molecules to test whether laser power could be doubled if exposure time was halved. In (E) – (G), curves were fitted with an exponential decay. In (D) – (G), number of molecules n are indicated.

4.2.7. Overview of DNA replication visualization and assay design

The goal was to visualize DNA replication using the substrates presented in section 4.2.4 as templates. In chapter 5, we will compare replication on different substrates. This section details the basic concept of the replication assay conducted in the linear flow cell, providing essential information for understanding how we monitor DNA synthesis and protein dynamics at the single-molecule level.

We visualized single T7 replisomes in real time using a single-molecule fluorescence assay, allowing us to monitor DNA synthesis and protein dynamics simultaneously. For this experiment, we used the unconstrained substrate. DNA was attached and prepared as described in section 4.2.4. To initiate DNA synthesis, gp4 (helicase/primase), gp5/trx (polymerase/processivity factor), gp2.5 (single-strand binding protein), ATP, CTP, and the four dNTPs were flushed into the flowcell and incubated (Hamdan et al., 2009; Tanner et al., 2009). The T7 replisome loaded onto the pre-assembled replication fork, with gp4 encircling the exposed lagging strand (Figure 4-5A).

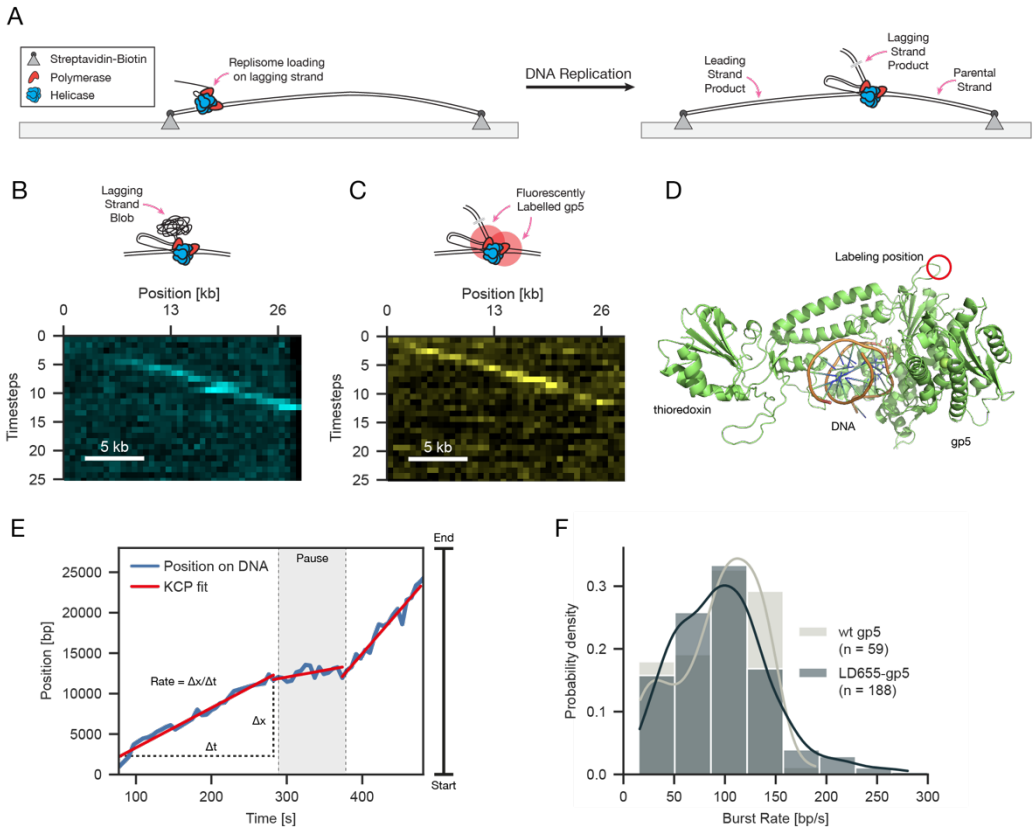


Figure 4-5 Readout methods for single-molecule replication assay.
(**A**) Schematic of single-molecule replication assay with T7 replisome. (**B**) Representative kymograph of a replication reaction following the lagging strand blob over time to visualize replication. (**C**) Representative kymograph of a replication reaction following labeled gp5 (polymerase) over time to visualize replication. Same molecule was used to create kymograph in (B) and (C). (**D**) PyMOL rendering of gp5 in complex with thioredoxin and DNA. The labeling position was indicated with the red circle showing an exposed loop. (**E**) Representative trace of a replication reaction. Trace was fitted using kinetic change point algorithm. A pause event was indicated in grey. (**F**) Replication rate distribution for replication reaction with wt gp5 (unlabeled) or LD655-gp5 (labeled).

The dsDNA was stained with SYTOX Orange and imaged in real-time without buffer flow. As DNA synthesis progressed, the lagging strand product began to form a DNA blob, which moved unidirectionally along with the replisome along the template DNA (Figure 4-5B). Over time, the intensity of the DNA blob increased, indicating the generation of the lagging strand product. The region ahead of the replisome represented the parental template, while the DNA behind it corresponded to the leading strand. Since SYTOX Orange intercalates between base pairs, it loses signal if the DNA is single-stranded. Therefore, we could infer coordinated DNA replication (coupled leading and lagging strand synthesis) as the DNA molecules exhibited an equal amount of DNA signal both ahead of and behind the replisome.

The accumulating lagging strand product was tracked using sub-pixel Gaussian fitting in single frames to detect the peak, which was then linked across consecutive frames to create time-series data for each molecule. The tracking was performed using the 'Peak Tracker' in Mars, and the molecules were transformed to the reference frame of the individual DNA molecules. The resulting time traces provided information about the position of the replisome over time (Figure 4-5E). Additionally, replication dynamics, such as pause positions, pause timing, replication duration, and processivity, were extracted from the data. The replication rate for each time trace was determined using kinetic change point analysis (Hill et al., 2018).

Additionally, we labelled the polymerase (gp5) of the replisome complex, as multiple copies are present at the replisome (Geertsema et al., 2014). Furthermore, tracking changes in the number of polymerases provides insights into the compositional dynamics of the replisome (see section 5.1.8). To label gp5, we integrated a ybbr tag (DSLEFIASKLA) between the residues isoleucine 464 and threonine 465. This labeling position was chosen because it is an exposed loop region, with no apparent interaction with the active site of gp5 or the cofactor trx (Figure 4-5D). The protein was labeled with LD655 CoA, which emits in the red fluorescent channel, thereby keeping the DNA and protein signals in separate channels. The polymerase signal moved in a manner similar to the DNA blob signal along the template (Figure 4-5C).

This labeling position had not been reported before. To test whether it influenced the replication rate, we compared the rate of wildtype gp5 (unlabeled) and LD655-gp5 (labeled) by tracking the lagging strand product. In the replication reaction with unlabeled gp5, the median replication rate was 100 ± 4 bp/s (Median ± SEM, n = 59).

In the replication reaction with LD655-gp5, the median replication rate was 94 ± 3 bp/s (Median ± SEM, n = 188). These rates were consistent within the margin of error and aligned with previously published results (Duderstadt et al., 2016; Lee et al., 2006; Tanner et al., 2009). Section 4.2.6 revealed that DNA remained nick-free for longer periods when OSS were present. However, no replication reactions were observed when the buffer contained OSS. It appeared that adding OSS to the buffer led to the aggregation of gp5, evidenced by large floating structures visible in the red channel. Due to this issue, the condition was not used in the study, and replication events occurring after a certain time threshold were excluded from the analysis.

4.3. Discussion

The aim of this chapter was to design a novel DNA toolkit to study the influence of topological strain on the replisome in a single-molecule assay. We explored the boundaries of the system through simple calculations and verified that almost all the substrates presented were properly formed. Our results demonstrate that we successfully created topologically constrained DNA, and we were able to estimate the proportion of unnicked molecules. Additionally, we developed an assay to quantify nicking events in a time-resolved manner, which allowed us to set a threshold for the experiments presented in the following chapter. We explored the single-molecule replication assay in detail and ensured proper replication with our labeled polymerase. Together, these efforts laid a strong foundation for the upcoming experiments.

Our simulation of the simplified cellular environment demonstrated that our research focus is a valid target for further investigation. The distribution of gyrase molecules alone is insufficient to maintain supercoiling levels low enough for undisturbed replication. The limited space for gyrase to act on the DNA amplifies the buildup of torsional strain, which must be relieved. Doubling the mean of the gyrase distribution helps reduce this effect, yet still results in pause behavior near the end of the DNA template. It is unspurprising that cells do not rely on an overabundance of enzymes to resolve topological problems, as these enzymes interact with DNA and could increase the risk of toxic collisions, potentially leading to DNA damage (Stracy et al., 2019). Although our simulation is relatively simple compared to more complex models of topological challenges (Sevier, 2020), it emphasizes that a more complex pathway to resolve topological strain must exist. The model does not account for the dynamic nature of replisome composition (Beattie et al., 2017), nor does it include other factors

that aid in overcoming topological strain. Despite these limitations, the simulation workflow is easily adaptable for exploring different scenarios and effectively highlights the knowledge gaps in this area of research.

The replication assay is based on the idea of tracking the replisome over time and measuring differences in the process such as changes in rate, processivity, and pausing frequency. The DNA substrates that we created are great for this assay since the similar length between substrates, makes them comparable. We estimated from a simple calculation, that even if the replisome would simply stall when encountering topological strain, we would still be able to resolve it on the microscope since the stalling event would occur at a distance of approximately 10% of the substrate. We calculated that the DNA molecules are not 100% stretched, thus the actual distance will look less than 10%. However, by using a labelled protein we can still detect movement and stalling events.

Most of the substrates we assembled exhibited a narrow distribution around the mean length, indicating successful assembly. However, the 'constrained leading' substrate was an exception, showing a wide range of lengths, with some being almost three times the crystallographic length (Figure 3-4F). This variability led us to exclude this substrate from the study until the purification protocol could be improved. We suspect that the separate ligation step used to attach the biotin handle to the fork may be the source of the problem. We propose an additional ligation and gel filtration step First, the fork is ligated with the sticky end matching NotI to the plasmid backbone and end piece, then the substrate is applied on a gel filtration column to remove excess unbound fork elements. The biotin handle with the NotI motif can be ligated to the fork in the final step. Although this might reduce the final yield of the substrate, it would increase the specificity between substrate components.

The current purification strategy seems effective for most substrates. Moreover, future projects could modify the DNA sequence of the plasmid backbone and oligos to integrate different sequence modifications or adapt to various replication systems. Previous studies have demonstrated the feasibility of such approaches, such as integrating a fluorescent label (Mueller et al., 2020) or adapting a pre-primed fork system to study the replication dynamics of the yeast replisome (Lewis et al., 2020). In its current state, our substrate could also be used to study the *E. coli* replisome. In

summary, this DNA toolkit expands the options for studying DNA replication under topological strain at the single-molecule level.

In our experimental design, we did not directly introduce a direct topological control element. Instead, the topological strain was inherently generated as the replisome unwound the DNA. This approach required us to thoroughly explore and validate the degree to which the DNA remained constrained over time. We used multiple biotin handles, inspired by magnetic tweezers and optical trap experiments, which allowed for multiple interactions between the surface and the DNA handle, each around 600 bp in length. We demonstrated that by ligating these handles to a linear DNA backbone, the DNA could be supercoiled with varying concentrations of intercalating dye, consistent with earlier studies (Ganji et al., 2016; Kim et al., 2018). Approximately 68% of the DNA molecules were topologically constrained, which is a strong result compared to other preparation methods (Table 4-3). Additionally, the substrate used in this study was longer than those in previous studies, making it more challenging to purify without causing damage.

We built on the ability to visualize plectonemes by developing an assay to detect single nick formation. This allowed us to systematically compare DNA stability under different imaging conditions, providing an estimate of DNA damage over time. We found that increasing laser power led to faster nicking of DNA molecules, which could be mitigated by adding an oxygen scavenging system (OSS) to reduce oxidative damage. OSS is commonly used in single-molecule experiments and clearly extends the lifespan of DNA, making it a valuable addition to the buffer (Aitken et al., 2008).

The assay we developed is a powerful tool for estimating and reducing DNA damage during single-molecule experiments on a fluorescent microscope. However, because it is conducted as a separate experiment, the results may vary when applied to different datasets due to potential changes in imaging conditions, such as variations in the TIRF angle. Thorough documentation of microscope parameters can help prevent significant variations between experiments, allowing the estimate to be effectively used for filtering datasets. Based on these findings, we established a threshold that was used to filter the molecules used for analysis in the next chapter, reducing the risk of artifacts resulting from the artificial release of tension due to nicking. This additional testing of imaging conditions could also benefit studies of molecular machines that are sensitive to single-strand breaks. For instance, a single

nick could lead to fork collapse, reducing the processivity of the studied molecular machine and potentially leading to a misinterpretation of the experiment (Vrtis et al., 2021). In conclusion, we successfully developed an assay to screen imaging conditions and set a filtering threshold to minimize potential biases in the data due to artifacts.

Lastly in this assay development, we introduced the readout methods for studying the replication reaction focusing on replisome and protein dynamics. We detailed the foundational aspects of the assay, which will be important in the next chapter. Our results showed that adding a fluorescent tag to gp5 did not alter the overall dynamics of the replisome, allowing us to confidently use this approach in our study. The combination of DNA and protein fluorescent signals provided complementary data, with the lagging strand blob offering a higher and more robust fluorescent signal for tracking than the labeled gp5.

We realized that the absence of a DNA blob at the beginning of the replication reaction poses a challenge, as it takes time for the lagging strand product to accumulate. The chosen DNA stain, SYTOX Orange, only intercalates with dsDNA, which confirms proper leading strand synthesis (Lewis et al., 2020). Fluorescent labeling can be used to track replication dynamics, similar to following the DNA blob. The intensity of the protein signal provides information about the number of polymerases at the fork, a technique previously employed to study protein dynamics (Geertsema et al., 2014; Lewis et al., 2020). The composition of the replisome may change in response to challenges such as DNA damage (Vrtis et al., 2021), altered buffer compositions (Geertsema et al., 2014), or variations in local protein concentration (Lewis, Spenkelink, Jergic, et al., 2017).

In the next chapter, we will investigate how the replisome responds to topological challenges. To summarize, we have laid a solid foundation for studying the effects of topological strain on the replisome by developing a modular DNA toolkit. We assessed the quality of our DNA preparation based on length distributions and selected imaging conditions using a single-molecule assay capable of detecting single nicks in the DNA. Based on these findings, we set thresholds to filter molecules, minimizing the risk of misinterpretation. Finally, we established the basic replication assay, which will be instrumental in uncovering the effects of topological strain on the replisome.

Chapter 5 - Single-Molecule studies of the replisome revealing fork rotation

In this final chapter, we utilized the insights gained from the control experiments to conduct a biological study of the T7 replisome as it encounters topological barriers. Additionally, we developed a more advanced imaging technique to resolve the dynamics at the replication fork revealing fork rotation for the first time. We also investigated the impact of topological strain on the composition of the replisome using labeled polymerases. Before presenting the results, we will briefly summarize the key points from the introduction.

Across all domains of life, the topological state of DNA is crucial for genome compaction, chromosome segregation, and regulatory processes such as the initiation of RNA synthesis and DNA replication (Wang et al., 2013). In bacteria, the chromosome is divided into topological domains where the topological state is confined to specific portions of the prokaryotic chromosome (Postow et al., 2004). Recent advancements suggest a highly complex network of fluctuating supercoil levels (Guo et al., 2021). Maintaining the balance between twist and writhe is essential, as imbalances can disrupt vital mechanisms and lead to detrimental effects. DNA replication, a key process for cell survival, both affects and is affected by the chromosome's topological state. During replication, the parental strand becomes overwound while the leading strand is simultaneously underwound, creating a significant topological challenge (section 1.3). Precatenanes form behind the replication fork and must be resolved before chromosome segregation.

Torsional strain introduced during these processes is resolved by a class of enzymes called topoisomerases (Vos et al., 2011) (Section 1.1.6). The primary function of these enzymes is to maintain the topological state of the chromosome and resolve disruptive imbalances caused by biological processes. By creating transient breaks in the DNA, topoisomerases can control the topological state. DNA gyrase, a type II topoisomerase from *E. coli*, can relax positive supercoils and introduce negative twists ahead of the replication fork, facilitating replisome progression and maintaining a superhelical density of -0.05, which is crucial for origin melting and the initiation of replication (Fulcrand et al., 2016). Recent studies have shown that the number of DNA gyrase

molecules, closely associated with the replisome, is insufficient to keep up with the rapid replication rate (Stracy et al., 2019) (Section 1.3.1).

This suggests the presence of intrinsic mechanisms within the replisome for coping with topological strain (Section 1.3.2). Additionally, there are scenarios where the space for topoisomerase activity is restricted, such as when the replisome approaches the end of a topological domain, encounters a head-on transcription-replication conflict, or when two replisomes converge during termination. One proposed mechanism for releasing torsional strain involves the detachment of parts of the protein complex, like the polymerases, which would allow the DNA to rotate freely. This concept is supported by the signal release model (Kurth et al., 2013).

