# Single-molecule studies of fully reconstituted CMG

Dissertation

for the purpose of obtaining the degree of doctor

at Delft University of Technology

by the authority of the Rector Magnificus Prof. dr. ir. T.H.J.J. van der Hagen

chair of the Board for Doctorates

to be defended publicly on

Wednesday 19 June 2024 at 15:00 o'clock

By

Daniel Fernando RAMÍREZ MONTERO

Master of Science in Biology

Massachusetts Institute of Technology, United States of America

born in Mexico City, Mexico

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

| | |
|---|---|
| Rector Magnificus | *chairperson* |
| Prof. dr. N.H. Dekker | Delft University of Technology/ University of Oxford, *promotor* |
| Prof. dr. C.J.A. Danelon | Delft University of Technology/ Toulouse Biotechnology Institute, *copromotor* |
| *Independent members:* | |
| Prof. dr. C. Joo | Delft University of Technology |
| Prof. dr. C. M. Kaiser | Utrecht University, The Netherlands |
| Prof. dr. K. Duderstadt | Technical University of Munich, Germany |
| Dr. M.H. Lamers | Leiden University, The Netherlands |
| Prof. dr. B. Rieger | Delft University of Technology, *reserve member* |
| *Other members:* | |
| Dr. J.F.X. Diffley | The Francis Crick Institute, United Kingdom |

Keywords: single-molecule, DNA replication, optical tweezers, TIRF microscopy

Printed by: Gildeprint

Cover: Daniel Ramírez Montero, inspired by the art of the Huichol people in Mexico



Casimir PhD series

ISBN: 978-94-6384-597-7

An electronic copy of this dissertation is available at: https://repository.tudelft.nl/

# Table of contents

# 1

# Introduction

## 1.1 Preface

In preparation for cell division, a cell needs to generate an exact copy of its entire genome to allow each of the resulting daughter cells to obtain a full copy and thus maintain genome stability. This process, known as DNA replication, is carried out by all known life forms (bacterial, archaeal, eukaryotic, and even viral) and is necessary for the survival of all organisms.

DNA replication is catalyzed by a large protein complex that moves along the DNA and whose function relies on the dynamic recruitment and release of different protein components along the way. Understanding all these dynamics is therefore key to understand how the replisome achieves its critical task, as well as to understand what goes wrong when it fails at it.

Our long-term goal is to use the precision of biophysics to obtain a deep quantitative understanding of all these dynamics. This is of course no easy task, and it will require the work of many researchers over many years to come. This thesis is my contribution to this goal.

## 1.2 DNA replication in eukaryotes

### 1.2.1 Basic architecture of the eukaryotic replisome

DNA replication is one of the most essential processes of life that allows the faithful transmission of genetic information between generations of cells. In eukaryotes, this Herculean task is carried out by a MDa-sized dynamic protein complex known as the replisome[1], which in general achieves this task in a peculiarly fast and accurate manner[2–11]. Nevertheless, the replisome does not always succeeds at its task, slowing down, stalling and/or introducing errors in the copied genome, which may result in genomic instability and even cell death[12–15].

Simply put, the replisome must catalyze two processes. First, it must unwind the two strands of the parental dsDNA, and then use the unwound parental ssDNA strands as templates to synthesize two complementary strands[1]. In eukaryotes, the unwinding of the parental DNA is carried out by the highly conserved ring-shaped replicative helicase Cdc45/Mcm2-7/GINS (CMG)[16–20]. This important helicase leads the way at the front of replication forks, separating the two parental DNA strands by translocating on ssDNA N-terminus first in a 3'-to-5' direction, and sterically excluding the non-translocation strand[18–23] (**Fig. 1.1**). In the wake of the CMG helicase, three polymerases named Pol α, Pol δ, and Pol ε use the unwound strands as templates to synthesize the DNA leading (synthesized by Pol ε) and lagging strands (synthesized by Pol δ) starting from RNA:DNA hybrid primers (synthesized by Pol α)[9,24–31].

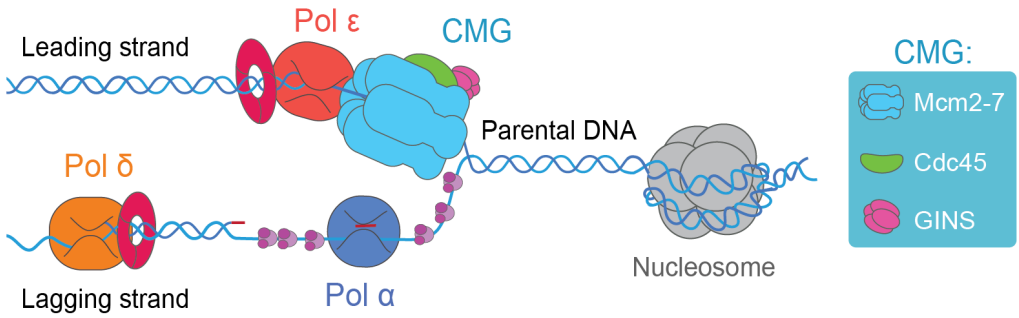



*Figure 1.1* Simplified diagram of a eukaryotic replisome. The replicative helicase CMG (consisting of Mcm2-7, Cdc45, and GINS) leads the way at the front of the replication fork. In the wake of CMG, three polymerases synthesize RNA:DNA primers (Pol α), leading (Pol ε), and lagging (Pol δ) strands. The replisome also coordinates the disassembly of nucleosomes ahead of the replisome and their deposition on the newly synthesized DNA strands.

Remarkably, despite the many different roadblocks that the replisome encounters along the way —including tightly bound DNA-binding proteins[32,33], covalent DNA-protein crosslinks[34–36], and highly thermally stable DNA secondary structures[37–40]— the replisome has evolved to be able to cruise through thousands of base pairs at an *in vivo* speed of 2-3 kb/min[2–6], all while keeping the mutation rate as low as ≤$10^{-10}$ mutations per base replicated per generation[7,8,10,11]. This is all the more impressive considering that the parental DNA is packaged into chromatin by wrapping around histone octamers containing epigenetic marks that must be preserved between generations of cells[41–44] (**Fig. 1.1**). This adds an extra layer of complication, as the replisome must also coordinate the disassembly of parental nucleosomes ahead of the replication fork, and their re-assembly on the newly synthesized DNA, both by recycling parental histones and by recruiting newly synthesized histones[44–47]. Furthermore, the replisome has also evolved to

efficiently deal with DNA lesions by recruiting proteins involved in replication-coupled DNA repair[40,48,49].

To successfully carry out all this functions, the replisome must exquisitely coordinate the recruitment and release of auxiliary proteins at different stages of DNA and chromatin replication. This not only highlights the dynamic nature of the replisome, but it also highlights the importance of deeply understanding these dynamics.

### 1.2.2 The eukaryotic replicative helicase CMG drives and organizes the replisome

One of the core components of the replisome is the replicative helicase CMG[1], which gets its name from its three constituent components: the hexameric AAA+ ATPase family helicase Mcm2-7, Cdc45, and the tetrameric GINS complex[16,19,20,50–52] (**Figure 1.1** (right)). This important helicase is not only thought to be the main molecular motor driving the progression of the replisome[20,53–55], but it also acts as an organizing center for the entire replisome, coordinating the recruitment of auxiliary factors involved in DNA replication[56] as well as in different replication-coupled processes, such as histone deposition[44,45] and DNA repair[40,48,49]. Therefore, to understand replisome dynamics and how cells achieve the complex task of faithfully replicating their entire genome once per cell cycle, we need a deep quantitative understanding of CMG as a molecular motor.

**This thesis focuses on the CMG helicase**. Further, it specifically focuses on the CMG ortholog from a relatively simple unicellular eukaryote: the budding yeast *Saccharomyces cerevisiae*. As we will describe below, important biochemical advancements[19,51,57] have rendered *S. cerevisiae* a very useful model organism to understand DNA replication with the great control that comes from biochemical reconstitution. Furthermore, as expected for such an important machinery, many components of the *S. cerevisiae* replisome —including CMG— are highly conserved across all eukaryotes[58] and many parallels have been found between the *S. cerevisiae* replisome and that of higher eukaryotes[9,20,51,52,59–62], further highlighting the potential reach of our findings.

### 1.2.3 *In vivo* assembly and activation of CMG

In vivo, eukaryotic DNA replication starts with the assembly and activation of the CMG helicase[1] (**Fig. 1.2**). Furthermore, in the specific case of the budding yeast, DNA replication starts at specific sequences known as origins of replication[63,64]. First, in the G1-phase of the cell cycle, the Origin Recognition Complex (ORC) binds

and scans the DNA by 1D diffusion in search of origins of replication[65,66]. At such sequences, in an ATP-hydrolysis-driven reaction[67,68], ORC together with Cdc6 and Cdt1 (collectively known as *loading factors*) load the first component of CMG, the Mcm2-7 hexamer, onto dsDNA[67–72]. While Mcm2-7 hexamers can be loaded as

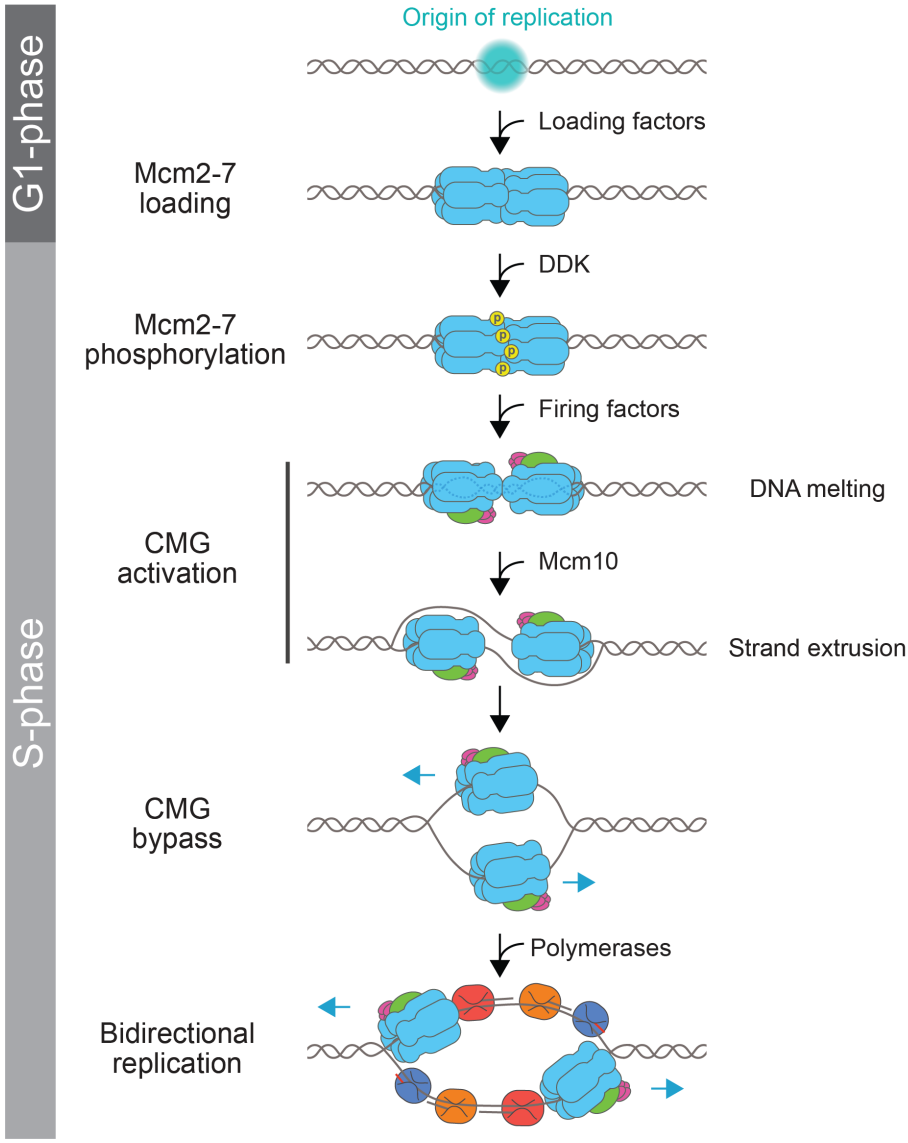



*Figure 1.2* Pictorial description of the *in vivo* assembly and activation of CMG and the establishment of bidirectional replication in yeast. In the G1-phase of the cell cycle double Mcm2-7 double hexamers are loaded at origins of replication in a head-to-head conformation. In the following S-phase, the selective phosphorylation of double hexamers by DDK facilitate the recruitment of the helicase-activating factors Cdc45 and GINS to form two sister CMGs still in a head-to-head conformation. Upon CMG assembly, each sister CMG melts 0.7 turns of dsDNA within their central channel. Finally, Mcm10 catalyzes the extrusion of one strand from each sister CMG, allowing the two CMGs to bypass one another and seed the formation of two diverging replisomes.

single or double hexamers[66,73–76], the basis of bidirectional DNA replication lies in the formation of double hexamers[62,69,73,77], in which the two Mcm2-7 hexamers are located in a head-to-head orientation[75].

Although Mcm2-7 constitutes the motor core of the CMG helicase, Mcm2-7 by itself is unable to unwind dsDNA[17]. To initiate DNA unwinding, Mcm2-7 must be matured into the fully active CMG, which takes place in the S-phase of the cell cycle[1,17,51] (**Fig. 1.2**). This maturation process starts by the selective *in trans* phosphorylation of Mcm2-7 double hexamers by the cell-cycle regulated Dpb4-Dependent Kinase (DDK)[78–80]. Once phosphorylated, Mcm2-7 double hexamers can then be converted into active CMGs[51,81]. This process, known as CMG activation, has two steps[19]. In the first step, a second set of proteins known as *firing factors* (comprising Sld3/7, Sld2, Pol ε, Dpb11, and S-phase cyclin-dependent kinase S-CDK[51]) mediates the recruitment of the helicase-activating factors Cdc45 and GINS to the phosphorylated Mcm2-7 double hexamers to form two sister CMGs in a head-to-head configuration[19,51,82]. Concomitantly with CMG assembly, ATP binding allows each of the two sister CMG helicases to melt 0.7 turns of dsDNA within their central channels[19,82] in preparation for DNA unwinding. In the second and final activation, through an as-of-yet poorly understood mechanism, the firing factor Mcm10 mediates the ATP-hydrolysis-driven extrusion of one ssDNA strand from the central channel of each sister CMG[19]. Then, the two sister CMG helicases translocate along ssDNA in an ATP hydrolysis-dependent manner, bypassing and then separating from one another[19,49]. The two diverging CMGs then seed the assembly of two replisomes, giving rise to bidirectional replication. The temporal separation of Mcm2-7 loading in G1 phase and CMG activation in S-phase ensures that DNA replication occurs only once per cell cycle.

## 1.3 Two *in vitro* reconstitution approaches to study CMG

*In vitro*, CMG has been studied with top-down studies in cell extracts, as well as with bottom-up biochemical reconstitution studies. Cell extract work has proven a great avenue for the discovery of novel protein factors involved at different states of DNA replication[34,40,49,55,83–86]. Nonetheless, in extract approaches lack the control of biochemical reconstitution, and the experiments that can be limited by the fact that the depletion of endogenous factors from extracts is always possible.

Bottom-up *in vitro* reconstitution studies of eukaryotic DNA replication, on the other hand, have given us detailed mechanistic insights into the structure and

function of CMG with unparalleled control. There have been two different *in vitro* reconstituted approaches to study CMG motion. The first approach came about in 2006 following the finding that Cdc45 co-purified with Mcm2-7 and GINS from *Drosophila melanogaster* embryo extracts, and that the purified complex supported DNA unwinding[16]. Subsequently, upon co-overexpression of Mcm2-7, Cdc45, and GINS, *S. cerevisiae* and *H. sapiens* CMG complexes were successfully purified and shown to have DNA unwinding activity [21,23,50,60,61,87–89], as well as to support leading and lagging strand replication when supplemented with polymerases[25–27,50,61]. We will refer to these approaches as *pre-formed CMG* (**Fig 1.3 a**). Notably, pre-formed CMG requires a stretch of ssDNA to bind DNA, which is typically added as a replication-fork-mimicking 3' ssDNA flap in the DNA substrate to be unwound/replicated[20,22,25,50,53,90–93]. Furthermore, pre-formed CMG-based unwinding and/or replication does not require an origin of replication and occurs unidirectionally.

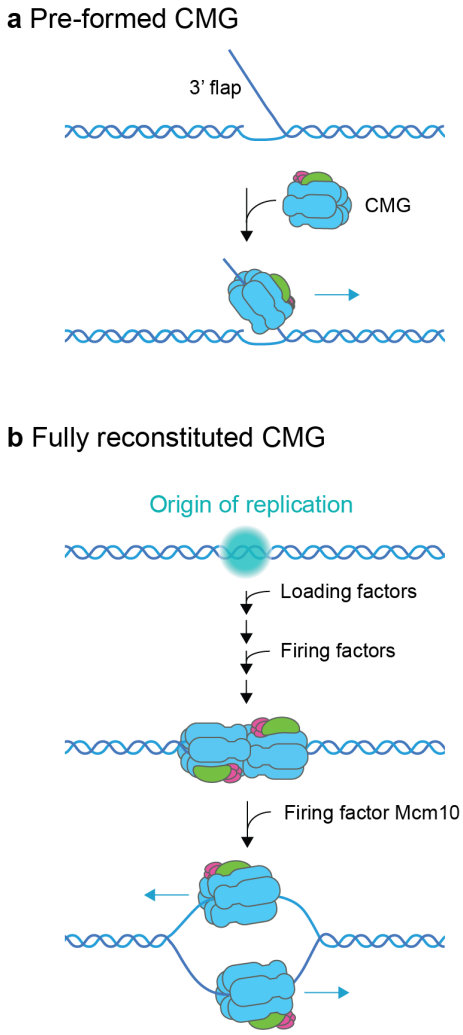



*Figure 1.3* Two in vitro reconstituted approaches to studying CMG motion. **a** Pre-formed CMG requires a stretch of ssDNA to bind DNA and does not need an origin of replication. The motion of pre-formed CMG is unidirectional. **b** Fully reconstituted CMG is assembled by reconstituted all the steps shown in Fig. 1.2, a subset of which is shown here for clarity. With fully reconstituted CMG, two head-to-head helicases are assembled onto dsDNA at an origin of replication. After their assembly, extrusion of one DNA strand from each initiates extensive unwinding.

The second *in vitro* reconstituted approach to study CMG came about in 2015 when origin-based DNA replication was fully reconstituted from purified *S. cerevisiae* protein components, and in 2018 when the mechanism of CMG activation was elucidated following the reconstitution of CMG activation and unwinding in the absence of replication (i.e. all the steps shown in **Fig. 1.2**). We will refer to this second approach as *fully reconstituted CMG* (**Fig 1.3 b**). Notably,

fully reconstituted CMG is replication origin-dependent, occurs bidirectionally and allows us to study the steps prior to CMG activation[9,19,51,69].

Although these two approaches to studying CMG have some differences, structural and biochemical work have shown no drastic differences between them[9,21,23,25,51,52,61,82]. Nevertheless, only the fully reconstituted approach allows the study of events prior to CMG activation.

## 1.4 Why study CMG motion one molecule at a time?

In this thesis, we set out to study CMG at the single-molecule level. But before we discuss how exactly we did achieve this, we will first address an important question that the reader might wonder: given all the prior structural and ensemble biochemical work studying CMG, why should we study it at the single-molecule level?

First, as opposed to ensemble experiments, which typically have a time resolution in the order of tens of seconds to minutes[19,22,90,91,93–95], single-molecule techniques can achieve time resolutions of milliseconds[96–98]. Furthermore, single-molecule studies of molecular motors can achieve sub-nanometer spatial resolutions[96,99]; this, combined with the higher time-resolutions not only allows us to observe many dynamic events and short-lived intermediates that would otherwise be hidden in an ensemble assay, but also allow us to quantify motion dynamics of molecular motors at a relevant spatial scale (for instance, a molecular motor that translocates on the DNA one base pair (0.34 nm) at a time[99,100]). In addition, while output of ensemble biochemical assays is the result of millions of asynchronous reactions, single-molecule experiments have the unique power to synchronize events *a posteriori*, uncovering otherwise inaccessible kinetic constants[72,74,92,101]. Lastly, ensemble studies only show the highest probability events happening in a reaction, as lower probability ones are averaged out in the ensemble output[102]. On the other hand, single-molecule analysis can unveil different behaviors within a given sample of a molecular motor, often uncovering different populations and giving us a more complete image of all the possible pathways that can take place in a reaction[102,103].

Altogether, single-molecule studies of CMG have the unique ability to allow us to understand and describe CMG as a molecular motor and better understand all the possible behaviors that it can exhibit, and the stochastic transitions that it can

undergo between different conformational states, giving us an unparalleled quantitative understanding of its function[53,99,102,103].

## 1.5 Single-molecule techniques used to study eukaryotic DNA replication

Four types of single-molecule techniques have been used so far to study eukaryotic DNA replication *in vitro*: magnetic tweezers[53], flow tweezers[92], total internal reflection fluorescence (TIRF) microscopy[22,25,85,93,101,104–106,49,62,71,72,74,81,83,84] and correlative dual-beam optical tweezers and confocal microscopy[66,107,108].

Magnetic tweezers and flow tweezers are force-based single-molecule techniques that allow for the high-throughput interrogation of molecular motors. Both techniques have in common that they use a linear DNA construct attached to a surface at one end and attached to a micron-sized bead (which must be magnetic in the case of optical tweezers) at the other[92,96]. A calibrated force is then applied to the bead (either through a magnetic field in the case of magnetic tweezers, or through flow in the case of flow tweezers) and changes of the position of the bead over time are used to infer changes in the length of the DNA tether as dsDNA is converted to ssDNA or vice versa[92,96]. In the field of DNA replication, magnetic tweezers experiments have been used to study the mechanism of dsDNA unwinding by pre-formed CMG[53]. Flow tweezers, on the other hand, have been used to study the rate of motion of a leading-strand replisome made up of pre-formed CMG in conjunction with Pol ε and some additional elongation factors[92]. Both magnetic and flow tweezers experiments have the advantage that they are label-free and allow for high-throughput imaging; nevertheless, the motion of CMG and the replisome are observed indirectly by monitoring length changes in the DNA tether[53,92], rendering the interrogation of the dynamics exchange of different replisome components very hard if not impossible to study.

On the other hand, single-molecule fluorescence techniques have the power of directly observing directly observing and quantifying the exchange dynamics of fluorescently labeled components of the replisome. Furthermore, some single-molecule fluorescence techniques also allow the direct observation of proteins as they move along the DNA substate, allowing to correlate the exchange dynamics of one component with the motion along the DNA of another component. Two single-molecule techniques with fluorescence outputs have been employed in the study of eukaryotic DNA replication: total internal reflection fluorescence microscopy[22,25,85,93,101,104–106,49,62,71,72,74,81,83,84] and correlative optical tweezers and

confocal microscopy[66,107–110]. These two techniques are also the two single-molecule techniques employed in this thesis, so we will describe them in more detail in the following sections.

### 1.5.1 Total Internal Reflection Fluorescence (TIRF) microscopy

Total internal reflection fluorescence (TIRF) microscopy is a fluorescence-based single-molecule technique based on the principle that the total internal reflection of incident laser light results in an evanescent field that only penetrates ~100 nm into a sample, selectively illuminating fluorescent molecules that are at or very close to the surface and excluding all others[111–113]. This in turn greatly increases the signal-to-noise ratio compared with standard epifluorescence microscopy, allowing for the observation of single fluorescent molecules when the fluorescent sample is up to low nM concentrations[97,113]. TIRF images are typically recorded on a sCMOS camera, which allows for a temporal resolution of up to 1 ms[97]. Furthermore, TIRF microscopy experiments typically involve several color lasers, allowing for the simultaneous observation of multiple components of the same reaction at the same time[114].

Two different modalities of TIRF microscopy have been used in the study of eukaryotic DNA replication (**Fig. 1.4**): colocalization-based kinetic studies[62,72,74,81,101,104,105] (**Fig 1.4 a,b**) and flow-stretched DNA protein motion studies[49,62,72,74,83–85,93,101] (**Fig 1.4 c,d**). In the first modality, a short (~1 kb) fluorescently labeled DNA construct is attached to a passivated glass surface (typically through biotin:avidin interactions), and a fluorescent protein/proteins of interest is/are then added . Upon fluorescent protein addition, the binding kinetics of the fluorescent protein(s) can be measured by monitoring the co-localization of different fluorophores with the DNA and with each other over time[114] (**Fig 1.4 a**). When used in this modality, TIRF microscopy can be used to study intra- and inter-molecular conformational changes by single-molecule Förster Resonance Energy Transfer[115].

Colocalization-based TIRF microscopy approaches have allowed for very precise measurements of binding kinetics and key intra- and inter- molecular interactions at different stages of Mcm2-7 loading[62,72,74,101,104,105] (**Fig 1.4 b**), CMG assembly and activation[81] (using the fully reconstituted system). These approaches have the advantage of having high-throughput; nevertheless, because of the short and singly tethered DNA substrates used, any protein motion along the DNA will be confined within a fluorescent diffraction-limited spot. Thus, when used in

colocalization mode, TIRF microscopy does not allow for the direct observation of protein motion along the DNA substrate.

Conversely, flow-stretched DNA TIRF microscopy studies allow for the direct observation of the motion of fluorescently labeled molecules along linear DNA molecules of tens of kb in length[49,62,72,74,83–85,93,101] (that is, 1-2 orders of magnitude longer than those used in co-localization TIRF microscopy experiments). These longer DNA substrates are typically modified with biotin moieties at both ends and stretched onto a streptavidin-coated surface with the help of laminar flow[116]. This

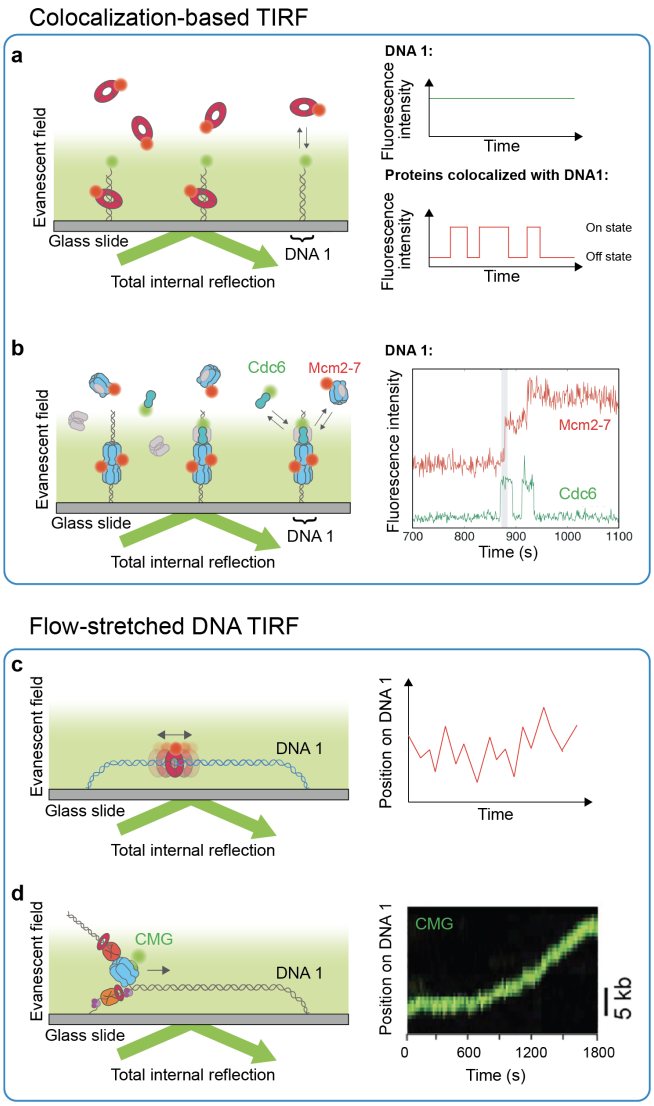



*Figure 1.4* Two modalities of total internal reflection fluorescence (TIRF) microscopy and their usage in the study of DNA replication. **a** Left: diagram of a colocalization-based TIRF experiment in which colocalization of a protein with DNA can be monitored over time. Right: mock single-molecule trace of the intensity of fluorescently labeled DNA 1 over time (top) and of fluorescently labeled proteins that associate with and dissociate from DNA 1 over time (bottom). Note that protein motion along the DNA cannot be detected with this modality of TIRF. **b** Left: diagram of colocalization-based TIRF experiment studying the association of fluorescently labeled Mcm2-7 and Cdc6 with origin-containing DNA during Mcm2-7 loading (adapted from Ref.[74]). Right: real single-molecule trace illustrating the loading of two Mcm2-7 hexamers (top) by two different Cdc6 molecules (bottom) (adapted from Ref.[74]). **c** Left: diagram of a flow-stretched DNA TIRF experiment with a fluorescently labeled protein diffusing along the DNA. Right: mock trace of the position along a DNA molecule of a protein diffusing on the DNA. **d** Left: diagram of flow-stretched DNA TIRF experiment studying the motion of a pre-formed CMG-based replisome (employing fluorescently labeled CMG) along the DNA (adapted from Ref.[25]). Right: real trace of the position along a DNA molecule of CMG moving along the DNA (adapted from Ref.[25])

results in the formation of micrometer-sized bridges along which the motion of fluorescent protein components can be visualized in the absence of flow (**Fig 1.4 c**). In addition, alternative tethering and stretching strategies have also been employed, either involving one attachment point to the imaging surface and another one to a bead[93] (in which case the DNA stretched with the help of flow), or using one attachment point to the imaging surface and extending the DNA through the application of flow during the experiment[106].

Flow-stretched DNA TIRF microscopy studies of CMG have been successfully used in top-down experiments in *Xenopus laevis* cell extracts, studying CMG disassembly after replication termination[84], the outcome of CMG collisions with DNA:protein crosslinks[49], the encounter of CMG with nicks on its translocation strand[83], and the outcome of CMG encountering nucleosomes[85].

In addition to these cell extracts studies, bottom-up *in vitro* reconstituted studies of CMG have also used flow-stretched DNA TIRF microscopy, studying the translocation of CMG in the presence of the ssDNA binding protein RPA[93], the exchange dynamics of the three replicative polymerases Pol α, Pol δ, and Pol ε during replisome progression[25], and the bypass of a protein-roadblock by a replisome[117]. It should be noted that these flow-stretched DNA TIRF bottom-up studies have so far only employed pre-formed CMG or pre-formed CMG-based replisomes[25,93,117] (**Fig 1.4 d**). To date, flow-stretched DNA TIRF microscopy has not been used to study fully reconstituted CMG.

### 1.5.2 Correlative optical tweezers and confocal microscopy

Optical tweezers are a single-molecule technique to exert forces on molecules of interest[98]. This powerful technique is based on the principle that the momentum of photons in a tightly focused laser beam can be transferred to a micron-sized dialectric diffractive object (typically a polystyrene bead), resulting in a net force that pulls the object towards the focus point of the laser[118]. Thus, by moving the beam, one can control the position of the bead in three dimensions, and if the bead is bound to a molecule of interest (for instance, a surface-bound nucleic acid molecule, a calibrated force can be exerted on it[98]. In addition, the trapping laser can be split into two orthogonally polarized beams, allowing to hold a molecule in place between two trapped beads and away from the surface[99,100,119], which can be highly beneficial in the study of biological molecules

To add even more power to this technique, dual-trap optical tweezers have been integrated with confocal microscopy and microfluidics[120,121] (**Fig. 1.5**). These additional functionalities add the possibility of exchanging buffer conditions *in situ*,

and the simultaneous interrogation of a biological molecule or complex with force and fluorescence spectroscopy, all while keeping the sample away from the surface[66,107,121–123].

In a typical correlative optical tweezers and confocal microscopy experiment, a long (tens of kb) linear DNA substrate is modified at both ends with biotin moieties and tethered between to two trapped beads[66,107,121,123]. The distance between the optical traps can then be adjusted so that the DNA substrate is held at a given tension, and then, the DNA can be incubated with fluorescently labeled proteins, which can then be visualized on the DNA over time with the confocal scanning laser[66,107,121,123].

Correlative dual-trap optical tweezers and confocal microscopy have been used to study the motion of pre-formed CMG and a pre-formed CMG-based replisome[107]. In both cases, the pre-formed CMG and/or the replisome were bound to the DNA after generating ssDNA/dsDNA junctions using high force[107].

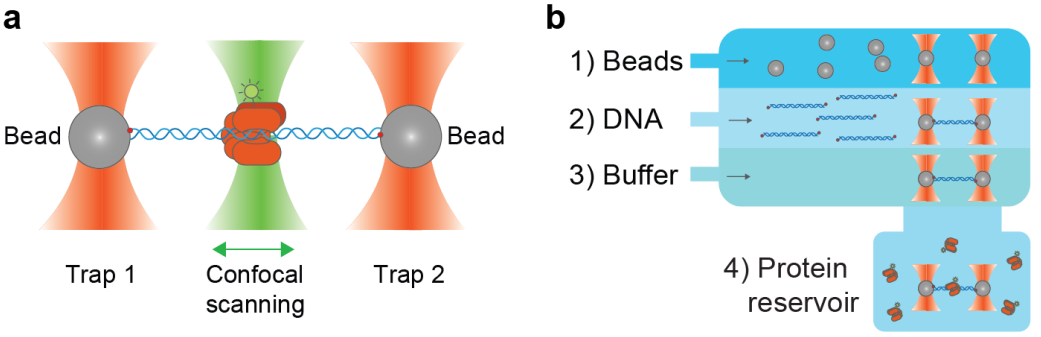



*Figure 1.5* **a** Diagram of a typical experiment using correlative dual-beam optical tweezers and confocal microscopy. A linear DNA molecule biotinylated at the ends (red spheres) is bound to two optically trapped beads. By adjusting the distance between both beads, a given tension can be applied to the DNA. A confocal scanning laser can then be used to scan the plane of the DNA and image the motion of a fluorescently labeled protein along the DNA. **b** A microfluidic flow cell that separates different buffer conditions using laminar flow helps to conduct the experiments. Beads are flowed into channel 1, after trapping two beads, the trapping lasers are then moved to channel 2 to bind a DNA molecule, which can then be moved to a reservoir containing fluorescently labeled protein (channel 4). Once bound, the motion of the protein along the DNA can be observed in the reservoir (channel 4), or in clean buffer (channel 3).

## 1.6 A new approach to study CMG: single-molecule studies of the motion of fully reconstituted CMG

As discussed in the previous section, single-molecule studies of CMG motion have thus far only employed pre-formed CMG. While these studies have given us

great insight into the dynamics of CMG and the replisome, employing pre-formed CMG makes it impossible to address some important questions. Firstly, because pre-formed CMG is already activated[16,50,61], this approach does not allow for the interrogation events leading up to the activation of CMG. Secondly, the unidirectional nature of pre-formed CMG approaches does not allow us to address questions regarding the bidirectional nature of DNA replication, such as whether sister CMGs that are activated together have correlated velocities or not. **To try to overcome these limitations, we set out to study the motion dynamics of CMG following the full reconstitution of its assembly and activation.**

Studying the motion of fully reconstituted CMG at the single-molecule level has several challenges compared to the pre-formed CMG approach. First, assembling and activating CMG alone requires 12 purified proteins/protein complexes, making it one of the most complex biochemical reactions ever studied at the single-molecule level[19,51]. This is as opposed to the pre-formed CMG approach, which requires only the purified CMG complex[16,20,22,53]. Secondly, the full reconstitution of CMG activation involves a cascade of biochemical reactions, requiring concentrations of the protein components involved in the range of 10-200 nM to maximize the efficiency of each step and thus maximize the overall efficiency[19,51,82]. These concentration ranges are on the high end of what most single-molecule techniques can tolerate, especially when using fluorescently labeled components[113]. Finally, CMG was evolved to cruise through thousands of base pairs in a cell[124–127]. Therefore, to study its motion at a biologically relevant spatial scale, one requires long DNA substrates (typically of lengths in the order of tens of kilobases). Using such long DNA substrates adds an extra layer of complication, as the longer the DNA substrate used, the more potential sites for non-specific adhesion of proteins and protein aggregates. This is of particular concern, as several of the loading and firing factors required for the assembly and activation of CMG contain intrinsically disordered regions[128] and are aggregation-prone.

To tackle these challenges, we had to re-think the way in which we performed the complex task of imaging the motion of fully reconstituted CMG at the single-molecule level. To this end, we had to develop novel hybrid methods that integrate cutting-edge biochemistry with single-molecule biophysics, as well as new data acquisition and analysis routines to best analyze the data. These novel assays allowed us to image for the first time the motion of fully reconstituted at the single-molecule level. We will describe each of these assays in detail in the following chapters.

## 1.7 Thesis description

In **Chapter 2**, we report the first single-molecule studies of the motion of fully reconstituted CMG using correlative optical tweezers and confocal microscopy. To achieve this, we developed a novel hybrid ensemble and single-molecule assay relying on the double functionalization of linear DNA with two orthogonal surface-attachment moieties. We found that CMG can move along the DNA in two different ways, by unidirectional translocation and by 1D diffusion, the latter of which is halted by nucleotide binding.

In **Chapter 3**, we report a new way of assembling fully custom-sequence long (tens of kb) linear DNA constructs for single-molecule experiments with modified ends for surface attachment, as well as internal extrahelical structures. As a proof-of-principle, we synthesized > 10 kb custom-sequence DNA substrates with either one or two internal 3' ssDNA flaps, and characterized them using correlative dual-beam optical tweezers and confocal microscopy. Finally, we functionally validated our internal 3' flaps by showing that pre-formed CMG can bind them and then translocate in the expected direction.

In our studies of fully reconstituted CMG using correlative dual-beam optical tweezers and confocal microscopy, we had to develop new ways of acquiring and analyzing single-molecule data. In **Chapter 4**, we report our recent developments in the study of DNA:protein interactions at the single-molecule level using correlative optical tweezers and confocal microscopy. These include the development of a special microfluidic flow cell, the developments of data acquisition automation routines, and the development of an analysis toolkit to streamline the analysis of correlative optical tweezers and confocal microscopy data in the form of a user-friendly graphical user interface.

One of the limitations of optical tweezers is their low-throughput. Thus, to complement our optical tweezers studies, we set out to develop a high-throughput way to image the motion of fully reconstituted CMG. In **Chapter 5**, we report how we adapted the hybrid ensemble and single-molecule assay developed in Chapter 2 to allow us to image the motion of fully reconstituted CMG in high-throughput using flow-stretched DNA-based TIRF microscopy. This was made possible through the development of a novel surface functionalization strategy that we developed and validated, as well as the design of a new microfluidic flow cell.

In **Chapter 6**, we provide a detailed description the hybrid ensemble and single-molecule assay developed in **Chapter 2** that allowed us to translate the complex biochemical reactions required to activate CMG from the ensemble to the

single-molecule level. In this chapter, we focus on the crucial parts of the protocol, important quality controls necessary for a successful experiment, and propose different ways in which the assay can be improved in future investigations.

Finally, in **Chapter 7**, we will revisit and interconnect the main findings of the other chapters. We will also discuss different directions in which the biological and technological advancements presented in this thesis can be taken in subsequent studies.

## 1.8 References


1. Bell, S. P. & Labib, K. Chromosome duplication in Saccharomyces cerevisiae. *Genetics* **203**, 1027–1067 (2016).
2. Méchali, M. Eukaryotic DNA replication origins: Many choices for appropriate answers. *Nat. Rev. Mol. Cell Biol.* **11**, 728–738 (2010).
3. Sekedat, M. D. *et al.* GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Mol. Syst. Biol.* **6**, 1–10 (2010).
4. Kaykov, A. & Nurse, P. The spatial and temporal organization of origin firing during the S-phase of fission yeast. *Genome Res.* **25**, 391–401 (2015).
5. Arbona, J. M., Goldar, A., Hyrien, O., Arneodo, A. & Audit, B. The eukaryotic bell-shaped temporal rate of DNA replication origin firing emanates from a balance between origin activation and passivation. *Elife* **7**, 1–13 (2018).
6. van den Berg, J., van Batenburg, V. & van Oudenaarden, A. Acceleration of genome replication uncovered by single-cell nascent DNA sequencing. *bioRxiv* (2022) doi:10.1101/2022.12.13.520365.
7. Kunkel, T. A. & Bebenek, K. DNA Replication Fidelity. *Annu. Rev. Biochem.* **69**, 497–529 (2000).
8. McCulloch, S. D. & Kunkel, T. A. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res.* **18**, 148–161 (2008).
9. Yeeles, J. T. P., Janska, A., Early, A. & Diffley, J. F. X. How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication. *Mol. Cell* **65**, 105–116 (2017).
10. Shcherbakova, P. V. *et al.* Unique Error Signature of the Four-subunit Yeast DNA Polymerase ε. *J. Biol. Chem.* **278**, 43770–43780 (2003).

11. Fortune, J. M. *et al.* Saccharomyces cerevisiae DNA Polymerase δ: High fidelity for base substitutions but lower fidelity for single-and multi-base deletions. *J. Biol. Chem.* **280**, 29980–29987 (2005).

12. Macheret, M. & Halazonetis, T. D. DNA replication stress as a hallmark of cancer. *Annu. Rev. Pathol. Mech. Dis.* **10**, 425–448 (2015).

13. Berti, M., Cortez, D. & Lopes, M. The plasticity of DNA replication forks in response to clinically relevant genotoxic stress. *Nat. Rev. Mol. Cell Biol.* **21**, 633–651 (2020).

14. Bellelli, R. & Boulton, S. J. Spotlight on the Replisome: Aetiology of DNA Replication-Associated Genetic Diseases. *Trends Genet.* **37**, 317–336 (2021).

15. Saxena, S. & Zou, L. Hallmarks of DNA replication stress. *Mol. Cell* **82**, 2298–2314 (2022).

16. Moyer, S. E., Lewis, P. W. & Botchan, M. R. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10236–10241 (2006).

17. Ilves, I., Petojevic, T., Pesavento, J. J. & Botchan, M. R. Activation of the MCM2-7 Helicase by Association with Cdc45 and GINS Proteins. *Mol. Cell* **37**, 247–258 (2010).

18. Abid Ali, F. *et al.* Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate. *Nat. Commun.* **7**, (2016).

19. Douglas, M. E., Ali, F. A., Costa, A. & Diffley, J. F. X. The mechanism of eukaryotic CMG helicase activation. *Nature* **555**, 265–268 (2018).

20. Eickhoff, P. *et al.* Molecular Basis for ATP-Hydrolysis-Driven DNA Translocation by the CMG Helicase of the Eukaryotic Replisome. *Cell Rep.* **28**, 2673-2688.e8 (2019).

21. Georgescu, R. *et al.* 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**, E697–E706 (2017).

22. Kose, H. B., Larsen, N. B., Duxin, J. P. & Yardimci, H. Dynamics of the Eukaryotic Replicative Helicase at Lagging-Strand Protein Barriers Support the Steric Exclusion Model. *Cell Rep.* **26**, 2113-2125.e6 (2019).

23. Yuan, Z. *et al.* DNA unwinding mechanism of a eukaryotic replicative CMG helicase. *Nat. Commun.* **11**, 1–10 (2020).

24. Nick McElhinny, S. A., Gordenin, D. A., Stith, C. M., Burgers, P. M. J. & Kunkel, T. A. Division of Labor at the Eukaryotic Replication Fork. *Mol. Cell* **30**, 137–144 (2008).

25. Lewis, J. S. *et al.* Tunability of DNA polymerase stability during eukaryotic DNA replication. *Mol. Cell* **77**, 1–9 (2020).

26. Georgescu, R. E. *et al.* Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat. Struct. Mol. Biol.* **21**, 664–670 (2014).

27. Georgescu, R. E. *et al.* Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. *Elife* **2015**, 1–20 (2015).

28. Yuan, Z. *et al.* Ctf4 organizes sister replisomes and Pol α into a replication factory. *bioRxiv* (2019) doi:10.1101/735746.

29. Zhou, J. C. *et al.* CMG-Pol epsilon dynamics suggests a mechanism for the establishment of leading-strand synthesis in the eukaryotic replisome. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4141–4146 (2017).

30. Taylor, M. R. G. & Yeeles, J. T. P. Dynamics of Replication Fork Progression Following Helicase–Polymerase Uncoupling in Eukaryotes. *J. Mol. Biol.* **431**, 2040–2049 (2019).

31. Guilliam, T. A. & Yeeles, J. T. P. Reconstitution of translesion synthesis reveals a mechanism of eukaryotic DNA replication restart. *Nat. Struct. Mol. Biol.* **27**, 450–460 (2020).

32. Gupta, M. K. *et al.* Protein-DNA complexes are the primary sources of replication fork pausing in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7252–7257 (2013).

33. Ragheb, M. & Merrikh, H. The enigmatic role of Mfd in replication-transcription conflicts in bacteria. *DNA Repair* vol. 81 1568–7864 (2019).

34. Duxin, J. P., Dewar, J. M., Yardimci, H. & Walter, J. C. Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell* **159**, 346–357 (2014).

35. Stingele, J., Schwarz, M. S., Bloemeke, N., Wolf, P. G. & Jentsch, S. A DNA-dependent protease involved in DNA-protein crosslink repair. *Cell* **158**, 327–338 (2014).

36. Nakano, T. *et al.* Translocation and stability of replicative DNA helicases upon encountering DNA-protein cross-links. *J. Biol. Chem.* **288**, 4649–4658 (2013).

37. Aguilera, A. & García-Muse, T. Causes of Genome Instability. *Annu. Rev. Genet.* **47**, 1–32 (2013).

38. Casas-Delucchi, C. S., Daza-Martin, M., Williams, S. L. & Coster, G. The mechanism of replication stalling and recovery within repetitive DNA. *Nat. Commun.* **13**, 1–20 (2022).

39. Williams, S. L. *et al.* Replication-induced DNA secondary structures drive fork uncoupling and breakage. *EMBO J.* **42**, 1–24 (2023).

40. Sato, K., Martin-Pintado, N., Post, H., Altelaar, M. & Knipscheer, P. Multistep mechanism of G-quadruplex resolution during DNA replication. *Sci. Adv.* **7**, (2021).

41. Kornberg, R. Chromatin Structure : A Repeating Unit of Histones and DNA Chromatin structure is based on a repeating unit of eight. *Science (80-. ).* **184**, 868–871 (1974).

42. Olins, A. L. & Olins, D. E. Spheroid chromatin units (v bodies). *Science (80-. ).* **183**, 330–332 (1974).

43. Oudet, P. Electron Microscopic and Biochemical evidence that Chromatin Structure Is a Repeating Unit*. *Cell* **4**, 281–300 (1975).

44. Stewart-Morgan, K. R., Petryk, N. & Groth, A. Chromatin replication and epigenetic cell memory. *Nature Cell Biology* vol. 22 361–371 (2020).

45. Escobar, T. M., Loyola, A. & Reinberg, D. Parental nucleosome segregation and the inheritance of cellular identity. *Nature Reviews Genetics* vol. 22 379–392 (2021).

46. Verreault, A., Kaufman, P. D., Kobayashi, R. & Stillman, B. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* **87**, 95–104 (1996).

47. Sauer, P. V *et al.* Mechanistic insights into histone deposition and nucleosome assembly by the chromatin assembly factor-1. *Nucleic Acids Res.* **46**, 9907–9917 (2018).

48. Cortez, D. Replication-Coupled DNA Repair. *Molecular Cell* vol. 74 866–876 (2019).

49. Sparks, J. L. *et al.* The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell* **176**, 167-181.e21 (2019).

50. Langston, L. D. *et al.* CMG helicase and DNA polymerase ε form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15390–15395 (2014).

51. Yeeles, J. T. P., Deegan, T. D., Janska, A., Early, A. & Diffley, J. F. X. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**, 431–435 (2015).

52. Baretić, D. *et al.* Cryo-EM Structure of the Fork Protection Complex Bound to CMG at a Replication Fork. *Mol. Cell* **78**, 926-940.e13 (2020).

53. Burnham, D. R., Kose, H. B., Hoyle, R. B. & Yardimci, H. The mechanism of DNA

unwinding by the eukaryotic replicative helicase. *Nat. Commun.* **10**, 1–14 (2019).

54. Goswami, P. *et al.* Structure of DNA-CMG-Pol epsilon elucidates the roles of the non-catalytic polymerase modules in the eukaryotic replisome. *Nat. Commun.* **9**, (2018).

55. Kose, H. B., Larsen, N. B., Duxin, J. P. & Yardimci, H. Dynamics of the Eukaryotic Replicative Helicase at Lagging-Strand Protein Barriers Support the Steric Exclusion Model. *Cell Rep.* **26**, 2113-2125.e6 (2019).

56. Lewis, J. S. *et al.* 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**, 10630–10635 (2017).

57. Kurat, C. F., Yeeles, J. T. P., Patel, H., Early, A. & Diffley, J. F. X. Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates. *Mol. Cell* **65**, 117–130 (2017).

58. Aves, S. J., Liu, Y. & Richards, T. A. Evolutionary Diversification of Eukaryotic DNA Replication Machinery. in *The Eukaryotic Replisome: a Guide to Protein Structure and Function* (ed. S., M.) 19–35 (Springer, 2012).

59. Jenkyn-Bedford, M. *et al.* A conserved mechanism for regulating replisome disassembly in eukaryotes. *Nature* **600**, 743–747 (2021).

60. Jones, M. L., Baris, Y., Taylor, M. R. G. & Yeeles, J. T. P. Structure of a human replisome shows the organisation and interactions of a DNA replication machine. *EMBO J.* **40**, 1–23 (2021).

61. Baris, Y., Taylor, M. R. G., Aria, V. & Yeeles, J. T. P. Fast and efficient DNA replication with purified human proteins. *Nature* **606**, 204–210 (2022).

62. Champasa, K., Blank, C., Friedman, L. J., Gelles, J. & Bell, S. P. A conserved Mcm4 motif is required for Mcm2-7 double-hexamer formation and origin DNA unwinding. *Elife* **8**, 1–23 (2019).

63. Hsiao, C. L. & Carbon, J. High-frequency transformation of yeast by plasmids containing the cloned yeast ARG4 gene. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 3829–3833 (1979).

64. Stinchcomb, D. T., Struhl, K. & Davis, R. W. Isolation and characterisation of a yeast chromosomal replicator. *Nature* **282**, 39–43 (1979).

65. Bell, S. P. & Stillman, B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**, 128–134 (1992).

66. Sánchez, H. *et al.* DNA replication origins retain mobile licensing proteins. *Nat. Commun.* **12**, (2021).

67. Frigola, J., Remus, D., Mehanna, A. & Diffley, J. F. X. ATPase-dependent quality control of DNA replication origin licensing. *Nature* **495**, 339–343 (2013).

68. Coster, G., Frigola, J., Beuron, F., Morris, E. P. & Diffley, J. F. X. Origin Licensing Requires ATP Binding and Hydrolysis by the MCM Replicative Helicase. *Mol. Cell* **55**, 666–677 (2014).

69. Remus, D. *et al.* Concerted Loading of Mcm2-7 Double Hexamers around DNA during DNA Replication Origin Licensing. *Cell* **139**, 719–730 (2009).

70. Coster, G. & Diffley, J. F. X. Bidirectional eukaryotic DNA replication is established by quasi-symmetrical helicase loading. *Science (80-. ).* **357**, 314–318 (2017).

71. Duzdevich, D. *et al.* The dynamics of eukaryotic replication initiation: Origin specificity, licensing, and firing at the single-molecule level. *Mol. Cell* **58**, 483–494 (2015).

72. Ticau, S. *et al.* Mechanism and timing of Mcm2-7 ring closure during DNA replication origin licensing. *Nat. Struct. Mol. Biol.* **24**, 309–315 (2017).

73. Gambus, A., Khoudoli, G. A., Jones, R. C. & Blow, J. J. MCM2-7 form double hexamers at licensed origins in Xenopus egg extract. *J. Biol. Chem.* **286**, 11855–11864 (2011).

74. Ticau, S., Friedman, L. J., Ivica, N. A., Gelles, J. & Bell, S. P. Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell* **161**, 513–525 (2015).

75. Abid Ali, F. *et al.* Cryo-EM structure of a licensed DNA replication origin. *Nat. Commun.* **8**, 1–10 (2017).

76. Lim, C. T. *et al.* Cell Cycle Regulation has Shaped Budding Yeast Replication Origin Structure and Function. *bioRxiv* 1–33 (2024).

77. Chistol, G. & Walter, J. C. Single-molecule visualization of MCM2-7 DNA loading: Seeing is Believing. *Cell* **161**, 429–430 (2015).

78. Sheu, Y. J. & Stillman, B. Cdc7-Dbf4 Phosphorylates MCM Proteins via a Docking Site-Mediated Mechanism to Promote S Phase Progression. *Mol. Cell* **24**, 101–113 (2006).

79. Sheu, Y. J. & Stillman, B. The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature* **463**, 113–117 (2010).

80. Randell, J. C. W. *et al.* Mec1 Is One of Multiple Kinases that Prime the Mcm2-7 Helicase for Phosphorylation by Cdc7. *Mol. Cell* **40**, 353–363 (2010).

81. de Jesús-Kim, L., Friedman, L. J., Ramsoomair, C., Gelles, J. & Bell, S. P. DDK regulates replication initiation by controlling the multiplicity of Cdc45-GINS

binding to Mcm2-7. *Elife* **10:e65471**, 1–30 (2021).

82. Lewis, J. S. *et al.* Mechanism of replication origin melting nucleated by CMG helicase assembly. *Nature* (2022) doi:10.1038/s41586-022-04829-4.

83. Vrtis, K. B. *et al.* Single-strand DNA breaks cause replisome disassembly. *Mol. Cell* **81**, 1309-1318.e6 (2021).

84. Low, E., Chistol, G., Zaher, M. S., Kochenova, O. V. & Walter, J. C. The DNA replication fork suppresses CMG unloading from chromatin before termination. *Genes Dev.* **34**, 1534–1545 (2020).

85. Gruszka, D. T., Xie, S., Kimura, H. & Yardimci, H. Single-molecule imaging reveals control of parental histone recycling by free histones during DNA replication. *Sci. Adv.* **6**, (2020).

86. Cvetkovic, M. A. *et al.* The structural mechanism of dimeric DONSON in replicative helicase activation. *Mol. Cell* **83**, 4017-4031.e9 (2023).

87. Georgescu, R. E. *et al.* Supplement- Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat. Struct. Mol. Biol.* **21**, 664–670 (2014).

88. Rzechorzek, N. J., Hardwick, S. W., Jatikusumo, V. A., Chirgadze, D. Y. & Pellegrini, L. CryoEM structures of human CMG-ATPγS-DNA and CMG-AND-1 complexes. *Nucleic Acids Res.* **48**, 6980–6995 (2020).

89. Kang, Y. H., Galal, W. C., Farina, A., Tappin, I. & Hurwitz, J. Properties of the human Cdc45/Mcm2-7/GINS helicase complex and its action with DNA polymerase ε in rolling circle DNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 6042–6047 (2012).

90. Langston, L. D. *et al.* Mcm10 promotes rapid isomerization of CMG-DNA for replisome bypass of lagging strand DNA blocks. *Elife* **6**, 1–21 (2017).

91. Mayle, R. *et al.* Mcm10 has potent strand-annealing activity and limits translocase-mediated fork regression. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 798–803 (2019).

92. Lewis, J. S. *et al.* 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**, 10630–10635 (2017).

93. Kose, H. B., Xie, S., Cameron, G., Strycharska, M. S. & Yardimci, H. Duplex DNA engagement and RPA oppositely regulate the DNA-unwinding rate of CMG helicase. *Nat. Commun.* **11**, 1–15 (2020).

94. McClure, A. W. & Diffley, J. F. X. Rad53 checkpoint kinase regulation of dna replication fork rate via mrc1 phosphorylation. *Elife* **10**, 1–24 (2021).

95. Langston, L. & O'Donnell, M. Action of CMG with strand-specific DNA blocks supports an internal unwinding mode for the eukaryotic replicative helicase. *Elife* **6**, 1–23 (2017).

96. Dulin, D. *et al.* High Spatiotemporal-Resolution Magnetic Tweezers: Calibration and Applications for DNA Dynamics. *Biophys. J.* **109**, 2113–2125 (2015).

97. Choi, J., Grosely, R., Puglisi, E. V. & Puglisi, J. D. Expanding single-molecule fluorescence spectroscopy to capture complexity in biology. *Curr. Opin. Struct. Biol.* **58**, 233–240 (2019).

98. Bustamante, C. J., Chemla, Y. R., Liu, S. & Wang, M. D. Optical tweezers in single-molecule biophysics. *Nat. Rev. Methods Prim.* **1**, (2021).

99. Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460–465 (2005).

100. Shaevitz, J. W., Abbondanzieri, E. A., Landick, R. & Block, S. M. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* **426**, 684–687 (2003).

101. Gupta, S., Friedman, L. J., Gelles, J. & Bell, S. P. A helicase- - tethered ORC flip enables bidirectional helicase loading. *Elife* **10:e74282**, 1–31 (2021).

102. Lewis, J. S., van Oijen, A. M. & Spenkelink, L. M. Embracing Heterogeneity: Challenging the Paradigm of Replisomes as Deterministic Machines. *Chem. Rev.* **123**, 13419–13440 (2023).

103. Dulin, D., Berghuis, B. A., Depken, M. & Dekker, N. H. Untangling reaction pathways through modern approaches to high-throughput single-molecule force-spectroscopy experiments. *Curr. Opin. Struct. Biol.* **34**, 116–122 (2015).

104. Amasino, A. L., Gupta, S., Friedman, L. J., Gelles, J. & Bell, S. P. Regulation of replication origin licensing by ORC phosphorylation reveals a two- step mechanism for Mcm2- 7 ring closing Audra. *Proc. Natl. Acad. Sci.* **120**, 1–12 (2023).

105. Zhang, A., Friedman, L. J., Gelles, J. & Bell, S. P. Changing protein–DNA interactions promote ORC binding- site exchange during replication origin licensing. *Proc. Natl. Acad. Sci.* **120**, 1–12 (2023).

106. Scherr, M. J., Wahab, S. A., Remus, D. & Duderstadt, K. E. Mobile origin-licensing factors confer resistance to conflicts with RNA polymerase ll ll Mobile origin-licensing factors confer resistance to conflicts with RNA polymerase. *CellReports* **38**, (2022).

107. Wasserman, M. R., Schauer, G. D., O'Donnell, M. E. & Liu, S. Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase. *Cell* **178**, 600-611.e16 (2019).

108. Sánchez, H. *et al.* A chromatinized origin reduces the mobility of ORC and MCM through interactions and spatial constraint. *Nat. Commun.* **14**, (2023).

109. Li, S. *et al.* Origin recognition complex harbors an intrinsic nucleosome remodeling activity. *Proc. Natl. Acad. Sci. U. S. A.* **119**, 1–10 (2022).

110. Li, S. *et al.* Nucleosome-directed replication origin licensing independent of a consensus DNA sequence. *Nat. Commun.* **13**, (2022).

111. Axelrod, D., Thompson, N. L. & Burghardt, T. P. Total internal reflection fluorescent microscopy. *J. Microsc.* **129**, 19–28 (1983).

112. Martin-Fernandez, M. L., Tynan, C. J. & Webb, S. E. D. A "pocket guide" to total internal reflection fluorescence. *J. Microsc.* **252**, 16–22 (2013).

113. White, D. S., Smith, M. A., Chanda, B. & Goldsmith, R. H. Strategies for Overcoming the Single-Molecule Concentration Barrier. *ACS Meas. Sci. Au* **3**, 239–257 (2023).

114. Friedman, L. J., Chung, J. & Gelles, J. Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence. *Biophys. J.* **91**, 1023–1031 (2006).

115. Lerner, E. *et al.* FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices. *Elife* **10**, 1–69 (2021).

116. Yardimci, H., Loveland, A. B., van Oijen, A. M. & Walter, J. C. Single-molecule analysis of DNA replication in Xenopus egg extracts. *Methods* **57**, 179–186 (2012).

117. Schauer, G. D. *et al.* Replisome bypass of a protein-based R-loop block by Pif1. *PNAS* **117**, 30354–30361 (2020).

118. Ashkin, A., Dziedzic, J. M., Bjorkholm, J. E. & Chu, S. Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Angular Momentum* **11**, 288–290 (1986).

119. Moffitt, J. R., Chemla, Y. R., Izhaky, D. & Bustamante, C. Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9006–9011 (2006).

120. Candelli, A., Wuite, G. J. L. & Peterman, E. J. G. Combining optical trapping, fluorescence microscopy and micro-fluidics for single molecule studies of DNA-protein interactions. *Phys. Chem. Chem. Phys.* **13**, 7263–7272 (2011).

121. Candelli, A. *et al.* Visualization and quantification of nascent RAD51 filament

formation at single-monomer resolution. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15090–15095 (2014).

122. Carcamo, C. C. *et al.* ATP binding enhances SWR1 target search by promoting 1D scanning on DNA. *Biophys. J.* **121**, 284a (2022).

123. Chua, G. N. L. & Liu, S. When Force Met Fluorescence: Single-Molecule Manipulation and Visualization of Protein–DNA Interactions. *Annu. Rev. Biophys.* **53**, 169–191 (2024).

124. Liachko, I. *et al.* A comprehensive genome-wide map of autonomously replicating sequences in a naive genome. *PLoS Genet.* **6**, 22 (2010).

125. Kapadia, N. *et al.* Processive Activity of Replicative DNA Polymerases in the Replisome of Live Eukaryotic Cells. *Mol. Cell* **80**, 114-126.e8 (2020).

126. Claussin, C., Vazquez, J. & Whitehouse, I. Single-molecule mapping of replisome progression. *Mol. Cell* **82**, 1372-1382.e4 (2022).

127. Polo Rivera, C. & Deegan, T. D. Replicon-seq: seeing is believing. *Trends Genet.* **38**, 987–988 (2022).

128. Parker, M. W. *et al.* A new class of disordered elements controls DNA replication through initiator self-assembly. *Elife* **8**, 1–35 (2019).

# 2

# Nucleotide binding halts diffusion of the eukaryotic replicative helicase during activation

*This chapter is published as: Ramírez Montero D., Sánchez H., van Veen E., van Laar T., Solano B., Diffley J.X., and Dekker, N.H. (2023). Nucleotide binding halts diffusion of the eukaryotic replicative helicase during activation. Nature Communications 14:2082.*

## 2.1 Abstract

The eukaryotic replicative helicase CMG centrally orchestrates the replisome and leads the way at the front of replication forks. Understanding the motion of CMG on the DNA is therefore key to our understanding of DNA replication. *In vivo*, CMG is assembled and activated through a cell-cycle-regulated mechanism involving 36 polypeptides that has been reconstituted from purified proteins in ensemble biochemical studies. Conversely, single-molecule studies of CMG motion have thus far relied on pre-formed CMG assembled through an unknown mechanism upon overexpression of individual constituents. Here, we report the activation of CMG fully reconstituted from purified yeast proteins and the quantification of its motion at the single-molecule level. We observe that CMG can move on DNA in two ways: by unidirectional translocation and by diffusion. We demonstrate that CMG preferentially exhibits unidirectional translocation in the presence of ATP, whereas it preferentially exhibits diffusive motion in the absence of ATP. We also demonstrate that nucleotide binding halts diffusive CMG independently of DNA melting. Taken together, our findings support a mechanism by which nucleotide binding allows newly assembled CMG to engage with the DNA within its central channel, halting its diffusion and facilitating the initial DNA melting required to initiate DNA replication.

## 2.2 Introduction

Eukaryotic DNA replication is catalyzed by a MDa-sized dynamic protein complex known as the replisome. The replisome is powered by the replicative helicase CMG (Cdc45/Mcm2-7/GINS), which centrally orchestrates the other components and leads the way at the front of replication forks[1]. Understanding the motion of CMG on DNA is therefore crucial to our understanding of how cells successfully replicate DNA. *In vivo*, loading and activation of CMG occur in temporally separated fashion. In *Saccharomyces cerevisiae* in particular, CMG loading occurs at specific sequences known as origins of replication[1]. First, in the G1-phase of the cell cycle, a set of proteins known as 'loading factors' scans the DNA until such origins of replication are located, at which inactive single and double Mcm2-7 hexamers are then loaded onto dsDNA[2–5]. In the subsequent S-phase, double Mcm2-7 hexamers are selectively phosphorylated by the cell cycle-regulated Dbf4-dependent kinase (DDK)[6]. Then, a set of proteins known as 'firing factors' facilitates the assembly of full CMG by recruiting the helicase-activating factors Cdc45 and GINS to the phosphorylated Mcm2-7 double hexamers[7,8]. Upon full assembly, CMG must transition from encircling dsDNA to encircling ssDNA, so that it can unwind dsDNA by steric exclusion of the non-translocation strand[9]. This transition is known as CMG activation and consists of two steps. In the first step, ATP binding allows each CMG in a double hexamer to melt 0.6-0.7 turns of dsDNA within its central channel[10,11]. In the second step, each CMG extrudes one strand of the double helix from its central channel; this final step requires ATP hydrolysis and the action of the firing factor Mcm10[10]. After having extruded one strand of DNA, each activated sister CMG translocates on ssDNA in a 3'-to-5' direction by hydrolyzing ATP[1,12,13], allowing the two helicases to bypass and move away from each other[10], and committing the cell to initiate DNA replication[1]. This entire process requires a minimal set of 36 polypeptides and has been fully reconstituted from purified *Saccharomyces cerevisiae* proteins in ensemble biochemical studies[10].

To date, the motion of fully reconstituted CMG has been studied in bulk biochemical assays[10,14] with a temporal resolution of minutes. These ensemble biochemical studies have provided us with important insights into the average behavior of CMG; nonetheless, lower probability behaviors are averaged out in the ensemble readouts. On the other hand, single-molecule studies have the power to isolate and study low probability events with higher temporal resolution[15]; nevertheless, single-molecule studies of CMG motion so far[16–18] have focused on pre-formed CMG assembled through a poorly understood mechanism that requires co-overexpression of individual subunits[19,20]. What is more, given that such pre-

formed CMG requires an artificial region of ssDNA to bind DNA[16,19], these studies could not access the intricacies of CMG activation, such as the role of CMG motion during this process. We therefore set out to assemble and activate CMG from purified *Saccharomyces cerevisiae* proteins and study its motion during this process at the single-molecule level.

## 2.3 Results

### 2.3.1 A hybrid ensemble and single-molecule assay to visualize fully reconstituted CMG

One of the biggest challenges of studying the motion of fully reconstituted CMG at the single-molecule level is to prevent all the proteins involved from aggregating onto the long (tens of kbp) DNA molecules needed to observe motion of diffraction-limited fluorescent spots[16,17,21]. To overcome this challenge, we developed a hybrid ensemble and single-molecule assay to 1) assemble and activate fully reconstituted fluorescent CMG onto ~24 kbp DNA molecules containing a natural yeast ARS1 replication origin in an aggregation-free manner; and 2) use a combination of dual optical trapping and confocal scanning microscopy[22] to image and quantify the motion of fluorescent CMG along DNA molecules held in an optical trap (**Fig. 2.1a-*i***, **S2.1a**, **Methods**). To this end, we added both digoxigenin and desthiobiotin moieties at each end of a linear 23.6 kb DNA containing a natural ARS1 origin of replication. We then bound the functionalized DNA to streptavidin-coated magnetic beads and used it to assemble and activate CMG (**Methods**). In short, we loaded Mcm2-7 hexamers onto the bead-bound DNA, phosphorylated double Mcm2-7 hexamers with DDK and washed the beads with a buffer solution containing 300 mM KCl. We then assembled *and* activated CMG for 15 min in the presence of fluorescently labeled Cdc45$^{LD555}$ (**Fig. S2.1a**), which supports DNA unwinding near WT levels (**Fig. S2.1b**). Following CMG assembly and activation, we washed the beads again with a buffer solution containing 300 mM KCl to select for fully mature CMG[7,8], and 'paused' the reaction by removing ATP. DNA:CMG complexes were then eluted from the magnetic beads by competing the desthiobiotin-streptavidin interaction with an excess of free biotin[23]. Following elution, DNA:CMG complexes were tethered between two optically-trapped anti-digoxigenin-coated polystyrene beads, and transferred into a buffer solution containing Mcm10, RPA and either ATP, no nucleotide, or the slowly hydrolyzable ATP analog ATPγS. We then scanned the DNA with a confocal scanning laser and observed fluorescent CMG helicases as diffraction-limited spots on the otherwise unlabeled DNA (**Fig. 2.1a-*ii***). Approximately a third of the trapped DNA molecules contained diffraction-limited

fluorescent CMG spots, typically a single one (**Fig. 2.1b**). We deduced the number of CMG per diffraction-limited spot by counting the photobleaching steps within each spot (**Fig. S2.3**, **Methods**). As this showed that most spots contained 1 CMG (**Fig. 2.1c**), it followed that most DNA molecules had a total of 1 CMG (**Fig. S2.1c**), where *a priori* one might have expected a total closer to 2 or multiples thereof. We consider it unlikely that our experimentally measured lower number is substantially influenced by the labeling efficiency of Cdc45$^{LD555}$, which we measured to be 85 ± 4 % (**Methods**); rather, we attribute it to (a potential combination of) other factors including loss of Cdc45 during the high salt washes and downstream handling, CMG dissociation at nicks[24] on the DNA during the ensemble activation, or CMG diffusing off the ends of the DNA during elution. Furthermore, it was recently shown that each Mcm2-7 in a double hexamer independently matures into CMG[8]. Thus, we cannot discard the possibility that in our system only one of the two Mcm2-7 hexamers is fully matured into CMG[25].

### 2.3.2 Mature CMG is preferentially assembled near origins of replication

We first looked at the initial positions of CMG on the DNA. Of note, because we cannot differentiate between the two possible orientations of the DNA in our experiments, we display the initial positions of CMG in plots showing the distance from the center of the DNA[2]. We observed a wide distribution of initial positions with a peak near or at the ARS1 origin (**Fig. 2.1d**). Furthermore, spots containing two CMG complexes were less widely distributed around the origin than spots containing one CMG (**Fig. 2.1e, f**). Taken together, these results are consistent with a preferential assembly of sister CMG helicases near the ARS1 origin, followed by the motion of individual activated helicases away from the origin during the 15-min ensemble activation reaction.