Another possibility is that the torsional strain ahead of the replisome is transferred behind it through a process known as fork rotation. However, this phenomenon has not yet been directly observed, with evidence only suggested by the accumulation of excess precatenanes formation in gel-based assays (Schalbetter et al., 2015) (Section 1.3.2). Given the helical structure of DNA, fork rotation could occur as the replisome follows the helical path presented by the DNA, similar to the rotation observed in RNA polymerase when fixed to a glass slide, as reported by Harada et al. (2001) and Kosuri et al. (2019) (Section 1.4.1). Interestingly, it has been shown that when RNA polymerase is not fixed to a surface, it introduces positive supercoils ahead and negative supercoils behind, without evidence of rotation or the exchange of topological content (Janissen et al., 2024). Despite these findings, no direct evidence has been reported for fork rotation of the replisome under topological strain in single-molecule assays. This raises the question of how the replisome responds to torsional strain in real-time and whether it can tolerate it. Additionally, if the replisome does rotate, what impact does this have on the composition and functionality of the replisome?

To investigate the dynamics of the replisome during DNA replication in real-time, we developed a single-molecule assay to study how the replisome reacts to torsional strain induced by its own activity. We chose the bacterial T7 replication system, which has been established as an effective model for studying the replisome, particularly in single-molecule experiments (Duderstadt et al., 2016; Geertsema et al., 2014; Hamdan et al., 2009; Lee et al., 2006) (Section 1.2.2). Moreover, the principles of this system can be compared to the *E. coli* and yeast replisomes (Gao et al., 2019). In this study, we compared two DNA substrates. The first substrate featured single biotin

attachments to the slide, allowing it to rotate freely ('Unconstrained'). The second substrate was constrained by multiple biotins ahead of the replication fork, creating a topological challenge similar to the condition where no DNA gyrase molecules are in close proximity ('Constrained'). Observations of visual differences in the kymographs led to the development of a transverse flow approach, enabling us to directly visualize the dynamics at the fork while simultaneously reading out the lengths of the leading, lagging, and parental strand. When comparing the two substrates using the transverse flow method, we observed direct evidence of rotational movement in the replisome when it was challenged with constrained parental DNA.

5.1. Results

5.1.1. Single-molecule visualization of DNA topology dynamics during replication

We developed a single-molecule fluorescence imaging platform to simultaneously visualize DNA synthesis and protein dynamics of individual T7 replisomes in real-time. Replication was observed on surface-immobilized, double-stranded DNA molecules, each ~27 kb in length, which were attached at both ends to a cover slip via biotin-streptavidin handles within a microfluidic flow cell. Replisomes were assembled at the replication fork, introduced to one end of the DNA molecules. T7 DNA replication on the DNA substrate was visualized using a micro-mirror total internal reflection fluorescence (mmTIRF) microscope (Larson et al., 2014) with SYTOX orange-stained DNA, in the absence of buffer flow (Figure 5-1A). Upon the addition of nucleotides and all necessary replication proteins for T7, the replication process accumulated lagging strand products, forming a visible blob of DNA that moved unidirectionally along the DNA molecules (Figure 5-2A). The rate and processivity of individual replication events were monitored by tracking the position of the lagging-strand products over time using subpixel gaussian fitting (Figure 5-1B, C).

Subpixel localization and tracking of the lagging-strand replication products revealed a burst replication rate of 94 ± 7 bp/s (median ± SEM, n = 189 molecules) on the single biotin-attached substrate ('Unconstrained') (Figure 5-1B gray), in agreement with previous studies (Duderstadt et al., 2016). The mean processivity was 24.5 ± 0.5 kb (mean ± SD) (Figure 5-1C gray), slightly higher than previously reported (Kulczyk et al., 2012; Lee et al., 2006), with 74% of the DNA molecules in which replication was

initiated successfully replicating the entire 27 kb DNA substrate (Figure 5-1C). Consistent with earlier findings (Hamdan et al., 2009; Lee et al., 2006), tracking revealed pausing events, with 12% of replisomes exhibiting a single pause and 1% showing a second pause (Figure 5-1D). The DNA molecules were approximately 68 ± 7% (mean ± SD) stretched.

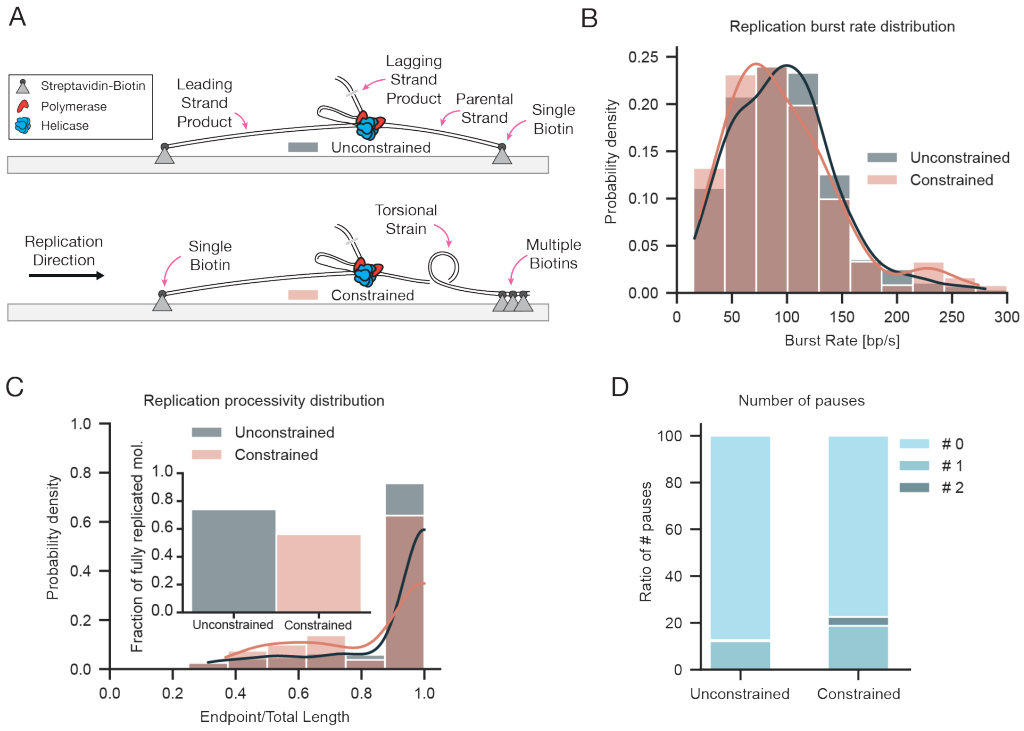


Figure 5-1 Replisomes tolerate topological strain.
(A) Schematic of the time-resolved single-molecule replication assay for the unconstrained and constrained DNA substrates. **(B)** Distribution of replication burst rates for the 'Unconstrained' and 'Constrained' substrates, measured in base pairs per second. **(C)** Distribution of replication processivity, with the x-axis representing the endpoint position in base pairs relative to the total substrate length. The inset bar graph shows the fraction of fully replicated molecules to the total number of replicating molecules. **(D)** Probability of replication pauses, comparing the 'Unconstrained' and 'Constrained' substrates.

Several measures were implemented to prevent photo-damage to DNA, which could lead to the formation of nicks that disrupt replication and hinder the observation of DNA topology dynamics (Sections 4.2.5 and 4.2.6). Different laser settings were tested by monitoring supercoil loss in topologically constrained DNA molecules, which were generated by varying the concentration of an intercalating dye (SYTOX orange), similar to the approach used by Ganji et al. (2016). When a nick was introduced into the DNA substrate, the DNA molecule exhibited a shape change (Section 4.2.5). This allowed

us to establish an optimal low-light imaging condition, where the DNA molecules remained intact within the 95% confidence interval for the selected molecules used in the analysis. Additionally, using the same DNA and biotin handle, we were able to create constrained DNA, demonstrating that torsional strain can indeed build up when the DNA is attached to microscopy slides (Section 4.2.5).

5.1.2. Replisomes tolerate topological strain

To evaluate the influence of topological challenges ahead of the replication fork, we designed a second substrate in which the parental strand was torsionally constrained through the introduction of multiple biotins integrated into the DNA sequence at the end opposite to the fork ('Constrained') (Figure 5-1A). This modification was possible because of our modularly designed DNA substrates, which allowed for the introduction of torsional constraints at specific locations (Section 4.2.2). In this configuration, when the helicase separates the two strands of the dsDNA to create ssDNA for the polymerases, the torsional constraint may lead to overwinding of the parental DNA ahead of the replisome, simulating a scenario where spatial limitations prevent DNA gyrase from binding.

Similar to the unconstrained DNA substrate, a visible blob of DNA formed and moved unidirectionally along the constrained DNA (Figure 5-2A). Tracking of the lagging-strand replication products unexpectedly revealed a remarkable tolerance by the replisome when faced with torsional constraints in the parental DNA. A modest 13% reduction in the median burst replication rate was observed, which measured 82 ± 2 bp/s (median ± SEM, n = 80 molecules) (Figure 5-1B). The number of molecules replicating the full length of the constrained substrate was reduced to 56%, while the mean processivity decreased slightly to 23.3 ± 0.6 kb (mean ± SD) (Figure 5-1C). However, pausing events became more frequent, increasing by 19%, with the fraction of molecules exhibiting a second pause rising to 4% (Figure 5-1D). Constraining the parental strand of the forked substrate resulted in a slightly slower replisome and a reduction in overall processivity, indicating a challenge. Despite these challenges, the replication process was not significantly impaired, with the replisome still able to withstand or resolve the additional strain and replicate nearly to the same extent as the unconstrained substrate.

5.1.3. Dynamics of the replication product revealing differences in shape and movement patterns

A closer inspection of the kymographs, generated from individual molecules, revealed greater variability in the shape and movement patterns of lagging-strand products during DNA synthesis when using the constrained substrate (Figure 5-2A). These shape changes were not observed in the unconstrained substrate, where the lagging-strand product consistently appeared as a single, uniform bundle moving unidirectionally (Figure 5-2A). To further illustrate this behavior a line profile was drawn through each pixel line of the kymograph, with the line plots representing consecutive time points that trace the movement of the replication fork (Figure 5-2B). In the absence of torsional constraints, the line profile displayed a narrow peak with increasing height, corresponding to the signal amplification of the DNA. However, when the parental strand was constrained, the peak of the line profile broadened and varied in intensity, indicating a spreading of the lagging-strand DNA signal along the substrate.

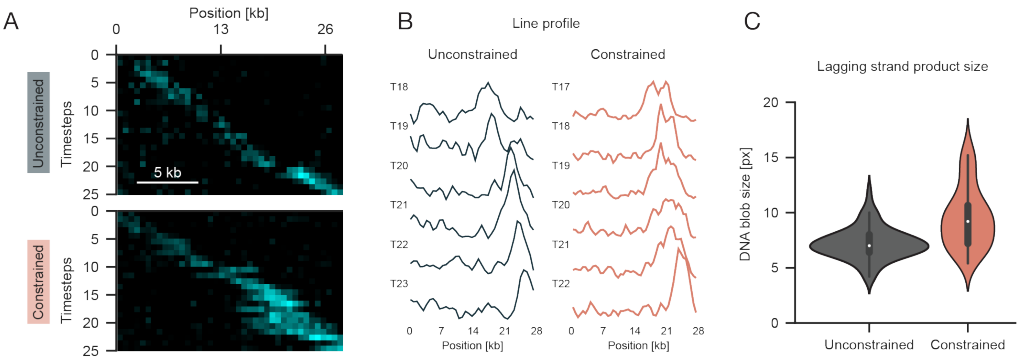


Figure 5-2 Dynamics of the DNA product shape.
(A) Representative kymographs showing the accumulation of lagging strand products for unconstrained (top) and constrained (bottom) DNA substrates. **(B)** Line profile of consecutive time steps of the pixel values from the corresponding kymographs. **(C)** Violin plot representing the size of the lagging strand product for unconstrained (n = 115) and constrained (n = 53) DNA substrates.

To quantify this behavior, the kymographs were analyzed using image thresholding, where pixels above an automatically set threshold were considered as a DNA signal, while the others were classified as background. This approach allowed us to measure the size of the DNA blob in pixels (px). We focused on a window of 5 timepoints near the end of the replication process, where the size changes were most pronounced. The mean size of the DNA blob increased from 7.2 px in the unconstrained condition (n = 115 molecules) to 9.5 px in the constrained condition (n = 53 molecules) (Figure 5-2C). Additionally, the fluctuations in size nearly doubled, with the standard deviation

increasing from 1.5 px (unconstrained) to 2.8 px (constrained). These findings suggest significant variability in the shape of the DNA blob when the parental strand is constrained, hinting at dynamic processes at the replication fork that are not fully resolved by the standard linear replication assay.

5.1.4. Transverse flow platform resolving spatial dynamics

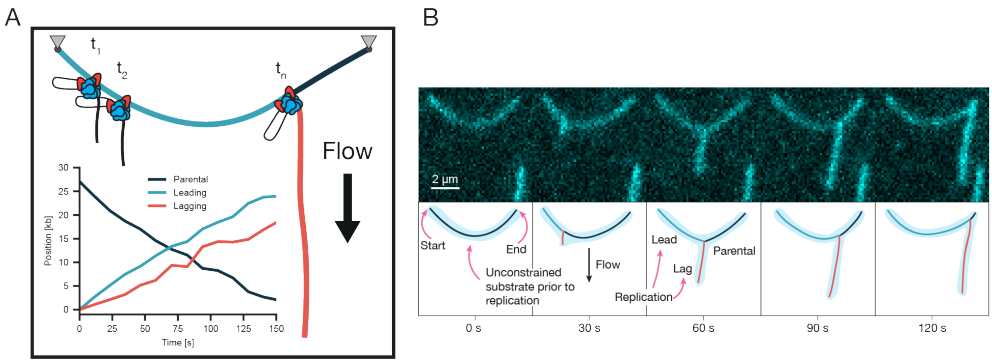


Figure 5-3 Transverse flow platform to study DNA topology dynamics.
(A) Schematic of the transverse flow replication assay. The transverse flow stretches out the lagging strand product, which increases in length over time (t_1, t_2, t_n). The different DNA strands are color-coded. The inset shows a time trace for each of the different strands over time. **(B)** Representative kymograph of an unconstrained replication event in the transverse flow configuration (top). The shape of the arched DNA molecule was fitted and overlaid with the segmentation result (bottom).

To more definitively resolve the spatial dynamics of the individual DNA products synthesized at the replication fork, we developed an alternative flow cell geometry that allowed replication to be monitored in the presence of a transverse flow. This approach enabled us to use the same experimental setup and components while also gaining additional information about the length distributions of the DNA products and the exerting force on the replisome. Instead of using a single flow lane yielding in linear attachment of the DNA, we employed a flow cell with two crossing flow lanes.

In the initial step, the DNA substrate was applied in the same manner as in the linear configuration. However, during the imaging process, a transverse flow was introduced, leading to the formation of arched shaped DNA molecules (Figure 5-3A,B). When the replisome synthesized the lagging strand product, it was stretched out by the transverse flow, rather than forming a DNA blob as previously observed. This setup allowed us to simultaneously measure the lengths of the parental, leading, and lagging

strand (Figure 5-3A,B). The lagging strand product appeared as a straight line that grew in length over time (Figure 5-3B).

As the lagging strand extended, it split the arch into leading and parental sections, which were defined by the direction of movement (Figure 5-3B). The lagging strand arm moved from one side of the arch to the other while being stretched by the flow. Upon reaching the end of the arch, the DNA arm disappeared. The extensions were segmented, and the measurements were converted to base pairs. These measured extensions can be used to estimate the forces acting on the DNA using the worm-like chain model (Section 1.12).

5.1.5. Estimating the force on the replisome

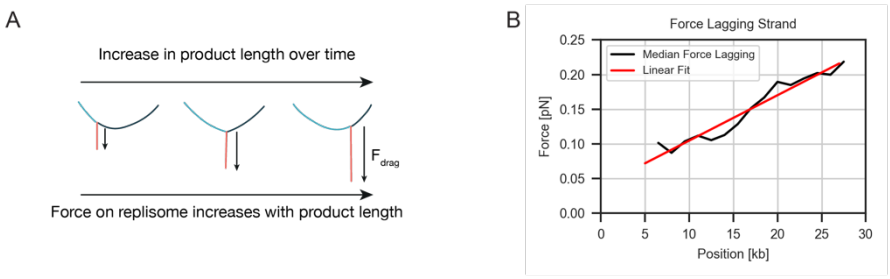


Figure 5-4 Estimating the force on the replisome.
(**A**) Schematic of the force ramp. The lagging strand product increases in length linearly exerting force on the replisome sitting at the junction. (**B**) The median force applied by the lagging strand plotted for different positions on the DNA. The median force was fitted using a linear fit.

Measuring the extension of the lagging strand allowed us to estimate the force exerted on the replisome using the worm-like chain model (Figure 5-4A) (Section 1.1.2). We assumed that the lagging strand was approximately equal in length to the leading strand at any given time during replication. By comparing the extension of the lagging strand to its contour length, as inferred from the leading strand's extension, we were able to estimate the forces involved. Median forces at various positions along all molecules were calculated and plotted (Figure 5-4B). The force increased linearly, reaching approximately 0.22 pN. This linear increase corresponded with the linear growth of the lagging strand product over time, effectively creating a force ramp experiment that applied a force on the replisome through the drag of the flow. Given that the positive supercoil formation is inhibited at forces of 6.5 pN (positive buckling transition) (Kriegel et al., 2017), the much lower forces measured here indicate that positive supercoiling ahead of the fork could potentially form.

5.1.6. Consistent replication dynamics across flow cell configurations

When comparing the two flow cell configurations, the burst median replication rate of the leading strand was identical for the unconstrained substrate in both setups (94 ± 7 bp/s; Median ± SEM, n = 24), indicating that the transverse flow approach did not affect the replication rate. The percentage of molecules completing replication decreased by 7%, likely due to the smaller sample size. Additionally, the mean endpoint of replication was 25.0 ± 0.4 kb (mean ± SD, n = 24), consistent with the results from the linear configuration. These findings suggest that the alternative flow cell geometry did not influence the measured outcomes.

This assay represents the first single-molecule replication assay where individual DNA strands are simultaneously resolved in real time, allowing for direct observation of differences in leading and lagging strand behavior. While the burst replication rate of the leading strand was identical to that observed in the linear configuration, the lagging strand exhibited a lower burst rate of approximately 69 ± 7 bp/s (Median ± SEM, n = 24) for the unconstrained substrate. This approach enables separate analysis of leading and lagging strand replication dynamics, offering a powerful method to monitor both synthesis rates simultaneously.

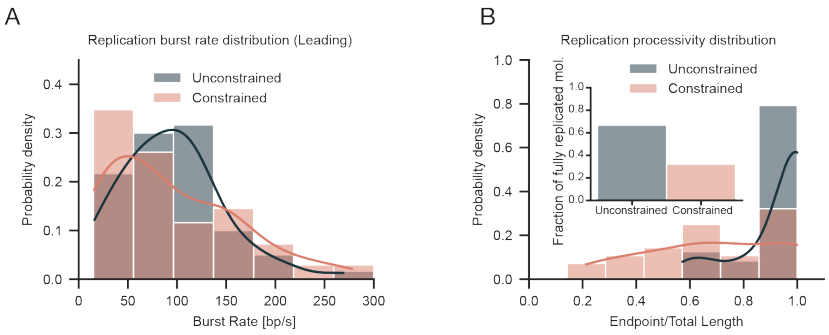


Figure 5-5 Replication rates and processivities of the replisome in the transverse flow assay.
(**A**) Distribution of replication burst rates (Leading strand) for the 'Unconstrained' and 'Constrained' substrates, measured in base pairs per second for the transverse flow configuration. (**B**) Replication processivity distribution for the transverse flow configuration. The x-axis shows the endpoint position in base pairs, normalized to the total length of the substrate. The bar graph insert represents the ratio of molecules that completed replication to the total number of replicating molecules.

Constraining the parental strand results in a median burst replication rate that is consistent within the margin of error (constrained: 82 ± 8 bp/s, n = 26), similar to the

rate observed in the linear configuration. The processivity was 19.2 ± 0.7 kb (mean ± SD, n = 26), with only 32% of molecules reaching the end of the substrate. The significantly reduced processivity in the constrained substrate could be due to the additional force exerted on the lagging strand, which scales with its extension. This added force may interfere with the mechanism that releases torsional strain, as revealed by the transverse flow assay. Overall, the replication behavior measured with the transverse flow approach was not significantly different from the linear configuration, yielding a similar pattern of results (Figure 5-5A, B).

5.1.7. Lagging strand dynamics revealing fork rotation

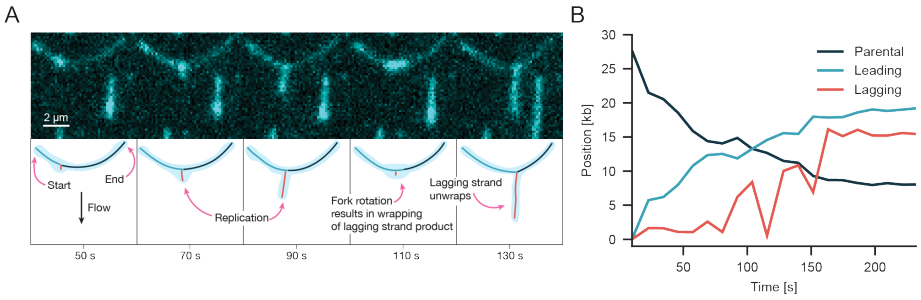


Figure 5-6 Example molecule of a constrained substrate.
(**A**) Representative kymograph (top) of a replication event on constrained DNA. The shape of the arched molecule was fitted and overlaid with the segmentation result (bottom). (**B**) Time traces of the molecule presented in (A). The parental, leading, and lagging strand lengths are plotted over time. A dip in extension in the lagging strand at around 110 s corresponds with wrapping of the lagging strand product in the kymograph.

Interestingly, by extending the lagging strand, dynamic features at the replication fork become apparent when comparing the substrates side by side. The unconstrained substrate shows an almost continuous increase in length over time (Figure 5-3B, 5-7A), whereas the constrained substrate displays significant fluctuations in length, including reductions where no extension of the lagging strand is visible between consecutive time points (Figure 5-6A, 5-7A). In the constrained substrate, when the extension shortens, more DNA appears to accumulate at the arch, indicated by a higher fluorescent signal. Comparing the measurements for each lagging strand extension over time, reveals a clear difference in appearance (Figure 5-7A). For the unconstrained substrate, the extensions steadily increase, while for the constrained substrate, the lagging strand length alternately extends and retracts, which is observed

across multiple traces. This behavior is indicated by changes in length approaching zero extension (Figure 5-7A).

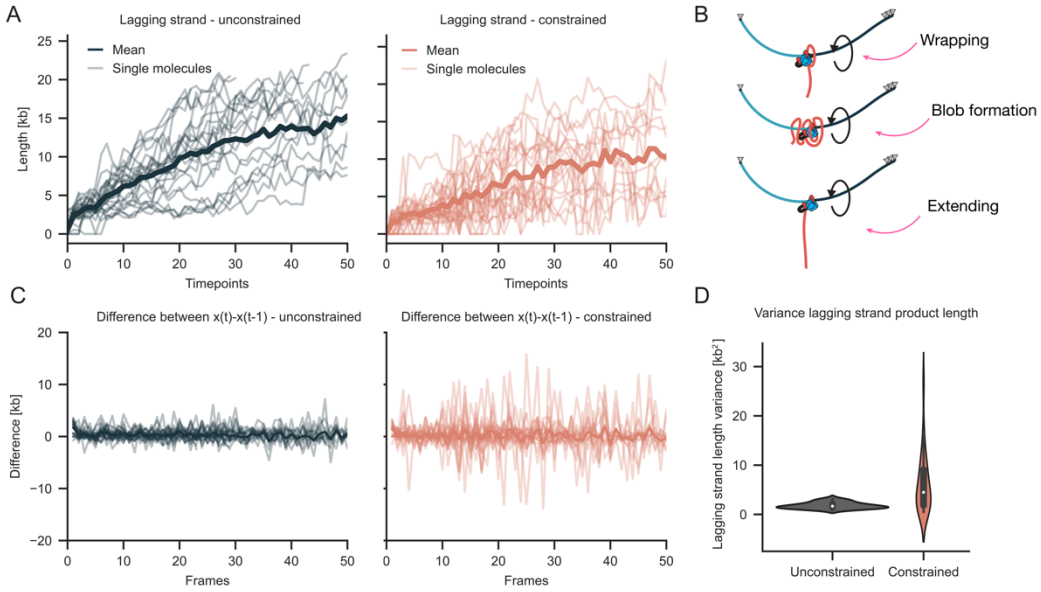


Figure 5-7 Lagging strand dynamics reveal fork rotation.
(**A**) Lagging strand extension over time for unconstrained and constrained conditions. The solid line represents the mean of all traces, aligned to the same starting point. Individual traces are shown with lower opacity. (**B**) Visual depiction of the wrapping of the lagging strand product around the DNA substrate. (**C**) Stationary function of the lagging strand extension over time, calculated by subtracting the extension at x(t) from the previous extension at x(t-1). The solid line represents the mean of all traces, with individual traces shown in lower opacity. (**D**) Violin plot showing the variance in lagging strand product length from traces in (C). (Unconstrained: n = 24, Constrained: n = 26).

To quantify the behavior of the lagging strand arm, the traces were transformed into a stationary function by calculating the difference between two consecutive extension values (x(t) - x(t-1)) (Figure 5-7C). We expected a linear relationship between product length and time (corresponding to the burst rate multiplied by time). If a steady growth was assumed, the change in length should be constant per timestep, resulting in differences around zero when consecutive changes in length are subtracted. The mean variance of all stationary traces more than tripled when comparing the unconstrained and constrained cases, increasing from 1.8 ± 0.6 kb to 6.1 ± 5.9 kb (Mean ± SD) (Figure 5-7D). The broad distribution of values in the constrained case also indicates that there are significant length fluctuations in some traces. Notably, the product length often returns to its initial extension, and the accumulation of DNA suggests that the DNA is

not cleaved, but instead wraps around the arched DNA substrate (Figure 5-7B). This fluctuation is absent in the unconstrained substrate, where no torsional tension accumulates due to the single attachment.

There are no noticeable length changes in the parental portion of the dsDNA arch, indicating that no supercoils are introduced. Such changes would typically manifest as alterations in the DNA's appearance and significant length variations due to supercoil formation (Ganji et al., 2016; van Loenhout et al., 2012). The observed fluctuations in product length for the constrained substrate could suggest a rotational movement of the fork, where torsional strain is transferred behind the replication fork. This mechanism allows the replisome to continue DNA synthesis even when the parental strand is under torsional constraint, explaining why the replication process is only slightly affected by a topological barrier ahead of the replisome. This finding suggests that we were able to visualize fork rotation for the first time in a single-molecule level by using the transverse flow assay.

5.1.8. Increased polymerase count suggesting replisome component detachment facing topological strain

We discovered that the replisome appears to rotate continuously during DNA synthesis when the DNA is a constrained ahead. At a synthesis rate of approximately 90 bp/s, this would result in the replisome rotating around 9 times per second. To capture this dynamic behavior, we employed faster imaging, taking images every 200 milliseconds. As expected, the replisome on the unconstrained substrate showed no signs of rotation, while it exhibited extensive rotational movement on the constrained substrate, repeatedly extending and contracting within hundreds of milliseconds (Figure 5-8). To fully wrap the lagging strand around, multiple rotations were required, as indicated by the stepwise shortening of the exposed lagging strand (Figure 5-8). This rapid rotation exerts significant stress on the replisome. To gain deeper insights into how this stress may result in changes in the composition of the replication machinery we utilized labeled polymerases (gp5). As demonstrated in chapter 4, the labeling did not affect replication rates, allowing us to use this method to add another layer of analysis. Here, our focus was on the compositional changes and the behavior of the polymerases under these conditions.

In the linear replication assay, we tracked the fluorescent signal of gp5 over time. We integrated the fluorescent signal within a radius larger than the expected size of the

replisome, allowing us to capture signals from polymerases that might detach from the replisome but remain bound to on one of the daughter strands. To estimate the number of polymerases, we identified proteins that were immobilized on the surface of the slide and monitored their photobleaching over time. We fitted the photobleaching steps and calculated the intensity difference before and after each bleaching event. This difference provided a median step size, representing the intensity of a single dye. We repeated this calculation separately for each experiment.

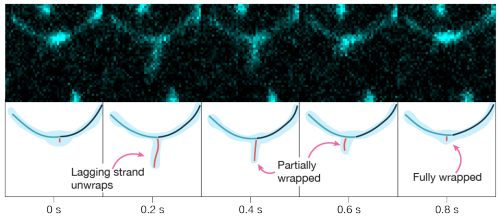


Figure 5-8 Replisome rotates at a millisecond timescale.
Representative kymograph of replication on constrained DNA imaged at a high frame rate (200 ms steps). At 0 seconds, the lagging strand DNA was wrapped around the arch. In the next frame, the lagging strand was extended, followed by a partial wrapping. At 0.6 seconds the lagging strand was wrapped further and at 0.8 seconds the lagging strand product was fully wrapped similar to 0 seconds.

For the unconstrained substrate, the median number of polymerases was 3.2 ± 0.1 (Median ± SEM, n = 75). The maximum number of polymerases that can attach to the helicase gp4 is six, given that it is a homohexamer where one monomer has a single binding site (Gao et al., 2019). Previous studies have reported up to six polymerases simultaneously associated with the replisome, depending on the frequency of lagging strand repriming (Geertsema et al., 2014). Under conditions similar to our reaction buffer, with 300 μM ATP and CTP, they measured an average of 2.4 gp5 molecules at the replisome. Our estimated number of polymerases aligns closely with these previously published results. It is important to note that the earlier study used autocorrelation between two fluorescent dyes to estimate the number of polymerases. In the representative kymograph (Figure 5-9A), the number of polymerases varied but generally remained around three over time. The fluctuations could be attributed to the dynamic nature of the lagging strand, which is synthesized discontinuously.

For the constrained substrate, the estimated number of polymerases was higher than the maximum number of binding sites. The median number of polymerases increased to 9.8 ± 0.2 (Median ± SEM, n = 56) (Figure 5-9C, D), three times greater than the

number observed in the unconstrained case. In the representative kymograph, the number of polymerases increased over time (Figure 5-9B). The distribution broadened significantly from a narrow peak to a much wider range for the number of polymerases when comparing the two substrates (Figure 5-9C). The observed polymerase count exceeded the binding capacity at the helicase. A possible explanation for this could be the frequent detachment of polymerases to facilitate the release of accumulating topological strain, as suggested by the 'signal-release' model (Kurth et al., 2013). Additionally, the high frequency of rotation and wrapping may contribute to the detachment. Previous studies have reported that unregulated rotation can cause extensive DNA damage in yeast (Schalbetter et al., 2015).

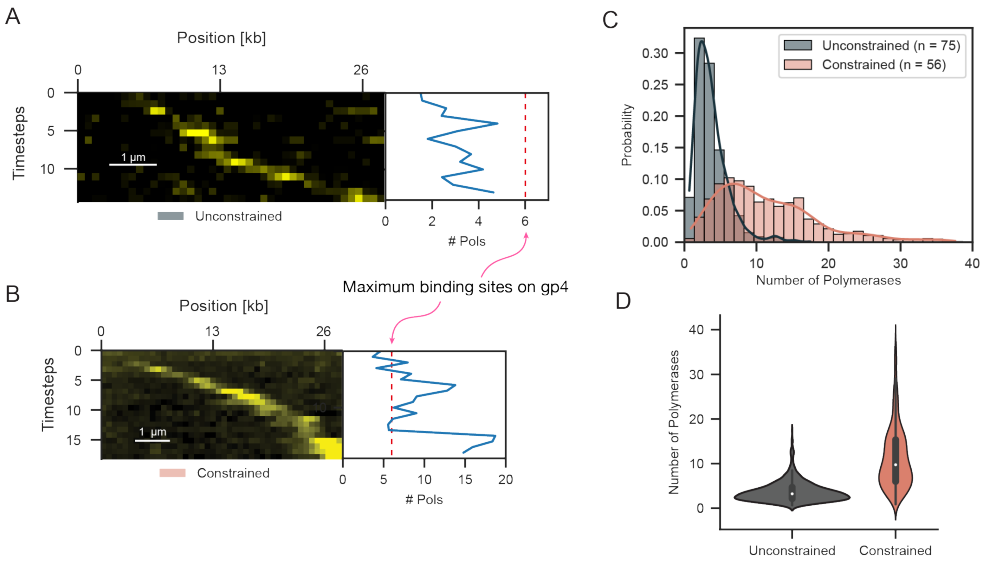


Figure 5-9 Increased numbers of polymerases in close proximity to the replisome.
(**A**) Representative kymograph showing the gp5 signal during replication of the unconstrained DNA substrate. The number of polymerases over time is plotted on the right, with six being the maximum number of polymerases that can bind to the helicase gp4. (**B**) Representative kymograph of the gp5 signal during replication of the constrained DNA substrate. The number of gp5 molecules in close proximity to the replisome increases beyond the maximum number of binding sites. (**C**) Distribution of gp5 molecules in close proximity to the replisome for both unconstrained and constrained DNA substrates. (**D**) Violin plot depicting the distribution of polymerase numbers for the two substrates (Unconstrained and Constrained).