### 2.3.3 Colocalization of fluorescent Cdc45 and fluorescent Mcm2-7 hexamers is DDK-dependent

Salt-resistant Cdc45 is considered a hallmark of mature CMG[7,8]. Nonetheless, if the Cdc45$^{LD555}$ spots that we observe are part of *bona fide* CMG, their presence on the DNA should be dependent on DDK[6–8]. To confirm this, we quantified the co-localization of red fluorescently labeled Mcm2-7$^{JF646\text{-}Mcm3}$ with green fluorescently labeled Cdc45$^{LD555}$ (shown to jointly support DNA unwinding (**Fig. S2.2a**)) in the presence and absence of DDK (**Fig. 2.1g-j**). While nearly 20% of Mcm2-7$^{JF646\text{-}Mcm3}$ spots colocalized with Cdc45$^{LD555}$ in the presence of DDK, we observed an ~8-fold decrease in this colocalization in the absence of DDK (**Fig. 2.1h-j**). The 20% colocalization that we observe in the presence of DDK is in agreement with

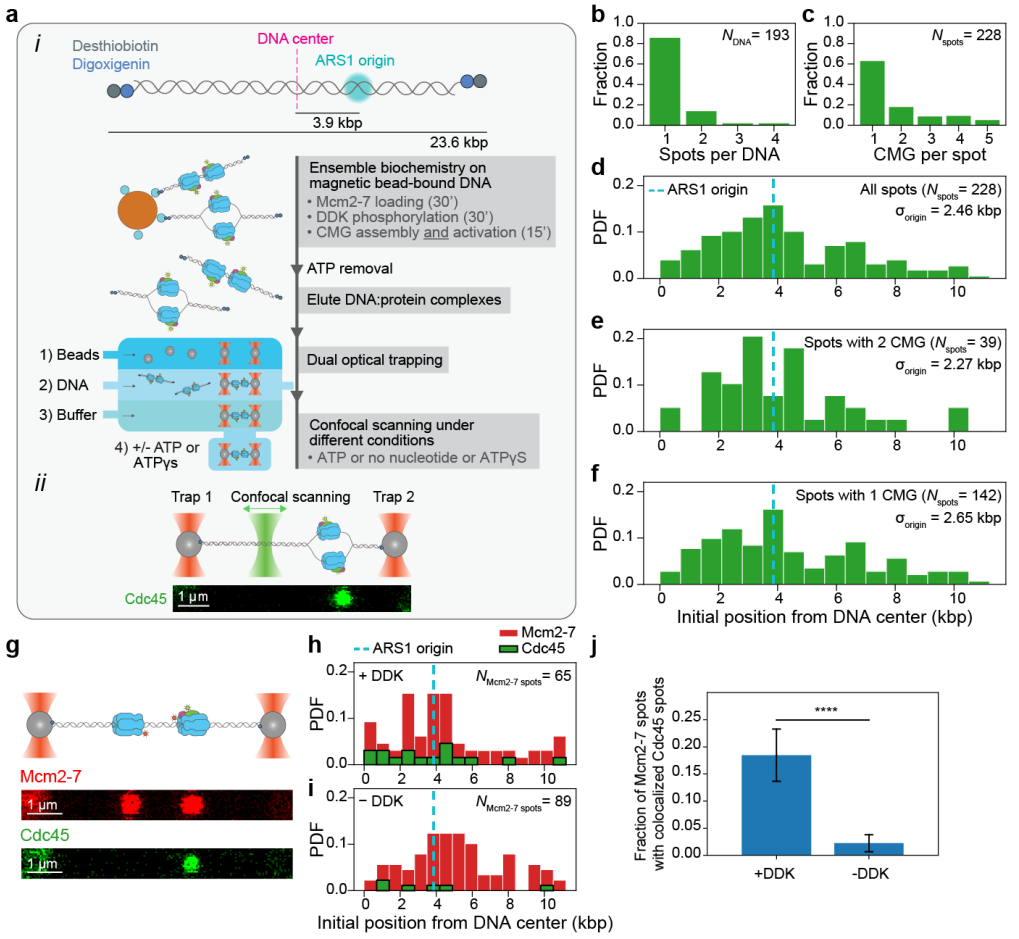



*Figure 2.1* Single-molecule imaging of fully reconstituted CMG. **a-i** Description of hybrid ensemble and single-molecule assay to image fully reconstituted CMG. **a-ii** Example scan of an optically trapped DNA molecule containing one CMG diffraction-limited spot. **b** Distribution of numbers of CMG diffraction-limited spots per DNA. **c** Distribution of numbers of CMG complexes within each diffraction-limited spot. **d-f** Distribution of initial positions on the DNA of **d** all CMG diffraction-limited spots, **e**, diffraction-limited spots containing 2 CMG complexes or **f** diffraction-limited spots containing 1 CMG complex; the ARS1 origin of replication is indicated by the dashed cyan line. **g** Example scans separately showing Mcm2-7$^{JF646}$ diffraction-limited spots (top) and Cdc45$^{LD555}$ diffraction-limited spots (bottom) on the same DNA molecule. **h-i** Distributions of initial positions of Mcm2-7$^{JF646}$ spots and Cdc45$^{LD555}$ spots on the DNA in the **h** presence or **i** absence of DDK. In each condition (with or without DDK), the histograms of Mcm2-7$^{JF646}$ and Cdc45$^{LD555}$ initial positions are weighted by the total number of Mcm2-7$^{JF646}$ spots. **j** Mean fraction of Mcm2-7$^{JF646}$ diffraction-limited spots with colocalized Cdc45$^{LD555}$ diffraction-limited spots in the presence ($N_{\text{Mcm2-7 spots}}$=65) or absence ($N_{\text{Mcm2-7 spots}}$=89) of DDK; error bars show the standard error of proportion. Statistical significance was obtained from a two-sided binomial test (p-value=2.2 × $10^{-8}$).

previous observations that the *in vitro* assembly of CMG is less efficient than the loading of Mcm2-7 double hexamers[10,11]. Taken together, these results show that the Cdc45$^{LD555}$ fluorescent spots in our images correspond to *bona fide* CMG. We attribute the residual colocalization of Mcm2-7$^{JF646\text{-}Mcm3}$ and Cdc45$^{LD555}$ in the absence of DDK to non-specific interactions (**Fig. S2.2h**) and/or to traces of phosphorylated Mcm2-7 in the protein preparation[8].

### 2.3.4 Fully reconstituted CMG exhibits two quantitatively distinct motion types

We next sought to quantify the motion of CMG in the presence of ATP. For this, we implemented a change-point algorithm (CPA) to fit linear segments through regions of the position-vs.-time plots of individual spots (**Fig. 2.2a-c**, **Methods**); the slopes of these segments then give us a noise-reduced value of the instantaneous velocities of individual fluorescent spots. To calibrate our analysis, we imaged dCas9$^{LD555}$ with the same imaging conditions that we used for CMG (**Fig. 2.2a**, **Fig. S2.3**); because dCas9$^{LD555}$ is static on the DNA, it provides us with a measure of the velocity error in our system. After drift correction, the distribution of instantaneous velocities of fluorescent dCas9$^{LD555}$ spots after the CPA fit is centered at 0 bp/s and has a width $\sigma_{dCas9}$ that reflects our experimental uncertainty in velocity measurement (**Fig. 2.2a inset**). For all CMG motion analysis, we defined a conservative velocity cutoff of 5 × $\sigma_{dCas9}$ (= 2.0 bp/s) to categorize fluorescent spots as static or mobile; we considered mobile any fluorescent spot with at least one CPA segment with a slope above this threshold, and all other spots static.

Following the approach described above, we determined that ~70% of CMG spots are mobile when imaged in a buffer solution containing RPA, Mcm10 and ATP (**Figs. 2.2b**, **2.2d**, and **S2.5a**). Unexpectedly, when we imaged CMG in a buffer solution containing RPA, Mcm10 and no ATP, we observed that ~40% of CMG spots were also mobile (**Figs. 2.2c, d** and **S2.5d**). Nonetheless, we noticed qualitative differences in the motion of CMG in the presence and absence of ATP: while CMG seemed to move unidirectionally in the presence of ATP (**Fig. 2.2b**), it appeared to move in a more random (e.g. diffusive) manner in the absence of ATP (**Fig. 2.2c**). To quantitatively characterize these two apparently distinct motion types, we employed two independent approaches. First, we looked at the CPA segments of all the traces in each condition, and calculated the probability that consecutive segments have the same direction (**Fig. 2,2e**); in the absence of noise, this probability should equal 1 for unidirectional motion, and 0.5 for random motion.

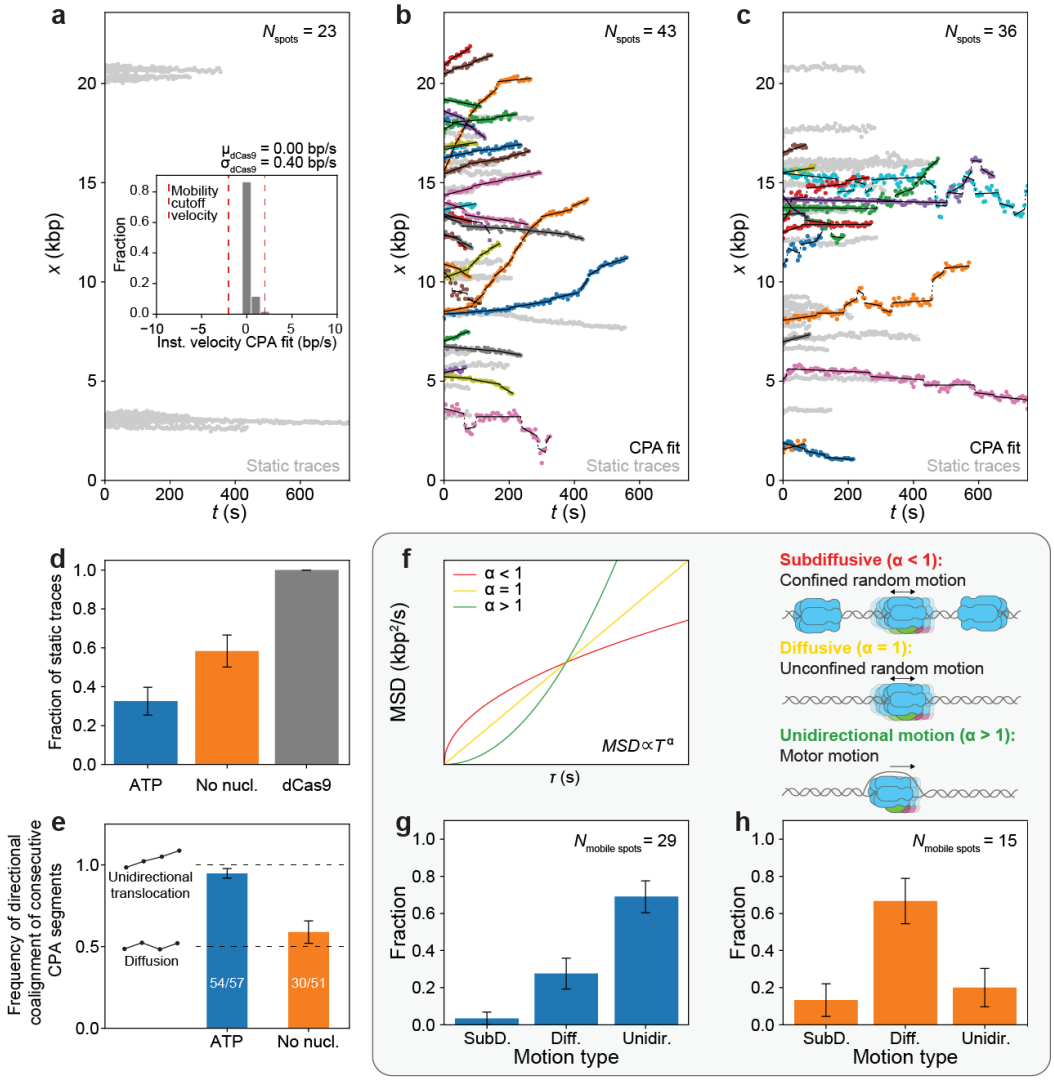



*Figure 2.2* Fully reconstituted CMG exhibits two different motion types. **a** Position vs. time plots of dCas9$^{LD555}$ spots; (inset) distribution of instantaneous velocities coming from the CPA fits of dCas9$^{LD555}$ spots; red lines show the instantaneous velocity cutoff ($5\sigma_{dCas9}$) used to separate CMG spots in **b** and **c** into static or mobile; CPA fits are not shown for clarity. **b, c** Position vs. time plots of CMG spots in the **b** presence of ATP or **c** absence of nucleotide; CPA fits are plotted in black, static traces are shown in light gray. **d** Ratio of static CMG traces in the presence of ATP ($N_{spots}$=43), absence of nucleotide ($N_{spots}$=36), and static dCas9 ($N_{spots}$=23) traces; error bars show the standard error of proportion. **e** Frequency of consecutive CPA segments with the same direction for CMG spots in the presence of ATP ($N_{mobile\ spots}$=29) or absence of nucleotide ($N_{mobile\ spots}$=15); inset diagrams illustrate expected segment directions of a unidirectionally moving spot (top) or a diffusive spot (bottom); error bars show the standard error of proportion. **f** (left panel) Idealized examples of MSD vs. delay time τ plots with an anomalous coefficients α < 1 (red), α = 1 (yellow) and α > 1 (green); (right panel) diagrams illustrating the types of CMG motion corresponding to each of these three cases: constrained diffusion (α << 1), free diffusion (α ≈ 1) or unidirectional motion (α >>1). **g, h** Fraction of mobile CMG traces classified into different motion types in the **g** presence of ATP or **h** absence of nucleotide; error bars show the standard error of proportion.

As seen in **Fig. 2.2e**, our measured probabilities closely match these expected values, providing quantitative underpinning of our initial observations. As an independent approach, we conducted anomalous diffusion analysis of the mobile traces in each condition (**Figs. 2.2f-h** and **S2.5b,e**). For each individual trace, we calculated the mean-squared displacement (MSD) as a function of the lag time τ, and then fitted the result to the equation $\mathrm{MSD}(\tau) \propto \tau^{\alpha}$ to extract the anomalous diffusion coefficient $\alpha$ (**Methods**). The value of $\alpha$ then allowed us to classify each trace into different motion types, as $\alpha \gg 1$ for unidirectionally moving molecules, $\alpha \approx 1$ for freely diffusive molecules, and $0 < \alpha \ll 1$ for molecules undergoing constrained diffusion (**Fig. 2.2f**). This anomalous diffusion analysis confirmed that unidirectional motion is most likely when ATP is present (**Fig. 2.2g**), whereas diffusive behavior is most likely when ATP is absent (**Fig. 2.2h**).

We note that we observed a small population of seemingly diffusive CMG spots in the presence of ATP, and a small population of seemingly unidirectionally moving CMG spots in the absence of ATP (**Fig. 2.2g, h**). We hypothesized that these subpopulations might have arisen from misclassification of short traces[26]. To test this hypothesis, we simulated two populations of single-molecule traces of varying lengths within the range of our experimental data: one population solely consisting of unidirectionally moving traces, and the other population solely consisting of freely diffusive traces (**Methods**). We then carried out the same anomalous diffusion analysis that we did on the experimental CMG data on both simulated data sets. We observed that the distribution of motion types for the simulated unidirectional traces looked very similar to that of the experimental mobile CMG traces in the presence of ATP (**Figs. S2.6a** and **2.2g**), whereas the distribution of motion types for the simulated diffusive traces looked very similar to that of the experimental mobile CMG traces in the absence of ATP (**Figs. S2.6b** and **2.2h**). Thus, the results of our simulations suggest that, overall, the mobile traces in the presence of ATP represent unidirectional motion, whereas the mobile traces in the absence of ATP represent diffusive motion.

### 2.3.5 Analysis of CMG motor motion

Following this identification of two distinct types of CMG mobility, we investigated both in further depth. We first investigated the unidirectional motor motion of CMG, the motion type that powers the replisome. We thus specifically analyzed the velocities of unidirectionally moving CMG spots in the presence of ATP, which yielded a distribution of instantaneous velocities with a peak at ~5 bp/s (**Fig. 2.3a**), consistent with previous single-molecule studies on pre-formed CMG in the presence of RPA[17]. This distribution has a long tail, reaching up to

instantaneous velocities of ~45 bp/s. These higher velocities are low in probability, suggesting that CMG can only achieve these high velocities in short time bursts. Consistent with this, when we calculated the time-averaged velocity of each CMG spot and examined the resulting distribution (**Fig. 2.3a inset**), we did not observe such high velocities. The distribution of time-averaged velocities has a peak at ~5 bp/s, which is consistent with ensemble biochemical studies of CMG motion[14]. Notably, we did not observe any noticeable backtracking of CMG (**Fig. 2.2b**), which is consistent with previous studies suggesting that RPA prevents CMG backtracking by keeping the lagging strand template out of the central channel of CMG[17].

When analyzing the unidirectional traces in the presence of ATP, we also observed a few instances of two unidirectionally translocating CMGs initially located within the same diffraction-limited spot, but that then split from one another and give rise to two diffraction-limited spots of half the intensity of the original spot (**Fig. 2.3c**).

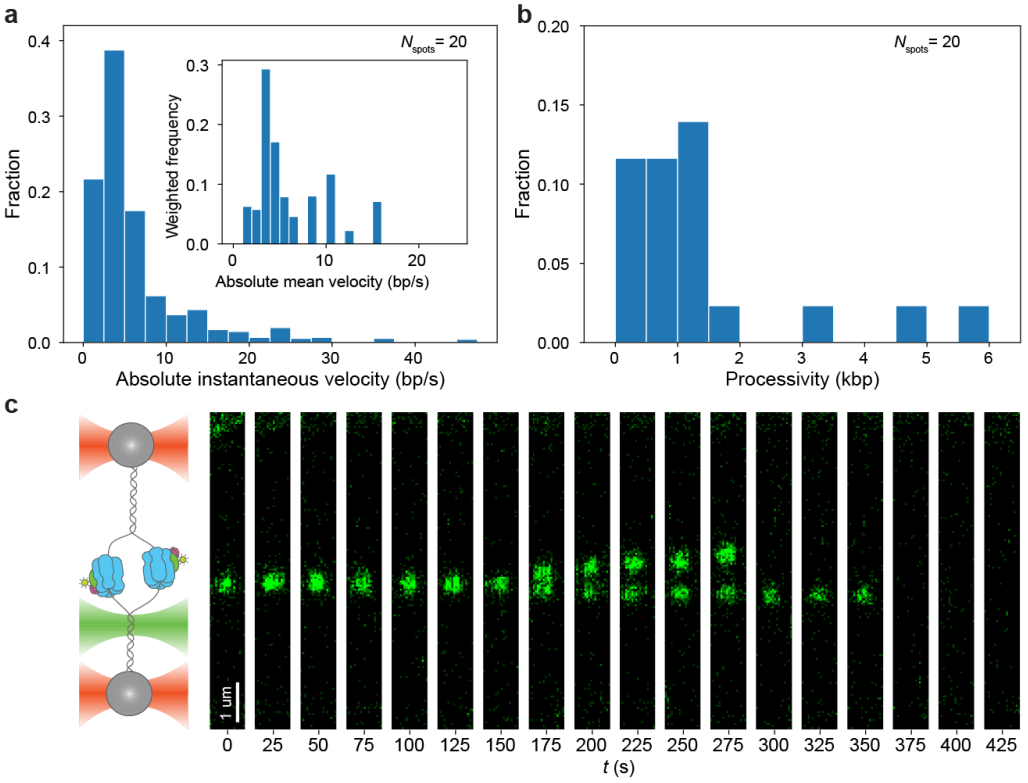



*Figure 2.3* Analysis of CMG motor motion. **a** Distribution of absolute instantaneous velocities of unidirectionally moving CMG spots in the presence of ATP; (inset) Distribution of absolute mean velocities of unidirectionally moving CMG spots in the presence of ATP normalized by the length of each trace. **b** Distribution of processivities of unidirectionally moving CMG spot in the presence of ATP. **c** Example kymograph of two unidirectionally moving CMG spots that start within the same diffraction-limited spot and split up into two distinct diffraction-limited spots that move along the DNA in opposite directions ($N_{\text{splitting events}}$=2).

These observations are consistent with *in vitro* biochemical studies of helicase activation[10] showing that sister CMGs move in opposite directions upon their activation. The low probability of these splitting events is to be expected because i) we allow CMG to become activated and translocate on the DNA for 15 min before imaging, and ii) because most DNA molecules contain 1 CMG (**Fig. S2.1c**).

### 2.3.6 Nucleotide binding halts diffusive CMG

Our data shows that CMG diffuses on DNA in the absence but not in the presence of ATP (**Fig. 2.2b, c**, **g, h**), suggesting that ATP is involved in stopping the diffusive motion of CMG. To investigate whether it was the binding or the hydrolysis of ATP that stopped the diffusive motion of CMG, we investigated CMG motion in a buffer solution supplemented with RPA, Mcm10 and the slowly hydrolysable ATP analog ATPγS. When we imaged CMG under these conditions, the vast majority of CMG spots were found to be static (**Figs. 2.4a, b** and **S2.5i**). Taken together with our data in the presence and absence of ATP (**Fig. 2.2b-d**), our results show that it is the nucleotide binding and not the hydrolysis that halts the diffusive motion of CMG. Furthermore, our results confirm that ATP hydrolysis is required for the unidirectional translocation of CMG[13].

Previous biochemical studies showed that ATP binding allows newly formed CMG to melt 0.6-0.7 turns of the DNA within its central channel[10], which was recently confirmed by cryo electron microscopy[11]. Comparing these previous observations with our single-molecule results led us to hypothesize that 1) the diffusive motion that we observed in the absence of ATP corresponded to CMG surrounding dsDNA, and that 2) the halting of such diffusive motion in the presence of ATPγS is due to CMG melting the DNA within its central channel.

To test whether CMG can diffuse on dsDNA in the absence of ATP, we developed an ensemble CMG sliding assay (**Fig. 2.4d, e**, **Methods**). Briefly, we synthesized two 1.4 kb linear DNA constructs biotinylated at one end and containing an ARS1 origin. The non-biotinylated end of the constructs was then either left as a free end or covalently crosslinked to a M.HpaII methyltransferase[27]. Because the crosslinked methyltransferase is too large to fit inside the central channel of CMG[28,29], it should stop CMG from diffusing off the end of the DNA, which would otherwise be free to diffuse off the free end. We then bound both DNA constructs to streptavidin-coated magnetic beads, and assembled CMG in the presence of fluorescent Cdc45$^{LD555}$ onto them; importantly, we omitted the firing factor Mcm10 from the activation reaction to prevent strand extrusion from the central channel of CMG and ensure that CMG is surrounding dsDNA[10]. After CMG assembly, we incubated the bead-bound DNA in a buffer solution with or without ATPγS, and monitored the amount of fluorescent Cdc45$^{LD555}$ present on the DNA

over time. As seen in **Fig. 2.4d and e**, we detected the fastest decay of Cdc45$^{LD555}$ signal in the DNA construct with a free end in the absence of ATPγS, whereas there was only a small decay in the same DNA construct when ATPγS was present. We also observed some decay, albeit smaller in magnitude, in the DNA construct with

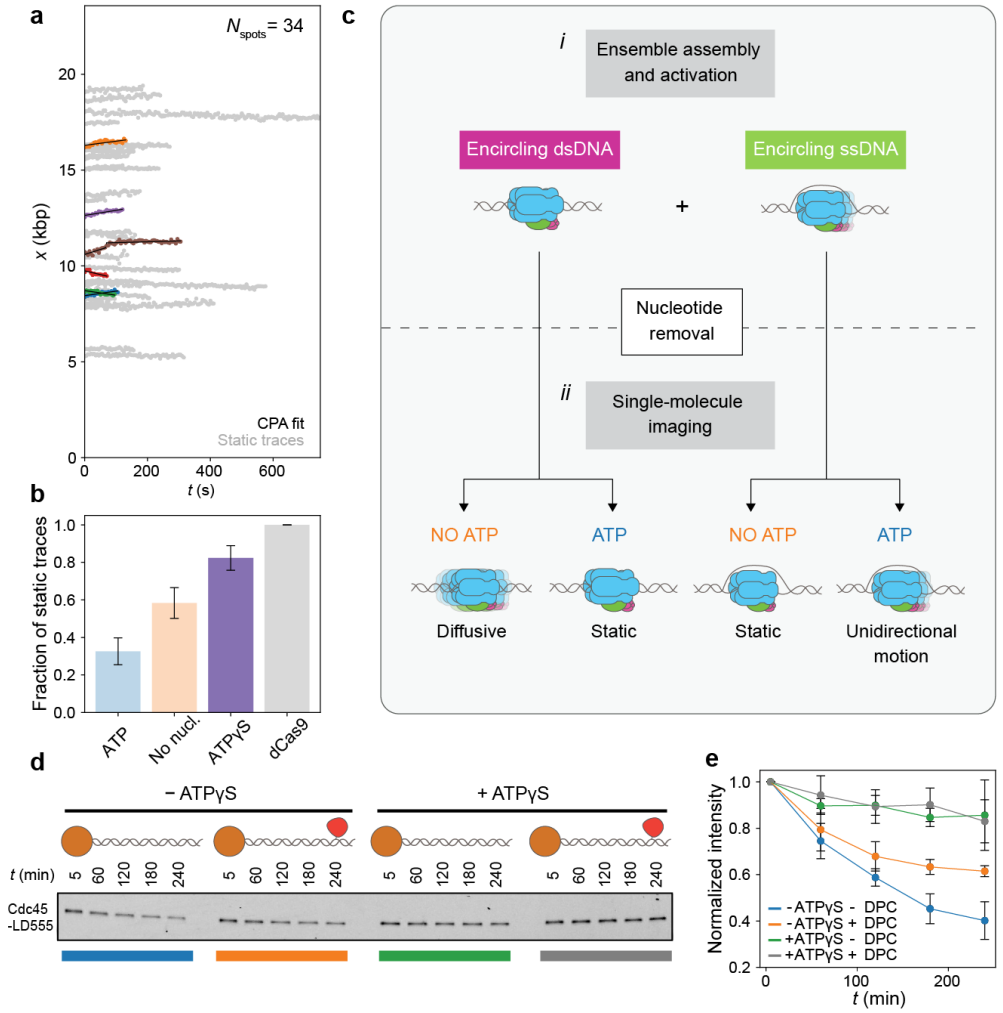



*Figure 2.4* Nucleotide binding halts CMG diffusion. **a** Position vs. time plots of CMG spots in the presence of ATPγS; CPA fits are plotted in black, static traces are shown in light gray and mobile traces are shown in all other colors. **b**, Fraction of static CMG spots in the presence of ATPγS ($N_{spots}$=34); the results from **Fig. 2.2d**, are shown as light bars for comparison; error bars show the standard error of proportion. **c** Model proposed to explain experimental motion results; **c-i** proposed two populations of CMG present in the ensemble CMG activation reaction; **c-ii** summary of experimental outcomes in **Fig. 2.2** and **Fig. 2.4a**, and proposed explanation of their origins. **d** Fluorescent scan of an SDS-PAGE gel showing the amount of Cdc45$^{LD555}$ left on linear DNA bound to magnetic beads at one end and containing either a free end or an end capped with a covalently crosslinked methyltransferase. **e** Densitometry quantification of the experiment shown in **d** showing the average normalized intensity of three replicates together with their standard deviation. Data points are connected by solid lines to guide the eye.

a capped end in the absence of ATPγS (**Fig. 2.4d, e**); this smaller decay may result from spontaneous opening of the Mcm2-7 ring of CMG in the absence of nucleotide[16,30] on such timescales. Thus, the results of our ensemble assay support our hypothesis that CMG diffuses on dsDNA in the absence of ATP, and that nucleotide binding halts this diffusive motion.

To test whether the nucleotide-binding-mediated halting of CMG diffusion was due to DNA melting within the central channel of CMG, we employed a recently reported Mcm2 mutant in which six residues directly involved in DNA melting are substituted with alanine (hereafter referred to as Mcm2$^{6A}$)[11]. Notably, Mcm2$^{6A}$ supports CMG assembly but does not support either DNA melting or strand extrusion[11], allowing us to separate the effect of DNA melting from DNA binding. Thus, if DNA melting by CMG is indeed what halts its diffusive motion, then CMG assembled with Mcm2$^{6A}$ should be fully diffusive in the presence of ATP. Nevertheless, the vast majority of CMG spots assembled with Mcm2$^{6A}$ are static in the presence of ATP (**Fig. S2.9a, f**). Furthermore, when we performed our ensemble CMG sliding assay with Mcm2$^{6A}$, we observed the same trends as with WT Mcm2 (**Fig. S2.9g, h**), providing further evidence that CMG diffuses on dsDNA and that DNA melting is not necessary for the halting of CMG diffusion after ATP-binding. Altogether, our single-molecule and ensemble biochemical data show that the halting of CMG diffusion in the presence of nucleotide is not due to DNA melting by CMG, but due to binding of CMG to the DNA *via* other Mcm2-7:DNA interactions[11].

## 2.4 Discussion

We report the single-molecule motion quantification of fully reconstituted CMG assembled at an origin of replication. To enable this, we developed a hybrid ensemble and single-molecule assay based on the double functionalization of DNA ends, which we validated by verifying that the number of fluorescent CMG helicases per diffraction-limited spot, the initial CMG positions on the DNA, and the DDK-dependent colocalization of fluorescently labeled Cdc45 and Mcm2-7 agree with the consensus established by previous biochemical observations (**Fig. 2.1**).

In our motion analysis, we observe a static and a mobile population of CMG both in the presence and in the absence of ATP (**Figs. 2.2** and **S2.5**). Nonetheless, we quantitatively show that the mobile population in the presence of ATP preferentially moves unidirectionally, as expected for a molecular motor, whereas the mobile population in the absence of ATP preferentially moves in a diffusive manner (**Fig. 2.2**). To explain these motion outcomes, we propose the model

summarized in **Fig. 2.4c**, which is based on the assumption that our ensemble activation reaction contains a mixture of two populations of CMG: one population encircling dsDNA, and another population encircling ssDNA (**Fig. 2.4c-*i***). In the first part of our model, we propose that the population of CMG encircling ssDNA moves unidirectionally in the presence of ATP (as the motor can then hydrolyze ATP to unwind DNA), but remains static in the absence of ATP (as the motor is unable to unwind DNA without ATP hydrolysis[9,10,16]) (**Fig. 2.4c, right half**). This part of the model is supported by the fact that we observe very similar instantaneous and average velocities of unidirectionally translocating CMG spots to those of previous single-molecule and ensemble biochemical studies of DNA unwinding[14,17,18]. We consider an alternative scenario in which at least part of the unidirectional translocation that we observe in the presence of ATP corresponds to CMG translocating on dsDNA (**Fig. S2.10**) – as was postulated to occur for CMG helicases that bypassed each other after the collision of two replication forks or after encountering a flush ss/dsDNA junction[31–33]– to be less likely. Consider, for example, the CMG splitting events that we observe (**Fig. 2.3c**). Because CMG translocation on dsDNA has the same 3'-to-5' polarity as on ssDNA[33], one could postulate that the splitting events represent two helicases that surround dsDNA following assembly at a single Mcm2-7 double hexamer, but such helicases would move *towards* each other (remaining within a single diffraction-limited spot), and not *away* from each other as our data shows. One could also postulate that the splitting events represent two helicases encircling dsDNA following assembly at distinct Mcm2-7 double hexamers within a single diffraction-limited spot. However, our analysis of Mcm2-7 spots shows such occurrences to be unlikely (**Fig. S2.2c**). Further studies will be required to assess whether and how the unidirectional motion of CMG differs according to whether it occurs on ssDNA or dsDNA.

In the second part of our model, we propose that the population of CMG encircling dsDNA loses its bound ATP when we remove ATP from the buffer following our ensemble activation (**Fig. 2.1a**); ATP dissociation then causes CMG to disengage from the DNA inside its central channel and thereby become diffusive. This diffusive population remains so in the absence of nucleotide (giving rise to the diffusive population we observe in the absence of nucleotide); nevertheless, upon ATP re-addition and re-binding, CMG is allowed to re-engage with the DNA inside its central channel and thus become static (giving rise to the static population in the presence of ATP). In support of this hypothesis, we showed that CMG can diffuse on dsDNA and that nucleotide binding halts this diffusion (**Fig. 2.4a, b**). Furthermore, we show that this halting occurs independent of the previously

observed DNA melting by CMG upon nucleotide binding[10,11] (**Fig. S2.9**), suggesting that DNA engagement and DNA melting by CMG need not occur concomitantly. We propose that the presence of ATP in the nucleus prevents newly assembled CMG from diffusing along the DNA, poising the helicase to catalyse the initial melting required to initiate replication. Further studies investigating which of the additional Mcm2-7:DNA contacts within CMG[11] are responsible for halting CMG diffusion upon nucleotide binding will shed further light into our observations.

Our measured diffusion coefficient of freely diffusive CMG under the ionic strength conditions of this study (250 mM K-glutamate) (**Fig. S2.5l**) is similar to our previously measured diffusion coefficient of single Mcm2-7 hexamers in higher ionic strength conditions (500 mM NaCl)[2]. This observation suggests that CMG diffuses more freely on the DNA than single Mcm2-7 hexamers, which in turn suggests that CMG has fewer contacts with the DNA than Mcm2-7 hexamers as recently confirmed by structural studies[11,34]. We also note that single-molecule studies with pre-formed *D. melanogaster* CMG showed no evidence of CMG diffusion in the presence of ATP[17,18], in agreement with our observations. However, single-molecule studies with pre-formed *S. cerevisiae* CMG reported extensive diffusive behavior in the presence of ATP[16]. Further studies will be required to investigate the reasons for this discrepancy, and to probe potential differences between fully reconstituted and pre-formed CMG, such as whether the phosphorylation state of Mcm2-7 is different, and whether it could influence the interactions that Mcm2-7, the substrate of DDK[1,6], have with the DNA.

ATP-dependent switching of motion modes, as we observe here for CMG, has been previously observed to act differently for other protein complexes, such as Type III restriction enzymes[35,36] and the yeast chromatin remodeler SWR1[37]. In the case of Type III restriction enzymes, ATP hydrolysis triggers diffusion on the DNA[35,36], whereas in the case of SWR1, ATP binding triggers its diffusion along the DNA[37]. Nonetheless, these behaviors are different than in the case of CMG, whose diffusion is halted and not promoted by nucleotide binding.

Given the complexity and the number of components required to fully reconstitute CMG assembly and activation, *in vitro* single-molecule studies quantifying CMG motion have so far relied on pre-formed CMG assembled through an unknown mechanism upon overexpression of individual constituents[19,20]. Although these studies have provided us with very important insights into how CMG works, it is unknown whether the assembly mechanism of pre-formed CMG is cell-cycle regulated. It is therefore also unknown whether pre-formed CMG has a

similar phosphorylation state to the one it has *in vivo*. In addition, pre-formed CMG requires an artificial region of ssDNA to bind DNA and thus does not allow us to access the intricacies of CMG assembly and activation. In this study, we quantify the motion of fully reconstituted CMG at the single-molecule level. We believe that the full reconstitution of CMG makes it more likely that the phosphorylation state of its constituents more closely mimics what happens *in vivo*; this in turn makes it more likely that fully reconstituted CMG can support all relevant interactions with proteins intrinsic and accessory to the replisome. Furthermore, the reliable assay we developed will allow us and others to address important questions regarding CMG motion, such as the mechanistic role of elongation factors, with unprecedented resolution. Finally, we anticipate that the hybrid assay we developed, based on the double functionalization of DNA ends with two orthogonal attachment types, will facilitate the study at the single-molecule level of similarly complex biochemical reactions involved in other nucleic acid:protein interactions.