5.2. Discussion

The aim of this study was to investigate whether the replisome can tolerate torsional strain and to understand how it manages this strain in the absence of topoisomerase.

To address this, we designed a DNA substrate that induces torsional strain during replication and then used this substrate to challenge the replisome during a single-molecule replication assay. Our findings demonstrate that the replisome can withstand torsional strain effectively. Additionally, we developed a novel imaging platform to enhance the resolution of replication dynamics, allowing simultaneous measurement of the leading, lagging, and parental strands. This was achieved by introducing a transverse flow on the DNA substrate, forming it into arched shape and stretching out the lagging strand. Through this approach, we were able to visualize fork rotation in a single-molecule context for the first time. This mechanism was previously observed only in bulk assays (Hiasa & Marians, 1994; Keszthelyi et al., 2016; Schalbetter et al., 2015; Snapka et al., 1988; Sundin & Varshavsky, 1980, 1981).

5.2.1. Transverse flow – simultaneous readout of the leading, lagging and parental strands

The transverse flow approach developed in this study provided significant advantages over standard single-molecule replication assays by allowing the simultaneous resolution of the parental, leading, and lagging strands. While the processivity and rate distributions obtained using this new method were consistent with results from the standard linear configuration, the added ability to observe different strands introduces a new dimension to replication studies, which was previously inaccessible. This innovation enabled detailed analysis of replication dynamics, such as differences in synthesis rates between strands or the visualization of gaps formed by incomplete lagging strand synthesis (Figure 5-10C). In addition to its unique capacity for visualizing strand-specific replication events, the transverse flow approach allowed for exerting forces on the replisome, a capability that is not possible in the standard linear configuration (Figure 5-10A). While the forces applied in this study were small, increasing the flow rate could elevate the drag forces exerted on the replisome, opening up new possibilities for force-based studies that are traditionally limited to techniques like optical traps or magnetic tweezers (Neuman & Nagy, 2008; Strick et al., 2003). The modular nature of our DNA toolkit further enhances the versatility of this approach. For example, attaching a bead to the DNA to increase the applied force could provide deeper insights into replisome mechanics (Duderstadt et al., 2016).

Although the T7 replisome was used as a proof of concept, the DNA substrates designed for this study are also compatible with other replication systems, such as the

E. coli replisome, which similarly loads on the lagging strand of the fork (Gao et al., 2019). These substrates could also be adapted for use with yeast or human replication systems, which feature distinct polymerases for each strand, by repositioning the biotin from the leading to the lagging strand in the fork structure (Baris et al., 2022; Lewis et al., 2020) (Figure 5-10B).

Moreover, the ability to differentiate between leading and lagging strand replication rates in more complex systems, like those of yeast or human, could reveal new insights into replication mechanics that were previously obscured (Figure 5-10C). This platform could also facilitate the study of chromatin dynamics, such as histone deposition in relation to DNA replication, with strand-specific resolution (Gan et al., 2018; Stewart-Morgan et al., 2020). In summary, the transverse flow approach significantly enhances the information obtained from single-molecule replication experiments by enabling simultaneous visualization of all three strands. This approach offers greater flexibility in experimental design, enabling the introduction of force as a variable and allowing more in-depth studies that reveal fork-specific dynamics by resolving strand-specific activities.

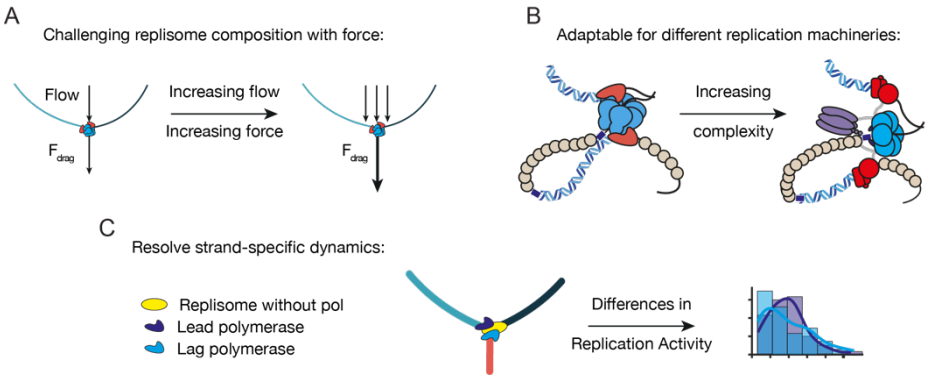


Figure 5-10 Possible applications of transverse flow assay.
(**A**) Higher forces can be applied to challenge the replisome composition by increasing the flow rate. (**B**) The assay is adaptable for studying other replication systems at the single-molecule level. (**C**) Strand-specific activities can be resolved, allowing for observation of differences in the rates of leading and lagging strand polymerases, as an example.

5.2.2. The replisome tolerates and overcomes topological strain

The main goal of this thesis was to investigate how the replisome responds to a topological challenge and to explore potential alternative pathways for releasing the

resulting strain. The topological challenge we examined was comparable to a scenario in which topoisomerase activity ahead of the replisome is entirely inhibited, similar to situations where there is limited space for DNA gyrase to function (Figure 5-11). In this setup, we observed only minor reductions in burst rate and processivity. In addition, by using our novel transverse flow imaging approach, we were able to directly visualize the strain release mechanism. For the first time, we captured images of fork rotation, a mechanism proposed in earlier publications (Champoux & Been, 1980; Postow et al., 2001), but never directly observed until now. However, while fork rotation appears to be a solution for overcoming topological challenges, it introduces new issues: The increased number of polymerases near the replication complex suggest possible detachment of polymerases, and there is increased potential of introducing DNA damage sites. Detaching and resuming replication downstream can result in gaps in the daughter strand that require postreplicative repair, similar to other DNA lesion bypass pathways (Quinet et al., 2021; Wong et al., 2021; Yeeles, 2014). Whether this is an integral part of the release mechanism or simply a consequence of extensive fork rotation in our experimental system remains unclear with the current assay.

Overall, our findings suggest that fork rotation acts as a pathway that is selectively used when the binding space for topoisomerases on the DNA is limited (Figure 5-11). This can be interpreted as follows: The replisome begins by separating the two strands of DNA to synthesize the daughter strands. As it progresses along the DNA, the additional twist generated ahead of the replisome is absorbed by the parental strand. The energy stored in the twists is inversely proportional to the length of the DNA (Equation 1.13), meaning that less work is needed to introduce twists when the DNA is longer. Fork rotation is likely triggered once a certain energy threshold is reached, making rotation or following the helical path more favorable than continuing to introduce twists. This results in the formation of intertwined daughter strands or precatenanes, a mechanism previously proposed based on interpretations of gel-based experiments (Hiasa & Marians, 1994; Keszthelyi et al., 2016; Schalbetter et al., 2015; Snapka et al., 1988; Sundin & Varshavsky, 1980, 1981). Furthermore, we did not observe any signs of positive supercoiling ahead of the replisome, suggesting that, instead of converting twist into writhe, the additional strain is immediately transferred behind the fork. High-speed imaging supports this conclusion, revealing a multi-step wrapping process observed in 200 ms intervals. Taken together, these findings suggest that fork rotation may occasionally be more favorable than continuing to introduce additional twists when strain is building up.

In our current setup, we created an extreme case of a topological barrier for the replisome. During normal replication conditions, there would be space for topoisomerases to act on the DNA. DNA gyrase, for example, has a footprint of about 140 base pairs (Baker et al., 2011; Gore et al., 2006). In addition, there are on average 12 gyrase molecules closely associated with the replisome supporting the replication process (Stracy et al., 2019). The complete lack of regulation of fork rotation observed in our study could explain the increased number of polymerases near the replication fork, possibly indicating restart events and potential DNA damage sites.

In more complex replication systems such as the yeast replisome, fork rotation is regulated by the fork protection complex Tof1/Csm3 (Schalbetter et al., 2015). When Tof1 is deleted and Top2 is depleted, signs of post-replication repair appear, likely due to excessive fork rotation and DNA precatenation. This could be due to gaps left behind by extensive fork rotation. Additionally, the high number of precatenanes could inhibit the maturation of Okazaki fragments and the activation of cohesin, a structural maintenance of chromosomes protein responsible for chromosome segregation. The human replication system also has homologs, Timeless and Tipin, similar to Tof1/Csm3, suggesting a comparable mechanism to that seen in yeast (Shyian et al., 2020; Vipat & Moiseeva, 2024; Westhorpe et al., 2020). Overall, these findings indicate that fork rotation is an immediate response when the replisome faces a topological challenge in special situations during replication. However, this rotation should be kept to a minimum to prevent potential DNA damage.

In this study, we focused on the relatively simple T7 replication system and successfully observed the physical phenomenon of fork rotation. The T7 replisome serves as a good mechanical mimic for both *E. coli* and yeast replisomes (Gao et al., 2019). As discussed in the context of the transverse flow approach, this system can be adapted to study *E. coli* and yeast replication as well. Recent advances in reconstituting human replication systems also make it feasible to apply our proposed assay to human replisomes (Baris et al., 2022). Such experiments would yield more detailed information on compositional changes within the replication machinery, which could be further explored using the multi-color TIRF setup employed in this study. We conducted our experiments under low-force conditions. Modifying the applied force could alter the dynamics of fork rotation and force the replisome to introduce twists when constrained instead of performing rotational movement. This approach would help determine whether the replisome follows the helical path immediately or if twists

must accumulate first before being transferred behind the replication fork. Our study was limited to only two types of DNA substrates. However, in chapter 4, we outlined other potential substrates for future studies. Furthermore, the presented topological barrier serves as a trigger for fork rotation, which we could detect with our experimental setup. This has laid the groundwork for future research using additional substrates.

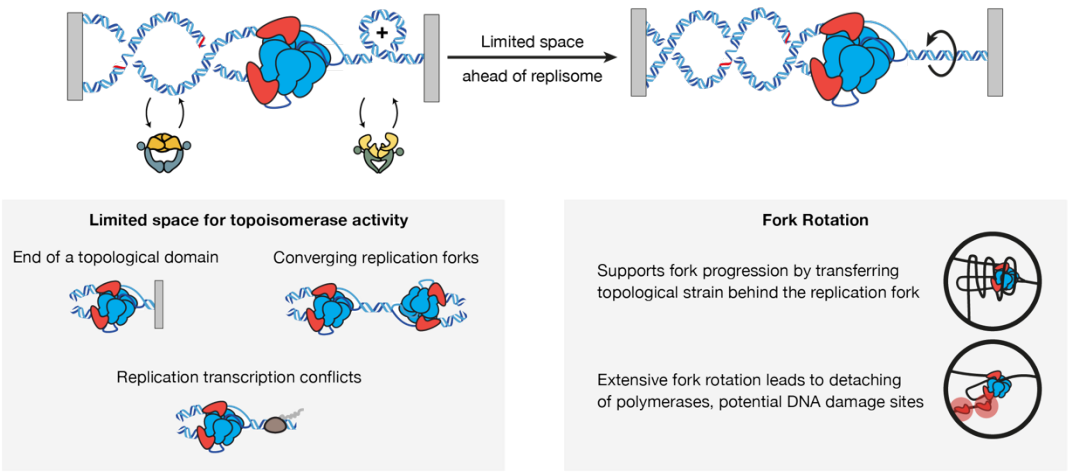


Figure 5-11 Fork rotation overcomes topological strain when space for topoisomerases is limited.
During DNA replication, excess torsional strain is typically relieved by topoisomerase activity ahead of and behind the replisome. However, when the replisome encounters situations where space for topoisomerase action becomes limited (indicated by the grey info box), the strain can be alleviated through fork rotation, thereby supporting the replication process. Fork rotation helps facilitate fork progression by transferring topological strain behind the replication fork. Nevertheless, excessive fork rotation may lead to the accumulation of DNA damage, which subsequently needs to be repaired.

In summary, the replisome can relieve topological strain ahead of the replication fork through fork rotation (Figure 5-11). This physical mechanism provides an immediate solution to the challenge of excess supercoiling during replication. However, excessive use of fork rotation can lead to DNA damage, posing another threat to chromosome integrity. Our novel imaging assay for replication, the transverse flow approach, enabled the first direct visualization of fork rotation at the single-molecule level. This study lays a foundation for future research aimed at better understanding the complex interplay between DNA topology, the replisome, and topoisomerases.

Appendix

Table 5-1 PDB IDs for fitting cryo-EM maps

Description	PDB ID	Publication
gp4 helicase	1E0J	Singleton et al. (2000)
Primase subunit of the helicase	1NUI	Kato et al. (2003)
gp5 Polymerase	1T8E	Brieba et al. (2004)

Table 5-2 EMDB Cryo maps used to create T7 replisome rendering

Description	EMDB ID	Publication
Helicase-Primase	EMD-0364	Gao et al. (2019)
Pol/trx lagging + Pol/trx apo	EMD-0386	Gao et al. (2019)
Pol/trx lagging strand	EMD-0387	Gao et al. (2019)
Pol/trx leading strand	EMD-0391	Gao et al. (2019)

Table 5-3 Mars commands.
Name and description of Fiji/ImageJ2 commands assisting the analysis of image-derived biomolecule data in Mars environment. Our documentation website (https://duderstadt-lab.github.io/mars-docs/) offers more detailed documentation of the different commands.

Command	Description
Image	
Peak Finder	Detects high-intensity pixel clusters (peaks) in an microscpe image. Additionally, 2D Gaussian fit can be utilized to find sub-pixel position.
DNA Finder	Identifies vertically aligned DNA molecules within an image and determines the sub-pixel positions of both ends of the molecule using a 2D Gaussian fit.
Peak Tracker	Detects, fits, and tracks peaks in images.
Object Tracker	Utilized for identifying unspecified objects in images through segmentation-based classification and tracking the position of the center of mass.
Molecule Integrator	Integrates the intensity of a peak over all frames.
Molecule Integrator (multiview)	Integrates the intensity of a peak over all frames in an image stack with multiview images.
Beam Profile Corrector	Corrects for image intensity deviations caused by the beam profile.

Gradient Calculator	Calculates the gradient between consecutive pixels either vertically or horizontally to identify long linear objects, such as DNA molecules.
Overlay channels	Combines multiple individual videos into a single video, with the information stored along the 'Channel (C)' dimension.
Molecule	
Open Archive	Opens a Molecule Archive.
Open Virtual Store	Opens a virtual Molecule Archive.
Build Archive from Table	Converts an opened table into a Molecule Archive. Needs a 'molecule' index column
Build DNA Archive	Constructs a DNA Molecule Archive from a single Molecule Archive and a list of DNA ROIs in the ROI Manager. It searches for molecules in the single Molecule Archive that overlap with parts of the specified DNA molecule locations.
Merge Archives	Merges multiple Molecule Archives into one. The archives need to be saved in a single folder.
Merge Virtual Stores	Same as 'Merge Archives' for virtual Molecule Archives.
Add Time	Converts time points (T) to real time values as specified by a user-defined time increment or in the metadata. Adds a column to the molecule tables.
Drift Corrector	Calculates and corrects for sample drift using a Molecule Archive and a tag identifying all immobile molecules in the dataset, and generates new columns for each molecule table.
Region Difference Calculator	Calculates the difference between the specified regions for all molecules in the Molecule Archive and adds the result as a molecule parameter.
Variance Calculator	Calculates the variance of a specified molecule table column and adds the result as a molecule parameter.
Table	
Open Table	Imports a comma or tab-delimited table to the MarsTable format.
Sort	Sorts a MarsTable based on values in a specified column.

Filter	Filters the rows of a MarsTable based on the specified criteria.
Import IJ1 Table	Imports any ImageJ1 table to the MarsTable format.
Import TableDisplay	Imports any SciJava table to the MarsTable format.
KCP	
Change Point Finder	Identifies linear regions or steps in single-molecule traces. This command generates molecule segment tables listing the endpoints and fits for the linear regions.
Single Change Point Finder	Identifies a single change point in a single-molecule trace. The output is a segments table that includes the endpoints and fit or position.
Sigma Calculator	Calculates the error value within a specified region of interest in all single-molecule traces, which can be used as input for the change point calculation commands.
ROI	
Transform ROIs	Transforms peak ROIs from one region of a multiview image to another.
Import	
LUMICKS h5	Opens optical tweezer data in HDF5 (h5) format collected with a LUMICKS instrument and converts it to Molecule Archive format.
Single-molecule dataset (SMD)	Opens SMD files in plaintext JSON format and converts the data to Molecule Archive format.

List of figures

List of tables

List of abbreviations

ATP – adenosine triphosphate

bp – base pair

CoSMoS – co-localization single-molecule spectroscopy

cryo-EM – Cryogenic electron microscopy

CTP – cytidine triphosphate

CV – column volume

DNA – deoxyribonucleic acid

dsDNA – double-stranded DNA

GTP – guanosine triphosphate

kb – kilobase pair

KCP – kinetic change point

kDa – kilodaltons

KDE – kernel density estimation

LB – Luria-Bertani

Mb – mega base

mmTIRF – micromirror TIRF

n – number of observations

nm – nanometer

NTP – nucleoside triphosphate

OF – Okazaki fragment

oriC – orgin of replication in *E. coli*

OSS – oxygen scavenging system

PCA – 3,4-Dihydroxybenzoic acid

PCD – Protocatechuate 3,4-Dioxygenase

PCR – polymerase chain reaction

PDB – Protein Data Bank

pN – pico Newton

Pol – polymerase

RNA – ribonucleic acid

RNAP – RNA polymerase

SD – standard deviation

SEM – standard error of the mean

SMC – structural maintenance of chromosomes

ssDNA – single-stranded DNA

TIRF – total internal reflection fluorescence

TIRFM – total internal reflection fluorescence microscopy

TOPcc – topoisomerase cleavage complexes

v/v – volume per volume

w/v – weight per volume

ACKNOWLEDGMENTS

I would like to express my gratitude to all the people whose support made this PhD journey possible—from its beginning, through its challenges, to its completion. Support comes in many forms, and the collective contributions of so many individuals have enabled me to finish this challenging and exciting journey.

Dear Karl, I cannot thank you enough for your unwavering support. When I approached you for a master thesis project nearly seven years ago, I could not have imagined that this conversation would mark the beginning of the most exciting and challenging period of my life. Your positive attitude and trust gave me the confidence to tackle every challenge that arose during my PhD. Your openness to ideas and suggestions made this experience truly enjoyable. Whenever I needed support, you were always there to help and find solutions. Thank you!

Thank you to my Thesis Advisory Committee members, Prof. Dr. Hannes Mutschler and Prof. Dr. Friedrich Simmel, for their valuable feedback and insightful discussions. I would also like to extend my gratitude to our collaborators, Prof. Dr. Samir Hamdan and Dr. Masateru Takahashi. Without their support and expertise, this project would not have been possible.

A big thank you goes out to all the past and present members of the Duderstadt Lab. Having such a supportive and vibrant environment made the PhD journey much more enjoyable. Special thanks to Matthias and Barbara for introducing me to the wonders of microscopy and biochemistry. Nadia, thank you for the great teamwork in realizing the Mars project. Anita and Rohit, I appreciate your supervision during my Master thesis project. To Hannah, Anna, Filip, Miriam, and Yu Chun, thank you for making the time in the lab so entertaining with your humor and random discussions. You brightened my mood every time. I especially would like to thank Lional for your endless patience and support since day one. I would always want to share an office with you and hearing about all your new hobbies. You always had an open ear for every question or request, no matter how small, and were always happy to discuss anything. Thank you!

The Max Planck Institute is filled with incredible people, and it is impossible to name everyone who had a positive impact on my PhD experience. I would especially like to thank Bianca for her support with any issue, whether it was related to contracts, travel, or the countless unexpected things that came up. Gerri and Chris, thank you both for your guidance, advice, and being there for coffee breaks – you helped me find clarity in a forest full of options. Lorenzo, the entire LM wing could not wish for a better colleague. You were the kindest and most helpful person, always ready for a joke.

I am deeply grateful to have been a part of the IMPRS graduate school. The graduate school brought together some of the most interesting people I have ever met and most of my favorite memories from this time came from IMPRS events. Throughout my PhD I had many worries but never about the support from IMPRS. Thank you to the former IMPRS-LS and the new IMPRS-ML coordinator office for always being there!

Two of the amazing people from IMPRS, who made a significant impact on my journey are Jacqueline and Pedro. Jacqueline, meeting you was the beginning of a wonderful friendship and I am incredibly grateful to have you as a friend. Our conversations about life in and outside the lab helped me navigate through difficult times. Pedro, you've been a part of my journey at the MPI since the start of my master thesis. Thank you for always being there, supporting me with your kind words and accepting me for who I am.

Thank you to all my close and long-distance friends. Alex, Andi, Flo, and Lukas, you remind me of a time when things were a bit simpler. I am incredibly grateful for our enduring friendship and your understanding.

Mama, Papa, Stefan, Laura und David: Danke, dass ihr immer für mich da wart. Ihr habt für Ablenkung gesorgt und mich manche Probleme für eine Weile vergessen lassen. Danke für euer Verständnis, dass meine Zeit manchmal knapp war, besonders in den letzten Monaten. Danke Papa, für die Retzer-Sturheit, die mich bis zum Ziel gebracht hat. Danke Mama, für die Gelassenheit, durch die ich in komplexen Situation Ruhe bewahren konnte. Irina, Valeri, Anton und Nele, danke für eure Unterstützung. Danke, dass ich Teil eurer Familie sein kann. Für die Pausen vom Alltag, die mich motiviert haben danach wieder durchzustarten.

Last but not least, I want to thank the most important person in my life. Marina, you have been my greatest supporter since day one. You believed in me when I couldn’t believe in myself. I cannot imagine my life without you. You are my best friend and your constant support has brought me to where I am today.

References

Abid Ali, F., Renault, L., Gannon, J., Gahlon, H. L., Kotecha, A., Zhou, J. C., Rueda, D., & Costa, A. (2016). Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate. *Nat Commun*, *7*, 10708. https://doi.org/10.1038/ncomms10708

Adachi, K., Oiwa, K., Nishizaka, T., Furuike, S., Noji, H., Itoh, H., Yoshida, M., & Kinosita, K., Jr. (2007). Coupling of rotation and catalysis in F(1)-ATPase revealed by single-molecule imaging and manipulation. *Cell*, *130*(2), 309-321. https://doi.org/10.1016/j.cell.2007.05.020

Agarwal, R., & Duderstadt, K. E. (2020). Multiplex flow magnetic tweezers reveal rare enzymatic events with single molecule precision. *Nat Commun*, *11*(1), 4714. https://doi.org/10.1038/s41467-020-18456-y

Aitken, C. E., Marshall, R. A., & Puglisi, J. D. (2008). An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys J*, *94*(5), 1826-1835. https://doi.org/10.1529/biophysj.107.117689

Akyuz, N., Altman, R. B., Blanchard, S. C., & Boudker, O. (2013). Transport dynamics in a glutamate transporter homologue. *Nature*, *502*(7469), 114-118. https://doi.org/10.1038/nature12265

Alberts, B., Heald, R., Johnson, A., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2022). *Molecular Biology of the Cell: Seventh Edition*. W. W. Norton & Company.

Allan, C., Burel, J. M., Moore, J., Blackburn, C., Linkert, M., Loynton, S., Macdonald, D., Moore, W. J., Neves, C., Patterson, A., Porter, M., Tarkowska, A., Loranger, B., Avondo, J., Lagerstedt, I., Lianas, L., Leo, S., Hands, K., Hay, R. T., . . . Swedlow, J. R. (2012). OMERO: flexible, model-driven data management for experimental biology. *Nat Methods*, *9*(3), 245-253. https://doi.org/10.1038/nmeth.1896

Amann, K. J., & Pollard, T. D. (2001). Direct real-time observation of actin filament branching mediated by Arp2/3 complex using total internal reflection fluorescence microscopy. *Proc Natl Acad Sci U S A*, *98*(26), 15009-15013. https://doi.org/10.1073/pnas.211556398

Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., Bao, B., Bell, P., Berard, D., Burovski, E., Chauhan, G., Chourdia, A., Constable, W., Desmaison, A., DeVito, Z., Ellison, E., Feng, W., Gong, J., Gschwind, M., . . . Chintala, S. (2024). *PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation* Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2,

Arias-Palomo, E., O'Shea, V. L., Hood, I. V., & Berger, J. M. (2013). The bacterial DnaC helicase loader is a DnaB ring breaker. *Cell*, *153*(2), 438-448. https://doi.org/10.1016/j.cell.2013.03.006

Arias-Palomo, E., Puri, N., O'Shea Murray, V. L., Yan, Q., & Berger, J. M. (2019). Physical Basis for the Loading of a Bacterial Replicative Helicase onto DNA. *Mol Cell*, *74*(1), 173-184 e174. https://doi.org/10.1016/j.molcel.2019.01.023

Arzt, M., Deschamps, J., Schmied, C., Pietzsch, T., Schmidt, D., Tomancak, P., Haase, R., & Jug, F. (2022). LABKIT: Labeling and Segmentation Toolkit

for Big Image Data. *Frontiers in Computer Science*, *4*. https://doi.org/10.3389/fcomp.2022.777728

Baek, I., Friedman, L. J., Gelles, J., & Buratowski, S. (2021). Single-molecule studies reveal branched pathways for activator-dependent assembly of RNA polymerase II pre-initiation complexes. *Mol Cell*, *81*(17), 3576-3588 e3576. https://doi.org/10.1016/j.molcel.2021.07.025

Baker, N. M., Weigand, S., Maar-Mathias, S., & Mondragon, A. (2011). Solution structures of DNA-bound gyrase. *Nucleic Acids Res*, *39*(2), 755-766. https://doi.org/10.1093/nar/gkq799

Baris, Y., Taylor, M. R. G., Aria, V., & Yeeles, J. T. P. (2022). Fast and efficient DNA replication with purified human proteins. *Nature*, *606*(7912), 204-210. https://doi.org/10.1038/s41586-022-04759-1

Bax, B. D., Murshudov, G., Maxwell, A., & Germe, T. (2019). DNA Topoisomerase Inhibitors: Trapping a DNA-Cleaving Machine in Motion. *J Mol Biol*, *431*(18), 3427-3449. https://doi.org/10.1016/j.jmb.2019.07.008

Baxter, J. (2015). "Breaking up is hard to do": the formation and resolution of sister chromatid intertwines. *J Mol Biol*, *427*(3), 590-607. https://doi.org/10.1016/j.jmb.2014.08.022

Beattie, T. R., Kapadia, N., Nicolas, E., Uphoff, S., Wollman, A. J., Leake, M. C., & Reyes-Lamothe, R. (2017). Frequent exchange of the DNA polymerase during bacterial chromosome replication. *Elife*, *6*. https://doi.org/10.7554/eLife.21763

Bell, N. A. W., & Molloy, J. E. (2022). Efficient golden gate assembly of DNA constructs for single molecule force spectroscopy and imaging. *Nucleic Acids Res*, *50*(13), e77. https://doi.org/10.1093/nar/gkac300

Berghuis, B. A., Dulin, D., Xu, Z. Q., van Laar, T., Cross, B., Janissen, R., Jergic, S., Dixon, N. E., Depken, M., & Dekker, N. H. (2015). Strand separation establishes a sustained lock at the Tus-Ter replication fork barrier. *Nat Chem Biol*, *11*(8), 579-585. https://doi.org/10.1038/nchembio.1857

Bermejo, R., Doksani, Y., Capra, T., Katou, Y. M., Tanaka, H., Shirahige, K., & Foiani, M. (2007). Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation. *Genes Dev*, *21*(15), 1921-1936. https://doi.org/10.1101/gad.432107

Bleichert, F., Botchan, M. R., & Berger, J. M. (2017). Mechanisms for initiating cellular DNA replication. *Science*, *355*(6327). https://doi.org/10.1126/science.aah6317

Brandao, H. B., Ren, Z., Karaboja, X., Mirny, L. A., & Wang, X. (2021). DNA-loop-extruding SMC complexes can traverse one another in vivo. *Nat Struct Mol Biol*, *28*(8), 642-651. https://doi.org/10.1038/s41594-021-00626-1

Breier, A. M., Weier, H. U., & Cozzarelli, N. R. (2005). Independence of replisomes in Escherichia coli chromosomal replication. *Proc Natl Acad Sci U S A*, *102*(11), 3942-3947. https://doi.org/10.1073/pnas.0500812102

Brieba, L. G., Eichman, B. F., Kokoska, R. J., Doublie, S., Kunkel, T. A., & Ellenberger, T. (2004). Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase. *EMBO J*, *23*(17), 3452-3461. https://doi.org/10.1038/sj.emboj.7600354

Brown, P. O., & Cozzarelli, N. R. (1979). A sign inversion mechanism for enzymatic supercoiling of DNA. *Science*, *206*(4422), 1081-1083. https://doi.org/10.1126/science.227059

Bryant, Z., Stone, M. D., Gore, J., Smith, S. B., Cozzarelli, N. R., & Bustamante, C. (2003). Structural transitions and elasticity from torque measurements on DNA. *Nature*, *424*(6946), 338-341. https://doi.org/10.1038/nature01810

Burmann, F., Funke, L. F. H., Chin, J. W., & Lowe, J. (2021). Cryo-EM structure of MukBEF reveals DNA loop entrapment at chromosomal unloading sites. *Mol Cell*, *81*(23), 4891-4906 e4898. https://doi.org/10.1016/j.molcel.2021.10.011

Burmann, F., & Lowe, J. (2023). Structural biology of SMC complexes across the tree of life. *Curr Opin Struct Biol*, *80*, 102598. https://doi.org/10.1016/j.sbi.2023.102598

Bush, N. G., Evans-Roberts, K., & Maxwell, A. (2015). DNA Topoisomerases. *EcoSal Plus*, *6*(2). https://doi.org/10.1128/ecosalplus.ESP-0010-2014

Bustamante, C., Marko, J. F., Siggia, E. D., & Smith, S. (1994). Entropic elasticity of lambda-phage DNA. *Science*, *265*(5178), 1599-1600. https://doi.org/10.1126/science.8079175

Carpenter, A. E., Kamentsky, L., & Eliceiri, K. W. (2012). A call for bioimaging software usability. *Nat Methods*, *9*(7), 666-670. https://doi.org/10.1038/nmeth.2073

Champoux, J. J., & Been, M. D. (1980). TOPOISOMERASES AND THE SWIVEL PROBLEM. In B. Alberts (Ed.), *Mechanistic Studies of DNA Replication and Genetic Recombination* (pp. 809-815). Academic Press. https://doi.org/10.1016/B978-0-12-048850-6.50072-7

Chen, J., Zhang, Z., Li, L., Chen, B. C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., Tjian, R., & Liu, Z. (2014). Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell*, *156*(6), 1274-1285. https://doi.org/10.1016/j.cell.2014.01.062

Chong, S., Chen, C., Ge, H., & Xie, X. S. (2014). Mechanism of transcriptional bursting in bacteria. *Cell*, *158*(2), 314-326. https://doi.org/10.1016/j.cell.2014.05.038

Cluzel, P., Lebrun, A., Heller, C., Lavery, R., Viovy, J. L., Chatenay, D., & Caron, F. (1996). DNA: an extensible molecule. *Science*, *271*(5250), 792-794. https://doi.org/10.1126/science.271.5250.792

Cobb, M., & Comfort, N. (2023). What Rosalind Franklin truly contributed to the discovery of DNA's structure. *Nature*, *616*(7958), 657-660. https://doi.org/10.1038/d41586-023-01313-5

Collin, F., Karkare, S., & Maxwell, A. (2011). Exploiting bacterial DNA gyrase as a drug target: current state and perspectives. *Appl Microbiol Biotechnol*, *92*(3), 479-497. https://doi.org/10.1007/s00253-011-3557-z

Crickard, J. B., Moevus, C. J., Kwon, Y., Sung, P., & Greene, E. C. (2020). Rad54 Drives ATP Hydrolysis-Dependent DNA Sequence Alignment during Homologous Recombination. *Cell*, *181*(6), 1380-1394 e1318. https://doi.org/10.1016/j.cell.2020.04.056

Curtis Rueden, E. E., Leon Yang, Michael Pinkert, Yang Liu, Mark Hiner, Jan Eglinger, Hadrien Mary, Macarse, Darío Hereñú, Philipp Hanslovsky, Wei Ouyang. (2021). Imagej/pyimagej: v1.0.2 (1.0.2). *Zenodo*. https://doi.org/10.5281/zenodo.5537065

Davidson, I. F., Bauer, B., Goetz, D., Tang, W., Wutz, G., & Peters, J. M. (2019). DNA loop extrusion by human cohesin. *Science*, *366*(6471), 1338-1345. https://doi.org/10.1126/science.aaz3418

Dequeker, B. J. H., Scherr, M. J., Brandao, H. B., Gassler, J., Powell, S., Gaspar, I., Flyamer, I. M., Lalic, A., Tang, W., Stocsits, R., Davidson, I. F., Peters, J. M., Duderstadt, K. E., Mirny, L. A., & Tachibana, K. (2022). MCM complexes are barriers that restrict cohesin-mediated loop extrusion. *Nature*, *606*(7912), 197-203. https://doi.org/10.1038/s41586-022-04730-0

Deweese, J. E., Osheroff, M. A., & Osheroff, N. (2008). DNA Topology and Topoisomerases: Teaching a "Knotty" Subject. *Biochem Mol Biol Educ*, *37*(1), 2-10. https://doi.org/10.1002/bmb.20244

Dey, S., Fan, C., Gothelf, K. V., Li, J., Lin, C., Liu, L., Liu, N., Nijenhuis, M. A. D., Saccà, B., Simmel, F. C., Yan, H., & Zhan, P. (2021). DNA origami. *Nature Reviews Methods Primers*, *1*(1). https://doi.org/10.1038/s43586-020-00009-8