## 2.5 Methods

### 2.5.1 Biological Materials and Preparation

#### 2.5.1.1 Molecular cloning and strain generation

**Cdc45-iS6-i2XFLAG:** to generate fluorescently labelled Cdc45, we introduced an 'S6' peptide[38] with an additional glycine linker (GDSLSWLLRLLNG) after residue E197 and immediately before the internal 2XFLAG tag of Cdc45. This was achieved by modifying plasmid pJY13 with a Q5® Site-Directed Mutagenesis Kit (NEB # E0554S) and mutagenic primers DRM_005 and DRM_006. The resulting plasmid pDRM19-01 was transformed into chemically competent One Shot™ TOP10 E. coli (Invitrogen # C404010). The sequence of Cdc45-iS6-i2XFLAG was verified by sequencing the entire ORF. S. cerevisiae strain yDRM2 overexpressing Cdc45-iS6-i2XFLAG was then generated by linearizing plasmid pDRM19-01 with NheI-HF® (NEB # R3131S) and transforming it into yeast strain yJF1 (MAT**a**, W303 background).

**6XHis-S6-dCas9-Halo:** to generate fluorescently labelled dCas9, we introduced an 'S6' peptide with an additional GSS linker (GDSLSWLLRLLNGGS) next to the N-terminal 6XHis tag. This was achieved by modifying plasmid pET302-6His-dCas9-halo with Q5® Site-Directed Mutagenesis Kit (NEB # E0554S) and mutagenic primers DRM_184 and DRM_185. The resulting plasmid, pDRM21-01 was transformed into NEB® 5-alpha Competent E. coli (High Efficiency) cells (NEB #

C2987I). The sequence of pDRM21-01 was then verified by sequencing the entire ORF. Plasmid pDRM21_01 was then transformed into E. coli BL21-Codon Plus (DE3)-RIL Competent Cells (Agilent # 230245).

**pGL50-ARS1:** the ARS1 sequence was amplified by PCR using plasmid p5.8kb-ARS1[7] as a template together with primers TL-033 and TL-034, which contained AscI restriction sites at both ends. This PCR product was then cloned into the dephosphorylated MluI site of the 21.2 kb plasmid pSupercos1-lambda 1,2[39]. The whole sequence of the ARS1 insert was confirmed by sequencing.

### 2.5.1.2 Protein purification

Cdc6, Mcm10, dCas9-Halo, S6-dCas9-Halo and Sfp phosphopantetheinyl transferase were expressed in *E. coli* BL21-Codon Plus (DE3)-RIL Cells (Agilent # 230245). GINS was expressed in *E. coli* Bl21 (DE3) Rosetta pLysS Cells (Novagen). Unless otherwise specified, cells were grown to OD600 = 0.40-0.60, induced with 400 µM IPTG for 16 h at 17 °C, and harvested by centrifugation. Pellets were resuspended in 40 mL of lysis buffer and sonicated on ice in a Qsonica Q500 sonicator for 2 min in 5 s on / 5 s off cycles and an amplitude of 40 %. Following sonification, the lysate was centrifuged at 8820 g for 20 min in a Beckman-Coulter Avanti JXN 26 centrifuge with rotor JA17, and the clarified supernatant was used for affinity pulldowns.

ORC, Mcm2-7/Cdt1, Pol ε, Dpb11, Sld2, Sld3/7, Cdc45, Cdc45-iS6, RPA, DDK, and S-CDK were expressed in *S. cerevisiae* strain yJF1 *MAT***a** Δpep1 Δbar1 from a bi-directional galactose-inducible promotor. Prior to expression, cells were inoculated at a density of 2 x $10^5$ cells/ml in YP medium supplemented with 2 % raffinose (Carbosynth # OR06197) and 50 mg/ml ampicillin (Merck-Sigma # A9518), and grown overnight at 30 °C to a density of 3-5 x $10^7$ cells/ml. For S-CDK, 5 µg/ml Nocodazole (Merck-Sigma # M1404), and 2 % galactose (Carbosynth # MG05201) were added to the medium and the cells were induced for 3 h before harvesting. For DDK, 2 % galactose (Carbosynth # MG05201) was added to the medium and the cells were induced for 3 h before harvesting. For all other proteins, cells were arrested in G1 phase for 3 h before induction with 100 ng/ml of α mating factor (Tebu-bio # 089AS-60221); cells were then induced with 2 % galactose (Carbosynth # MG05201) for 3 h before harvesting.

Cells were harvested by centrifugation and washed with lysis buffer. After centrifugation, cells were suspended in lysis buffer supplemented with protease

inhibitors (cOmpleteTM EDTA-free Protease Inhibitors (Merck-Sigma # 5056489001) and 0.3 mM phenylmethylsulfonyl fluoride (PMSF)), and dropwise dropped into liquid nitrogen. The frozen droplets were grounded in a 6875 SPEX freezer mill for six cycles (run time 2 min and cool time 1 min, with a rate of 15 cps). Following purification, all protein concentrations were determined with Bio-Rad Protein Assay Dye Reagent (Bio-Rad #5000006).

Purifications of ORC, Cdc6, Mcm2-7/Cdt1, Mcm2-7$^{Halo\text{-}Mcm3}$/Cdt1, Mcm2-7$^{Mcm2(6A)}$/Cdt1, and dCas9-Halo have been described previously[2,11].

**DDK:** DDK with a CBP-TEV tag on Dbf4 was purified from *S. cerevisiae* strain ySDK8[40]. Powder was suspended in DDK lysis buffer (25 mM HEPES-KOH pH 7.6. 0.05 % NP40 substitute, 10 % glycerol, 400 mM NaCl, and 1 mM DTT) supplemented with protease inhibitors. Lysate was cleared in a Beckman-Coulter ultracentrifuge (type Optima L90K with rotorTI45) for 1 h at 235,000 g and 4 °C. The cleared lysate was supplemented with $CaCl_2$ to a final concentration of 2 mM and incubated for 1 h at 4 °C with washed Sepharose 4B Calmodulin beads (GE Healthcare # 17-0529-01) in a spinning rotor. The beads were washed 10 times with 5 ml of DDK-binding buffer (25 mM HEPES-KOH pH 7.6, 0.05% NP40 substitute, 10% glycerol, 400 mM NaCl, 2 mM $CaCl_2$, and 1 mM DTT), and protein was eluted from the beads with DDK elution buffer (25 mM HEPES-KOH pH 7.6, 0.05 % NP40 substitute, 10 % glycerol, 400 mM NaCl, 2 mM EDTA, 2mM EGTA, and 1 mM DTT). DDK-containing fractions were pooled and dephosphorylated with 20000 units lambda phosphatase (NEB # P0753S) for 16 h at 4$^{o}$C. The sample was then concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024), and injected into a Superdex 200 Increase 10/300 GL column (Cytiva # 15182085) equilibrated in DDK GF buffer (25 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, and 200 mM K glutamate). Positive fractions were pooled and concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024). Aliquots were snap frozen and stored at −80 °C.

**GINS:** GINS with a N-terminal 6xHis**-**tag on Psf3 was purified from *E. coli Rosetta*. Cleared lysate was incubated with Ni-NTA agarose (Qiagen # 30210) for 1 h at 4 $^{o}$C in GINS lysis buffer (25 mM Tris-HCl pH 7.2, 0.02 % NP40 substitute, 10 % glycerol, 400 mM NaCl, 10 mM imidazole, and 1 mM DTT) supplemented with protease inhibitors. The beads were washed 5 times with 5 ml of GINS lysis buffer and 5 times with 5 ml of GINS wash buffer (25 mM Tris-HCl pH 7.2, 0.02 % NP40 substitute, 10 % glycerol, 100 mM NaCl, 15 mM imidazole and 1 mM DTT). Protein

was then eluted from the beads with GINS elution buffer (25 mM Tris-HCl pH 7.2, 0.02 % NP40 substitute, 10 % glycerol, 100 mM NaCl, 200 mM imidazole, and 1 mM DTT). GINS-containing fractions were pooled and dialyzed against GINS dialysis buffer I (25 mM Tris-HCl pH 7.2, 0.02 % NP40 substitute, 10 % glycerol, 100 mM NaCl, and 1 mM DTT). The sample was then flowed through a 0.2 micron filter and injected into a MonoQ 5/50 GL column (GE Healthcare # 17-5166-01) equilibrated in GINS MonoQ buffer A (25 mM Tris-HCl pH 7.2, 0.02 % NP40 substitute, 10 % glycerol, 100 mM NaCl, and 1 mM DTT). GINS was eluted from the column in a 30 CV NaCl gradient from 100 mM to 500 mM. Positive fractions were pooled, concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024), and injected into a Superdex 200 Increase 10/300 GL column (Cytiva # 15182085) equilibrated in GINS GF buffer (25 mM Tris-HCl pH 7.2, 0.02 % NP40 substitute, 10 % glycerol, 150 mM NaCl, and 1 mM DTT). Positive fractions were pooled and concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024). Aliquots were snap frozen and stored at −80°C.

**Pol ε:** Pol ε with a C-terminal CBP-TEV tag on Dpb4 was purified from *S. cerevisiae* strain yAJ2[7]. Powder was suspended in Pol ε lysis buffer (25 mM HEPES-KOH pH 7.6, 10 % glycerol, 400 mM KOAc, and 1 mM DTT) supplemented with protease inhibitors. Lysate was cleared in a Beckman-Coulter ultracentrifuge (type Optima L90K with rotor TI45) for 1 h at 235,000 g and 4 °C. The cleared lysate was supplemented with $CaCl_2$ to a final concentration of 2 mM and incubated for 1 h at 4°C with washed Sepharose 4B Calmodulin beads (GE Healthcare # 17-0529-01) in a spinning rotor. The beads were washed 10 times with 5 ml of Pol ε binding buffer (25 mM HEPES-KOH pH 7.6, 10 % glycerol, 400 mM KOAc, 2 mM $CaCl_2$, and 1 mM DTT) and protein was eluted from the beads with Pol ε elution buffer (25 mM HEPES-KOH pH 7.6, 10 % glycerol, 400 mM KOAc, 2 mM EDTA, 2 mM EGTA, and 1 mM DTT). Positive fractions were pooled, passed through a 0.2 micron filter and injected into a 1-ml heparin column (GE Healthcare # 17-0406-01) equilibrated in Pol ε heparin buffer A1 (25 mM HEPES-KOH pH 7.6, 10 % glycerol, 400 mM KOAc, and 1 mM DTT). The column was washed with Pol ε heparin buffer A2 (25 mM HEPES-KOH pH 7.6, 10 % glycerol, 450 mM KOAc, and 1 mM DTT) and Pol ε was eluted from the column in a 30 CV KOAc gradient from 450 mM to 1000 mM. Peak fractions were pooled, concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024), and injected into a Superdex 200 Increase 10/300 GL column (Cytiva # 15182085) equilibrated in Pol ε GF buffer (25 mM HEPES-KOH pH 7.6, 10 % glycerol, 500 mM KOAc, and 1 mM DTT). Peak fractions were pooled and concentrated in an Amicon Ultra-4 Ultracell 30 kDa

centrifugal filter (Merck-Millipore # UFC803024). Aliquots were snap frozen and stored at −80°C.

**S-CDK**: S-CDK with an N-terminal CBP-TEV tag on Clb5 Δ1-100 was purified from *S. cerevisiae* (Lucy Drury, Francis Crick Institute, unpublished). Powder was suspended in S-CDK lysis buffer (40 mM HEPES-KOH pH 7.6, 0.01 % NP40 substitute, 10 % glycerol, and 300 mM KOAc) supplemented with protease inhibitors. The lysate was cleared in a Beckman-Coulter ultracentrifuge (type Optima L90K with rotorTI45) for 1 h at 235,000 g and 4°C. The cleared lysate was supplemented with $CaCl_2$ to a final concentration of 2 mM and incubated for 1 h at 4 °C with washed Sepharose 4B Calmodulin beads (Agilent # 214303-52) in a spinning rotor. The beads were washed 5 times with 5 ml of S-CDK binding buffer (40 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, 300 mM KOAc, and 2 mM $CaCl_2$), and 5 times with 5 ml of S-CDK-TEV buffer (40 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, and 2 mM $CaCl_2$). S-CDK was then cleaved from the beads by incubating overnight with TEV protease at 4 °C in S-CDK TEV buffer. Released S-CDK was then concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024), and injected into a Superdex 200 Increase 10/300 GL column (Cytiva # 15182085) equilibrated in S-CDK GF buffer (40 mM HEPES-KOH pH 7.6, 0.01 % NP40 substitute, 10 % glycerol, and 300 mM KOAc). Peak fractions were pooled and concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024). Aliquots were snap frozen and stored at −80 °C.

**Dpb11**: Dpb11 with a C-terminal 3xFlag-tag was purified from S. *cerevisiae* strain yJY26[7]. Powder was suspended in Dpb11 lysis buffer (25 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, 500 mM KCl, 1 mM EDTA, and 1 mM DTT) supplemented with protease inhibitors. The lysate was cleared in a Beckman-Coulter ultracentrifuge (type Optima L90K with rotor TI45) for 1 h at 235,000 g and 4 °C. The cleared lysate was incubated for 1 h at 4 °C with washed M2 anti-flag affinity beads (Merck-Sigma # A2220; 1-2 mL of beads per 40 mL of lysate) in a spinning rotor. The beads were washed 10 times with 5 ml Dpb11 lysis buffer and Dpb11 was eluted from the beads by incubation with 3xFLAG peptides (Merck-Sigma # F4799). The eluate was dialysed against Dpb11 dialysis buffer I (25 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, 150 mM KCl, 1 mM EDTA, and 1 mM DTT), passed through a 0.2-micron filter and injected into a 1-ml MonoS 5/50 GL column (GE Healthcare # 17-5168-01) equilibrated in Dpb11 MonoS buffer A (25 mM HEPES-KOH pH 7.6. 0.02 % NP40 substitute, 10 % glycerol, 150 mM KCl, 1 mM EDTA, and 1 mM DTT). The column was washed with Dpb11

MonoS buffer A, and Dpb11 was eluted from the column in a 20 CV KCl gradient from 150 mM to 1000 mM. Peak fractions were pooled, concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024), and injected into a Superdex 200 Increase 10/300 GL column (Cytiva # 15182085) equilibrated in Dpb11 GF buffer (25 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, 300 mM KOAc, 1 mM EDTA, and 1 mM DTT). Peak fractions were pooled and concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024). Aliquots were snap frozen and stored at −80 °C.

**Sld2**: Sld2 with a C-terminal 3xFLAG-tag was purified from *S. cerevisiae* strain yTD8[7]. Powder was suspended in Sld2 lysis buffer (25 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, 500 mM KCl, 1 mM EDTA, and 1 mM DTT) supplemented with 0.3 mM PMSF, 7.5 mM benzamidine, 0.5 mM AEBSF, 1 mM leupeptin, 1 mM pepstatin A, and 1 µg/ml aprotinin. The lysate was cleared in a Beckman-Coulter ultracentrifuge (type Optima L90K with rotorTI45) for 1 h at 235,000 g and 4 °C. Solid ammonium sulphate was added to the cleared lysate up to a saturation of 32 %. After 15 min of tumbling at 4 °C, the lysate was cleared by centrifugation at 27000 g for 20 min. Then, solid ammonium sulphate was added to the supernatant up to a saturation of 48 %. After 15 min of tumbling at 4 °C, the lysate was cleared by centrifugation at 27000 g for 20 min. The pellet was dissolved in Sld2 lysis buffer supplemented with 0.3 mM PMSF, 7.5 mM benzamidine, 0.5 mM AEBSF, 1 mM leupeptin, 1 mM pepstatin A, and 1 µg/ml aprotinin, and incubated for 30 min at 4 °C with washed M2 anti-FLAG beads (Merck-Sigma # A2220; 1-2 mL of beads per 40 mL of lysate) in a spinning rotor. The beads were washed 8 times with 5 ml of Sld2 lysis buffer, tumbled for 10 min at 4 °C with 10 ml of FLAG resuspension buffer (25 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, 500 mM KCl, 1 mM ATP, 10 mM MgOAc, and 1 mM DTT) supplemented with 0.3 mM PMSF, 7.5 mM benzamidine, 0.5 mM AEBSF, 1 mM leupeptin, 1 mM pepstatin A, and 1 µg/ml aprotinin, and 8 more times with 5 ml of Sld2 lysis buffer. Sld2 was eluted from the beads by incubation with 3xFLAG peptides (Merck-Sigma # F4799). The eluate was dialyzed for 45 min at 4 °C against Sld2 dialysis buffer I (25 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, 280 mM KCl, 1 mM EDTA, and 1 mM DTT), passed through a 0.2 micron filter, and injected into a 1-ml HiTrap SPFF column (GE Healthcare # 17-5054-01) equilibrated in Sld2 SPFF buffer A (25 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, 250 mM KCl, 1 mM EDTA, and 1 mM DTT). The column was washed with Sld2 SPFF buffer A, and Sld2 was eluted from the column in a 20-CV KCl gradient from 280 mM to 1000 mM. Peak fractions were pooled, dialysed for 45 min at 4 °C against

Sld2 dialysis buffer II (25 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 40 % glycerol, 350 mM KCl, 1 mM EDTA, and 1 mM DTT). Aliquots were snap frozen and stored at −80 °C.

**Sld3/7**: Sld3/7 with a C-terminal CTP tag on Sld3 was purified from *S. cerevisiae* strain yTD6[7]. Powder was suspended in Sld3/7 lysis buffer (25 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, 500 mM KCl, 1 mM EDTA, and 1 mM DTT) supplemented with protease inhibitors. The lysate was cleared in a Beckman-Coulter ultracentrifuge (type Optima L90K with rotor TI45) for 1 h at 235,000 g and 4 °C, and incubated for 40 min at 4 °C with washed IgG Sepharose 6 Fast Flow (GE Healthcare cat # 17-0969-01) in a spinning rotor. The beads were washed with 15CV Sld3/7 wash buffer (25 mM HEPES-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, 500 mM KCl, 0,5 mM EDTA, and 1 mM DTT), and the protein complex was cleaved from the beads by overnight treatment at 4 °C with TEV protease in Sld3/7 lysis buffer. Sld3/7 was eluted from the column, concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024), and injected into a Superdex 200 Increase 10/300 GL column (Cytiva # 15182085) equilibrated in Sld3/7 GF buffer (25 mM hepes-KOH pH 7.6, 0.02 % NP40 substitute, 10 % glycerol, 500 mM KCl, 1 mM EDTA, and 1 mM DTT). Peak fractions were pooled and concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024). Aliquots were snap frozen and stored at −80 °C.

**Cdc45 and Cdc45-iS6**: Cdc45 with an internal 2xFLAG tag and Cdc45 with an internal 2xFLAG + S6 tag were purified from S. *cerevisiae* strains yJY13[7] and yDRM2 (this study), respectively. Powder was suspended in Cdc45 lysis buffer (25 mM HEPES-KOH pH 7.6, 10 % glycerol, 500 mM KOAc, 1 mM EDTA, and 1 mM DTT) supplemented with protease inhibitors. The lysate was cleared in a Beckman-Coulter ultracentrifuge (type Optima L90K with rotor TI45) for 1 h at 235,000 g and 4 °C. The cleared lysate was incubated for 1 h with washed M2 anti-flag affinity beads (Merck-Sigma # A2220; 1-2 mL of beads per 40 mL of lysate) at 4 °C in a spinning rotor. The beads were washed 10 times with 5 ml of Cdc45 lysis buffer, and Cdc45 was eluted from the beads by incubation with 3xFLAG peptides (Merck-Sigma # F4799). The eluate was dialyzed against Cdc45 dialysis buffer I (20 mM K phosphate pH 7.6, 10 % glycerol, 150 mM KOAc, and 0.5 mM DTT) and injected into a 2-ml hydroxyapatite Bio gel HTP column (Biorad # 130-0420) equilibrated in Cdc45 HTP equilibration buffer (20 mM K phosphate pH 7.6, 10 % glycerol, 150 mM KOAc, and 0.5 mM DTT). Cdc45 bound to the hydroxyapatite Bio gel for 45 min at 4 °C with tumbling. The column was then washed with Cdc45 wash buffer A (80

mM K phosphate pH 7.6, 10 % glycerol, 150 mM KOAc, and 0.5 mM DTT), and Cdc45 was eluted with Cdc45 HTP elution buffer (250 mM K phosphate pH 7.6, 10 % glycerol, 150 mM KOAc and 0.5 mM DTT). Positive fractions were pooled and dialyzed against Cdc45 dialysis buffer I (25 mM HEPES-KOH pH 7.6, 10 % glycerol, 300 mM KOAc, 1 mM EDTA, and 1 mM DTT). Finally, Cdc45 was concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024). Aliquots were snap frozen and stored at −80 °C.

**RPA:** RPA with a CBP-TEV tag on Rfa1 was purified from *S. cerevisiae* strain yAE31[7]. Powder was suspended in RPA lysis buffer (25 mM Tris-HCl pH 7.2, 10 % glycerol, 500 mM NaCl, and 1 mM DTT) supplemented with protease inhibitors. The lysate was cleared in a Beckman-Coulter ultracentrifuge (type Optima L90K with rotorTI45) for 1 h at 235,000 g and 4 °C. The cleared lysate was diluted with an equal volume of RPA dilution buffer I (25 mM Tris-HCl pH 7.2, 10% glycerol, and 1 mM DTT), supplemented with $CaCl_2$ to a final concentration of 2 mM, and incubated for 1 h at 4 °C with washed Sepharose 4B Calmodulin beads (GE Healthcare # 17-0529-01) in a spinning rotor. The beads were washed 10 times with 5 ml of RPA binding buffer (25 mM Tris-HCl pH 7.2, 10 % glycerol, 200 mM NaCl, 2 mM $CaCl_2$, and 1 mM DTT), and the protein complex was eluted from the beads with RPA elution buffer (25 mM Tris-HCl pH 7.2, 10 % glycerol, 200 mM NaCl, 2 mM EDTA, 2mM EGTA, and 1 mM DTT). Positive fractions were pooled and diluted with an equal volume of RPA dilution buffer II (25 mM Tris-HCl pH 7.2, 10 % glycerol, 1 mM EDTA, and 1 mM DTT) and twice dialyzed for 1 h against RPA dialysis buffer (25 mM Tris-HCl pH 7.2, 10 % glycerol, 50 mM NaCl, 1 mM EDTA, and 1 mM DTT). The sample was then loaded onto a 1-ml Hi Trap heparin HP column (GE Healthcare # 17-0406-01) equilibrated with buffer RPA heparin A (25 mM Tris-HCl pH 7.2, 10 % glycerol, 50 mM NaCl, 1 mM EDTA, and 1 mM DTT). The column was washed with RPA heparin buffer A, and RPA was eluted from the column in a 30 CV NaCl gradient from 50 mM to 1000 mM. Subsequently, the sample was concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024), and injected into a Superdex 200 Increase 10/300 GL column (Cytiva # 15182085) equilibrated in RPA GF buffer (25 mM Tris-HCl pH 7.2, 10 % glycerol, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT). Peak fractions were pooled and concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024). Aliquots were snap frozen and stored at −80 °C.

**Mcm10:** Mcm10 with a N-terminal 6xHis-tag and a C-terminal 3xFLAG tag (Max Douglas, Francis Crick Institute, unpublished) was purified from *E. coli* BL21-Codon Plus (DE3)-RIL. Cleared lysate was incubated for 1 h at 4 °C with washed M2 anti-

flag affinity beads (Merck-Sigma # A2220; 1-2 mL of beads per 40 mL of lysate) in a spinning rotor. The beads were washed 10 times with 5 ml of Mcm10 lysis buffer (25 mM Tris-HCl pH 7.2, 10 % glycerol, 500 mM NaCl, and 0,01 % NP40 substitute) and 5 times with 5 ml of Mcm10 lysis buffer with 300 mM NaCl (25 mM Tris-HCl pH 7.2, 10 % glycerol, 300 mM NaCl, and 0,01 % NP40 substitute). Mcm10 was eluted from the beads by incubation with 3xFLAG peptides (Merck-Sigma # F4799) and incubated with Ni-NTA agarose (Qiagen # 30210) for 1 h at 4 °C. The beads were washed 5 times with 5 ml of Mcm10 wash buffer II (25 mM Tris-HCl pH 7.2, 10 % glycerol, 500 mM NaCl, and 0,01 % NP40 substitute) and 5 times with 5 ml of Mcm10 wash buffer III (25 mM Tris-HCl pH 7.2, 10 % glycerol, 500 mM NaCl, 0,01 % NP40 substitute, and 20 mM Imidazole). Then the protein complex was eluted from the beads with RPA elution buffer (25 mM Tris-HCl pH 7.2, 10 % glycerol, 500 mM NaCl, 0,01 % NP40 substitute, and 200 mM Imidazole). Positive fractions were pooled and dialyzed against Mcm10 dialysis buffer (25 mM HEPES-KOH pH 7.6, 10 % glycerol, 200 mM KOAc, 1 mM EDTA, and 0.01 % NP40 substitute). After dialysis, Mcm10 was concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024). Aliquots were snap frozen and stored at −80 °C.

**S6-dCas9-Halo**: S6-dCas9-Halo with an N-terminal 6xHis-tag (this study) was purified from *E. coli* BL21-Codon Plus (DE3)-RIL. Cleared lysate was incubated with Ni-NTA agarose (Qiagen # 30210) for 1 h at 4 °C in Cas9 lysis buffer (50 mM Na-phosphate pH 7.0, and 300 mM NaCl) supplemented with protease inhibitors. Beads were washed 5 times with 5 ml of dCas9 wash buffer I (50 mM Na phosphate pH 7.0, 300 mM NaCl, and 20 mM imidazole). Then the protein complex was eluted from the beads with dCas9 elution buffer I (50 mM Na phosphate pH 7.0, 300 mM NaCl, and 150 mM imidazole). Positive fractions were pooled and dialyzed against dCas9 dialysis buffer I (50 mM HEPES-KOH pH 7.6, 100 mM KCl, and 1 mM DTT). Then the sample was passed through a 0.2 micron filter and injected into a 1-ml HiTrap SP HP column (GE Healthcare # 17-1151-01) equilibrated in dCas9 SPHP buffer A (50 mM HEPES-KOH pH 7.6, 100 mM KCl, and 1 mM DTT). S6-dCas9 was eluted from the column in a 30 CV KCl gradient from 100 mM to 1000 mM. Positive fractions were concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck Millipore # UFC803024), and injected into a Superose 6 Increase 10/300 GL column (GE Healthcare # 29-0915-96) equilibrated in dCas9 GF buffer (50 mM HEPES-KOH pH 7.6, 150 mM KCl, 20 % glycerol, and 1 mM DTT). Peak fractions were pooled and concentrated in an Amicon Ultra-4 Ultracell 30 kDa centrifugal filter (Merck-Millipore # UFC803024). Aliquots were snap frozen and stored at −80 °C.

**Sfp phosphopantetheinyl transferase:** Sfp phosphopantetheinyl transferase with a C-terminal 6xHis-tag (Addgene # 75015) was purified from *E. coli* BL21-Codon Plus (DE3)-RIL. Cleared lysate was incubated with Ni-NTA agarose (Qiagen # 30210) for 30 min at 4 °C in Sfp lysis buffer (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, and 10 mM imidazole) supplemented with protease inhibitors The beads were washed with 100 mL of Sfp wash buffer (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, and 30 mM imidazole), and protein was eluted with Sfp elution buffer (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, and 250 mM imidazole). Positive fractions were pooled and dialyzed overnight against Sfp dialysis buffer I (50 mM HEPES-KOH, pH 7.6, 100 mM KCl, and 50 % glycerol). Dialyzed Sfp transferase was then concentrated 12-fold in a 3 kDa Amicon® Ultra-15 Centrifugal Filter Units (Millipore # UFC9003). Aliquots were snap frozen and stored at −80 °C.

### 2.5.1.3 Protein labelling

**dCas9$^{LD555}$:** S6-dCas9-Halo was fluorescently labelled by incubating S6-dCas9 with Sfp phosphopantetheinyl transferase and LD555-CoA (Lumidyne Technologies, custom synthesis) in a 1:2:10 molar ratio in dCas9 gel filtration buffer (50 mM HEPES/KOH, pH 7.6, 150 mM KCl, 20 % glycerol, and 1 mM DTT) supplemented with 10 mM $MgCl_2$ at RT for 1 h. dCas9$^{LD555}$ was separated from unincorporated dye and Sfp phosphopantetheinyl transferase by gel filtration in a Superdex 200 Increase 10/300 GL column (Cytiva # 15182085).

**dCas9$^{JF646}$**: labeling of dCas9-Halo with fluorescent dye JF646 was carried out as previously described[2].

**Mcm2-7$^{JF646\text{-}Mcm3}$:** labeling of Mcm2-7$^{Halo\text{-}Mcm3}$ with fluorescent dye JF646-HaloTag ligand was carried out as previously described[2]. The labeling efficiency was determined to be at least 80% after estimating protein and fluorophore concentrations relative to known standards. Using this method, we obtained a distribution of number of Mcm2-7 complexes per diffraction spots similar to the one obtained in previous studies in which the labeling efficiency was measured at the single-molecule level[5].

**Cdc45$^{LD555}$:** Cdc45-iS6 was fluorescently labelled by incubating Cdc45-iS6 with Sfp phosphopantetheinyl transferase and LD555-CoA (Lumidyne Technologies, custom synthesis) in a 1:1:5 molar ratio in Cdc45 gel filtration buffer (250 K-phosphate, pH 7.6, 150 mM KOAc, 10 % glycerol, and 0.5 mM DTT) supplemented with 10 mM $MgCl_2$ at RT for 1 h. Cdc45$^{LD555}$ was separated from unincorporated dye

and Sfp phosphopantetheinyl transferase by gel filtration in a Superdex 200 Increase 10/300 GL column (Cytiva #15182085). After gel filtration, positive fractions were pooled and concentrated in an Amicon Ultra-4 centrifugal filter Ultracel 30 k (Millipore # UFC803024). Labelling efficiency was measured to be 85 ± 4 % (measured value ± instrumental error) by measuring the absorption at 280 nm and 555 nm.

## 2.5.2 Single-molecule instrumentation and imaging

### 2.5.2.1 Buffers

**Buffer A:** 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, and 1 M NaCl.
**Buffer B:** 10 mM HEPES-KOH pH 7.6, 1 mM EDTA, and 1 M KOAc.
**Buffer C:** 10 mM HEPES-KOH pH 7.6, and 1 mM EDTA.
**Loading Buffer:** 25 mM HEPES-KOH pH 7.6, 100 mM K glutamate, 10 mM MgOAc, 0.02 % NP40 substitute, 10 % glycerol, 2 mM DTT, 100 µg/ml BSA, and 5 mM ATP.
**HSW Buffer:** 25 mM HEPES-KOH pH 7.6, 300 mM KCl, 10 mM MgOAc, 0.02 % NP40 substitute, 10 % glycerol, 1 mM DTT, and 400 µg/ml BSA.
**CMG Buffer:** 25 mM HEPES-KOH pH 7.6, 250 mM K glutamate, 10 mM MgOAc, 0.02 % NP40 substitute, 10 % glycerol, 1 mM DTT, and 400 µg/ml BSA.
**Elution Buffer:** CMG buffer supplemented with 10 mM biotin.
**Imaging Buffers:** CMG buffer supplemented with 2 mM 1,3,5,7 cyclooctatetraene, 2 mM 4-nitrobenzylalchohol, and 2 mM Trolox.

### 2.5.2.2 DNA functionalization and binding to magnetic beads

20 µg of 23.6 kb plasmid pGL50-ARS1 containing a natural ARS1 origin were linearized overnight with AflII (NEB # R0520L). The resulting 4-nt overhangs TTAA at both ends of the linear DNA were functionalized by incorporating desthiobiotinylated dATP (Jena Bioscience # NU-835-Desthiobio) and digoxigenylated dUTP (Jena Bioscience # NU-803-DIGXL) with Klenow Fragment (3'→5' exo-) (NEB # M0212L); unincorporated nucleotides were removed with Microspin™ S-400 HR spin columns (GE Healthcare # GE27-5140-01). The functionalized DNA was bound overnight at 4 °C to 4 mg of Dynabeads™ M-280 Streptavidin magnetic beads (Invitrogen # 11205D) in Buffer A. After binding, beads were washed twice with Buffer B, twice with Buffer C, and stored at 4 °C in Buffer C. The amount of bound DNA was measured by comparing the concentration of DNA in the supernatant before and after binding, yielding a binding efficiency of 2.3-2.9 mg DNA (~150-190 fmol)/mg beads.

### 2.5.2.3 Hybrid ensemble and single-molecule assay

#### *2.5.2.3.1 Ensemble CMG assembly and activation*

CMG assembly and activation reactions were carried out in two stages: Mcm2-7 loading and phosphorylation, and CMG assembly and activation. Unless otherwise specified, each step of the reaction was conducted at 30 °C with 800 rpm shaking:

**Mcm2-7 loading and phosphorylation**: 1 mg of magnetic DNA-bound magnetic beads was washed with 200 µl of Loading Buffer, and resuspended in 75 µl of Loading Buffer. To load Mcm2-7 hexamers onto the origin-containing DNA, 35.7 nM ORC, 50 nM Cdc6, and 100 nM Mcm2-7/Cdt1 (or Mcm2-7$^{JF646\text{-}Halo\text{-}Mcm3}$/Cdt1 or Mcm2-7$^{Mcm2(6A)}$/Cdt1) were incubated with the beads for a total of 30 min, but added to the reaction at 0 min, 5 min and 10 min, respectively. Subsequently, 100 nM DDK was added and the reaction incubated for 30 min. The supernatant was then removed, and the beads were washed once with 200 µl of HSW buffer and once with 200 µl of CMG Buffer.

**CMG assembly and activation**: after washing, beads were resuspended in 50 µl of CMG Buffer supplemented with 5 mM ATP. Then, 50 nM Dpb11, 200 nM GINS, 30 nM Polε, 20 nM S-CDK, 50 nM Cdc45$^{LD555}$, 30 nM Sld3/7, 55 nM Sld2, and 10 nM Mcm10 were added to the reaction and incubated for 15 min; For this step, a master mix of all the proteins was made immediately before and incubated on ice. After CMG assembly and activation, the supernatant was removed, and the beads were washed three times with 200 µl of HSW Buffer and once with 100 µl of CMG Buffer. After washing, the assembled DNA-protein complexes were eluted by resuspending the magnetic beads in 200 µl of Elution Buffer, and incubated at RT for 1 h with 800 rpm shaking. The supernatant was then removed and diluted by the addition of 1400 µl of CMG Buffer, and divided into two 700 µl samples for single-molecule imaging.

#### *2.5.2.3.2 Single-molecule imaging*

In general, single-molecule experiments were performed simultaneously on two instruments that combine optical tweezers and confocal microscopy (C-Trap and Q-Trap, LUMICKS); The only exceptions to this are the two-color colocalization experiments and the experiments with the 6A mutant, which were carried out solely in the C-Trap (LUMICKS). Both instruments use a microfluidic chip with five inlets and one outlet. Three of these channels are injected from the left and used for bead trapping and DNA-protein complex-trapping. The other two channels were used as protein reservoirs and buffer exchange locations (**Fig. 2.1a**). Prior to each

experiment, the microfluidic chips of both instruments were passivated for at least 30 min with 1 mg/mL bovine serum albumin (BSA, NEB # B9000S) followed by 0.5% Pluronic® F-127 (Sigma # P2443).