Dillon, S. C., & Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol*, *8*(3), 185-195. https://doi.org/10.1038/nrmicro2261

Duchi, D., Bauer, D. L., Fernandez, L., Evans, G., Robb, N., Hwang, L. C., Gryte, K., Tomescu, A., Zawadzki, P., Morichaud, Z., Brodolin, K., & Kapanidis, A. N. (2016). RNA Polymerase Pausing during Initial Transcription. *Mol Cell*, *63*(6), 939-950. https://doi.org/10.1016/j.molcel.2016.08.011

Duderstadt, K. E., & Berger, J. M. (2013). A structural framework for replication origin opening by AAA+ initiation factors. *Curr Opin Struct Biol*, *23*(1), 144-153. https://doi.org/10.1016/j.sbi.2012.11.012

Duderstadt, K. E., Chuang, K., & Berger, J. M. (2011). DNA stretching by bacterial initiators promotes replication origin opening. *Nature*, *478*(7368), 209-213. https://doi.org/10.1038/nature10455

Duderstadt, K. E., Geertsema, H. J., Stratmann, S. A., Punter, C. M., Kulczyk, A. W., Richardson, C. C., & van Oijen, A. M. (2016). Simultaneous Real-Time Imaging of Leading and Lagging Strand Synthesis Reveals the Coordination Dynamics of Single Replisomes. *Mol Cell*, *64*(6), 1035-1047. https://doi.org/10.1016/j.molcel.2016.10.028

Dulin, D., Lipfert, J., Moolman, M. C., & Dekker, N. H. (2013). Studying genomic processes at the single-molecule level: introducing the tools and applications. *Nat Rev Genet*, *14*(1), 9-22. https://doi.org/10.1038/nrg3316

Edelstein, A., Amodaj, N., Hoover, K., Vale, R., & Stuurman, N. (2010). Computer control of microscopes using microManager. *Curr Protoc Mol Biol*, *Chapter 14*, Unit14 20. https://doi.org/10.1002/0471142727.mb1420s92

Eliceiri, K. W., Berthold, M. R., Goldberg, I. G., Ibanez, L., Manjunath, B. S., Martone, M. E., Murphy, R. F., Peng, H., Plant, A. L., Roysam, B., Stuurman, N., Swedlow, J. R., Tomancak, P., & Carpenter, A. E. (2012). Biological imaging software tools. *Nat Methods*, *9*(7), 697-710. https://doi.org/10.1038/nmeth.2084

Elshenawy, M. M., Jergic, S., Xu, Z. Q., Sobhy, M. A., Takahashi, M., Oakley, A. J., Dixon, N. E., & Hamdan, S. M. (2015). Replisome speed determines the efficiency of the Tus-Ter replication termination barrier. *Nature*, *525*(7569), 394-398. https://doi.org/10.1038/nature14866

Erkens, G. B., Hanelt, I., Goudsmits, J. M., Slotboom, D. J., & van Oijen, A. M. (2013). Unsynchronised subunit motion in single trimeric sodium-coupled aspartate transporters. *Nature*, *502*(7469), 119-123. https://doi.org/10.1038/nature12538

Erlich, Y., & Zielinski, D. (2017). DNA Fountain enables a robust and efficient storage architecture. *Science*, *355*(6328), 950-954. https://doi.org/10.1126/science.aaj2038

Ershov, D., Phan, M. S., Pylvanainen, J. W., Rigaud, S. U., Le Blanc, L., Charles-Orszag, A., Conway, J. R. W., Laine, R. F., Roy, N. H., Bonazzi, D., Dumenil, G., Jacquemet, G., & Tinevez, J. Y. (2022). TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. *Nat Methods*, *19*(7), 829-832. https://doi.org/10.1038/s41592-022-01507-1

Fernandez-Leiro, R., Conrad, J., Scheres, S. H., & Lamers, M. H. (2015). cryo-EM structures of the E. coli replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and tau. *Elife*, *4*. https://doi.org/10.7554/eLife.11134

Forth, S., Deufel, C., Sheinin, M. Y., Daniels, B., Sethna, J. P., & Wang, M. D. (2008). Abrupt buckling transition observed during the plectoneme formation of individual DNA molecules. *Phys Rev Lett*, *100*(14), 148301. https://doi.org/10.1103/PhysRevLett.100.148301

Franklin, R. E., & Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. *Nature*, *171*(4356), 740-741. https://doi.org/10.1038/171740a0

Fulcrand, G., Dages, S., Zhi, X., Chapagain, P., Gerstman, B. S., Dunlap, D., & Leng, F. (2016). DNA supercoiling, a critical signal regulating the basal expression of the lac operon in Escherichia coli. *Sci Rep*, *6*, 19243. https://doi.org/10.1038/srep19243

Fuller, F. B. (1971). The writhing number of a space curve. *Proc Natl Acad Sci U S A*, *68*(4), 815-819. https://doi.org/10.1073/pnas.68.4.815

Gan, H., Serra-Cardona, A., Hua, X., Zhou, H., Labib, K., Yu, C., & Zhang, Z. (2018). The Mcm2-Ctf4-Polalpha Axis Facilitates Parental Histone H3-H4 Transfer to Lagging Strands. *Mol Cell*, *72*(1), 140-151 e143. https://doi.org/10.1016/j.molcel.2018.09.001

Ganji, M., Kim, S. H., van der Torre, J., Abbondanzieri, E., & Dekker, C. (2016). Intercalation-Based Single-Molecule Fluorescence Assay To Study DNA Supercoil Dynamics. *Nano Lett*, *16*(7), 4699-4707. https://doi.org/10.1021/acs.nanolett.6b02213

Ganji, M., Shaltiel, I. A., Bisht, S., Kim, E., Kalichava, A., Haering, C. H., & Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. *Science*, *360*(6384), 102-105. https://doi.org/10.1126/science.aar7831

Gao, L., Shao, L., Higgins, C. D., Poulton, J. S., Peifer, M., Davidson, M. W., Wu, X., Goldstein, B., & Betzig, E. (2012). Noninvasive imaging beyond the diffraction limit of 3D dynamics in thickly fluorescent specimens. *Cell*, *151*(6), 1370-1385. https://doi.org/10.1016/j.cell.2012.10.008

Gao, Y., Cui, Y., Fox, T., Lin, S., Wang, H., de Val, N., Zhou, Z. H., & Yang, W. (2019). Structures and operating principles of the replisome. *Science*, *363*(6429). https://doi.org/10.1126/science.aav7003

Garcia, H. G., Grayson, P., Han, L., Inamdar, M., Kondev, J., Nelson, P. C., Phillips, R., Widom, J., & Wiggins, P. A. (2007). Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. *Biopolymers*, *85*(2), 115-130. https://doi.org/10.1002/bip.20627

Geertsema, H. J., Kulczyk, A. W., Richardson, C. C., & van Oijen, A. M. (2014). Single-molecule studies of polymerase dynamics and stoichiometry at the bacteriophage T7 replication machinery. *Proc Natl Acad Sci U S A*, *111*(11), 4073-4078. https://doi.org/10.1073/pnas.1402010111

Georgescu, R., Yuan, Z., Bai, L., de Luna Almeida Santos, R., Sun, J., Zhang, D., Yurieva, O., Li, H., & O'Donnell, M. E. (2017). Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. *Proc Natl Acad Sci U S A*, *114*(5), E697-E706. https://doi.org/10.1073/pnas.1620500114

Gervasio, J., da Costa Oliveira, H., da Costa Martins, A. G., Pesquero, J. B., Verona, B. M., & Cerize, N. N. P. (2024). How close are we to storing data in DNA? *Trends Biotechnol*, *42*(2), 156-167. https://doi.org/10.1016/j.tibtech.2023.08.001

Goldberg, I. G., Allan, C., Burel, J. M., Creager, D., Falconi, A., Hochheiser, H., Johnston, J., Mellen, J., Sorger, P. K., & Swedlow, J. R. (2005). The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biol*, *6*(5), R47. https://doi.org/10.1186/gb-2005-6-5-r47

Gore, J., Bryant, Z., Stone, M. D., Nollmann, M., Cozzarelli, N. R., & Bustamante, C. (2006). Mechanochemical analysis of DNA gyrase using rotor bead tracking. *Nature*, *439*(7072), 100-104. https://doi.org/10.1038/nature04319

Greenfeld, M., van de Meent, J. W., Pavlichin, D. S., Mabuchi, H., Wiggins, C. H., Gonzalez, R. L., Jr., & Herschlag, D. (2015). Single-molecule dataset (SMD): a generalized storage format for raw and processed single-molecule data. *BMC Bioinformatics*, *16*(1), 3. https://doi.org/10.1186/s12859-014-0429-4

Gudnason, H., Dufva, M., Bang, D. D., & Wolff, A. (2007). Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature. *Nucleic Acids Res*, *35*(19), e127. https://doi.org/10.1093/nar/gkm671

Guo, M. S., Kawamura, R., Littlehale, M. L., Marko, J. F., & Laub, M. T. (2021). High-resolution, genome-wide mapping of positive supercoiling in chromosomes. *Elife*, *10*. https://doi.org/10.7554/eLife.67236

Ha, T., Rasnik, I., Cheng, W., Babcock, H. P., Gauss, G. H., Lohman, T. M., & Chu, S. (2002). Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature*, *419*(6907), 638-641. https://doi.org/10.1038/nature01083

Hamdan, S. M., Loparo, J. J., Takahashi, M., Richardson, C. C., & van Oijen, A. M. (2009). Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. *Nature*, *457*(7227), 336-339. https://doi.org/10.1038/nature07512

Harada, Y., Ohara, O., Takatsuki, A., Itoh, H., Shimamoto, N., & Kinosita, K., Jr. (2001). Direct observation of DNA rotation during transcription by Escherichia coli RNA polymerase. *Nature*, *409*(6816), 113-115. https://doi.org/10.1038/35051126

Helgesen, E., Saetre, F., & Skarstad, K. (2021). Topoisomerase IV tracks behind the replication fork and the SeqA complex during DNA replication in Escherichia coli. *Sci Rep*, *11*(1), 474. https://doi.org/10.1038/s41598-020-80043-4

Hernandez, A. J., & Richardson, C. C. (2019). Gp2.5, the multifunctional bacteriophage T7 single-stranded DNA binding protein. *Semin Cell Dev Biol*, *86*, 92-101. https://doi.org/10.1016/j.semcdb.2018.03.018

Hiasa, H., & Marians, K. J. (1994). Topoisomerase III, but not topoisomerase I, can support nascent chain elongation during theta-type DNA replication. *Journal of Biological Chemistry*, *269*(51), 32655-32659. https://doi.org/10.1016/s0021-9258(18)31684-3

Hill, F. R., van Oijen, A. M., & Duderstadt, K. E. (2018). Detection of kinetic change points in piece-wise linear single molecule motion. *J Chem Phys*, *148*(12), 123317. https://doi.org/10.1063/1.5009387

Hiner, M. C., Rueden, C. T., & Eliceiri, K. W. (2016). SCIFIO: an extensible framework to support scientific image formats. *BMC Bioinformatics*, *17*(1), 521. https://doi.org/10.1186/s12859-016-1383-0

Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. *Nat Rev Mol Cell Biol*, *7*(5), 311-322. https://doi.org/10.1038/nrm1909

Holmes, V. F., & Cozzarelli, N. R. (2000). Closing the ring: links between SMC proteins and chromosome partitioning, condensation, and supercoiling. *Proc Natl Acad Sci U S A*, *97*(4), 1322-1324. https://doi.org/10.1073/pnas.040576797

Huisjes, N. M., Retzer, T. M., Scherr, M. J., Agarwal, R., Rajappa, L., Safaric, B., Minnen, A., & Duderstadt, K. E. (2022). Mars, a molecule archive suite for reproducible analysis and reporting of single-molecule properties from bioimages. *Elife*, *11*, e75899. https://doi.org/10.7554/eLife.75899

Ingargiola, A., Laurence, T., Boutelle, R., Weiss, S., & Michalet, X. (2016). Photon-HDF5: Open Data Format and Computational Tools for Timestamp-based Single-Molecule Experiments. *Proc SPIE Int Soc Opt Eng*, *9714*. https://doi.org/10.1117/12.2212085

Itsathitphaisarn, O., Wing, R. A., Eliason, W. K., Wang, J., & Steitz, T. A. (2012). The hexameric helicase DnaB adopts a nonplanar conformation during translocation. *Cell*, *151*(2), 267-277. https://doi.org/10.1016/j.cell.2012.09.014

Janissen, R., Barth, R., Polinder, M., van der Torre, J., & Dekker, C. (2024). Single-molecule visualization of twin-supercoiled domains generated during transcription. *Nucleic Acids Res*, *52*(4), 1677-1687. https://doi.org/10.1093/nar/gkad1181

Jergic, S., Horan, N. P., Elshenawy, M. M., Mason, C. E., Urathamakul, T., Ozawa, K., Robinson, A., Goudsmits, J. M., Wang, Y., Pan, X., Beck, J. L., van Oijen, A. M., Huber, T., Hamdan, S. M., & Dixon, N. E. (2013). A direct proofreader-clamp interaction stabilizes the Pol III replicase in the polymerization mode. *EMBO J*, *32*(9), 1322-1333. https://doi.org/10.1038/emboj.2012.347

Johnson, D. E., & Richardson, C. C. (2003). A covalent linkage between the gene 5 DNA polymerase of bacteriophage T7 and Escherichia coli thioredoxin, the processivity factor: fate of thioredoxin during DNA synthesis. *J Biol Chem*, *278*(26), 23762-23772. https://doi.org/10.1074/jbc.M301366200

Joyeux, M. (2020). Bacterial Nucleoid: Interplay of DNA Demixing and Supercoiling. *Biophys J*, *118*(9), 2141-2150. https://doi.org/10.1016/j.bpj.2019.09.026

Jun, S., & Wright, A. (2010). Entropy as the driver of chromosome segregation. *Nat Rev Microbiol*, *8*(8), 600-607. https://doi.org/10.1038/nrmicro2391

Kapadia, N., El-Hajj, Z. W., & Reyes-Lamothe, R. (2021). Bound2Learn: a machine learning approach for classification of DNA-bound proteins from single-molecule tracking experiments. *Nucleic Acids Res*, *49*(14), e79. https://doi.org/10.1093/nar/gkab186

Kapadia, N., El-Hajj, Z. W., Zheng, H., Beattie, T. R., Yu, A., & Reyes-Lamothe, R. (2020). Processive Activity of Replicative DNA Polymerases in the Replisome of Live Eukaryotic Cells. *Mol Cell*, *80*(1), 114-126 e118. https://doi.org/10.1016/j.molcel.2020.08.014

Kato, M., Ito, T., Wagner, G., Richardson, C. C., & Ellenberger, T. (2003). Modular architecture of the bacteriophage T7 primase couples RNA primer synthesis to DNA synthesis. *Mol Cell*, *11*(5), 1349-1360. https://doi.org/10.1016/s1097-2765(03)00195-3

Kavenoff, R., & Ryder, O. A. (1976). Electron microscopy of membrane-associated folded chromosomes of Escherichia coli. *Chromosoma*, *55*(1), 13-25. https://doi.org/10.1007/BF00288323

Keszthelyi, A., Minchell, N. E., & Baxter, J. (2016). The Causes and Consequences of Topological Stress during DNA Replication. *Genes (Basel)*, *7*(12). https://doi.org/10.3390/genes7120134

Khan, M. I., Ferdous, S. F., & Adnan, A. (2021). Mechanical Behavior of Axonal Actin, Spectrin, and Their Periodic Structure: A Brief Review. *Multiscale Science and Engineering*, *3*(3-4), 185-204. https://doi.org/10.1007/s42493-021-00069-2

Kim, S. H., Ganji, M., Kim, E., van der Torre, J., Abbondanzieri, E., & Dekker, C. (2018). DNA sequence encodes the position of DNA supercoils. *Elife*, *7*. https://doi.org/10.7554/eLife.36557

Kolbeck, P. J., Tisma, M., Analikwu, B. T., Vanderlinden, W., Dekker, C., & Lipfert, J. (2024). Supercoiling-dependent DNA binding: quantitative modeling and applications to bulk and single-molecule experiments. *Nucleic Acids Res*, *52*(1), 59-72. https://doi.org/10.1093/nar/gkad1055

Koster, D. A., Crut, A., Shuman, S., Bjornsti, M. A., & Dekker, N. H. (2010). Cellular strategies for regulating DNA supercoiling: a single-molecule perspective. *Cell*, *142*(4), 519-530. https://doi.org/10.1016/j.cell.2010.08.001

Kosuri, P., Altheimer, B. D., Dai, M., Yin, P., & Zhuang, X. (2019). Rotation tracking of genome-processing enzymes using DNA origami rotors. *Nature*, *572*(7767), 136-140. https://doi.org/10.1038/s41586-019-1397-7