In all experiments, the channels contained the following solutions:

- ***Channel 1*:** 2.06 µm anti-digoxigenin coated polystyrene beads (Spherotech # DIGP-20-2) diluted 1:50 in PBS.
- ***Channel 2*:** CMG-containing DNA eluted from magnetic beads.
- ***Channel 3*:** Imaging Buffer.
- ***Channels 4 and 5*:** Imaging Buffer and Imaging Buffer supplemented with 5 mM ATP or 5 mM ATPγS (Roche # 11162306001).

Prior to each experiment, the trapping laser power was adjusted to achieve a stiffness of 0.3 pN/nm in both traps[16,41]. Then, individual DNA-molecules were trapped between two beads in channel 2, and the tethering of single DNA molecules was confirmed by analysing the force-extension curve[42]. The DNA was then transferred to either channel 4 or channel 5. The distance between both beads was then fixed to achieve a tension of 2 pN, and the DNA was imaged without flow. In all single-color experiments, fluorescent dye LD555 was illuminated with a 561 nm laser at a power of 4 µW as measured at the objective, and the fluorescence was detected on a single-photon counting detector. 2D confocal scans were performed over an area of 160 x 18 pixels, which covered the entire DNA stretched at a tension of 2 pN and the edges of both beads. Pixel size was set to 50 x 50 nm, illumination time per pixel was set to 0.2 ms, and the frame rate was set to 5 s.

Dual-color experiments were carried out almost identically, with the following differences: 1) fluorescent dyes LD555 and JF646 were simultaneously illuminated with a 561 nm laser at a power of 4 µW and a 638 nm laser at a power of 12.5 µW, and 2) the frame rate was set to 0.7 s. The microscopes output HDF5 files that store the confocal scan data, as well as force data and bead location data monitored during the scan.

## 2.5.3 Ensemble assays

### 2.5.3.1 CMG sliding assay

#### *2.5.3.1.1 DNA template generation*

Both 1.4 kb DNA constructs used had one biotinylated end and the same overall sequence containing an ARS1 origin and an *Hpa*II methyltransferase recognition

site (CCGG) at the other end. However, only one of the constructs contained a 5-fluoro-2'-deoxy-cytosine within this recognition site to covalently trap the methyltransferase[27,43]. Both constructs were synthesized by PCR using gBlock™ DRM8 (IDT, custom synthesis) as a template and primer pairs DRM_222 and DRM_220 (for the construct without a protein crosslink), or DRM_222 and DRM_218 (for the construct with a protein crosslink). Both reactions were run on a 0.8 % agarose gel, and the appropriate bands were purified from the gel and stored at -20 °C.

#### *2.5.3.1.2 DNA:protein crosslink formation and binding to magnetic beads*

2.5 µg of each DNA construct were incubated at 37 °C overnight with *Hpa*II methyltransferase in a 50:1 protein:DNA molar ratio in CutSmart™ buffer (NEB) supplemented with 10 µM S-adenosylmethionine (NEB #B9003S). Then, each reaction was bound to 1.5 mg of Dynabeads™ M-280 Streptavidin magnetic beads (Invitrogen #11205D) in Buffer A for 1 h at 37 °C and 1000 rpm shaking. After binding, beads were washed twice with Buffer B, twice with Buffer C, and stored at 4 °C in Buffer C.

#### *2.5.3.1.3 Sliding assay*

Unless otherwise specified, every step of the reaction was carried out at 30 °C with 1250 rpm shaking. For each condition, 250 µg of DNA-bound magnetic beads were washed with 50 µl of Loading Buffer. Then, 35.7 nM ORC, 50 nM Cdc6, and 100 nM Mcm2-7/Cdt1 (or Mcm2-7$^{Mcm2(6A)}$/Cdt1) were added and incubated with the DNA-bound beads for 30 min; a master mix of all the proteins was made immediately before addition and incubated on ice. Subsequently, 100 nM DDK was added, and the reaction incubated for 30 min. The supernatant was removed, and the beads were washed once with 50 µl of HSW Buffer 2 (25 mM HEPES-KOH pH 7.6, 500 mM NaCl, 10 mM MgOAc, 0.02 % NP40 substitute, 10 % glycerol, 1 mM DTT, and 400 µg/ml BSA) and once with 50 µl of CMG Buffer. Beads were then resuspended in 50 µl of CMG Buffer supplemented with 5 mM ATP, 50 nM Dpb11, 200 nM GINS, 30 nM Polε, 20 nM S-CDK, 50 nM Cdc45$^{LD555}$, 30 nM Sld3/7, and 55 nM Sld2 and incubated for 5 min; a master mix of all the proteins was made immediately before and incubated on ice. After CMG assembly, the supernatant was removed, and the beads were washed once with 50 µl of HSW Buffer 2 and once with 50 µl of CMG Buffer. After washing, beads were resuspended in 110 µl of HSW Buffer (containing 300 mM KCl) with or without 5 mM ATPγS, and incubated at 30 °C with 1250 rpm shaking. At the indicated time points, a 20 µl sample was taken from each reaction, and beads were washed with 40 µl of CMG buffer; beads were then resuspended in 15 µl of MNase Elution Buffer (45 mM HEPES-KOH pH 7.6, 300 mM KOAc, 5 mM

MgOAc, 2 mM $CaCl_2$, and 10 % glycerol) supplemented with 0.45 µl of Micrococcal nuclease (NEB # M0247S), and incubated at 30 °C for 2 min without shaking. The supernatant was then collected and run on a 4-12 % Bis-Tris polyacrylamide gel. To monitor the amount of bound fluorescent Cdc45, gels were scanned with a green laser on an Amersham Typhoon. Densitometry was performed on ImageJ.

#### 2.5.3.2 Unwinding assay

All ensemble unwinding assays were carried out as previously described[10].

### 2.5.4 Data analysis

#### 2.5.4.1 Software and code

We used Python 3.8 with several libraries for image processing. We used the Laplacian of Gaussian detector from Python's "scipy" for spot detection. We used the Linear Assignment Problem method[44] and the "scipy" solver "linear_sum_assignment" to do spot tracking. Bleaching trace analysis was done with the "ruptures" library. A full list of the exact python libraries and their versions: numpy==1.19.5; matplotlib==3.2.2; lumicks-pylake==0.7.1; streamlit==0.74.1; scipy==1.6.1; scikit-image==0.16.2; scikit-learn==0.23.1; pyyaml==5.3.1; pandas==1.0.5; pillow==7.2.0; tifffile==2021.1.11; jupyterlab==2.1.5; notebook==6.0.3; ruptures==1.1.6; pykalman==0.9.5.

#### 2.5.4.2 Overview of data analysis

After taking confocal scans, the resulting raw image data was processed to generate a table containing the spot detections in each frame. These spot detections are connected between frames to produce traces that contain location and intensity information over time.
During the subsequent motion analysis, we fit linear segments to each location trace. The resulting velocities are used to determine whether a trace is static or not. Finally, we determine the type of motion of each non-static trace using anomalous diffusion analysis.

### 2.5.4.3 Acquiring trace data from raw images

#### *2.5.4.3.1 Spot detection and tracking*

For spot detection we use the scikit-image implementation of a Laplacian of Gaussian (LoG) blob detector[45]. We set the detection radius $r_{LoG}$ to 5 pixels (250 nm); the LoG sigma parameter is then given by $\sigma_{LoG} = r_{LoG} / \sqrt{2}$. The detection threshold is set to 0.5 ADU/pixel[2]. Detected spots are localized with subpixel resolution by performing Gaussian profile fits on spot intensity projections in both x- and y-directions. For frame-by-frame tracking of the spots, we use our own implementation of the Linear Assignment Problem (LAP) framework[44] with a maximum spot linking distance of 6 pixels (300 nm) and a maximum frame gap of 3 frames (15 s). Spots are considered colocalized if they are less than 2 pixels (100 nm) apart. Spot intensities are given by the total photon count within the detection radius.

#### *2.5.4.3.2 Location and fluorophore intensity calibration*

We use the fluorescent dCas9 data (**Fig. S2.3j**) to calibrate spot locations and expected fluorophore bleaching step sizes. Because the location of the dCas9 on the DNA is known, a pixel-to-base-pair map can be made for the confocal images by mapping the mean location of all dCas9 spots (on the left and the right side of the DNA) to the corresponding locations in base pairs. Moreover, because dCas9 is labelled with the same fluorophores as the fluorescent proteins used in the CMG experiments, and because dCas9 spots contain one dCas9 molecule, we can find the fluorophore bleaching step size mean $\mu_{\Delta I}$ and standard deviation $\sigma_{\Delta I}$. The minimum bleaching step size, needed for bleaching step fitting, is set to $\Delta I_{min} \leq \mu_{\Delta I} - 2\,\sigma_{\Delta I}$ to capture at least 95% of all bleaching events.

#### *2.5.4.3.3 Determination of number of fluorophores per diffraction-limited spot*

To determine the number of fluorescently labelled proteins within each diffraction-limited spot, we count the number of photobleaching step within each spot. For experiments with multiple laser colors (in this case red ($r$) and green ($g$)), we first correct spot intensities for crosstalk by using the equation $I_{r,\,corrected} = I_r - I_g \cdot \mu_{\Delta I,\,r\,(crosstalk)} / \mu_{\Delta I,\,g}$. Then, we fit bleaching traces to a piecewise constant function using Change-Point Analysis (CPA) (we use a Python implementation called 'ruptures'[46]). We use an L2 cost function to detect mean-shifts in the signal, with a minimum segment length of 2 and a penalty term of $\Delta I_{min}^2$. If any steps smaller than $\Delta I_{min}$ are detected, these are pruned starting at the smallest step, until only steps larger than $\Delta I_{min}$ remain.

#### 2.5.4.4 Data filtering

The resulting data table of traces with number of fluorescent proteins per spot was filtered in order to reduce noise, outliers, and data that is not suitable for further motion analysis:

1. While the distance between the optical traps is constant, the force between the traps can fluctuate; jumps in the force signal could indicate, for instance, DNA 'slipping' from the beads, or a protein aggregate landing on a bead, which makes the location signal inaccurate. Hence, if the force signal exhibits a jump larger than $2\sigma_F$ after fitting with CPA, where $\sigma_F$ = 0.1 pN is the force fluctuation of a clean trace, only the part of the trace before that jump is used for motion analysis.
2. Diffraction-limited spots containing more than 5 fluorescent proteins, likely aggregates, are filtered out.
3. Any traces starting or ending within 1 kbp from a bead are filtered out to prevent any proteins likely stuck to a bead from entering the dataset.
4. Any traces starting after frame 3 are also filtered away because we do not expect any fluorescent protein to land on the DNA during the scan.
5. The last frame of each trace is omitted for motion analysis because photobleaching often happens while that frame is being taken, resulting in a distorted spot with an incorrect position.
6. Finally, in order to perform reliable motion analysis, only traces with a length of 14 frames or more are retained and used for motion analysis (**Fig. S2.6c**).

#### 2.5.4.5 Positional analysis

In all positional plots, we report the average position of the first three frames of each trace as the initial position of CMG. The bin size of the initial position histograms was set to 700 bp to be close to the diffraction limit while having the ARS1 origin positioned near the center of its corresponding bin.

#### 2.5.4.6 Motion analysis

##### *2.5.4.6.1 Trace segment fitting*

To reduce noise before we fit segments to each trace, we first apply a Kalman filter with expectation-maximization (using the pykalman Python library

https://github.com/pykalman/pykalman). Then, we fit linear segments with CPA[46], this time using a linear cost function to fit a multiple linear regression model to the trace. The minimum segment size is set to 3 and the penalty term is set to 0.3; halving or doubling the penalty term does not give a significant change in results, showing that the results are robust. After this procedure, each spot detection has associated with it a CPA-fitted velocity $v_{CPA}$.

The CPA fit makes sense for static traces and traces exhibiting piecewise linear motion, but not for spots undergoing diffusive motion. For the latter, we expect to see CPA segments with randomly alternating directions, and random velocities from some distribution with a variance dictated by the diffusion constant. A detailed description of the analysis of diffusive spots can be found below.

#### *2.5.4.6.2 Motion calibration*

The distribution of $v_{CPA}$ for fluorescent dCas9 gives us two values to calibrate the motion analysis. Firstly, the mean velocity $\mu_v$ = 0.38 bp/s provides a drift correction value. Secondly, the standard deviation $\sigma_v$ = 0.40 bp/s gives us a cutoff value to determine whether a diffraction-limited spot is static or not; we set this cut-off at the conservative value of $5\sigma_v$ = 2.0 bp/s.

Another value we need for further analysis is the location measurement error $\sigma_x$. This error is given by the standard deviation of detected dCas9 locations around their mean, after drift correction, which is found to be $\sigma_x$ = 72 bp (≈ 24 nm) (**Fig. S2.3e**).

#### *2.5.4.6.3 Anomalous diffusion analysis*

First, we correct each trace for drift with $x_{corrected}(t) = x(t) - t \cdot \mu_v$. Then we use mean squared displacement (MSD) analysis[26] to fit an anomalous diffusion exponent α, which characterizes the motion type of each mobile trace. The MSD has the form:

$$MSD(\tau) = D_\alpha \tau^\alpha + 2\sigma_x^2 \quad \textbf{(Eq. 1)},$$

where $D_\alpha$ is the anomalous diffusion constant and $\tau$ is the lag time. For spots undergoing confined diffusion, $\alpha \ll 1$; for freely diffusive spots $\alpha \approx 1$, and for traces exhibiting unidirectional motion $\alpha \gg 1$. The fit is performed through the logarithm of the measurement error corrected MSD:

$$\log(MSD(\tau) - 2\sigma_x^2) = \log(D_\alpha) + \alpha \log(\tau) \quad \textbf{(Eq. 2)}.$$

We use least squares to fit up to a maximum lag time $\tau_M$ of 33% of the total length of the trace, with a minimum $\tau_M$ of 5 frames and a maximum of 50 frames. The value

of $\alpha$ is constrained between 0 and 2. The trace is then placed into one of three categories using the fitted value of $\alpha$, with confined diffusive spots $0 \leq \alpha < 0.5$, freely diffusive spots $0.5 \leq \alpha < 1.5$, and unidirectionally moving spots $1.5 \leq \alpha \leq 2$. Because we expect populations around $\alpha \approx 1$ and $\alpha \approx 2$, we need the error in alpha $\sigma_\alpha$ to be < 0.5 in order to ensure statistically significant results.

#### *2.5.4.6.4 Calculation of diffusion coefficients*

For traces that are found to be diffusive, we calculate the diffusion coefficient by redoing the MSD fit, setting $\alpha = 1$, and using a previously published appropriate range of delay times[47].

#### *2.5.4.6.5 Anomalous diffusion exponent error determination*

In order to study the error in $\alpha$ as a function of minimum trace length, we have run the same analysis on 512 simulated diffusive traces and 512 simulated traces with a constant speed (with $\alpha = 1$ and $\alpha = 2$, respectively), with representative values for the diffusion constant $D = 1.5 \times 10^{-3}$ kb$^2$/s and speed $v = 5$ bp/s. We use the experimentally determined measurement error ($\sigma_x = 72$ bp) and mean fluorophore lifetime (25 frames). These simulations show that we need a minimum trace length of 14 frames for the error in alpha, $\sigma_\alpha$, to be < 0.5, justified by the motion classification cutoffs discussed above.

In all plots we use experimental means and standard deviations whenever possible. On population bar plots we use the statistical error, i.e., the standard error of proportion, given by $\sqrt{p(1-p)/n}$, with $p$ the measured proportion and $n$ the sample size.

### 2.5.4.7 Bin size selections

In general, the bin size of all the histograms in this manuscript were chosen to be larger but in the order of magnitude of the error of the random variable being displayed. Specifically:

- The bin size of the initial position histograms was set to 700 bp to be close to the diffraction limit while having the ARS1 origin positioned near the center of its corresponding bin.
- The bin size of the absolute instantaneous velocities histograms was set to 2.5 bp/s, which is ~ 6 X the velocity noise (**Fig. 2.2a inset**)
- The bin size of the absolute mean velocities histograms was set to 1.0 bp/s, which is ~ 2 X the velocity noise (**Fig. 2.2a inset**)

- The bin size of instantaneous velocities from CPA histograms was set to 1.0 bp/s, which is ~ 2 X the velocity noise (**Fig. 2.2a inset**)
- The bin size of the processivities histograms was set to 0.5 kbp, which is ~ 7 X the location error (**Fig. S2.3e**)
- The bin sizes of the histograms of step sizes and location errors of dCas9 (**Fig. S2.3a-b,e**) are irrelevant because we only use the means and the standard deviations of the underlying distribution for our analysis.
- The bin sizes of the histograms of anomalous diffusion exponents $\alpha$ are set to 0.25, which is ~ 1/2 the error in $\alpha$ (**Fig. S2.6c**). These histograms, however, were only an intermediate in our analysis. In the final analysis, a bin size of 0.5 ~ the error in $\alpha$ (**Fig. S2.6c**, **Methods**) was used to classify motion types.

## 2.6 Data availability

Raw and processed ensemble and single-molecule data generated in this study have been deposited in the 4TU.ResearchData repository and can be found at https://doi.org/10.4121/19948253. The repository contains:

- A table with an overview of experiments.
- Spot position and intensity tables sorted by experimental condition and by setup.
- Filtered spot tracking tables, with connected spot detections for each frame in each scan. Each row has a scan_id and trace_id.
- Motion analysis summary tables, containing motion information for each trace. The scan_id, trace_id fields link this table to the table containing the full trace information.
- Example TIFF files

## 2.7 Code availability

All the code used in this manuscript is available at https://gitlab.tudelft.nl/nynke-dekker-lab/public/cmg-activation.

## 2.8 References

1. Bell, S. P. & Labib, K. Chromosome duplication in Saccharomyces cerevisiae. *Genetics* **203**, 1027–1067 (2016).
2. Sánchez, H. *et al.* DNA replication origins retain mobile licensing proteins. *Nat. Commun.* **12**, (2021).

3. Remus, D. *et al.* Concerted Loading of Mcm2-7 Double Hexamers around DNA during DNA Replication Origin Licensing. *Cell* **139**, 719–730 (2009).
4. Duzdevich, D. *et al.* The dynamics of eukaryotic replication initiation: Origin specificity, licensing, and firing at the single-molecule level. *Mol. Cell* **58**, 483–494 (2015).
5. Ticau, S., Friedman, L. J., Ivica, N. A., Gelles, J. & Bell, S. P. Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell* **161**, 513–525 (2015).
6. Greiwe, J. F. *et al.* Structural mechanism for the selective phosphorylation of DNA-loaded MCM double hexamers by the Dbf4-dependent kinase. *Nat. Struct. Mol. Biol.* **29**, (2021).
7. Yeeles, J. T. P., Deegan, T. D., Janska, A., Early, A. & Diffley, J. F. X. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**, 431–435 (2015).
8. de Jesús-Kim, L., Friedman, L. J., Ramsoomair, C., Gelles, J. & Bell, S. P. DDK regulates replication initiation by controlling the multiplicity of Cdc45-GINS binding to Mcm2-7. *Elife* **10:e65471**, 1–30 (2021).
9. Kose, H. B., Larsen, N. B., Duxin, J. P. & Yardimci, H. Dynamics of the Eukaryotic Replicative Helicase at Lagging-Strand Protein Barriers Support the Steric Exclusion Model. *Cell Rep.* **26**, 2113-2125.e6 (2019).
10. Douglas, M. E., Ali, F. A., Costa, A. & Diffley, J. F. X. The mechanism of eukaryotic CMG helicase activation. *Nature* **555**, 265–268 (2018).
11. Lewis, J. S. *et al.* Mechanism of replication origin melting nucleated by CMG helicase assembly. *Nature* (2022) doi:10.1038/s41586-022-04829-4.
12. Abid Ali, F. *et al.* Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate. *Nat. Commun.* **7**, (2016).
13. Eickhoff, P. *et al.* Molecular Basis for ATP-Hydrolysis-Driven DNA Translocation by the CMG Helicase of the Eukaryotic Replisome. *Cell Rep.* **28**, 2673-2688.e8 (2019).
14. McClure, A. W. & Diffley, J. F. X. Rad53 checkpoint kinase regulation of dna replication fork rate via mrc1 phosphorylation. *Elife* **10**, 1–24 (2021).
15. Choi, J., Grosely, R., Puglisi, E. V. & Puglisi, J. D. Expanding single-molecule fluorescence spectroscopy to capture complexity in biology. *Curr. Opin. Struct. Biol.* **58**, 233–240 (2019).
16. Wasserman, M. R., Schauer, G. D., O'Donnell, M. E. & Liu, S. Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase. *Cell* **178**, 600-611.e16 (2019).
17. Kose, H. B., Xie, S., Cameron, G., Strycharska, M. S. & Yardimci, H. Duplex DNA engagement and RPA oppositely regulate the DNA-unwinding rate of CMG helicase. *Nat. Commun.* **11**, 1–15 (2020).
18. Burnham, D. R., Kose, H. B., Hoyle, R. B. & Yardimci, H. The mechanism of DNA unwinding by the eukaryotic replicative helicase. *Nat. Commun.* **10**, 1–14 (2019).
19. Moyer, S. E., Lewis, P. W. & Botchan, M. R. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication

fork helicase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10236–10241 (2006).
20. Langston, L. D. *et al.* CMG helicase and DNA polymerase ε form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15390–15395 (2014).
21. Lewis, J. S. *et al.* Tunability of DNA polymerase stability during eukaryotic DNA replication. *Suppl. Inf.* **77**, 602086 (2019).
22. Candelli, A., Wuite, G. J. L. & Peterman, E. J. G. Combining optical trapping, fluorescence microscopy and micro-fluidics for single molecule studies of DNA-protein interactions. *Phys. Chem. Chem. Phys.* **13**, 7263–7272 (2011).
23. Hirsch, J. D. *et al.* Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: Uses for protein labeling, detection, and isolation. *Anal. Biochem.* **308**, 343–357 (2002).
24. Vrtis, K. B. *et al.* Single-strand DNA breaks cause replisome disassembly. *Mol. Cell* **81**, 1309-1318.e6 (2021).
25. Champasa, K., Blank, C., Friedman, L. J., Gelles, J. & Bell, S. P. A conserved Mcm4 motif is required for Mcm2-7 double-hexamer formation and origin DNA unwinding. *Elife* **8**, 1–23 (2019).
26. Kepten, E., Weron, A., Sikora, G., Burnecki, K. & Garini, Y. Guidelines for the fitting of anomalous diffusion mean square displacement graphs from single particle tracking experiments. *PLoS One* **10**, 1–10 (2015).
27. Duxin, J. P., Dewar, J. M., Yardimci, H. & Walter, J. C. Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell* **159**, 346–357 (2014).
28. Sparks, J. L. *et al.* The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell* **176**, 167-181.e21 (2019).
29. Miller, T. C. R., Locke, J., Greiwe, J. F., Diffley, J. F. X. & Costa, A. Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM. *Nature* **575**, 704–710 (2019).
30. Costa, A. *et al.* The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat. Struct. Mol. Biol.* **18**, 471–479 (2011).
31. Langston, L. & O'Donnell, M. Action of CMG with strand-specific DNA blocks supports an internal unwinding mode for the eukaryotic replicative helicase. *Elife* **6**, 1–23 (2017).
32. Jenkyn-Bedford, M. *et al.* A conserved mechanism for regulating replisome disassembly in eukaryotes. *Nature* **600**, 743–747 (2021).
33. Low, E., Chistol, G., Zaher, M. S., Kochenova, O. V. & Walter, J. C. The DNA replication fork suppresses CMG unloading from chromatin before termination. *Genes Dev.* **34**, 1534–1545 (2020).
34. Abid Ali, F. *et al.* Cryo-EM structure of a licensed DNA replication origin. *Nat. Commun.* **8**, 1–10 (2017).
35. Schwarz, F. W. *et al.* The helicase-like domains of Type III restriction enzymes trigger long-range diffusion along DNA. *Science (80-. ).* **340**, 353–356 (2013).
36. Gupta, Y. K., Chan, S. H., Xu, S. Y. & Aggarwal, A. K. Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nat. Commun.* **6**, (2015).

37. Carcamo, C. C. *et al.* ATP binding enhances SWR1 target search by promoting 1D scanning on DNA. *Biophys. J.* **121**, 284a (2022).
38. Zhou, Z. *et al.* Genetically Encoded Short Peptide Tags for Orthogonal Protein Labeling by Sfp and AcpS Phosphopantetheinyl Transferases. *ACS Chem. Biol.* **2**, 337–346 (2007).
39. Van Loenhout, M. T. J., De Grunt, M. V. & Dekker, C. Dynamics of DNA supercoils. *Science (80-. ).* **338**, 94–97 (2012).
40. On, K. F. *et al.* Prereplicative complexes assembled in vitro support origin-dependent and independent DNA replication. *EMBO J.* **33**, 605–620 (2014).
41. Candelli, A. *et al.* Visualization and quantification of nascent RAD51 filament formation at single-monomer resolution. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15090–15095 (2014).
42. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic Elasticity of lambda-Phage DNA. *Science (80-. ).* **265**, 1599–1600 (1994).
43. Chen, L. *et al.* Direct Identification of the Active-Site Nucleophile in a DNA (Cytosine-5)-methyltransferase. *Biochemistry* **30**, 11018–11025 (1991).
44. Jaqaman, K. *et al.* Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* **5**, 695–702 (2008).
45. Van Der Walt, S. *et al.* Scikit-image: Image processing in python. *PeerJ* 1–18 (2014) doi:10.7717/peerj.453.
46. Truong, C., Oudre, L. & Vayatis, N. Selective review of offline change point detection methods. *Signal Processing* **167**, 107299 (2020).
47. Michalet, X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **82**, 1–13 (2010).

## 2.9 Acknowledgements

We thank Anne Early, Lucy Drury, and Max Douglas for providing the yeast strains for the overexpression of unlabeled proteins, Jacob Lewis for providing us with purified M.HpaII methyltransferase, and N.D. lab members Anuj Kumar, Katinka Ligthart, and Julien Gros for assistance in purifying loading proteins and/or DDK. We also thank Kaley McCluskey, Dorian Mikolajczak, Joseph Yeeles, Jacob Lewis, Alessandro Costa, Hasan Yardimci and Taekjip Ha for scientific discussions. D.R.M. acknowledges funding from a Boehringer Ingelheim Fonds PhD Fellowship, and N.D. acknowledges funding from the Netherlands Organisation for Scientific Research (NWO) through Top grant 714.017.002, and from the European Research Council through an Advanced Grant (REPLICHROMA; grant number 789267).

## 2.10 Author contributions statement

D.R.M., H.S., J.D., and N.D. conceived the study, and N.D. supervised the study. D.R.M. developed the ensemble hybrid and single-molecule study with input from H.S., J.D. and N.D. D.R.M. carried out the ensemble biochemical assays and the did ensemble part of the hybrid ensemble and single-molecule assays. D.R.M. and H.S. acquired the single-molecule data. J.D. provided the strains for the overexpression of unlabeled proteins and advised on biochemical conditions. D.R.M. and T.v.L. carried out the cloning. D.R.M. generated the strains for the overexpression of labeled proteins. D.R.M. and T.v.L. purified the proteins. D.R.M. and H.S. labeled the proteins. E.v.V. designed and wrote data analysis routines with input from all authors as well as conceived and performed the simulations. D.R.M. and E.v.V. performed the data analysis. B.S. contributed to project planning and instrument acquisition. All authors were involved in the discussion of the data. D.R.M. and N.D. wrote the manuscript with input from all the authors.

## 2.11 Competing interests statement

The authors declare no competing financial interests.

## 2.12 Materials & correspondence

Correspondence and requests for materials should be addressed to N.D. (n.h.dekker@tudelft.nl) or J.D. (john.diffley@crick.ac.uk)

## 2.13 Supplementary figures

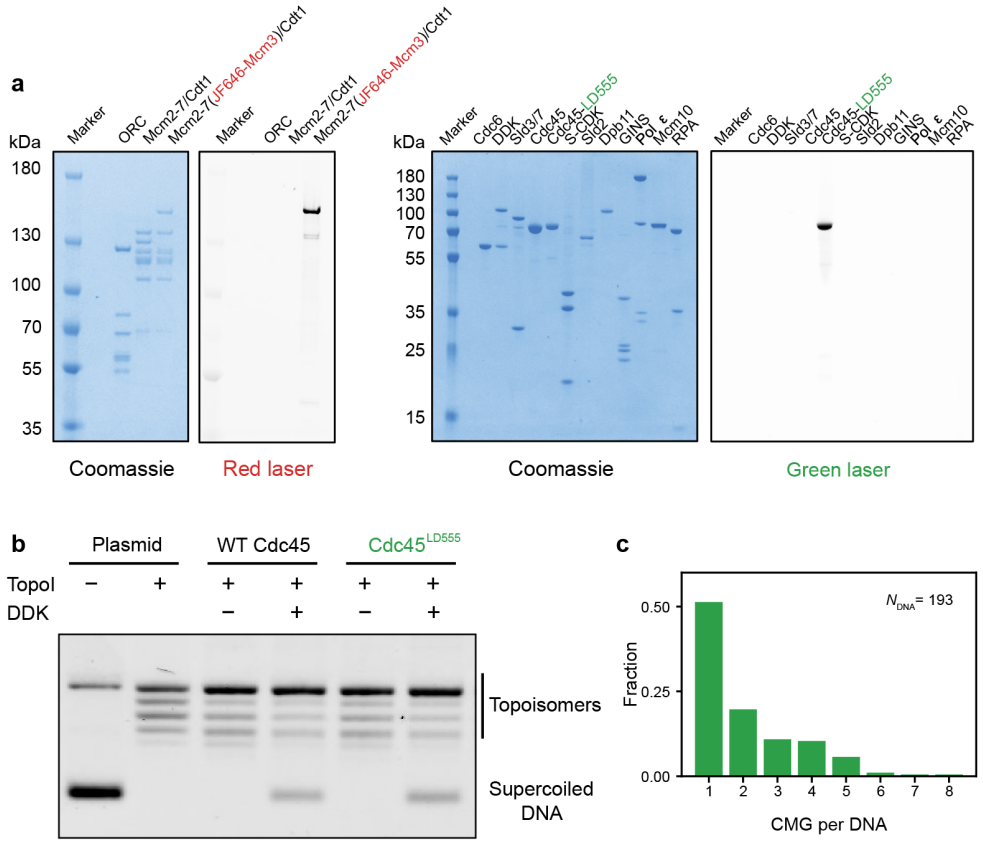



*Figure S2.1* Hybrid ensemble and single-molecule assay and reagent validation. **a** SDS-PAGE showing the minimal set of purified proteins required for the reconstitution of CMG assembly and activation; the gels were stained with Coomassie Blue Stain and fluorescently scanned with either a red or a green laser, to show the fluorescently labeled proteins in either color. **b** Ensemble unwinding assay showing that Cdc45$^{LD555}$ supports DNA unwinding to near WT levels ($N$=2 biological replicates). **c** Distribution of total numbers of fluorescent CMG complexes per DNA, obtained by combining the total number of CMG diffraction limited spots per DNA (Fig. 2.1b) with the number of CMG complexes within each spot (Fig. 2.1c).

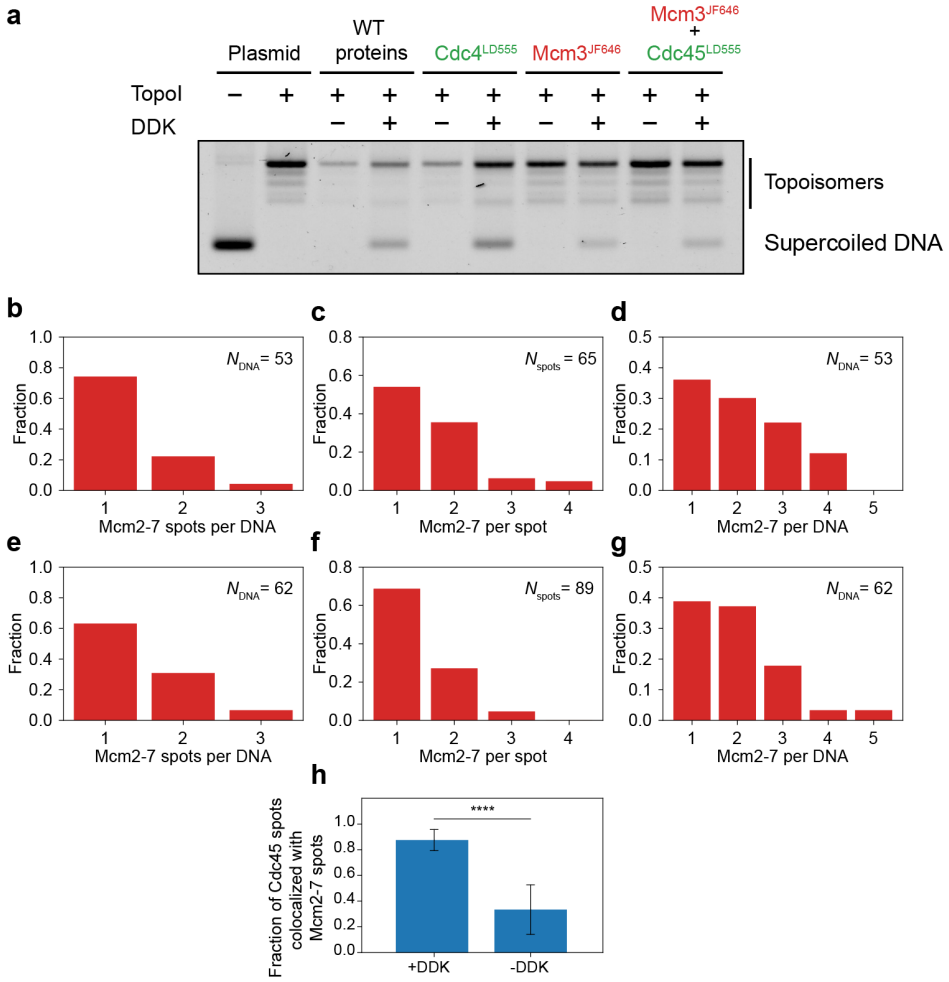
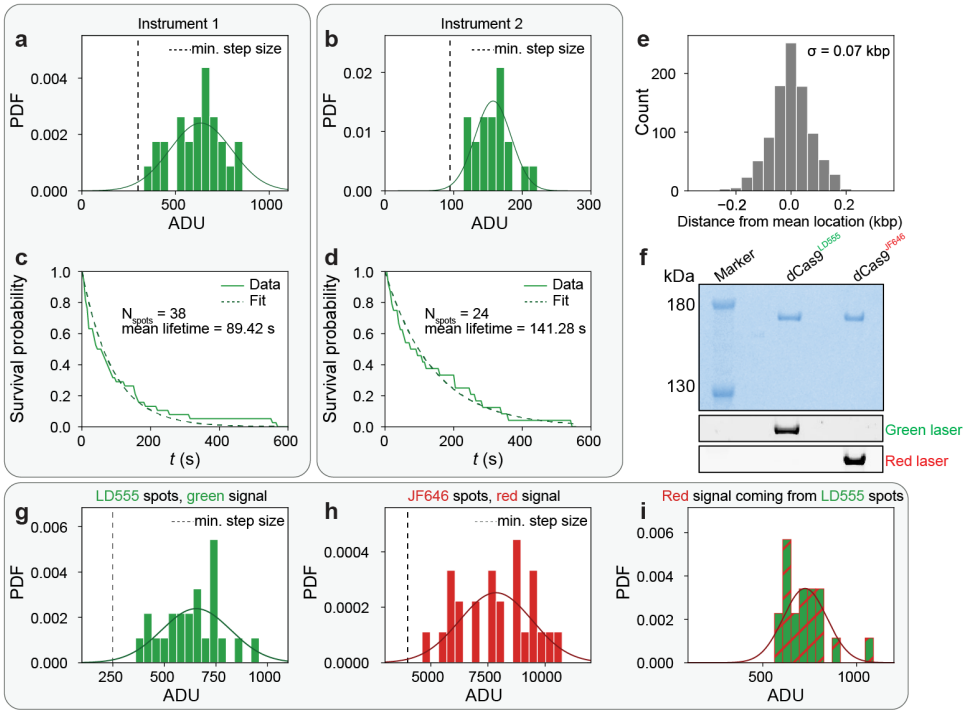



*Figure S2.2* Reagent validation, distribution of number of Mcm2-7 spots and distribution of Mcm2-7 complexes within each spot. **a** Ensemble unwinding assay showing that Mcm2-7$^{JF646}$ supports DNA unwinding alone and in conjunction with Cdc45$^{LD555}$ ($N$=1 biological replicate). **b** Distribution of the number of Mcm2-7 diffraction-limited spots per DNA in the presence of DDK. **c** Distribution of the number of Mcm2-7 complexes within each diffraction-limited spot in the presence of DDK. **d** Distribution of the total number of Mcm2-7 complexes per DNA molecule in the presence of DDK, obtained by combining data from **b** and **c**. **e** Distribution of the number of Mcm2-7 diffraction-limited spots per DNA in the absence of DDK. **f** Distribution of the number of Mcm2-7 complexes within each diffraction-limited spot in the absence of DDK. **g** Distribution of the total number of Mcm2-7 complexes per DNA molecule in the absence of DDK, obtained by combining data from **e** and **f**. **h** Mean fraction of Cdc45$^{LD555}$ diffraction-limited spots that are colocalized with Mcm2-7$^{JF646}$ diffraction-limited spots in the presence ($N_{Cdc45\ spots}$=16) or absence ($N_{Cdc45\ spots}$=6) of DDK; error bars show the standard error of proportion. Statistical significance was obtained from a two-sided binomial test (p-value=$1.2 \times 10^{-5}$).




j

| Instrument | fluorophore | signal | $\Delta I_{min}$ (ADU) | $\mu_{\Delta I}$ (ADU) | $\sigma_{\Delta I}$ (ADU) | $N_{spots}$ |
|---|---|---|---|---|---|---|
| 1 | LD555 | g | 300 | 634 | 166 | 27 |
| 2 | LD555 | g | 95 | 157 | 26.3 | 20 |
| 1 | LD555 | g | 250 | 655 | 167 | 24 |
| 1 | LD555 | r (crosstalk) | - | 725 | 117 | 20 |
| 1 | JF646 | r | $4.00 \times 10^3$ | $7.85 \times 10^3$ | $1.58 \times 10^3$ | 25 |
| 1 | JF646 | g (crosstalk) | - | negligible | negligible | - |

*Figure S2.3* Fluorescently labeled dCas9 proteins as standards for determination of number of proteins per diffraction-limited spot and localization accuracy. **a,b** Distribution of photobleaching step sizes of fluorescently labeled dCas9$^{LD555}$ imaged under the same imaging conditions as fluorescent CMG in the single-color experiments; **a** and **b** correspond to the two instruments used in this study; both distributions were fitted to a normal distribution; μ – 2σ was used as the minimum step size in the single-color CMG experiments to capture at least 95 % of bleaching events. **c-d** Distribution of times to photobleaching of fluorescently labeled dCas9$^{LD555}$ imaged under the same imaging conditions as fluorescent CMG; **c** and **d** correspond to the two instruments used in this study; both distributions were fitted to a single exponential decay. **e** distribution of positional measurements of fluorescently labeled dCas9$^{LD555}$; as dCas9$^{LD555}$ is expected to be static, the standard deviation of this distribution gives us the localization error in our experiments. **f** SDS-PAGE of dCas9 with fluorescently labeled with dyes LD555, and JF646, respectively; the gel was stained with Coomassie Blue stain and fluorescently scanned with a red, green laser, respectively. **g** Distribution of photobleaching step sizes of fluorescently labeled dCas9$^{LD555}$ when simultaneously excited with the green and red lasers in instrument 1, as done in the Mcm2-7 and Cdc45 colocalization experiments; the distribution was fitted to a normal distribution; μ – 2σ was used as the minimum step size in the dual-color CMG experiments to capture at least 95 % of bleaching events. **h** Distribution of photobleaching step sizes of fluorescently labeled dCas9$^{JF646}$ when simultaneously excited with the green and red lasers in instrument 1, as done in the Mcm2-7 and Cdc45 colocalization experiments; the distribution was fitted to a normal distribution; μ – 2σ was used as the minimum step size in the dual-color CMG experiments to capture at least 95 % of bleaching events. (*Continues on the next page*)

*Figure S2.3 (continued).* **i** Distribution of red signal coming from green fluorescently labeled dCas9$^{LD555}$ when simultaneously excited with the green and red lasers in instrument 1, as done in the Mcm2-7 and Cdc45 colocalization experiments. The distribution was fitted to a normal distribution and the mean value was used for crosstalk corrections. **j** Summary table of all the parameters obtained from **a-e**, and **g-i**.

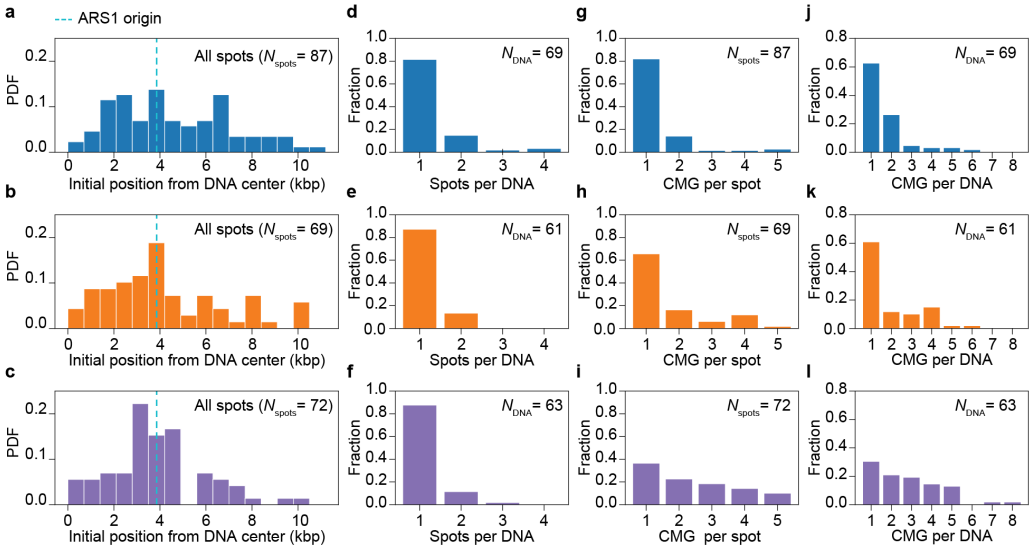



*Figure S2.4* Distribution of initial positions, numbers of CMG spots and numbers of CMG complexes within each spot for the different biochemical conditions tested. **a-c** Distribution of initial positions on the DNA of all Cdc45 diffraction-limited spots for DNA molecules imaged in **a** the presence of ATP, **b** the absence of nucleotide, or **c** the presence of ATPγS. **d-f** Distribution of numbers of CMG diffraction-limited spots for DNA molecules imaged in **d** the presence of ATP, **e** the absence of nucleotide, or **f** the presence of ATPγS. **g-i** Distribution of numbers of CMG complexes within each diffraction limited spot on DNA molecules imaged in **g** the presence of ATP, **h** the absence of nucleotide, or **i** the presence of ATPγS. **j-l** Distribution of numbers of CMG complexes per DNA for DNA molecules imaged in **j** the presence of ATP, **k** the absence of nucleotide, or **l** the presence of ATPγS.

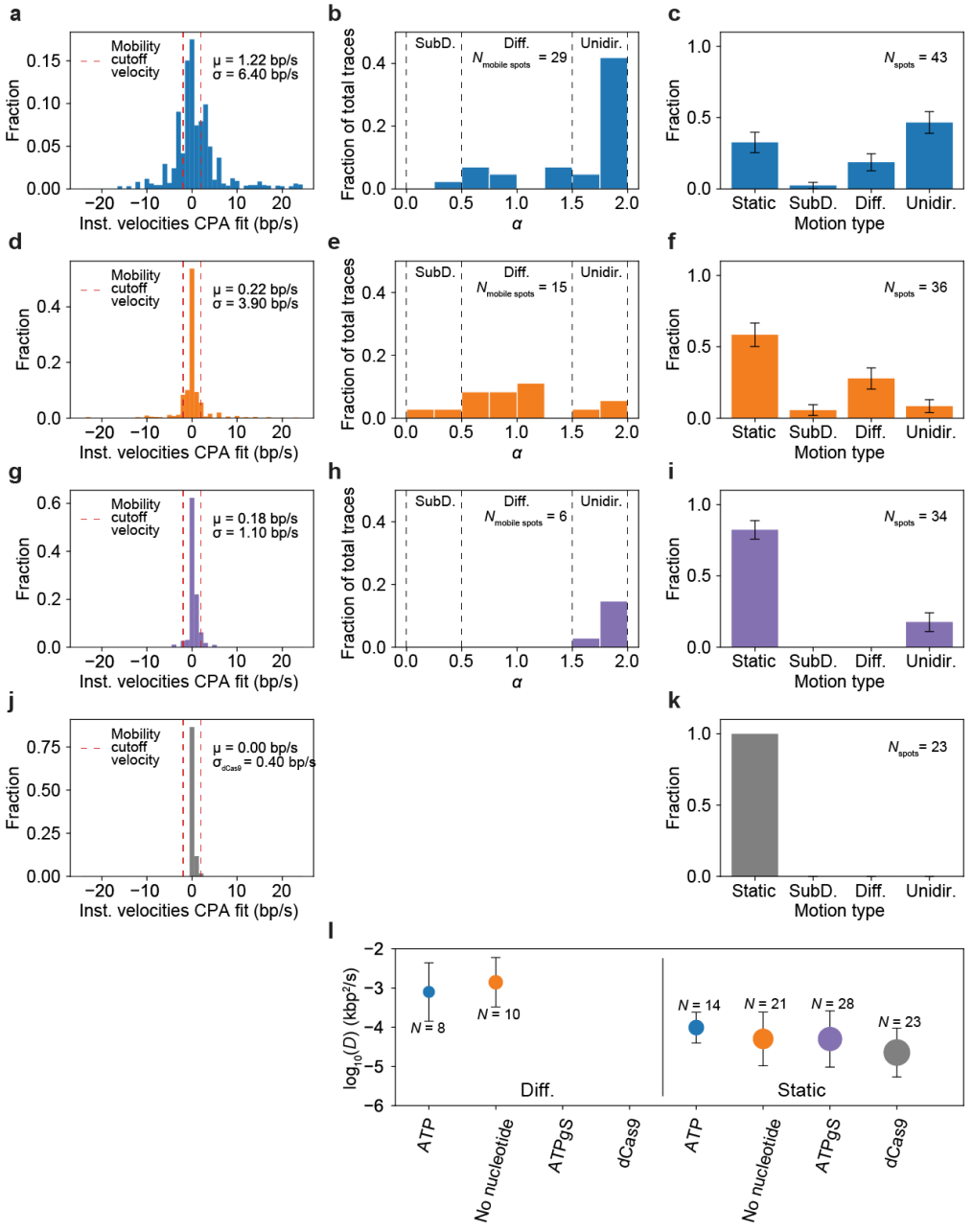



*Figure S2.5* Mobility determination and motion classification of fluorescent spots imaged under different biochemical conditions. **a** Distribution of instantaneous velocities coming from the CPA fits of CMG spots in the presence of ATP; red lines show the instantaneous velocity cutoff ($5\sigma_{dCas9}$) used to separate CMG spots into static or mobile. **b** Distribution of anomalous coefficients α of mobile CMG spots in the presence of ATP. **c** Fraction of CMG spots imaged in the presence of ATP classified into static, subdiffusive, diffusive or unidirectionally moving ($N_{spots}$=43); error bars show the standard error of proportion. **d** Distribution of instantaneous velocities coming from the CPA fits of CMG spots in the absence of nucleotide; red lines show the instantaneous velocity cutoff ($5\sigma_{dCas9}$) used to separate CMG spots into static or mobile. (*Continues on the next page*).

*Figure S2.5* (*continued*). **e** Distribution of anomalous coefficients α of mobile CMG spots in the absence of nucleotide. **f** Fraction of CMG spots imaged in the absence of nucleotide classified into static, subdiffusive, diffusive or unidirectionally moving ($N_{spots}$=36); error bars show the standard error of proportion. **g** Distribution of instantaneous velocities coming from the CPA fits of CMG spots in the presence of ATPγS; red lines show the instantaneous velocity cutoff ($5\sigma_{dCas9}$) used to separate CMG spots into static or mobile. **h** Distribution of anomalous coefficients α of mobile CMG spots in the presence of ATPγS. **i** Fraction of CMG spots imaged in the presence of ATPγS classified into static, subdiffusive, diffusive or unidirectionally moving ($N_{spots}$=34); error bars show the standard error of proportion. **j** (*same as inset in Fig. 2.2a*) Distribution of instantaneous velocities coming from the CPA fits of $dCas9^{LD555}$ spots; red lines show the instantaneous velocity cutoff ($5\sigma_{dCas9}$) used to separate CMG spots into static or mobile. **k** Fraction of $dCas9^{LD555}$ spots classified into static, subdiffusive, diffusive or unidirectionally moving ($N_{spots}$=23). **l** (left half) Diffusion constants of spots classified as diffusive for the different biochemical conditions tested (mean $D$ +/− standard deviation); (right half) Diffusion constants of spots classified as static for the different biochemical conditions tested (mean $D$ +/− standard deviation).

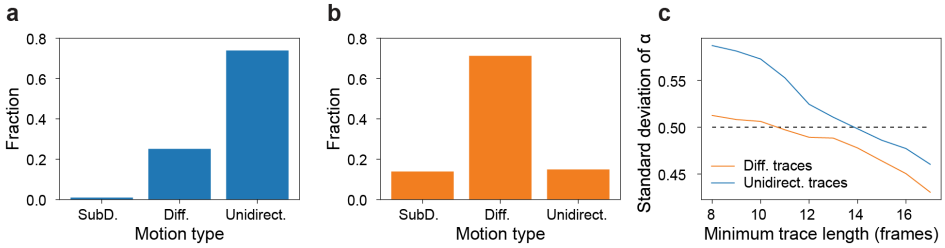



*Figure S2.6* Motion classification of simulated unidirectional or diffusive traces and anomalous diffusion exponent error determination. Motion classification of simulated **a** unidirectionally translocating traces with a representative velocity (5 bp/s) and **b** diffusive traces with a representative diffusion coefficient ($1.5 \times 10^{-3}$ $kb^2/s$). **c** Error determination of the anomalous diffusion exponent α as a function of the minimum trace length; the error falls below 0.5 for a minimum trace length of 14 frames. We start with 512 traces of each motion type with a minimum trace length of 8 pulled from a population with a mean fluorophore lifetime of 25 frames, and gradually increase the trace length filtering. The traces used in **a-b**, are those with a minimum trace length of 14, to mirror the motion analysis done on experimentally obtained CMG spots.

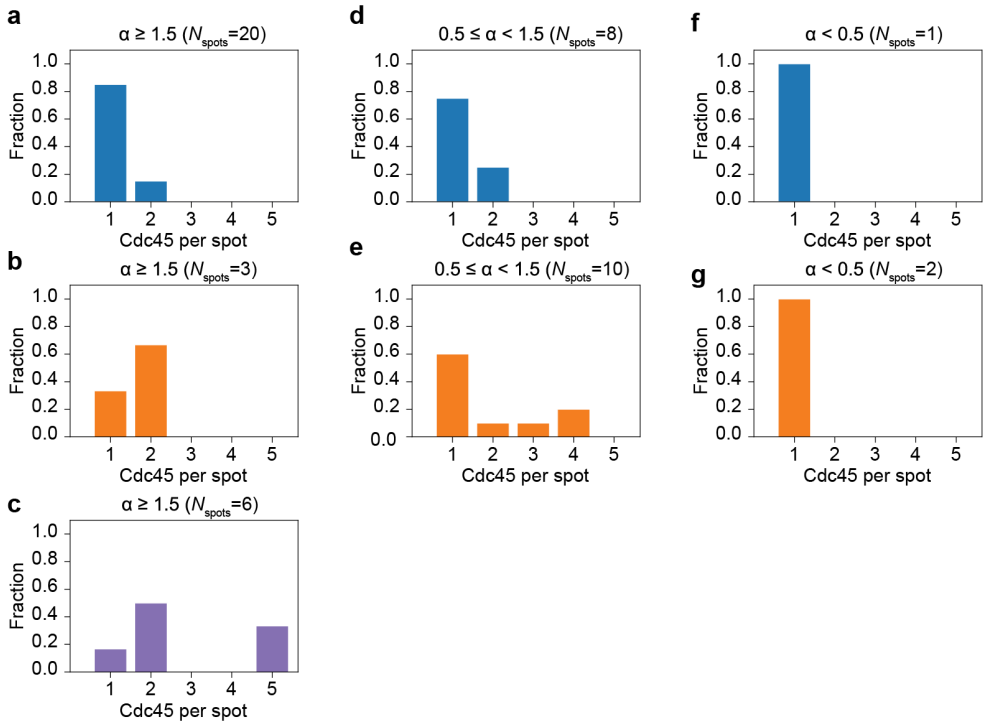



*Figure S2.7* Distribution of number of Cdc45 molecules per mobile diffraction-limited spot for the different biochemical conditions tested. **a-c** Distribution of number of Cdc45 molecules within diffraction-limited spots classified as unidirectionally moving in the **a** presence of ATP, **b** absence of nucleotide, or **c** presence of ATPγS. **d-e** Distribution of number of Cdc45 molecules within diffraction-limited spots classified as diffusive in the **d** presence of ATP, or **e** absence of nucleotide. **f-g** Distribution of number of Cdc45 molecules within diffraction-limited spots classified as subdiffusive in the **f** presence of ATP, or **g** absence of nucleotide.

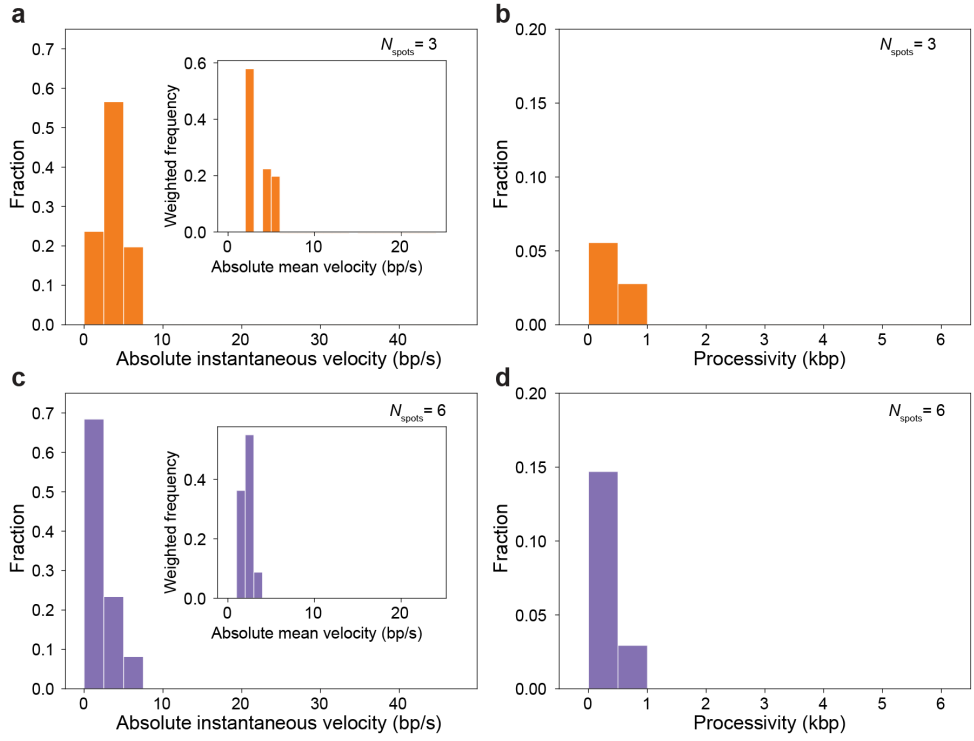



*Figure S2.8* Analysis of unidirectionally moving CMG under different biochemical conditions. **a** Distribution of absolute instantaneous velocities of unidirectionally moving CMG spots in the absence of nucleotide; (inset) Distribution of absolute mean velocities of unidirectionally moving CMG spots in the absence of nucleotide normalized by the length of each trace. **b** Distribution of processivities of unidirectionally moving CMG spots in the absence of nucleotide. **c** Distribution of absolute instantaneous velocities of unidirectionally moving CMG spots in the presence of ATPγS; (inset) Distribution of absolute mean velocities of unidirectionally moving CMG spots in the presence of ATPγS normalized by the length of each trace. **d** Distribution of processivities of unidirectionally moving CMG spots in the presence of ATPγS.

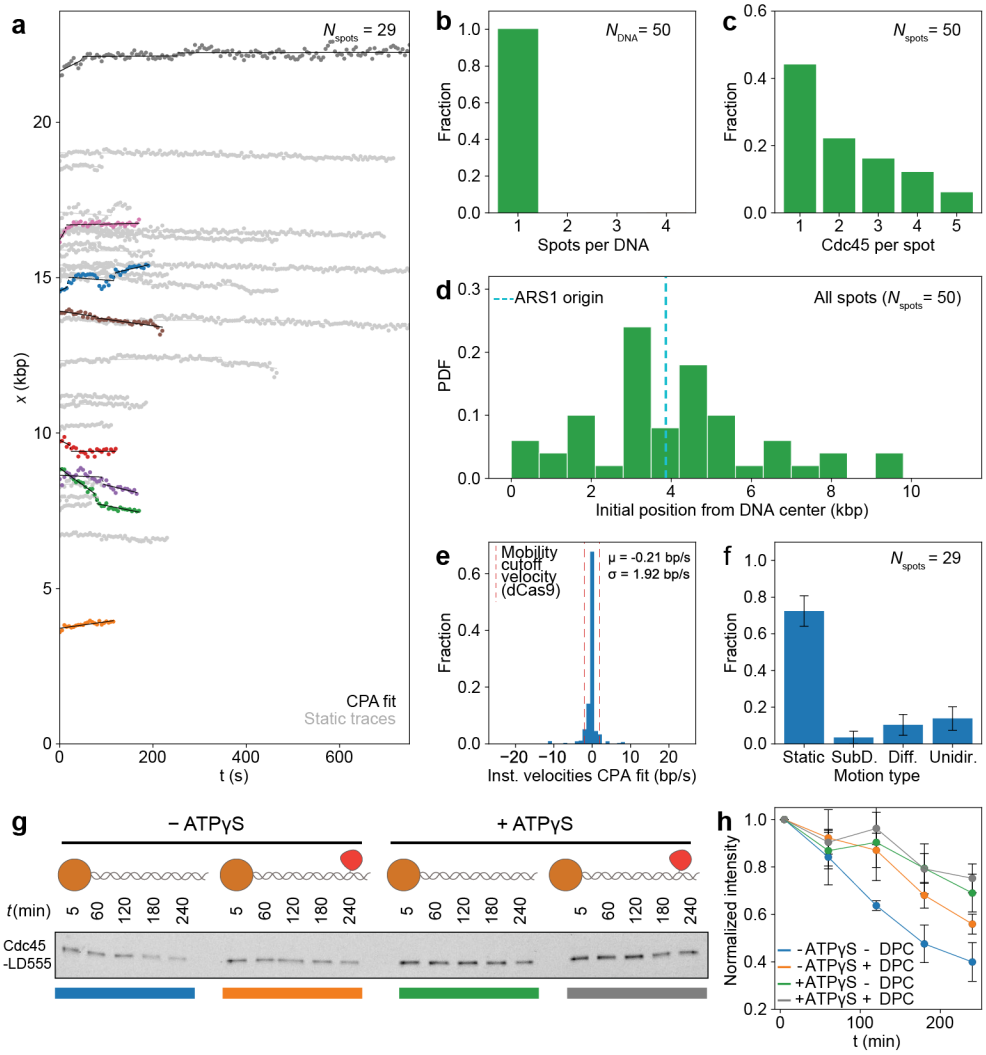



*Figure S2.9* Nucleotide binding halts CMG diffusion independently of DNA melting. **a** Position vs. time plots of CMG$^{Mcm2(6A)}$ spots in the presence of ATP; CPA fits are plotted in black, static traces are shown in light gray and mobile traces are shown in all other colors. **b** Distribution of numbers of CMG$^{Mcm2(6A)}$ diffraction-limited spots per DNA. **c** Distribution of numbers of CMG$^{Mcm2(6A)}$ complexes within each diffraction-limited spot. **d** Distribution of initial positions on the DNA of all CMG$^{Mcm2(6A)}$ diffraction-limited spots. **e** Distribution of instantaneous velocities coming from the CPA fits of CMG$^{Mcm2(6A)}$ spots in the presence of ATP; red lines show the instantaneous velocity cutoff ($5\sigma_{dCas9}$) used to separate CMG$^{Mcm2(6A)}$ spots into static or mobile. **f** Fraction of CMG$^{Mcm2(6A)}$ spots imaged in the presence of ATP classified into static, subdiffusive, diffusive or unidirectionally moving ($N_{spots}$=29); error bars show the standard error of proportion. **g** Fluorescent scan of an SDS-PAGE gel showing the amount of Cdc45$^{LD555}$ left on linear DNA bound to magnetic beads at one end and containing either a free end or an end capped with a covalently crosslinked methyltransferase. **h** Densitometry quantification of the experiment shown in **g** showing the average normalized intensity of three replicates together with their standard deviation. Data points are connected by solid lines to guide the eye.

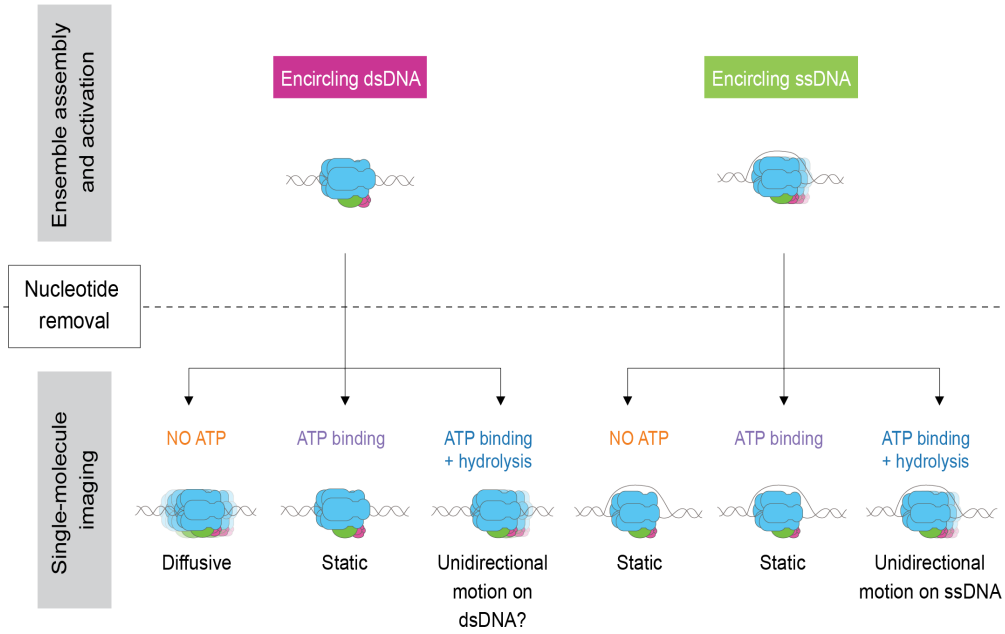



*Figure S2.10* Final model. Model showing all the experimental outcomes observed in this study with different potential explanations.

## 2.14 Supplementary tables

*Table S1.1* Oligos and primers used in this study.

| Name | 5' to 3' sequence |
|---|---|
| DRM_005 | GCTGCGCCTGCTGAACGGTGATTATAAAGATGATGATGGG |
| DRM_006 | AGCCAGCTCAGGCTATCGCCCTCGTCTGTGACTTCATC |
| DRM_184 | ACGGCTGTTAAATGGGGGGAGTGATAAGAAATACTCAATAGGC |
| DRM_185 | AATAACCAACTTAATGAATCCCCCACGTGATGATGATGATG |
| TL_033 | GCGCGCCAATTGGAGCTCCACCGCGG |
| TL_034 | GGCGCGCCGGAAACAGCTATGACCATGATTACGCC |
| DRM_218 | ATACTTTAGATTGATTTC[5-Fluoro-2'-dC]GGCTTCACCTG |
| DRM_220 | ATACTTTAGATTGATTTCCGGCTTCACCTG |
| DRM_222 | Biotin-CTAGTGGATCCCCAGGGCT |

*Table S1.2* gBlocks™ used in this study.

| gBlock DRM8: |
|---|
| TCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCTGGCGTCAATACGGGATAATACCGCGCCACATA<br>GCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGA<br>TCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGC<br>AAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTT<br>TTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACA<br>AATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTA<br>AATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCATAATCCCTTATAAATCAAAAGAATAGACCG<br>AGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCG<br>AAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGT<br>AAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCCGCGAACGTGGCGAGA<br>AAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCAC<br>CACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTGCTGAGGCGCAACTGTTGGGAAGG<br>GCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGT<br>AACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAA<br>TTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCAGGGCTGCAGGAATTCGAGCTCGGTACC<br>CACAATCAATCAAAAAGCCAAATGATTTAGCATTATCTTTACATCTTGTTATTTTACAGATTTTATGTTTAGATCTTTTA<br>TGCTTGCTTTTCAAAAGGCCTGCAGGCAAGTGCACAAACAATACTTAAATAAATACTACTCAGTAATAACCTATTTCTT<br>AGCATTTTTGACGAAATTTGCTATTTTGTTAGAGTGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTATCG<br>ATACCGTCGACCTCGAGGGGGGGCACGGTACCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTTCGAGCTTGGCGTA<br>ATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCTGAAGCATAAA<br>GTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACAACCTCAGCTTGCGCTCACTGCCCGCTTTCCAGTCGG<br>GAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT<br>CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG<br>TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAA<br>CCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCA<br>AGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTC<br>CTGTTCCGACCCTGCCGCTTAACGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTC<br>ACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAAAACAGGATTAGCAGA<br>GCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTG<br>GTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATTCGGCAAACATACCAACGCT<br>GGTAGCGGTAGTATTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGATGATCCTTTGATCTTT<br>TCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATATTGGTCATGAGATTATCAAAATGGATCTT<br>CACCTAGATCCTTTTAAATTACAGGTGAAGCCGGAAATCAATCTAAAGTAT |

# 3

# *De novo* fabrication of custom-sequence plasmids for the synthesis of long DNA constructs with extrahelical features

*This chapter is published as: Ramírez Montero D., Liu Z., and Dekker, N.H. (2024). De novo fabrication of custom-sequence plasmids for the synthesis of long DNA constructs with extrahelical features. Biophysical Journal 123:1, 31-41.*

## 3.1 Abstract

DNA constructs for single-molecule experiments often require specific sequences and/or extrahelical/non-canonical structures to study DNA-processing mechanisms. The precise introduction of such structures requires extensive control of the sequence of the initial DNA substrate. A commonly used substrate in the synthesis of DNA constructs is plasmid DNA. Nevertheless, the controlled introduction of specific sequences and extrahelical/non-canonical structures into plasmids often requires several rounds of cloning on pre-existing plasmids whose sequence one cannot fully control. Here, we describe a simple and efficient way to synthesize 10.1 kb plasmids *de novo* using synthetic gBlocks that provides full control of the sequence. Using these plasmids, we developed a 1.5-day protocol to assemble 10.1 kb linear DNA constructs with end and internal modifications. As a proof-of-principle, we synthesize two different DNA constructs with biotinylated ends and one or two internal 3' ssDNA flaps, characterize them using single-molecule force and fluorescence spectroscopy, and functionally validate them by showing that the eukaryotic replicative helicase Cdc45/Mcm2-7/GINS (CMG) binds the 3' ssDNA flap and translocates in the expected direction. We anticipate that our approach can be used to synthesize custom-sequence DNA constructs for a variety of force and fluorescence single-molecule spectroscopy experiments to interrogate DNA replication, DNA repair and transcription.

## 3.2 Introduction

Single-molecule studies of many DNA-processing mechanisms often require DNA constructs with specific sequence features important for the biological process being studied (e.g., origins of replication, promoter regions, and nucleosome-positioning sequences)[1–8]. Furthermore, such studies may also require extrahelical and/or non-canonical structural features at defined positions within the DNA construct[1,9–15], the controlled incorporation of which often requires specific restriction sites at defined locations within the starting DNA substrate[1,11,14,16–20]. Controlling the sequence of the starting DNA substrate used to synthesize DNA constructs for single-molecule studies is therefore of utmost importance.

Several essential DNA-processing mechanisms, such as DNA replication, transcription, and DNA compaction, are carried out by protein complexes evolved to cruise through thousands of base pairs[5,8,21–29]. Therefore, to study such processes in a biologically relevant spatial scale, single-molecule experiments often require DNA constructs of several kb in length[4,5,8,11,13,14,18,21,24–26,29–32]. In such long constructs, extrahelical and/or non-canonical DNA structures can be used as a specific binding site for the protein complex being studied[10–13], or to study the outcome of their encounter by the protein complex after having established a baseline behavior of the complex on long stretches of duplex DNA[1,14,15].