Kreuzer, K. N., & Cozzarelli, N. R. (1979). Escherichia coli mutants thermosensitive for deoxyribonucleic acid gyrase subunit A: effects on deoxyribonucleic acid replication, transcription, and bacteriophage growth. *J Bacteriol*, *140*(2), 424-435. https://doi.org/10.1128/jb.140.2.424-435.1979

Kriegel, F., Ermann, N., Forbes, R., Dulin, D., Dekker, N. H., & Lipfert, J. (2017). Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque tweezers. *Nucleic Acids Res*, *45*(10), 5920-5929. https://doi.org/10.1093/nar/gkx280

Kuhn, H., Demidov, V. V., & Frank-Kamenetskii, M. D. (2002). Rolling-circle amplification under topological constraints. *Nucleic Acids Res*, *30*(2), 574-580. https://doi.org/10.1093/nar/30.2.574

Kuhn, W. (1934). Über die Gestalt fadenförmiger Moleküle in Lösungen. *Kolloid-Zeitschrift*, *68*(1), 2-15. https://doi.org/10.1007/bf01451681

Kulczyk, A. W., Akabayov, B., Lee, S. J., Bostina, M., Berkowitz, S. A., & Richardson, C. C. (2012). An interaction between DNA polymerase and helicase is essential for the high processivity of the bacteriophage T7 replisome. *J Biol Chem*, *287*(46), 39050-39060. https://doi.org/10.1074/jbc.M112.410647

Kulczyk, A. W., Moeller, A., Meyer, P., Sliz, P., & Richardson, C. C. (2017). Cryo-EM structure of the replisome reveals multiple interactions coordinating DNA synthesis. *Proc Natl Acad Sci U S A*, *114*(10), E1848-E1856. https://doi.org/10.1073/pnas.1701252114

Kulczyk, A. W., & Richardson, C. C. (2016). The Replication System of Bacteriophage T7. *Enzymes*, *39*, 89-136. https://doi.org/10.1016/bs.enz.2016.02.001

Kurth, I., Georgescu, R. E., & O'Donnell, M. E. (2013). A solution to release twisted DNA during chromosome replication by coupled DNA polymerases. *Nature*, *496*(7443), 119-122. https://doi.org/10.1038/nature11988

Kvilekval, K., Fedorov, D., Obara, B., Singh, A., & Manjunath, B. S. (2010). Bisque: a platform for bioimage analysis and management. *Bioinformatics*, *26*(4), 544-552. https://doi.org/10.1093/bioinformatics/btp699

Larson, J., Kirk, M., Drier, E. A., O'Brien, W., MacKay, J. F., Friedman, L. J., & Hoskins, A. A. (2014). Design and construction of a multiwavelength, micromirror total internal reflectance fluorescence microscope. *Nat Protoc*, *9*(10), 2317-2328. https://doi.org/10.1038/nprot.2014.155

Leake, M. C. (2021). Correlative approaches in single-molecule biophysics: A review of the progress in methods and applications. *Methods*, *193*, 1-4. https://doi.org/10.1016/j.ymeth.2021.06.012

Lebel, P., Basu, A., Oberstrass, F. C., Tretter, E. M., & Bryant, Z. (2014). Gold rotor bead tracking for high-speed measurements of DNA twist, torque and extension. *Nat Methods*, *11*(4), 456-462. https://doi.org/10.1038/nmeth.2854

Lee, J. B., Hite, R. K., Hamdan, S. M., Xie, X. S., Richardson, C. C., & van Oijen, A. M. (2006). DNA primase acts as a molecular brake in DNA replication. *Nature*, *439*(7076), 621-624. https://doi.org/10.1038/nature04317

Lee, S. J., & Richardson, C. C. (2011). Choreography of bacteriophage T7 DNA replication. *Curr Opin Chem Biol*, *15*(5), 580-586. https://doi.org/10.1016/j.cbpa.2011.07.024

Leng, F., Amado, L., & McMacken, R. (2004). Coupling DNA supercoiling to transcription in defined protein systems. *J Biol Chem*, *279*(46), 47564-47571. https://doi.org/10.1074/jbc.M403798200

Lerner, E., Barth, A., Hendrix, J., Ambrose, B., Birkedal, V., Blanchard, S. C., Borner, R., Sung Chung, H., Cordes, T., Craggs, T. D., Deniz, A. A., Diao, J., Fei, J., Gonzalez, R. L., Gopich, I. V., Ha, T., Hanke, C. A., Haran, G., Hatzakis, N. S., . . . Weiss, S. (2021). FRET-based dynamic structural

biology: Challenges, perspectives and an appeal for open-science practices. *Elife*, *10*. https://doi.org/10.7554/eLife.60416

Lewis, J. S., Spenkelink, L. M., Jergic, S., Wood, E. A., Monachino, E., Horan, N. P., Duderstadt, K. E., Cox, M. M., Robinson, A., Dixon, N. E., & van Oijen, A. M. (2017). Single-molecule visualization of fast polymerase turnover in the bacterial replisome. *Elife*, *6*. https://doi.org/10.7554/eLife.23932

Lewis, J. S., Spenkelink, L. M., Schauer, G. D., Hill, F. R., Georgescu, R. E., O'Donnell, M. E., & van Oijen, A. M. (2017). Single-molecule visualization of Saccharomyces cerevisiae leading-strand synthesis reveals dynamic interaction between MTC and the replisome. *Proc Natl Acad Sci U S A*, *114*(40), 10630-10635. https://doi.org/10.1073/pnas.1711291114

Lewis, J. S., Spenkelink, L. M., Schauer, G. D., Yurieva, O., Mueller, S. H., Natarajan, V., Kaur, G., Maher, C., Kay, C., O'Donnell, M. E., & van Oijen, A. M. (2020). Tunability of DNA Polymerase Stability during Eukaryotic DNA Replication. *Mol Cell*, *77*(1), 17-25 e15. https://doi.org/10.1016/j.molcel.2019.10.005

Linkert, M., Rueden, C. T., Allan, C., Burel, J. M., Moore, W., Patterson, A., Loranger, B., Moore, J., Neves, C., Macdonald, D., Tarkowska, A., Sticco, C., Hill, E., Rossner, M., Eliceiri, K. W., & Swedlow, J. R. (2010). Metadata matters: access to image data in the real world. *J Cell Biol*, *189*(5), 777-782. https://doi.org/10.1083/jcb.201004104

Liu, L. F., & Wang, J. C. (1987). Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A*, *84*(20), 7024-7027. https://doi.org/10.1073/pnas.84.20.7024

Loparo, J. J., Kulczyk, A. W., Richardson, C. C., & van Oijen, A. M. (2011). Simultaneous single-molecule measurements of phage T7 replisome composition and function reveal the mechanism of polymerase exchange. *Proc Natl Acad Sci U S A*, *108*(9), 3584-3589. https://doi.org/10.1073/pnas.1018824108

Ma, J., Bai, L., & Wang, M. D. (2013). Transcription under torsion. *Science*, *340*(6140), 1580-1583. https://doi.org/10.1126/science.1235441

Makowska-Grzyska, M., & Kaguni, J. M. (2010). Primase directs the release of DnaC from DnaB. *Mol Cell*, *37*(1), 90-101. https://doi.org/10.1016/j.molcel.2009.12.031

Marko, J. F., & Siggia, E. D. (1995). Stretching DNA. *Macromolecules*, *28*(26), 8759-8770. https://doi.org/10.1021/ma00130a008

Martín Abadi, A. A., Paul Barham, Eugene Brevdo,, Zhifeng Chen, C. C., Greg S. Corrado, Andy Davis,, Jeffrey Dean, M. D., Sanjay Ghemawat, Ian Goodfellow,, Andrew Harp, G. I., Michael Isard, Rafal Jozefowicz, Yangqing Jia,, Lukasz Kaiser, M. K., Josh Levenberg, Dan Mané, Mike Schuster,, Rajat Monga, S. M., Derek Murray, Chris Olah, Jonathon Shlens,, Benoit Steiner, I. S., Kunal Talwar, Paul Tucker,, Vincent Vanhoucke, V. V., Fernanda Viégas,, Oriol Vinyals, P. W., Martin Wattenberg, Martin Wicke,, & Yuan Yu, a. X. Z. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://doi.org/10.5281/zenodo.12726004

McGlynn, P., Lloyd, R. G., & Marians, K. J. (2001). Formation of Holliday junctions by regression of nascent DNA in intermediates containing stalled replication forks: RecG stimulates regression even when the DNA

is negatively supercoiled. *Proc Natl Acad Sci U S A*, *98*(15), 8235-8240. https://doi.org/10.1073/pnas.121007798

McKie, S. J., Neuman, K. C., & Maxwell, A. (2021). DNA topoisomerases: Advances in understanding of cellular roles and multi-protein complexes via structure-function analysis. *Bioessays*, *43*(4), e2000286. https://doi.org/10.1002/bies.202000286

Meijering, E., Carpenter, A. E., Peng, H., Hamprecht, F. A., & Olivo-Marin, J. C. (2016). Imagining the future of bioimage analysis. *Nat Biotechnol*, *34*(12), 1250-1255. https://doi.org/10.1038/nbt.3722

Monachino, E., Ghodke, H., Spinks, R. R., Hoatson, B. S., Jergic, S., Xu, Z. Q., Dixon, N. E., & van Oijen, A. M. (2018). Design of DNA rolling-circle templates with controlled fork topology to study mechanisms of DNA replication. *Anal Biochem*, *557*, 42-45. https://doi.org/10.1016/j.ab.2018.07.008

Monachino, E., Jergic, S., Lewis, J. S., Xu, Z. Q., Lo, A. T. Y., O'Shea, V. L., Berger, J. M., Dixon, N. E., & van Oijen, A. M. (2020). A Primase-Induced Conformational Switch Controls the Stability of the Bacterial Replisome. *Mol Cell*, *79*(1), 140-154 e147. https://doi.org/10.1016/j.molcel.2020.04.037

Mueller, S. H., Spenkelink, L. M., van Oijen, A. M., & Lewis, J. S. (2020). Design of customizable long linear DNA substrates with controlled end modifications for single-molecule studies. *Anal Biochem*, *592*, 113541. https://doi.org/10.1016/j.ab.2019.113541

Mulcair, M. D., Schaeffer, P. M., Oakley, A. J., Cross, H. F., Neylon, C., Hill, T. M., & Dixon, N. E. (2006). A molecular mousetrap determines polarity of termination of DNA replication in E. coli. *Cell*, *125*(7), 1309-1319. https://doi.org/10.1016/j.cell.2006.04.040

Nasmyth, K., & Haering, C. H. (2005). The structure and function of SMC and kleisin complexes. *Annu Rev Biochem*, *74*, 595-648. https://doi.org/10.1146/annurev.biochem.74.082803.133219

Nelson, D. L., & Cox, M. (2021). *Lehninger Principles of Biochemistry* (Vol. Eighth Edition). Macmillan Learning.

Nelson, P. C., Radosavljević, M., & Bromberg, S. . (2014). *Biological Physics : Energy, Information, Life / with new art by David Goodsell*. W. H. Freeman and Company New York.

Neuman, K. C., & Nagy, A. (2008). Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Methods*, *5*(6), 491-505. https://doi.org/10.1038/nmeth.1218

Neylon, C., Kralicek, A. V., Hill, T. M., & Dixon, N. E. (2005). Replication termination in Escherichia coli: structure and antihelicase activity of the Tus-Ter complex. *Microbiol Mol Biol Rev*, *69*(3), 501-526. https://doi.org/10.1128/MMBR.69.3.501-526.2005

Niekamp, S., Stuurman, N., Zhang, N., & Vale, R. D. (2021). Three-color single-molecule imaging reveals conformational dynamics of dynein undergoing motility. *Proc Natl Acad Sci U S A*, *118*(31). https://doi.org/10.1073/pnas.2101391118

Noji, H., Yasuda, R., Yoshida, M., & Kinosita, K., Jr. (1997). Direct observation of the rotation of F1-ATPase. *Nature*, *386*(6622), 299-302. https://doi.org/10.1038/386299a0

Nolivos, S., Upton, A. L., Badrinarayanan, A., Muller, J., Zawadzka, K., Wiktor, J., Gill, A., Arciszewska, L., Nicolas, E., & Sherratt, D. (2016). MatP regulates the coordinated action of topoisomerase IV and MukBEF in chromosome segregation. *Nat Commun*, *7*, 10466. https://doi.org/10.1038/ncomms10466

Nollmann, M., Crisona, N. J., & Arimondo, P. B. (2007). Thirty years of Escherichia coli DNA gyrase: from in vivo function to single-molecule mechanism. *Biochimie*, *89*(4), 490-499. https://doi.org/10.1016/j.biochi.2007.02.012

Nollmann, M., Stone, M. D., Bryant, Z., Gore, J., Crisona, N. J., Hong, S. C., Mitelheiser, S., Maxwell, A., Bustamante, C., & Cozzarelli, N. R. (2007). Multiple modes of Escherichia coli DNA gyrase activity revealed by force and torque. *Nat Struct Mol Biol*, *14*(4), 264-271. https://doi.org/10.1038/nsmb1213

O'Donnell, M., Langston, L., & Stillman, B. (2013). Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb Perspect Biol*, *5*(7). https://doi.org/10.1101/cshperspect.a010108

Ouyang, W., & Zimmer, C. (2017). The imaging tsunami: Computational opportunities and challenges. *Current Opinion in Systems Biology*, *4*, 105-113. https://doi.org/10.1016/j.coisb.2017.07.011

Paik, D. H., Roskens, V. A., & Perkins, T. T. (2013). Torsionally constrained DNA for single-molecule assays: an efficient, ligation-free method. *Nucleic Acids Res*, *41*(19), e179. https://doi.org/10.1093/nar/gkt699

Papini, F. S., Seifert, M., & Dulin, D. (2019). High-yield fabrication of DNA and RNA constructs for single molecule force and torque spectroscopy experiments. *Nucleic Acids Res*, *47*(22), e144. https://doi.org/10.1093/nar/gkz851

Perkins, T. T., Smith, D. E., & Chu, S. (1994). Direct observation of tube-like motion of a single polymer chain. *Science*, *264*(5160), 819-822. https://doi.org/10.1126/science.8171335

Peter, B. J., Ullsperger, C., Hiasa, H., Marians, K. J., & Cozzarelli, N. R. (1998). The structure of supercoiled intermediates in DNA replication. *Cell*, *94*(6), 819-827. https://doi.org/10.1016/s0092-8674(00)81740-7

Pietzsch, T., Preibisch, S., Tomancak, P., & Saalfeld, S. (2012). ImgLib2--generic image processing in Java. *Bioinformatics*, *28*(22), 3009-3011. https://doi.org/10.1093/bioinformatics/bts543

Pietzsch, T., Saalfeld, S., Preibisch, S., & Tomancak, P. (2015). BigDataViewer: visualization and processing for large image data sets. *Nat Methods*, *12*(6), 481-483. https://doi.org/10.1038/nmeth.3392

Pommier, Y. (2013). Drugging topoisomerases: lessons and challenges. *ACS Chem Biol*, *8*(1), 82-95. https://doi.org/10.1021/cb300648v

Pommier, Y., Nussenzweig, A., Takeda, S., & Austin, C. (2022). Human topoisomerases and their roles in genome stability and organization. *Nat Rev Mol Cell Biol*, *23*(6), 407-427. https://doi.org/10.1038/s41580-022-00452-3

Pommier, Y., Sun, Y., Huang, S. N., & Nitiss, J. L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat Rev Mol Cell Biol*, *17*(11), 703-721. https://doi.org/10.1038/nrm.2016.111

Porschke, D. (1991). Persistence length and bending dynamics of DNA from electrooptical measurements at high salt concentrations. *Biophys Chem*, *40*(2), 169-179. https://doi.org/10.1016/0301-4622(91)87006-q

Postow, L., Crisona, N. J., Peter, B. J., Hardy, C. D., & Cozzarelli, N. R. (2001). Topological challenges to DNA replication: conformations at the fork. *Proc Natl Acad Sci U S A*, *98*(15), 8219-8226. https://doi.org/10.1073/pnas.111006998

Postow, L., Hardy, C. D., Arsuaga, J., & Cozzarelli, N. R. (2004). Topological domain structure of the Escherichia coli chromosome. *Genes Dev*, *18*(14), 1766-1779. https://doi.org/10.1101/gad.1207504

Pradhan, B., Kanno, T., Umeda Igarashi, M., Loke, M. S., Baaske, M. D., Wong, J. S. K., Jeppsson, K., Bjorkegren, C., & Kim, E. (2023). The Smc5/6 complex is a DNA loop-extruding motor. *Nature*, *616*(7958), 843-848. https://doi.org/10.1038/s41586-023-05963-3

Quinet, A., Tirman, S., Cybulla, E., Meroni, A., & Vindigni, A. (2021). To skip or not to skip: choosing repriming to tolerate DNA damage. *Mol Cell*, *81*(4), 649-658. https://doi.org/10.1016/j.molcel.2021.01.012