A commonly used starting substrate to synthesize DNA constructs for single-molecule experiments given its length of ~48.5 kb and commercially availability is λ phage DNA[1,11,14,18,30,33–35]. Nonetheless, engineering the sequence of the λ phage genome requires complex molecular cloning that is typically low in efficiency or requires the purification of specialized proteins[36–38]. Furthermore, the incorporation of internal modifications into the λ phage genome can only be done in dispensable genomic regions[1,11]. An alternative starting substrate to synthesize constructs for single-molecule studies is PCR-synthesized DNA, which combined with oligonucleotide-based structures has been used to successfully synthesize DNA constructs containing internal non-canonical/extrahelical structures such as hairpins[20], ssDNA flaps[39], and Holiday junctions[15]. These PCR-based methods, however, either require multiple low-efficiency ligation steps, thus having low overall yields of the desired final product[20,39], or require the incorporation of abasic sites into the final construct[15], which can affect the behavior of DNA-binding proteins [40]. An alternative starting substrate for single-molecule DNA constructs is plasmid DNA, which is commonly used given the ease with which large amounts of it can be generated by bacterial propagation[4,5,8,13,14,19,32,41–44]. Although using

standard molecular cloning techniques plasmids are easier to engineer than phage genomes, controlling the number and the position of restriction enzyme sites within the plasmid (which are required to introduce extrahelical modifications or non-canonical DNA structures) often involves several rounds of multi-day molecular cloning to either add desired restriction sites or to remove unwanted ones. Furthermore, commonly used plasmid substrates for single-molecule experiments are based on λ phage DNA[4,8,19,42–44], which can make cloning particularly difficult as these plasmids are typically very large in size and contain repetitive sequences. We currently lack a way to fully control of the sequence of the initial plasmid DNA substrate used to synthesize DNA constructs for single-molecule studies, which is particularly important for the site-specific incorporation of extrahelical and/or non-canonical structures.

One type of extrahelical structure commonly incorporated into single-molecule constructs is ssDNA flaps, which can be used to load helicases and other proteins onto the DNA as it mimics a DNA unwinding intermediate[10,12,13,31,39]. To date, ssDNA flaps are usually incorporated into DNA constructs in the form of pre-folded oligo-based Y-shaped structure ligated at the end of long (10-50 kb) DNA molecules[10,13,19,31]. Nonetheless, in correlative optical tweezer and fluorescence single-molecule experiments[45], having the binding site for e.g. a fluorescently labeled helicase at the end of the DNA is non-ideal, as the end of the DNA is close to the trapping laser, which decreases the lifetime of fluorophores[46]. Additionally, the end of the DNA is in direct contact with the protein-coated beads used for optical trapping experiments, which may 1) decrease the accessibility of the helicase binding site by steric hindrance and 2) result in non-specific adhesion of the helicase to the beads. Previous PCR-based approaches have incorporated ssDNA flaps internally in shorter (~3.7 kb) DNA constructs via a time-consuming method involving several rounds of ligations[39]. Alternatively, a nicking-based approach[16,17] has successfully introduced an internal ssDNA flap into λ phage genomic DNA[11]. This approach, however, requires complex phage genome engineering, and the incorporation of extrahelical structures is limited to specific dispensable regions in the λ phage genome[11]. We therefore lack a method to synthesize long (≥ 10 kb) linear DNA constructs for single-molecule studies with modified ends for surface attachment, as well as internal extrahelical and/or non-canonical structures that 1) provides full control of the DNA sequence and 2) takes less than a couple of days to complete.

Here we describe an effective way to assemble two different 10.1 kb plasmids *de novo* from synthetic fragments with fully custom-made sequences. This provides full control of the sequence without having to modify pre-existing plasmids; the *de*

*novo* synthesis and validation of these plasmids takes 3-4 days. With this in hand, we develop a 1.5-day long method that uses these custom-sequence plasmids to synthesize 10.1 kb linear DNA constructs with end and internal modifications, suitable for single-molecule experiments. As a proof-of-principle, we synthesize two linear 10.1 kb DNA constructs with three biotins at each end and with either one or two internal 3' ssDNA flaps, and characterize them in bulk and with force and fluorescence single-molecule spectroscopy. Finally, we functionally validate our synthesized constructs at the single-molecule level by showing that the eukaryotic replicative helicase Cdc45/Mcm2-7/GINS (CMG) can specifically bind the internal 3' ssDNA flap and then translocate in the expected 3'-to-5' direction. Although we focus here on introducing 3' flaps, our fully custom-sequence and versatile approach to the synthesis of DNA constructs for single-molecule studies can be easily generalized to constructs with several other extrahelical and/or non-canonical structures at desired sequences.

## 3.3 Results

### 3.3.1 Synthesis of fully custom-sequence 10.1 kb plasmids for single-molecule experiments

To have full control of the sequence of the plasmids used to synthesize DNA constructs for single-molecule experiments, we designed two sets of four custom-sequence 2.5 kb gBlocks. Notably, these gBlocks were designed to contain overlapping sequences of 25-28 bp and a $T_m$ of ~ 60-63 °C at both ends (**Fig. 3.1a**). These overlapping end sequences allowed for the unidirectional and scarless assembly of both sets of four gBlocks into two different 10.1 kb plasmids by Gibson Assembly[54] (**Fig. 3.1a**, **Methods**), which we named pDRM1 and pDRM2, respectively. Both these plasmids contain a bacterial origin of replication as well as an ampicillin resistance gene to allow for the bacterial propagation of the newly synthesized plasmids. The assembly reactions were directly transformed into *E. coli,* clones were grown in selection medium, and each plasmid was isolated giving yields of 300-900 ng DNA/mL of bacterial culture (**Methods**). The most important feature of both custom-sequence plasmids is the number and location of specific restriction enzyme sites which are used to introduce specific features into the single-molecule constructs (**Fig. 3.1b, d**): pDRM1 contains a unique XhoI restriction site to allow for linearization of the plasmid, 5 tandem Nb.BbvCI nicking sites separated by 16 bp spacers for the incorporation of one 3' ssDNA flap, and two BsaI sites separated by 100 bp for the optional cloning of additional DNA sequences

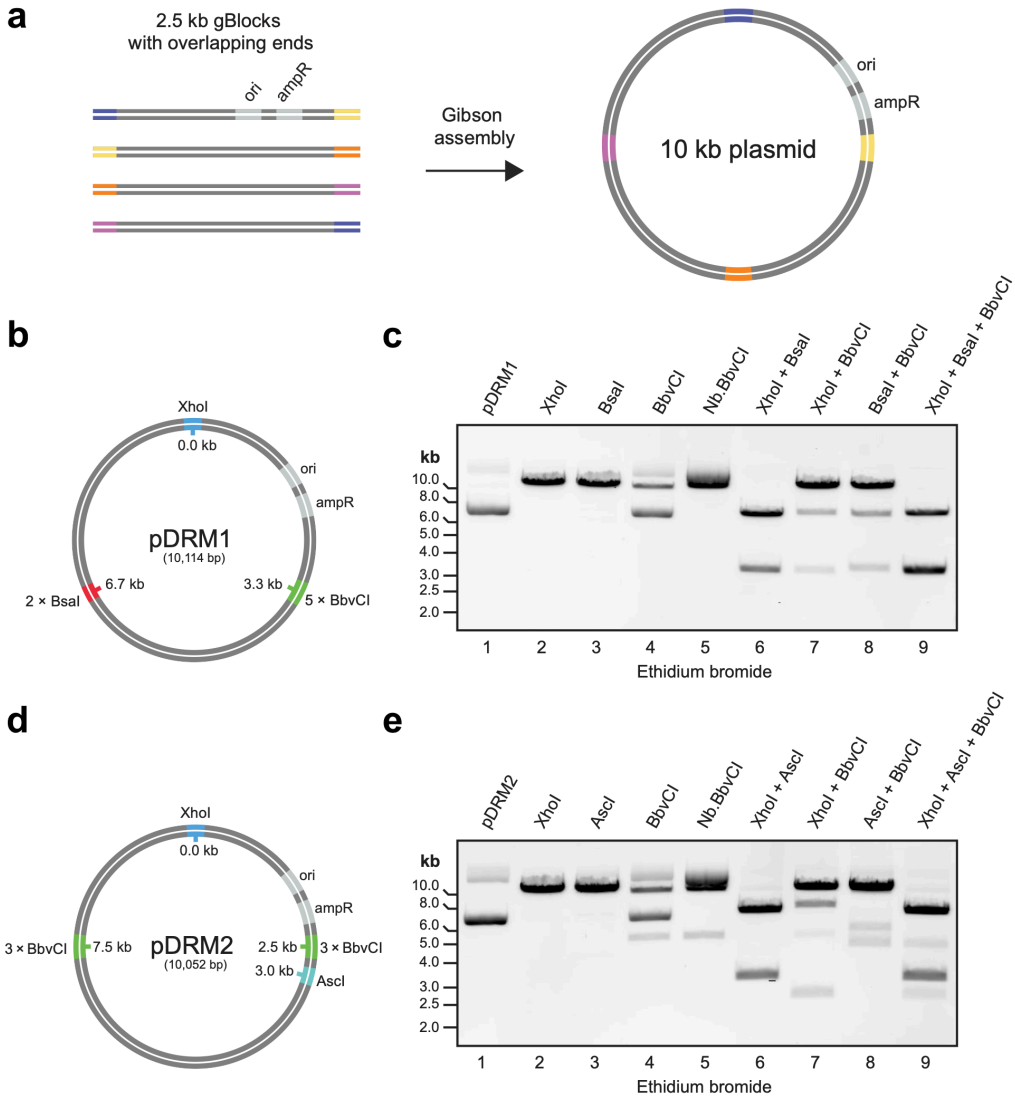



*Figure 3.1* Synthesis of fully custom sequence plasmids from gBlocks. **a** Illustration of the plasmid synthesis procedure. Four fully custom sequence 2.5 kb gBlocks were designed to have pairwise overlapping sequences at their ends (represented in purple, yellow, orange, and pink) to ensure their ordered and scarless assembly into a 10.1 kb plasmid via Gibson Assembly. To ensure bacterial propagation, one of the gBlocks must contain a bacterial origin of replication and an ampicillin resistance gene. **b** Diagram of plasmid pDRM1 synthesized as described in panel **a** highlighting the position of restriction enzyme sites to be used for the synthesis of DNA constructs for single-molecule studies. **c** Single, double, and triple test restriction digestions of plasmid pDRM1 corroborate the correct order of assembly of the gBlocks used to synthesize it. **d** Diagram of plasmid pDRM2 synthesized as described in panel **a** highlighting the position of restriction enzyme sites to be used for the synthesis of DNA constructs for single-molecule studies. **e** Single, double, and triple test restriction digestions of plasmid pDRM2 corroborate the correct order of assembly of the gBlocks used to synthesize it.

into the plasmid via Golden Gate assembly[20] (**Fig. 1b**). pDRM2 also contains a unique XhoI restriction site to allow for linearization of the plasmid, but it contains two sets of 3 tandem Nb.BbvCI nicking sites separated by 16 bp spacers for the incorporation of two 3' ssDNA flaps[1,16,17] (**Fig. 1d**). To confirm that both sets of gBlocks were assembled scarlessly and in the right order, we sequenced the overlapping regions of the gBlocks as well as the restriction enzyme sites in both plasmids. Furthermore, we carried out test digestions of the plasmids with the restriction enzymes described above and saw either the expected band patterns or bands that can be explained by the incomplete cleavage by some of the enzymes (**Fig. 1c, e**). A full assignment of all the bands can be found in **Fig. S3.1**. Taken together, our data shows that the two sets of designed gBlocks were successfully and scarlessly assembled in the right order, generating two different 10.1 kb fully custom-sequence plasmids. The whole *de novo* assembly and plasmid validation procedure takes 3-4 days.

### 3.3.2 Synthesis of linear 10.1 kb DNA constructs containing an internal ssDNA flap for single-molecule experiments

After confirming the successful assembly of the desired plasmids, we developed a fast (1.5-day long) and efficient method to use these plasmids to synthesize two different 10.1 kb linear DNA constructs containing three biotin moieties at each end for surface attachment, and either one (in the case of pDRM1) or two (in the case of pDRM2) internally located ssDNA flaps for helicase binding (**Methods**, **Fig. 3.2**). This method, which we describe step-by-step in **Fig. 3.2a, c** is as follows: we first linearized plasmids pDRM1 and pDRM2 with the restriction enzyme XhoI (**Fig. 3.2b**, lanes 1-2; **Fig. 3.2d**, lanes 1-2). Digesting the plasmids with XhoI generates 4 nt 5'overhangs at both ends of the linearized DNA. These overhangs were then biotinylated via a blunting reaction with Klenow fragment DNA polymerase in the presence of biotinylated nucleotides (**Fig. 3.2b**, lane 3; **Fig. 3.2d**, lane 3), which resulted in the incorporation of three biotin moieties at each end of the DNA. After biotinylation, excess nucleotides were removed by chromatography (**Methods**) and the biotinylated DNA constructs were nicked with the nicking enzyme Nb.BbvCI. Following the nicking reaction, we detected some unexpected bands in addition to the expected 10.1 kb band (**Fig. 3.2b**, lane 4; **Fig. 3.2d**, lane 4). We hypothesized that these additional bands corresponded to products of star endonuclease activity of the nicking enzyme Nb.BbvCI, as their sizes matched what we would expect from full endonuclease activity of Nb.BbvCI: two bands of ~3.3 kb and ~6.8 kb in the case of the construct derived from pDRM1, and three bands of ~2.5 kb, ~5.0 kb and ~7.5 kb in the case of the construct derived from pDRM2. To test this hypothesis, we digested both biotinylated DNA constructs

with full BbvCI endonuclease and found the same band patterns that we found when we nicked the biotinylated DNA with Nb.BbvCI (**Fig. 3.2b**, lanes 4-5; **Fig. 3.2d**, lanes 4-5). Following nicking, we heated up the DNA substrates to remove the 16-bp ssDNA spacers between the nicking sites, and slowly cooled down the reaction (**Methods**) in the presence of an excess of 3'Cy3-labeled ssDNA oligo to replace the nicked spacers[11,16,17]; these Cy3-labeled oligos are fully complementary to the nicked regions but contain an additional 40 nt 3' poly-dT to create ssDNA flaps. Of note, to confirm the incorporation of the fluorescent oligos, we scanned the DNA gel with a green laser and saw fluorescent bands at the expected locations

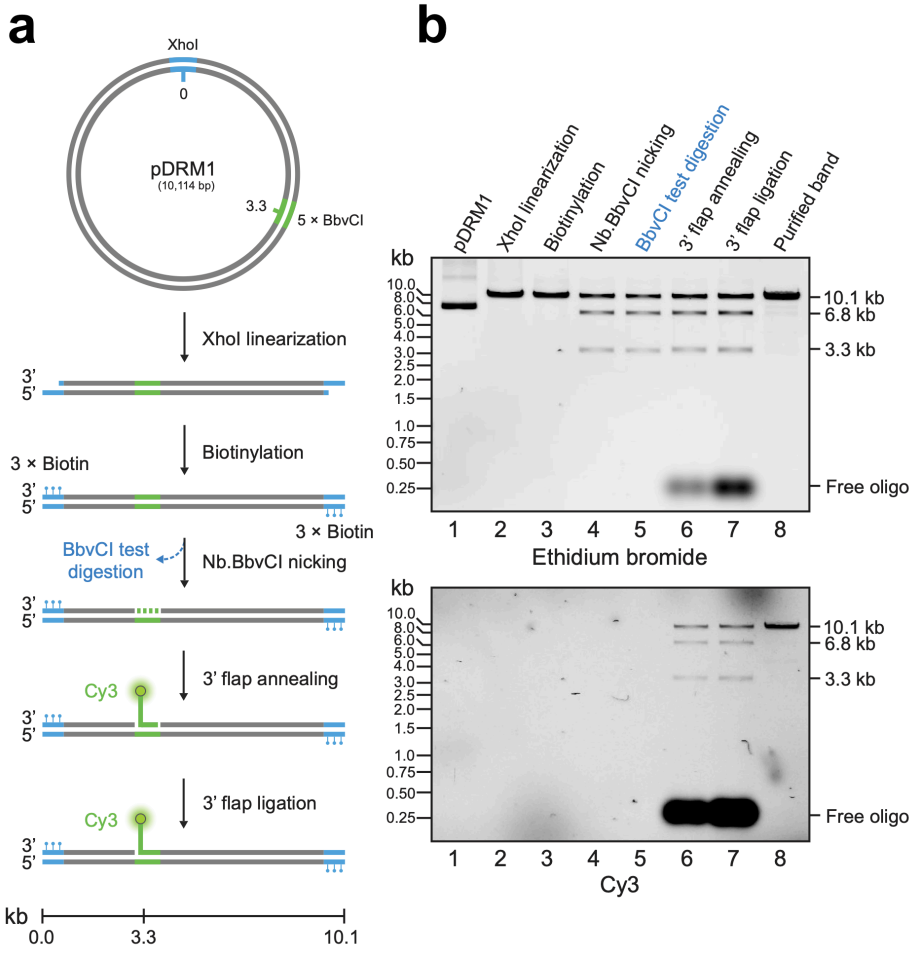



*Figure 3.2* Synthesis of linear 10.1 kb DNA constructs containing an internal ssDNA flap for single-molecule experiments. **a** Step-by-step pictorial description of the synthesis of construct C-DRM1 using pDRM1 as a starting substrate. **b** Agarose gel monitoring each of the steps described in panel **a** showing either total DNA stained with ethidium bromide (top) or Cy3-labeled DNA illuminated with a green laser (bottom). *(Figure continues on the next page)*.

(**Fig. 3.2b**, lane 6; **Fig. 3.2d**, lane 6). After the annealing reaction, we ligated the 5' end of the fluorescent ssDNA flap(s) overnight with T4 ligase (**Fig. 3.2b**, lane 7; **Fig. 3.2d**, lane 7), and purified the 10.1 kb bands (**Fig. 3.2b**, lane 8; **Fig. 3.2d**, lane 8). Finally, in the case of the DNA construct derived from pDRM2 (**Fig. 3.2c**), we further confirmed the incorporation of both fluorescent ssDNA flaps by digesting the purified 10.1 kb construct with AscI. The AscI restriction site was engineered to be between the location of both fluorescent forks but off-centered, thus yielding

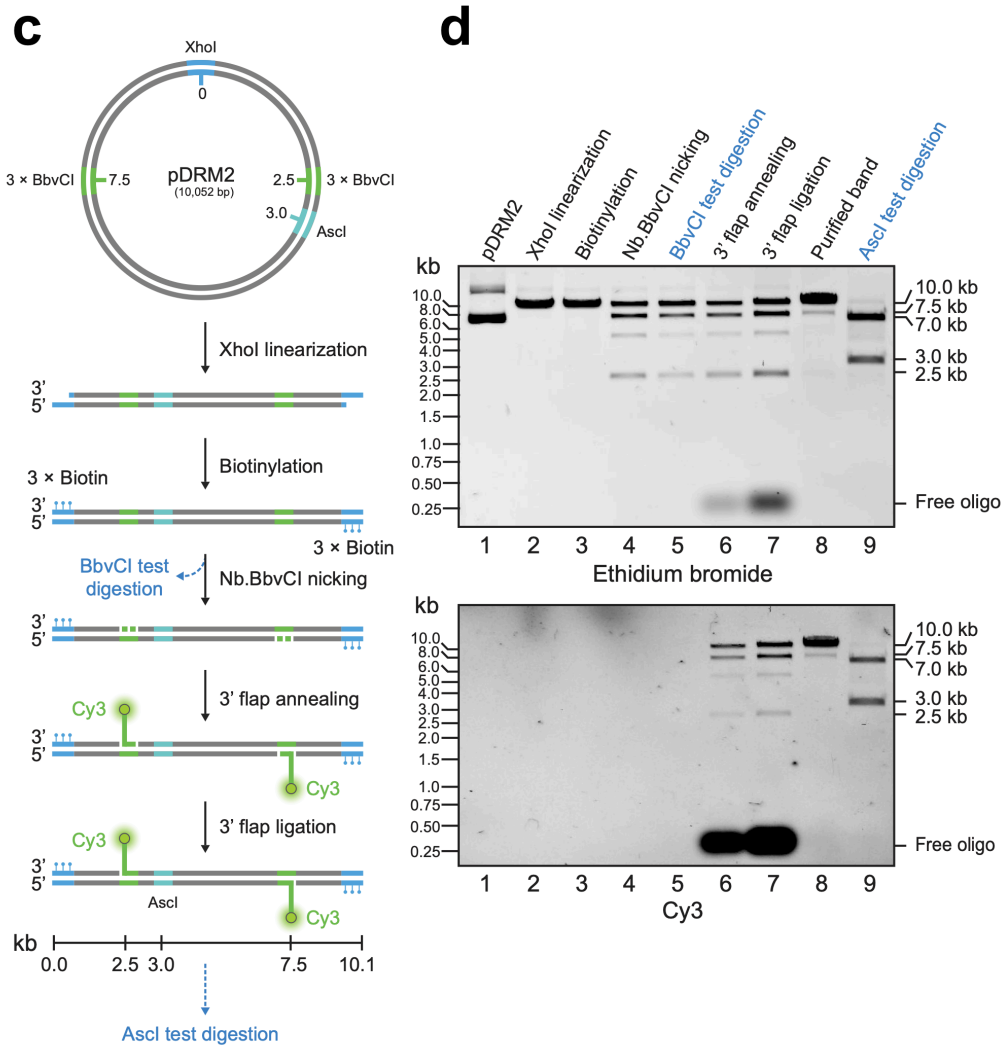



*Figure 3.2 (continued).* **c** Step-by-step pictorial description of the synthesis of construct C-DRM2 using pDRM2 as a starting substrate. **d** Agarose gel monitoring each of the steps described in panel **c** showing either total DNA stained with ethidium bromide (top) or Cy3-labeled DNA illuminated with a green laser (bottom). Steps colored in blue are control test restriction digestions that are not part of the synthesis procedure.

two bands of different sizes (~3 kb and ~7.1 kb). As seen in **Fig. 3.2d**, lane 9, both bands contained the fluorescent ssDNA flap, showing that both forks were successfully incorporated into the construct derived from pDRM2. After the final purification, we consistently had yields of ~10-15% relative the initial amount of plasmid DNA, which are significantly higher than typical multi-ligation-based protocols. We henceforth call the synthesized constructs C-DRM1 and C-DRM2, respectively.

### 3.3.3 Single-molecule characterization of DNA C-DRM1 and C-DRM2

We then proceeded to characterize both synthesized DNA constructs at the single-molecule level. For this, we used a combination of dual optical trapping and confocal scanning microscopy[45] to monitor the presence of fluorescent ssDNA flap diffraction-limited spots on single DNA molecules held in an optical trap (**Fig. 3.3a, h**; **Methods**). First, we monitored the efficiency of fluorescent ssDNA flap incorporation by counting the number of fluorescent spots in each DNA molecule. We found that in the case of C-DRM1, ~82% of the trapped DNA molecules contained 1 fluorescent spot (**Fig. 3.3b**), and in the case of construct C-DRM2, ~79% of the trapped DNA molecules contained 2 fluorescent spots (**Fig. 3.3i**). These fork incorporation efficiencies likely reflect the actual ssDNA flap incorporation efficiency, as our measured labeling efficiencies of the flap oligos were ~100% (**Methods**). Of note, we counted the number of photobleaching steps within each diffraction-limited spot and found that all the detected spots photobleached in one step (**Fig. S3.2a, b**), confirming that we incorporated one ssDNA flap at each nicking location (**Fig. 3.2a, c**). We next proceeded to analyze the position of the fluorescent ssDNA flaps along the DNA. Notably, as we cannot differentiate between the two possible orientations that the DNA can have in the optical tweezers, we display the position of the fluorescent spots as distances from the DNA center[4,8]. This analysis confirmed that the forks were incorporated at the expected locations on the DNA (**Fig. 3.3c, j**). We also obtained force-extension curves for both DNA constructs[50] and fitted the obtained curves with an extensible worm-like chain model[53] (**Fig. 3.3d, k**). From this model, we obtained three parameters: contour length, persistence length, and stretch modulus[53]. The average contour lengths obtained for both constructs agreed with the expected values within experimental error (**Fig. 3.3e, l**), and the average persistence length and stretch modulus for each of the two DNA constructs in the buffer conditions employed agreed with previously reported values[55,56] (**Fig. 3.3f, g, m, n**). This single-molecule characterization confirmed that we synthesized the expected DNA constructs in an efficient manner.

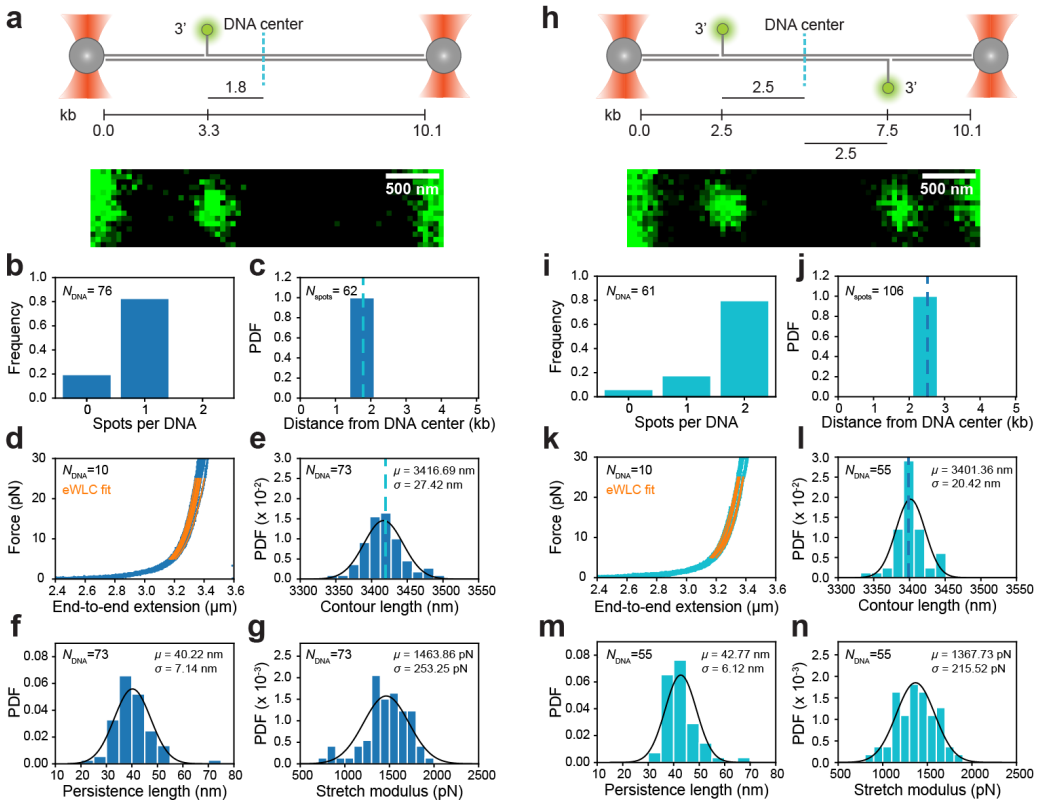



*Figure 3.3* Single-molecule characterization of DNA constructs C-DRM1 and C-DRM2. **a** Diagram of construct C-DRM1 (top) together with an example confocal scan of a C-DRM1 molecule held in place in an optical trap (bottom). **b** Distribution of numbers of Cy3 diffraction-limited spots per molecule of construct C-DRM1. **c** Probability density function of positions from the DNA center of Cy3 diffraction-limited spots in C-DRM1. Dotted cyan line indicates the expected position from the DNA center of the fluorescent ssDNA flap. **d** Example force-extension curves of construct C-DRM1 (blue) together with the fitted extensible worm-like chain (eWLC) model plotted in the force range used for fitting (orange). **e-g.** Probability density function of **e** contour lengths, **f** persistence lengths, and **g** stretch moduli obtained from the extensible worm-like chain model fits of force-extension curves of C-DRM1; black lines show a Gaussian fit of the data, and μ and σ are the mean and the standard deviation of the data, respectively. Dotted cyan line in panel **e** shows the expected contour length of C-DRM1. **h** Diagram of construct C-DRM2 (top) together with an example confocal scan of a C-DRM2 molecule held in place in an optical trap (bottom). **i** Distribution of numbers of Cy3 diffraction-limited spots per molecule of construct C-DRM2. **j** Probability density function of positions from the DNA center of Cy3 diffraction-limited spots in C-DRM2. Dotted blue line indicates the expected position from the DNA center of the fluorescent ssDNA flaps. **k** Example force-extension curves of construct C-DRM2 (cyan) together with the fitted extensible worm-like chain (eWLC) model plotted in the force range used for fitting (orange). **l-n.** Probability density function of **l** contour lengths, **m** persistence lengths, and **n** stretch moduli obtained from the extensible worm-like chain model fits of force-extension curves of C-DRM2; black lines show a Gaussian fit of the data, and μ and σ are the mean and the standard deviation of the data, respectively. Dotted blue line in panel **l** shows the expected contour length of C-DRM2.

### 3.3.4 Demonstration of ssDNA flap binding and unwinding by a helicase

Having validated the DNA synthesis method, we sought to assess whether the incorporated 3' ssDNA flap can be bound by the 3'-to-5' eukaryotic replicative helicase CMG[13,23,31,57], and whether the helicase could then translocate in the expected direction. To this end, we used pDRM1 as a template and cloned into it a ~ 6.5 kb duplex region (**Fig. 3.1b**, **Methods**), yielding a longer version of plasmid pDRM1 that we named pZL7. Following the same method described above, we used plasmid pZL7 to synthesize DNA construct C-ZL7 (**Fig. 3.4a**). Of note, to prevent any artifacts caused by the presence of a 3' fluorophore in the binding site of the helicase, C-ZL7 contains a non-fluorescently labeled ssDNA flap. We first characterized C-ZL7 by force spectroscopy[50,53] (**Fig. S3.3a**) in the same way that we did for C-DRM1 and C-DRM2, and obtained an average value

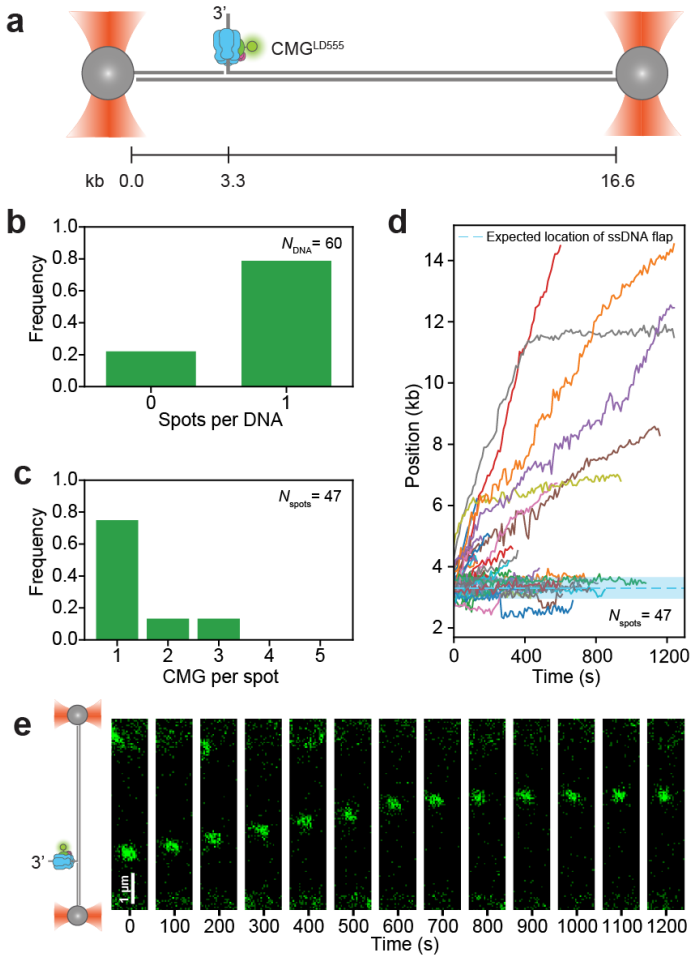



*Figure 3.4* Functional validation of the introduced 3' ssDNA flaps. **a** Diagram of construct C-ZL7 held in place in an optical trap and containing a fluorescently labeled CMG helicase bound to the 3' ssDNA flap. **b** Distribution of numbers of fluorescently labeled CMG diffraction-limited spots per molecule of construct C-ZL1. **c** Distribution of numbers of fluorescently labeled CMG complexes within each diffraction-limited spot. **d** Position along DNA construct C-ZL1 versus time of fluorescently labeled CMG spots. For clarity, the traces have been compensated for DNA orientation, i.e., they have been rotated so that the initial position of all traces are within the same half of the DNA. Note that this rotation does not automatically place the start of the trace at the expected location of the ssDNA flap. **e** Example kymograph of a fluorescently labeled CMG helicase initially bound at the expected location of the ssDNA flap which translocates unidirectionally in a 3'-to-5' direction upon the addition of ATP.

for the contour length that agreed with the expected value within experimental error (**Fig. S3.3b**), and average persistence length and stretch modulus values (**Fig. S3.3c, d**) that agree with previously reported values in the buffer conditions employed[55,56]. We then incubated construct C-ZL7 with fluorescently labeled $CMG^{Cdc45\text{-}LD555}$. We first incubated $CMG^{Cdc45\text{-}LD555}$ with C-ZL7 in bulk in the presence of the slowly hydrolyzable ATP analog ATPγS to allow $CMG^{Cdc45\text{-}LD555}$ to bind to the 3' ssDNA flap in C-ZL7 without unwinding it[12,31] (**Methods**). We then diluted the reaction and flowed it into our optical tweezers. Notably, our optical tweezers are equipped with a microfluidic flow cell that allows us to buffer exchange the trapped DNA:protein complex *in situ*[4,8,45]. Using this flow cell, we moved the trapped DNA:$CMG^{Cdc45\text{-}LD555}$ complexes into buffer solution containing ATP and immediately started to image CMG motion. We found that 78.3% of the trapped DNA contained $CMG^{Cdc45\text{-}LD555}$ diffraction-limited spots (**Fig. 3.4b**), which is consistent with our measured ssDNA flap incorporation efficiency in pDRM1 (**Fig. 3.3b**). Of these spots, 74.5% contained one CMG complex (**Fig. 3.4c**). As can be seen in **Fig. 3.4d**, the vast majority of the CMG spots were initially located at the expected position of the ssDNA flap. Furthermore, as seen in **Fig. 3.4d** and **e**, we detected motion of $CMG^{Cdc45\text{-}LD555}$ that started at the expected 3' ssDNA flap location and proceeded in the expected 3'-to-5'direction, thereby showing that our method to synthesize DNA constructs results in a 3' ssDNA flap that can be bound and translocated from by a helicase. The observed long-range translocation also shows that the ssDNA flap is successfully ligated, as CMG has been previously reported to dissociate from DNA at nicks[32].

## 3.4 Discussion

We describe a method to synthesize fully custom-sequence DNA constructs for single-molecule studies. For this, we use Gibson Assembly[54] to assemble fully custom-sequence gBlocks into two different 10.1 kb plasmids (**Fig. 3.1**) using a simple and fast *in vitro* protocol that takes 3-4 days. This approach allowed us to generate large amounts of plasmid by propagation in standard *E. coli* strains (**Methods**). In this work, we assembled two plasmids of 10.1 kb in length, as this length sufficed for the downstream production of linear DNA molecules suited to single-molecule experiments. However, Gibson Assembly has also been used to efficiently clone genomic fragments of up to 100 kb in length into plasmids using overlapping sequences of similar lengths to the ones used in this study[58], suggesting that the *de novo* synthesis of much larger plasmids may be easily achievable. Further studies to determine the length limit in these assembly reactions will be of

interest. The use of Gibson Assembly to assemble fully custom sequence plasmids from gBlocks alone, as done in this study, provides a novel methodology in the field of single-molecule biophysics. We believe that the protocol of plasmid assembly we describe will allow other scientists working on a wide variety of DNA-processing mechanisms to design more biologically relevant DNA constructs, as well as to investigate the role of DNA sequence on many of these processes with unprecedented control.