Revyakin, A., Liu, C., Ebright, R. H., & Strick, T. R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*, *314*(5802), 1139-1143. https://doi.org/10.1126/science.1131398

Reyes-Lamothe, R., Possoz, C., Danilova, O., & Sherratt, D. J. (2008). Independent positioning and action of Escherichia coli replisomes in live cells. *Cell*, *133*(1), 90-102. https://doi.org/10.1016/j.cell.2008.01.044

Rezende, L. F., Hollis, T., Ellenberger, T., & Richardson, C. C. (2002). Essential amino acid residues in the single-stranded DNA-binding protein of bacteriophage T7. Identification of the dimer interface. *J Biol Chem*, *277*(52), 50643-50653. https://doi.org/10.1074/jbc.M207359200

Robinow, C., & Kellenberger, E. (1994). The bacterial nucleoid revisited. *Microbiol Rev*, *58*(2), 211-232. https://doi.org/10.1128/mr.58.2.211-232.1994

Rueden, C. T., Ackerman, J., Arena, E. T., Eglinger, J., Cimini, B. A., Goodman, A., Carpenter, A. E., & Eliceiri, K. W. (2019). Scientific Community Image Forum: A discussion forum for scientific image software. *PLoS Biol*, *17*(6), e3000340. https://doi.org/10.1371/journal.pbio.3000340

Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., & Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, *18*(1), 529. https://doi.org/10.1186/s12859-017-1934-z

Rybenkov, V. V., Herrera, V., Petrushenko, Z. M., & Zhao, H. (2014). MukBEF, a chromosomal organizer. *J Mol Microbiol Biotechnol*, *24*(5-6), 371-383. https://doi.org/10.1159/000369099

Sarkar, R., & Rybenkov, V. V. (2016). A Guide to Magnetic Tweezers and Their Applications. *Frontiers in Physics*, *4*. https://doi.org/10.3389/fphy.2016.00048

Sawitzke, J. A., & Austin, S. (2000). Suppression of chromosome segregation defects of Escherichia coli muk mutants by mutations in topoisomerase I. *Proc Natl Acad Sci U S A*, *97*(4), 1671-1676. https://doi.org/10.1073/pnas.030528397

Schakenraad, K., Biebricher, A. S., Sebregts, M., Ten Bensel, B., Peterman, E. J. G., Wuite, G. J. L., Heller, I., Storm, C., & van der Schoot, P. (2017). Hyperstretching DNA. *Nat Commun*, *8*(1), 2197. https://doi.org/10.1038/s41467-017-02396-1

Schalbetter, S. A., Mansoubi, S., Chambers, A. L., Downs, J. A., & Baxter, J. (2015). Fork rotation and DNA precatenation are restricted during DNA replication to prevent chromosomal instability. *Proc Natl Acad Sci U S A*, *112*(33), E4565-4570. https://doi.org/10.1073/pnas.1505356112

Scherr, M. J., Safaric, B., & Duderstadt, K. E. (2018). Noise in the Machine: Alternative Pathway Sampling is the Rule During DNA Replication. *Bioessays*, *40*(2). https://doi.org/10.1002/bies.201700159

Scherr, M. J., Wahab, S. A., Remus, D., & Duderstadt, K. E. (2022). Mobile origin-licensing factors confer resistance to conflicts with RNA polymerase. *Cell Rep*, *38*(12), 110531. https://doi.org/10.1016/j.celrep.2022.110531

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods*, *9*(7), 676-682. https://doi.org/10.1038/nmeth.2019

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*, *9*(7), 671-675. https://doi.org/10.1038/nmeth.2089

Senavirathne, G., Lopez, M. A., Jr., Messer, R., Fishel, R., & Yoder, K. E. (2018). Expression and purification of nuclease-free protocatechuate 3,4-dioxygenase for prolonged single-molecule fluorescence imaging. *Anal Biochem*, *556*, 78-84. https://doi.org/10.1016/j.ab.2018.06.016

Seol, Y., & Neuman, K. C. (2011). Single-molecule measurements of topoisomerase activity with magnetic tweezers. *Methods Mol Biol*, *778*, 229-241. https://doi.org/10.1007/978-1-61779-261-8_15

Seol, Y., & Neuman, K. C. (2016). The dynamic interplay between DNA topoisomerases and DNA topology. *Biophys Rev*, *8*(Suppl 1), 101-111. https://doi.org/10.1007/s12551-016-0240-8

Sevier, S. A. (2020). Mechanical properties of DNA replication. *Physical Review Research*, *2*(2). https://doi.org/10.1103/PhysRevResearch.2.023280

Shutt, T. E., & Gray, M. W. (2006). Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet*, *22*(2), 90-95. https://doi.org/10.1016/j.tig.2005.11.007

Shyian, M., Albert, B., Zupan, A. M., Ivanitsa, V., Charbonnet, G., Dilg, D., & Shore, D. (2020). Fork pausing complex engages topoisomerases at the replisome. *Genes Dev*, *34*(1-2), 87-98. https://doi.org/10.1101/gad.331868.119

Singleton, M. R., Sawaya, M. R., Ellenberger, T., & Wigley, D. B. (2000). Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell*, *101*(6), 589-600. https://doi.org/10.1016/s0092-8674(00)80871-5

Smith, S. B., Finzi, L., & Bustamante, C. (1992). Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, *258*(5085), 1122-1126. https://doi.org/10.1126/science.1439819

Snapka, R. M., Powelson, M. A., & Strayer, J. M. (1988). Swiveling and decatenation of replicating simian virus 40 genomes in vivo. *Mol Cell Biol*, *8*(2), 515-521. https://doi.org/10.1128/mcb.8.2.515-521.1988

Sowa, Y., Rowe, A. D., Leake, M. C., Yakushi, T., Homma, M., Ishijima, A., & Berry, R. M. (2005). Direct observation of steps in rotation of the bacterial flagellar motor. *Nature*, *437*(7060), 916-919. https://doi.org/10.1038/nature04003

Spenkelink, L. M., Lewis, J. S., Jergic, S., Xu, Z. Q., Robinson, A., Dixon, N. E., & van Oijen, A. M. (2019). Recycling of single-stranded DNA-binding protein by the bacterial replisome. *Nucleic Acids Res*, *47*(8), 4111-4123. https://doi.org/10.1093/nar/gkz090

Stephan Saalfeld, I. P., Philipp Hanslovsky, Andrew Champion, Curtis Rueden, John Bogovic, Mark Kittisopikul, & jakirkham. (2022). saalfeldlab/n5: n5-2.5.1. *Zenodo*. https://doi.org/10.5281/zenodo.6578232

Stewart-Morgan, K. R., Petryk, N., & Groth, A. (2020). Chromatin replication and epigenetic cell memory. *Nat Cell Biol*, *22*(4), 361-371. https://doi.org/10.1038/s41556-020-0487-y

Stracy, M., Wollman, A. J. M., Kaja, E., Gapinski, J., Lee, J. E., Leek, V. A., McKie, S. J., Mitchenall, L. A., Maxwell, A., Sherratt, D. J., Leake, M. C., & Zawadzki, P. (2019). Single-molecule imaging of DNA gyrase activity in living Escherichia coli. *Nucleic Acids Res*, *47*(1), 210-220. https://doi.org/10.1093/nar/gky1143

Strick, T. R., Allemand, J. F., Bensimon, D., Bensimon, A., & Croquette, V. (1996). The elasticity of a single supercoiled DNA molecule. *Science*, *271*(5257), 1835-1837. https://doi.org/10.1126/science.271.5257.1835

Strick, T. R., Allemand, J. F., Bensimon, D., & Croquette, V. (1998). Behavior of supercoiled DNA. *Biophys J*, *74*(4), 2016-2028. https://doi.org/10.1016/S0006-3495(98)77908-1

Strick, T. R., Dessinges, M. N., Charvin, G., Dekker, N. H., Allemand, J. F., Bensimon, D., & Croquette, V. (2003). Stretching of macromolecules and proteins. *Reports on Progress in Physics*, *66*(1), 1-45. https://doi.org/10.1088/0034-4885/66/1/201

Sundin, O., & Varshavsky, A. (1980). Terminal stages of SV40 DNA replication proceed via multiply intertwined catenated dimers. *Cell*, *21*(1), 103-114. https://doi.org/10.1016/0092-8674(80)90118-x

Sundin, O., & Varshavsky, A. (1981). Arrest of segregation leads to accumulation of highly intertwined catenated dimers: dissection of the final stages of SV40 DNA replication. *Cell*, *25*(3), 659-669. https://doi.org/10.1016/0092-8674(81)90173-2

Tanner, N. A., Loparo, J. J., Hamdan, S. M., Jergic, S., Dixon, N. E., & van Oijen, A. M. (2009). Real-time single-molecule observation of rolling-circle DNA replication. *Nucleic Acids Res*, *37*(4), e27. https://doi.org/10.1093/nar/gkp006

Thomen, P., Lopez, P. J., Bockelmann, U., Guillerez, J., Dreyfus, M., & Heslot, F. (2008). T7 RNA polymerase studied by force measurements varying cofactor concentration. *Biophys J*, *95*(5), 2423-2433. https://doi.org/10.1529/biophysj.107.125096

Thomsen, J., Sletfjerding, M. B., Jensen, S. B., Stella, S., Paul, B., Malle, M. G., Montoya, G., Petersen, T. C., & Hatzakis, N. S. (2020). DeepFRET, a software for rapid and automated single-molecule FRET data

classification using deep learning. *Elife*, *9*. https://doi.org/10.7554/eLife.60404

Ticau, S., Friedman, L. J., Ivica, N. A., Gelles, J., & Bell, S. P. (2015). Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell*, *161*(3), 513-525. https://doi.org/10.1016/j.cell.2015.03.012

Tinevez, J. Y., Perry, N., Schindelin, J., Hoopes, G. M., Reynolds, G. D., Laplantine, E., Bednarek, S. Y., Shorte, S. L., & Eliceiri, K. W. (2017). TrackMate: An open and extensible platform for single-particle tracking. *Methods*, *115*, 80-90. https://doi.org/10.1016/j.ymeth.2016.09.016

Tokunaga, M., Imamoto, N., & Sakata-Sogawa, K. (2008). Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat Methods*, *5*(2), 159-161. https://doi.org/10.1038/nmeth1171

van der Valk, T., Pecnerova, P., Diez-Del-Molino, D., Bergstrom, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J. A., Dehasque, M., Saglican, E., Fidan, F. R., Barnes, I., Liu, S., Somel, M., Heintzman, P. D., Nikolskiy, P., Shapiro, B., Skoglund, P., Hofreiter, M., . . . Dalen, L. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. *Nature*, *591*(7849), 265-269. https://doi.org/10.1038/s41586-021-03224-9

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T., & the scikit-image, c. (2014). scikit-image: image processing in Python. *PeerJ*, *2*, e453. https://doi.org/10.7717/peerj.453

van Loenhout, M. T., de Grunt, M. V., & Dekker, C. (2012). Dynamics of DNA supercoils. *Science*, *338*(6103), 94-97. https://doi.org/10.1126/science.1225810

Vilfan, I. D., Lipfert, J., Koster, D. A., Lemay, S. G., & Dekker, N. H. (2009). Magnetic Tweezers for Single-Molecule Experiments. In P. Hinterdorfer & A. Oijen (Eds.), *Handbook of Single-Molecule Biophysics* (pp. 371-395). Springer US. https://doi.org/10.1007/978-0-387-76497-9_13

Vipat, S., & Moiseeva, T. N. (2024). The TIMELESS Roles in Genome Stability and Beyond. *J Mol Biol*, *436*(1), 168206. https://doi.org/10.1016/j.jmb.2023.168206

Vizcay-Barrena, G., Webb, S. E., Martin-Fernandez, M. L., & Wilson, Z. A. (2011). Subcellular and single-molecule imaging of plant fluorescent proteins using total internal reflection fluorescence microscopy (TIRFM). *J Exp Bot*, *62*(15), 5419-5428. https://doi.org/10.1093/jxb/err212

Vologodskii, A. V., & Marko, J. F. (1997). Extension of torsionally stressed DNA by external force. *Biophys J*, *73*(1), 123-132. https://doi.org/10.1016/S0006-3495(97)78053-6

Vos, S. M., Tretter, E. M., Schmidt, B. H., & Berger, J. M. (2011). All tangled up: how cells direct, manage and exploit topoisomerase function. *Nat Rev Mol Cell Biol*, *12*(12), 827-841. https://doi.org/10.1038/nrm3228

Vrtis, K. B., Dewar, J. M., Chistol, G., Wu, R. A., Graham, T. G. W., & Walter, J. C. (2021). Single-strand DNA breaks cause replisome disassembly. *Mol Cell*, *81*(6), 1309-1318 e1306. https://doi.org/10.1016/j.molcel.2020.12.039

Wang, X., Montero Llopis, P., & Rudner, D. Z. (2013). Organization and segregation of bacterial chromosomes. *Nat Rev Genet*, *14*(3), 191-203. https://doi.org/10.1038/nrg3375

Wang, X., Reyes-Lamothe, R., & Sherratt, D. J. (2008). Modulation of Escherichia coli sister chromosome cohesion by topoisomerase IV. *Genes Dev*, *22*(17), 2426-2433. https://doi.org/10.1101/gad.487508

Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737-738. https://doi.org/10.1038/171737a0

Westhorpe, R., Keszthelyi, A., Minchell, N. E., Jones, D., & Baxter, J. (2020). Separable functions of Tof1/Timeless in intra-S-checkpoint signalling, replisome stability and DNA topological stress. *Nucleic Acids Res*, *48*(21), 12169-12187. https://doi.org/10.1093/nar/gkaa963

White, J. H. (1969). Self-Linking and the Gauss Integral in Higher Dimensions. *American Journal of Mathematics*, *91*(3). https://doi.org/10.2307/2373348

Wilkins, M. H., Stokes, A. R., & Wilson, H. R. (1953). Molecular structure of deoxypentose nucleic acids. *Nature*, *171*(4356), 738-740. https://doi.org/10.1038/171738a0

Williams, E., Moore, J., Li, S. W., Rustici, G., Tarkowska, A., Chessel, A., Leo, S., Antal, B., Ferguson, R. K., Sarkans, U., Brazma, A., Salas, R. E. C., & Swedlow, J. R. (2017). The Image Data Resource: A Bioimage Data Integration and Publication Platform. *Nat Methods*, *14*(8), 775-781. https://doi.org/10.1038/nmeth.4326

Wong, R. P., Petriukov, K., & Ulrich, H. D. (2021). Daughter-strand gaps in DNA replication - substrates of lesion processing and initiators of distress signalling. *DNA Repair (Amst)*, *105*, 103163. https://doi.org/10.1016/j.dnarep.2021.103163

Worcel, A., & Burgi, E. (1972). On the structure of the folded chromosome of Escherichia coli. *J Mol Biol*, *71*(2), 127-147. https://doi.org/10.1016/0022-2836(72)90342-7

Wu, F., Swain, P., Kuijpers, L., Zheng, X., Felter, K., Guurink, M., Solari, J., Jun, S., Shimizu, T. S., Chaudhuri, D., Mulder, B., & Dekker, C. (2019). Cell Boundary Confinement Sets the Size and Position of the E. coli Chromosome. *Curr Biol*, *29*(13), 2131-2144 e2134. https://doi.org/10.1016/j.cub.2019.05.015

Xu, Z. Q., & Dixon, N. E. (2018). Bacterial replisomes. *Curr Opin Struct Biol*, *53*, 159-168. https://doi.org/10.1016/j.sbi.2018.09.006

Yao, N. Y., & O'Donnell, M. (2010). SnapShot: The replisome. *Cell*, *141*(6), 1088, 1088 e1081. https://doi.org/10.1016/j.cell.2010.05.042

Yeeles, J. T. (2014). Discontinuous leading-strand synthesis: a stop-start story. *Biochem Soc Trans*, *42*(1), 25-34. https://doi.org/10.1042/BST20130262

Yim, S. S., McBee, R. M., Song, A. M., Huang, Y., Sheth, R. U., & Wang, H. H. (2021). Robust direct digital-to-biological data storage in living cells. *Nat Chem Biol*, *17*(3), 246-253. https://doi.org/10.1038/s41589-020-00711-4

Yin, J., Lin, A. J., Golan, D. E., & Walsh, C. T. (2006). Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat Protoc*, *1*(1), 280-285. https://doi.org/10.1038/nprot.2006.43

Zawadzki, P., Stracy, M., Ginda, K., Zawadzka, K., Lesterlin, C., Kapanidis, A. N., & Sherratt, D. J. (2015). The Localization and Action of

Topoisomerase IV in Escherichia coli Chromosome Segregation Is Coordinated by the SMC Complex, MukBEF. *Cell Rep*, *13*(11), 2587-2596. https://doi.org/10.1016/j.celrep.2015.11.034

Zlatanova, J., & van Holde, K. (2006). Single-molecule biology: what is it and how does it work? *Mol Cell*, *24*(3), 317-329. https://doi.org/10.1016/j.molcel.2006.10.017