Using our *de novo* assembled plasmids, we developed a fast (1.5-day long) and efficient approach to synthesize two different linear constructs containing end and internal modifications for single-molecule force and fluorescence spectroscopy experiments (**Fig. 3.2**). As a proof-of-principle, we incorporated either one or two internal ssDNA flaps into our constructs and characterized them in bulk (**Fig. 3.2**) and at the single-molecule level (**Fig. 3.3**). Finally, we functionally validated our approach at the single-molecule level by showing that the incorporated ssDNA flaps can be bound by the eukaryotic replicative helicase CMG, which then translocates in the expected direction (**Fig. 3.4**). Notably, the pre-formed CMG that we employed in this study has been previously studied in correlative optical tweezers and fluorescence experiments after loading it onto λ phage DNA in a non-specific manner after generating ssDNA regions in the otherwise dsDNA construct by the use of high force[26]; this approach does not only lack control of the loading site of the helicase, but also lacks control of the number of helicases that can be loaded at the long stretches of ssDNA generated, and could also have force-related artefacts. The new approach that we developed in this study will not only allow us and others to site-specifically bind pre-formed CMG to the DNA without the use of high force, but also allow us and others to study the motion of CMG along a DNA substrate with a fully controlled sequence.

Although in this proof-of-concept study we introduced 3'ssDNA flaps, the directionality of the ssDNA flaps can easily be reversed without modifying the DNA sequence through the use of the complementary nicking enzyme to the one used in this study[16,17]. Additionally, we expect/anticipate that the same protocol can be used to easily incorporate other DNA modifications relevant to the study of DNA repair and DNA secondary structure processing, such as ssDNA/dsDNA junctions, hairpins and G-quadruplexes[16,18,59–61]. Furthermore, our synthesis strategy can easily be used to incorporate non-canonical nucleotides in the form of modified oligos[1,16,17], allowing for the introduction of a wide range of chemical moieties that in turn facilitate the site-specific covalent and non-covalent attachment of other molecules of interest (e.g., protein crosslinks, streptavidin, fluorophores etc.)[1,12,14,16,17,62] to the DNA, all within a fully controlled sequence context. Finally,

even without the introduction of extrahelical/non-canonical structures, we anticipate that our method will benefit the field by allowing full control of the sequence background used to study a wide range of DNA:protein interactions.

## 3.5 Materials and Methods

### 3.5.1 Plasmid synthesis

#### 3.5.1.1 gBlock design

All gBlocks used in this study were purchased from Integrated DNA Tecnhologies (**Supplementary table 3.1**). gBlocks were designed to contain overlapping sequences of 25-28 bp and a $T_m$ of ~ 60-63 °C at both ends for their assembly into custom-sequence plasmids. Notably, at least one of the gBlocks used to assemble a plasmid *de novo* must contain a bacterial origin of replication for bacterial propagation of the plasmid, as well as an antibiotic resistance marker.

#### 3.5.1.2 *De novo* assembly of pDRM1 and pDRM2 from custom-made gBlocks

pDRM1 and pDRM2 were assembled by incubating 0.016 pmol of gBlocks 1, 2, 3, and 4 (for pDRM1) or gBlocks 3, 5, 6, and 7 (for pDRM2) (**Supplementary table 3.1**) in 20 µl of 1X NEBuilder® HiFi DNA Assembly Master Mix (NEB # E2621S) and incubated at 50 °C for 120 min. The assembly reactions were then directly transformed into NEB® 5-alpha Competent *E. coli* (High Efficiency) cells (NEB # 02987). Transformed cells were plated on Luria-Bertani medium (LB) agar plates supplemented with 50 µg/mL ampicillin and incubated at 37 °C overnight. Individual clones were selected, grown overnight at 30 °C in LB medium supplemented with 50 µg/mL ampicillin with 180 rpm shaking, and stored at -80 °C in LB medium supplemented with 14% glycerol. A swab of the glycerol stocks was inoculated into 200 mL of LB medium supplemented with 50 µg/mL ampicillin and grown overnight at 30 °C. Plasmids were then purified using a Macherey-Nagel™ NucleoBond™ Xtra Midi Plus (Bioké # 740412.50). The overlapping ends of the gBlocks used to assemble pDRM1 as well as key regions in pDRM1 were sequenced with oligos DRM_140, DRM_141, DRM_142, DRM_143, DRM_144, and DRM_145. The overlapping ends of the gBlocks used to assemble pDRM2 as well as key regions in pDRM2 were sequenced with oligos DRM_140, DRM_142, DRM_143,

DRM_144, DRM_177, DRM_180, DRM_181, and DRM_182 (**Supplementary table 3.2**).

#### 3.5.1.3 Test digestions of pDRM1 and pDRM2

400 ng of either pDRM1 or pDRM2 were digested for 2 h at 37 °C in a final volume of 25 µl of 1 X CutSmart™ buffer (NEB # B7204) with 0.8 µl of one, two or three of the following enzymes: XhoI (NEB # R0146S), BsaI-HF®v2 (NEB # R3733L), BbvCI (NEB # R0601S), Nb.BbvCI (NEB # R0631L) and AscI (NEB # R0558S). The reaction mixtures were then incubated at 80 °C for 20 min to inactivate the restriction enzymes, and then run on a 0.8% agarose gel stained with ethidium bromide.

#### 3.5.1.4 Synthesis of pZL7

pDRM1 was linearized by PCR using primers Vector.FOR and Vector.REV, and a 393 bp fragment was inserted using NEBuilder® HiFi DNA Assembly Master Mix (NEB # E2621S) to generate plasmid pER1. Then, an NdeI, AscI, and NcoI restriction enzyme sites were inserted into pER1 by amplifying the plasmid by PCR using primers ZL3 and ZL4, followed by PCR amplification of the resulting plasmid using primers ZL7 and ZL8, and PCR amplification of the resulting plasmid using primers ZL13 and ZL14, to generated plasmid pZL1. A 3.1 kb fragment of pZL1 was duplicated twice into the same plasmid using a previously reported protocol [47]. Briefly, pZL1 was digested with NcoI-HF (NEB # R3193S) to construct the plasmid backbone for fragment insertion. Then, a 3.1 kb insert was generated by the digestion of pZL1with NdeI (NEB # R0111S) and AscI (NEB # R0558S). The plasmid backbone was mixed with 1.2-fold molar excess of the 3.1 kb insert in 1X NEBuilder® HiFi DNA Assembly Master Mix (NEB # E2621S). The reaction mixture was incubated at 50 °C for 15 min, and subsequently transformed into NEB® 5-alpha Competent *E. coli* (High Efficiency) cells (NEB # 02987) to propagate the assembled plasmid product, named pZL3. The aforementioned digestion and insertion protocol was repeated using pZL3 as the plasmid backbone and the 3.1 kb fragment from pZL1, to generate pZL7.

### 3.5.2 Synthesis of single-molecule constructs from pDRM1 and pDRM2

#### 3.5.2.1 Plasmid linearization

10 µg of either pDRM1, pDRM2 or pZL5 were linearized overnight with 50 units of XhoI (NEB # R0146S) in 1X CutSmart™ buffer (NEB # B7204) in a final volume of

50 µl. The reaction mixture was then incubated at 65 °C for 20 min to inactivate XhoI.

#### 3.5.2.2 Biotinylation of DNA ends

50 µl XhoI linearization reactions were supplemented with 33 µM dGTP (Invitrogen # 10218014), 33 µM Biotin-14-dATP (Invitrogen # 19524-016), 33 µM Biotin-14-dCTP (Invitrogen # 19518-018), 33 µM Biotin-16-dUTP (Jena Bioscience # NU-803-BIO16), 15 units of Klenow Fragment (3'→5' exo-) (NEB # M0212L), 4 µl of 10 X NEBuffer™ 2 (NEB # B7002S) and water to a final volume of 90 µl, and incubated at 37 °C for 30 min. EDTA was then added at a final concentration of 10 mM to terminate the reaction, and the reaction mixture was incubated at 75 °C for 20 min to inactivate the Klenow fragment. Unincorporated nucleotides were removed with a Microspin™ S-400 HR spin columns equilibrated in Tris/EDTA (TE) buffer (GE Healthcare # GE27-5140-01) as follows: columns were vortexed and centrifuged for 1 min at 0.8 g, then 100 µl of biotinylation reaction was then loaded onto the packed resin, centrifuged for 2 additional min at 0.8 g, and the flowthrough collected in a new tube.

#### 3.5.2.3 Nicking of biotinylated DNA

Biotinylated DNA (~ 90 µl) was supplemented with 11 µl of 10X CutSmart™ buffer (NEB # B7204), 50 units of nicking enzyme Nb.BbvCI (NEB # R0631L) and water to a final volume of 110 µl, and incubated at 37 °C for 5 h. The reaction mixture was then incubated at 80 °C for 20 min to inactivate Nb.BbvCI.

#### 3.5.2.4 Fluorescent flap oligo annealing

Flap oligo DRM_186 (for construct C-DRM1), or DRM_159 (for construct C-DRM2), or ZL_15 (for construct C-ZL7) was added in 100-fold molar excess to the nicked DNA. The mixture was then heated to 80 °C for 10 min and cooled down to 20 °C at a rate of −1 °C/min in a thermocycler.

#### *3.5.2.5* Ligation and purification

The annealed oligo was ligated overnight at 16 °C in 1X CutSmart™ buffer (NEB # B7204) supplemented with 1.5 mM ATP (Thermo Scientific # R0441) and 1600 units of T4 DNA ligase (NEB # M0202L) in a final volume of 120 µl. The ligation reaction was then run on a 0.8% agarose gel pre-stained with SYBR™ Safe

(Invitrogen # S33102) and the final product band was excised and purified using Promega's Wizard® SV Gel and PCR Clean-Up System (Promega # A9282).

#### 3.5.2.6 Oligo labelling efficiency measurements

Labelling efficiencies of flap oligos DRM_186 and DRM_159 were measured to be 103% ± 5% and 105% ± 5%, respectively, by measuring the absorption at 550 nm.

### 3.5.3 Protein purification and labeling

**$CMG^{Cdc45\text{-}S6}$ purification:** ySMG14 S6-iFLAG-CDC45 CBP-MCM3 strain was a gift from the laboratory of John Diffley (Francis Crick institute, U.K.). CMG was then expressed and purified as previously reported [26].

**$CMG^{Cdc45\text{-}S6}$ labeling:** $CMG^{Cdc45\text{-}S6}$ protein was fluorescently labeled by adding equimolar Sfp transferase and 4.5-fold molar excess of LD555-CoA fluorophore. The reaction mixture was incubated at 25 °C for 1 h, followed by purification using a Superose 6 3.2/300 column (Cytiva) equilibrated in CMG buffer (25 mM HEPES-KOH pH 7.6, 2 mM magnesium acetate, 0.02% Tween-20, 100 mM NaCl, and 10% glycerol).

### 3.5.4 Single-molecule imaging and force-spectroscopy

#### *3.5.4.1* Single-molecule fluorescence imaging and force spectroscopy

Single-molecule experiments were conducted on an instrument on that combines optical tweezers and confocal microscopy (C-Trap, LUMICKS). This instrument is equipped with a microfluidic flow cell with four inlets and one outlet. Three of these channels are injected from the left and used for bead trapping, DNA/DNA:protein complex trapping, and imaging, respectively. The other channel is used as a buffer exchange location. Prior to each experiment, the microfluidic flow cell and the tubing of the instrument were passivated for at least 30 min with 1 mg/mL bovine serum albumin (BSA, NEB # B9000S) followed by 0.5% Pluronic® F-127 (Sigma # P2443).

The channels contained the following solutions:

***Channel 1:*** 1.76 µm diameter streptavidin-coated polystyrene beads (Spherotech) diluted 1:1000 in PBS.

***Channel 2:*** 5 pM DNA in PBS (in the cases of C-DRM1 and C-DRM2) or 10 pM DNA prebound by CMG in CMG imaging buffer (250 mM potassium glutamate, 25 mM HEPES-KOH pH 7.6, 10 mM magnesium acetate, 0.02% NP40, 1 mM DTT, 0.1

mg/mL BSA, 2 mM 1,3,5,7 cyclooctatetraene, 2 mM 4-nitrobenzyl alcohol, 2 mM Trolox, 0.2 mM ATPγS, and 10% glycerol).

***Channel 3:*** PBS supplemented with 2 mM Trolox (in the cases of C-DRM1 and C-DRM2) or CMG imaging buffer in the case of CMG.

***Channel 4:*** CMG unwinding buffer (250 mM potassium glutamate, 25 mM HEPES, 10 mM magnesium acetate, 0.02% NP40, 1 mM DTT, 0.1 mg/mL BSA, 2 mM 1,3,5,7 cyclooctatetraene, 2 mM 4-nitrobenzyl alcohol, 2 mM Trolox, 5 mM ATP, and 10% glycerol).

A detailed description of the single-molecule data acquisition has been recently published by our group [48]. At the beginning of each experiment, the trapping laser power was adjusted to achieve a stiffness of 0.3 pN/nm in both traps [26,49]. Then, two beads were trapped in channel 1. Subsequently, individual DNA molecules/DNA:protein complexes were trapped between two beads in channel 2, and we confirmed the tethering of individual DNA molecules by analyzing the force-extension curve [50]. The DNA was then transferred to channel 3, the distance between both beads fixed to achieve a tension of 2 pN, and the was DNA illuminated without flow with a 561 nm laser at a power of 7.8 μW (in the case of C-DRM1 and C-DRM2) or 2 μW (in the case of CMG experiments) as measured at the sample plane. Fluorescence signal was detected using a single-photon counting detector. 2D confocal scans were obtained over an area of 80 × 20 pixels, which covered the entire DNA and the edges of both beads [48]. Pixel size was set to 50 × 50 nm, illumination time per pixel was set to 0.2 ms, and the frame rate was set to 600 ms. The microscope outputs HDF5 files storing the confocal scan data, force data, and bead location data monitored during the scan, as well as the force-extension curves. In the case of DNA:CMG complex experiments, the confocal scan sizes were 124× 24 pixels and the pixel size was 50 × 50 nm. Illumination time per pixel was set to 0.2 ms, and the frame rate was set to 10 s.

#### 3.5.4.2 Data acquisition automation

All single-molecule data were acquired in an automated manner. We used the Lumicks Harbor experiment automation scripts (https://harbor.lumicks.com/scripts) as a starting point (specifically, Joep Vanlier's automation script for trapping beads, trapping DNA, and making force-extension curves) and added the functionality to acquire confocal images after successfully trapping DNA. A detailed description of the automation code steps has been recently published by our group[48].

## 3.5.5 Data analysis

### 3.5.5.1 Software and code

We used Python 3.8 with several libraries for image processing. We used the Laplacian of Gaussian detector from Python's "scipy" for spot detection. We used the Linear Assignment Problem method[51] and the "scipy" solver "linear_sum_assignment" to track spots[48]. Bleaching trace analysis was done with the "ruptures" library[48]. The exact python libraries used and their versions are as follows: numpy==1.19.5; matplotlib==3.2.2; lumicks-pylake==0.7.1; streamlit==0.74.1; scipy==1.6.1; scikit-image==0.16.2; scikit-learn==0.23.1; pyyaml==5.3.1; pandas==1.0.5; pillow==7.2.0; tifffile==2021.1.11; jupyterlab==2.1.5; notebook==6.0.3; ruptures==1.1.6; pykalman==0.9.5.

### 3.5.5.2 Overview of data analysis

After acquiring confocal scans, raw image data was processed to generate a table containing the spot detections in each frame. Spot detections are then connected between frames to produce traces that contain location and intensity information over time[48]. A detailed description of the data analysis steps has been recently published by our group[48].

### 3.5.5.3 Spot detection and tracking

To detect fluorescence spots, we employ the scikit-image implementation of a Laplacian of Gaussian (LoG) blob detector[48,52]. We set the detection radius $r_{LoG}$ to 5 pixels (250nm); the LoG sigma parameter is given by $\sigma_{LoG} = r_{LoG}/\sqrt{2}$. We set the detection threshold to 0.3 ADU/pixel (for experiments with C-DRM1 and C-DRM2) or 0.5 ADU/pixel (for experiments with with C-ZL7 and CMG$^{LD555}$)[4,8]. Spots are then localized with subpixel resolution by fitting a Gaussian profile on spot intensity projections in both x- and y-directions. To track spots frame-by-frame, we use our own implementation of the Linear Assignment Problem (LAP) framework[51] using a maximum spot linking distance of 6 pixels (300 nm) and a maximum frame gap of 3 frames (for experiments with C-DRM1 and C-DRM2) or a maximum spot linking distance of 10 pixels (500 nm) and a maximum frame gap of 1 frame (for experiments with C-ZL7 and CMG$^{LD555}$).

#### 3.5.5.4 Determination of number of fluorophores per diffraction-limited spot

To determine the number of fluorophores contained within each diffraction-limited spot, we count the number of photobleaching step within each spot [48].

#### 3.5.5.5 Data filtering

The resulting data tables of traces with number of fluorescent proteins per spot was filtered in order to reduce noise, outliers, and data that is not suitable for further motion analysis:

- Diffraction-limited spots containing more than 5 fluorophores, likely aggregates, are filtered out.
- Traces starting or ending within 1 kb from the beads are filtered out to exclude fluorophores likely stuck to the beads from entering the dataset.
- Traces starting after frame 3 (in the case of experiments with C-DRM1 and C-DRM2) or 1 (in the case of experiments with CMG) are also filtered out, as we do not expect any fluorescently labeled molecule to bind the DNA during the scan.

#### 3.5.5.6 Positional analysis

In all plots that show the positions of diffraction-limited fluorescent spots, we plot the average position of the first three frames of each trace. The bin size of the position histograms was set to 700 bp, a value close to the diffraction limit that takes into account the absolute error in the position coordinate that results from the offset between the brightfield and the confocal channels and the relative error in the position coordinate as measured with a static fluorescent standard [8].

#### 3.5.5.7 Force-distance analysis

The force-extension curves of individual DNA molecules were fitted in the force range of 5-25 pN using the following extensible worm-like chain model [53]:

$$\langle x \rangle = L_c \left[ 1 - \frac{1}{2} \left( \frac{k_B T}{F L_p} \right)^{\frac{1}{2}} + \frac{F}{S} \right], \qquad \textbf{(Eq. 3.1)}$$

where $\langle x \rangle$ is the end-to-end distance of the DNA, $L_p$ is the persistence length of the DNA, $L_c$ is the contour length of the DNA, $S$ is the stretch modulus of the DNA, $k_B$ is

Boltzmann's constant, $T$ is temperature, and $F$ is the pulling force. From the fit, we obtained the parameters $L_c$, $L_p$, and $S$ (**Fig. 3.3d-g**, **k-n** and **Fig. S3.3a-d**).

## 3.6 Data availability

Raw and processed ensemble and single-molecule data supporting the findings of this study have been deposited in the 4TU.ResearchData repository and can be found at https://doi.org/10.4121/fe506b56-0ae0-4b89-9a5d-654b8b222cb0. The repository contains a table with an overview of experiments; force-distance tables sorted by experimental condition; spot position and intensity tables sorted by experimental condition; filtered spot tracking tables, with connected spot detections for each frame in each scan, each row having a scan_id and trace_id; and example TIFF files.

## 3.7 Code availability

All the code used in the current study is available at https://gitlab.tudelft.nl/nynke-dekker-lab/public/flapped_DNA_synthesis.

## 3.8 Funding

DRM acknowledges funding from a Boehringer Ingelheim Fonds PhD fellowship. Z.L. acknowledges funding from an EMBO Postdoctoral Fellowship (ALTF 484-2022). N.D. acknowledges funding from the Netherlands Organisation for Scientific Research (NWO) through TOP grant 714.017.002, from the 'BaSyC—Building a Synthetic Cell' Gravitation grant (024.003.019) of the Netherlands Ministry of Education, Culture and Science (OCW), and from the European Research Council through an Advanced Grant (REPLICHROMA; grant number 789267).

## 3.9 Author contributions

D.R.M. conceived the study, and N.D. supervised it. D.R.M. designed, assembled, and characterized plasmids pDRM1 and pDRM2. D.R.M. developed the synthesis strategy for the single-molecule constructs used in this study, and synthesized constructs C-DRM1 and C-DRM2. Z.L. assembled and characterized plasmid pZL7, and synthesized construct C-ZL7. D.R.M. conducted the single-molecule

experiments with DNA constructs C-DRM1 and C-DRM2 with input from Z.L. Z.L. expressed, purified, and fluorescently labeled CMG, and conducted the single-molecule experiments with C-ZL7 and CMG. Z.L. wrote automation and force spectroscopy analysis scripts. D.R.M. analyzed the data. All authors were involved in the discussion of the data. D.R.M. wrote the manuscript with input from Z.L. and N.D.

## 3.10 Declaration of interests

The authors declare no competing interests.

## 3.11 Acknowledgments

We thank N.D. lab members Theo van Laar, Julien Gros, Katinka Ligthart, Nerea Murugarren, Pang Yen Wang, and Humberto Sánchez, as well as Miloš Cvetković, Alessandro Costa, Samson Glaser, and John Diffley for their help with the expression, purification, and labeling of CMG. We also thank Vincent Kruit and Elena Radul for their help with the initial characterization of plasmid pDRM2 and the assembly of plasmid pZL7.

## 3.12 References

1. Yardimci, H. *et al.* Bypass of a protein barrier by a replicative DNA helicase. *Nature* **492**, 205–209 (2012).

2. Ticau, S., Friedman, L. J., Ivica, N. A., Gelles, J. & Bell, S. P. Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell* **161**, 513–525 (2015).

3. Ticau, S. *et al.* Mechanism and timing of Mcm2-7 ring closure during DNA replication origin licensing. *Nat. Struct. Mol. Biol.* **24**, 309–315 (2017).

4. Sánchez, H. *et al.* DNA replication origins retain mobile licensing proteins. *Nat. Commun.* **12**, (2021).

5. Scherr, M. J., Wahab, S. A., Remus, D. & Duderstadt, K. E. Mobile origin-licensing factors confer resistance to conflicts with RNA polymerase ll ll Mobile origin-licensing factors confer resistance to conflicts with RNA polymerase. *CellReports* **38**, (2022).

6. Díaz-Celis, C. *et al.* Assignment of structural transitions during mechanical unwrapping of nucleosomes and their disassembly products. *Proc. Natl.*

*Acad. Sci. U. S. A.* **119**, 1–11 (2022).

7. Safaric, B. *et al.* The fork protection complex recruits FACT to reorganize nucleosomes during replication. *Nucleic Acids Res.* **50**, 1317–1334 (2022).

8. Ramírez Montero, D. *et al.* Nucleotide binding halts diffusion of the eukaryotic replicative helicase during activation. *Nat. Commun.* **14**, (2023).

9. Zhang, B. *et al.* G-quadruplex and G-rich sequence stimulate Pif1p-catalyzed downstream duplex DNA unwinding through reducing waiting time at ss/dsDNA junction. *Nucleic Acids Res.* **44**, 8385–8394 (2016).

10. Lewis, J. S. *et al.* 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**, 10630–10635 (2017).

11. Kim, Y., De La Torre, A., Leal, A. A. & Finkelstein, I. J. Efficient modification of λ-DNA substrates for single-molecule studies. *Sci. Rep.* **7**, 1–14 (2017).

12. Kose, H. B., Larsen, N. B., Duxin, J. P. & Yardimci, H. Dynamics of the Eukaryotic Replicative Helicase at Lagging-Strand Protein Barriers Support the Steric Exclusion Model. *Cell Rep.* **26**, 2113-2125.e6 (2019).

13. Lewis, J. S. *et al.* Tunability of DNA polymerase stability during eukaryotic DNA replication. *Mol. Cell* **77**, 1–9 (2020).

14. Sparks, J. L. *et al.* The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell* **176**, 167-181.e21 (2019).

15. Kaczmarczyk, A. *et al.* Search and Processing of Holliday Junctions within Long DNA by Junction-Resolving Enzymes. *Nat. Commun.* **13:5921**, (2022).

16. Luzzietti, N. *et al.* Efficient preparation of internally modified single-molecule constructs using nicking enzymes. *Nucleic Acids Res.* **39**, (2011).

17. Luzzietti, N., Knappe, S., Richter, I. & Seidel, R. Nicking enzyme-based internal labeling of DNA at multiple loci. *Nat. Protoc.* **7**, 643–653 (2012).

18. Belan, O. *et al.* Article Single-molecule analysis reveals cooperative stimulation of Rad51 filament nucleation and growth by mediator proteins Article Single-molecule analysis reveals cooperative stimulation of Rad51 filament nucleation and growth by mediator proteins. *Mol. Cell* **81**, 1–16 (2021).

19. Mueller, S. H., Spenkelink, L. M., van Oijen, A. M. & Lewis, J. S. Design of customizable long linear DNA substrates with controlled end modifications for single-molecule studies. *Anal. Biochem.* **592**, 113541 (2020).

20. Bell, N. A. W. & Molloy, J. E. Efficient golden gate assembly of DNA constructs for single molecule force spectroscopy and imaging. *Nucleic Acids Res.* **50**,

E77 (2022).

21. Pradhan, B. *et al.* The Smc5/6 complex is a DNA loop-extruding motor. *Nature* **616**, 843–848 (2023).

22. Wang, F. *et al.* The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion. *Nat. Struct. Mol. Biol.* **20**, 174–181 (2013).

23. Lewis, J. S. *et al.* 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**, 10630–10635 (2017).

24. Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science (80-. ).* **360**, 102–105 (2018).

25. Lewis, J. S. *et al.* Tunability of DNA polymerase stability during eukaryotic DNA replication. *Suppl. Inf.* **77**, 602086 (2019).

26. Wasserman, M. R., Schauer, G. D., O'Donnell, M. E. & Liu, S. Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase. *Cell* **178**, 600-611.e16 (2019).

27. Kapadia, N. *et al.* Processive Activity of Replicative DNA Polymerases in the Replisome of Live Eukaryotic Cells. *Mol. Cell* **80**, 114-126.e8 (2020).

28. Lee, C. Y. & Myong, S. Probing steps in DNA transcription using single-molecule methods. *J. Biol. Chem.* **297**, 101086 (2021).

29. Tanasie, N. L., Gutiérrez-Escribano, P., Jaklin, S., Aragon, L. & Stigler, J. Stabilization of DNA fork junctions by Smc5/6 complexes revealed by single-molecule imaging. *Cell Rep.* **41**, (2022).

30. Silverstein, T. D., Gibb, B. & Greene, E. C. Visualizing protein movement on DNA at the single-molecule level using DNA curtains. *DNA Repair (Amst).* **20**, 94–109 (2014).

31. Kose, H. B., Xie, S., Cameron, G., Strycharska, M. S. & Yardimci, H. Duplex DNA engagement and RPA oppositely regulate the DNA-unwinding rate of CMG helicase. *Nat. Commun.* **11**, 1–15 (2020).

32. Vrtis, K. B. *et al.* Single-strand DNA breaks cause replisome disassembly. *Mol. Cell* **81**, 1309-1318.e6 (2021).

33. Newton, M. D. *et al.* DNA stretching induces Cas9 off-target activity. *Nat. Struct. Mol. Biol.* **26**, 185–192 (2019).

34. Low, E., Chistol, G., Zaher, M. S., Kochenova, O. V. & Walter, J. C. The DNA replication fork suppresses CMG unloading from chromatin before termination. *Genes Dev.* **34**, 1534–1545 (2020).

35. Gruszka, D. T., Xie, S., Kimura, H. & Yardimci, H. Single-molecule imaging reveals control of parental histone recycling by free histones during DNA replication. *Sci. Adv.* **6**, (2020).

36. Thorpe, H. M. & Smith, M. C. M. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the-resolvase/invertase family. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5505–5510 (1998).

37. Thomason, L. C., Oppenheim, A. B. & Court, D. L. Modifying Bacteriophage λ with Recombineering. *Methods Mol. Biol.* **501**, 239–251 (2009).

38. Bell, J. C., Plank, J. L., Dombrowski, C. C. & Kowalczykowski, S. C. Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA. *Nature* **491**, 274–278 (2012).

39. Burnham, D. R., Kose, H. B., Hoyle, R. B. & Yardimci, H. The mechanism of DNA unwinding by the eukaryotic replicative helicase. *Nat. Commun.* **10**, 1–14 (2019).

40. Naegeli, H., Bardwell, L. & Friedberg, E. C. Inhibition of Rad3 DNA Helicase Activity by DNA Adducts and Abasic Sites: Implications for the Role of a DNA Helicase in Damage-Specific Incision of DNA. *Biochemistry* **32**, 613–621 (1993).

41. von Hippel, P. H. & Marcus, A. H. The many roles of binding cooperativity in the control of DNA replication. *Biophys. J.* **117**, 2043–2046 (2019).

42. Van Loenhout, M. T. J., De Grunt, M. V. & Dekker, C. Dynamics of DNA supercoils. *Science (80-. ).* **338**, 94–97 (2012).

43. Whinn, K. S. *et al.* Nuclease dead Cas9 is a programmable roadblock for DNA replication. *Sci. Rep.* **9**, 1–9 (2019).

44. Schauer, G. D. *et al.* Replisome bypass of a protein-based R-loop block by Pif1. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 30354–30361 (2020).

45. Candelli, A., Wuite, G. J. L. & Peterman, E. J. G. Combining optical trapping, fluorescence microscopy and micro-fluidics for single molecule studies of DNA-protein interactions. *Phys. Chem. Chem. Phys.* **13**, 7263–7272 (2011).

46. Van Dijk, M. A., Kapitein, L. C., Van Mameren, J., Schmidt, C. F. & Peterman, E. J. G. Combining optical trapping and single-molecule fluorescence spectroscopy: Enhanced photobleaching of fluorophores. *J. Phys. Chem. B* **108**, 6479–6484 (2004).

47. Spakman, D., King, G. A., Peterman, E. J. G. & Wuite, G. J. L. Constructing arrays of nucleosome positioning sequences using Gibson Assembly for single-molecule studies. *Sci. Rep.* **10**, 1–11 (2020).

48. Liu, Z. *et al.* A biophysics toolbox for reliable data acquisition and processing

in integrated force-confocal fluorescence microscopy. *ACS Photonics* (2024).

49. Candelli, A. *et al.* Visualization and quantification of nascent RAD51 filament formation at single-monomer resolution. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15090–15095 (2014).

50. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic Elasticity of lambda-Phage DNA. *Science (80-. ).* **265**, 1599–1600 (1994).

51. Jaqaman, K. *et al.* Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* **5**, 695–702 (2008).

52. Van Der Walt, S. *et al.* Scikit-image: Image processing in python. *PeerJ* 1–18 (2014) doi:10.7717/peerj.453.

53. Odijk, T. Stiff Chains and Filaments under Tension. *Macromolecules* **28**, 7016–7018 (1995).

54. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

55. Wang, M. D., Yin, H., Landick, R., Gelles, J. & Block, S. M. Stretching DNA with optical tweezers. *Biophys. J.* **72**, 1335–1346 (1997).

56. Baumann, C. G., Smith, S. B., Bloomfield, V. A. & Bustamante, C. Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 6185–6190 (1997).

57. Moyer, S. E., Lewis, P. W. & Botchan, M. R. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10236–10241 (2006).

58. Jiang, W. *et al.* Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun.* **6**, 1–8 (2015).

59. Dhakal, S. *et al.* Structural and mechanical properties of individual human telomeric G-quadruplexes in molecularly crowded solutions. *Nucleic Acids Res.* **41**, 3915–3923 (2013).

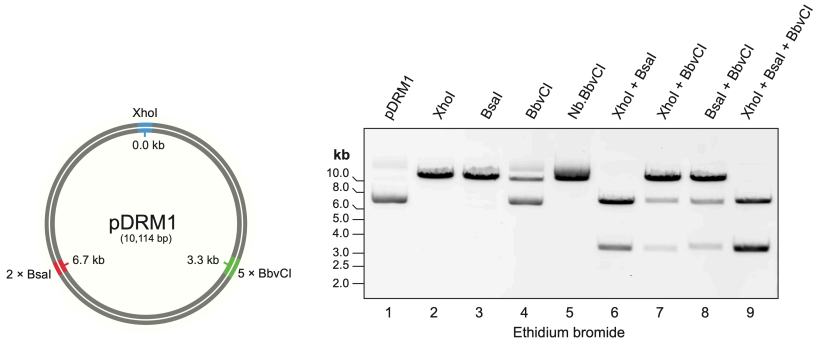
60. Parks, J. W. & Stone, M. D. Single-Molecule Studies of Telomeres and Telomerase. *Annu. Rev. Biophys.* **46**, 357–377 (2017).

61. Belan, O. *et al.* Generation of versatile ss-dsDNA hybrid substrates for single-molecule analysis. *STAR Protoc.* **2**, 100588 (2021).

62. Duxin, J. P., Dewar, J. M., Yardimci, H. & Walter, J. C. Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell* **159**, 346–357 (2014).

## 3.13 Supplementary figures

**a**

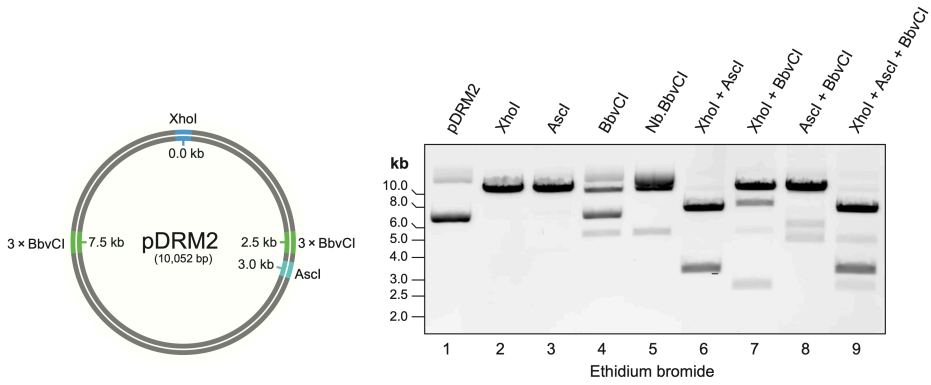



**b**

| Lane 1 | |
|---|---|
| **Band size (kb)** | **Assignment** |
| >10 | Nicked plasmid |
| 7 | Supercoiled plasmid |
| **Lane 2** | |
| **Band size (kb)** | **Assignment** |
| 10.1 | XhoI cut |
| **Lane 3** | |
| **Band size (kb)** | **Assignment** |
| 10.1 | BsaI cut |
| **Lane 4** | |
| **Band size (kb)** | **Assignment** |
| >10 | Nicked plasmid |
| 10.1 | BbvCI cut |
| 7 | Supercoiled plasmid |
| **Lane 5** | |
| **Band size (kb)** | **Assignment** |
| >10 | Nb.BbvCI nicked plasmid |
| 10.1 | Nb.BbvCI cut |
| **Lane 6** | |
| **Band size (kb)** | **Assignment** |
| 6.7 | XhoI cut + BsaI cut |
| 3.4 | XhoI cut + BsaI cut |
| **Lane 7** | |
| **Band size (kb)** | **Assignment** |
| 10.1 | XhoI cut or BbvCI cut |
| 6.8 | XhoI cut + BbvCI cut |
| 3.3 | XhoI cut + BbvCI cut |
| **Lane 8** | |
| **Band size (kb)** | **Assignment** |
| 10.1 | BsaI cut or BbvCI cut |
| 6.7 | BsaI cut + BbvCI cut |
| 3.4 | BsaI cut + BbvCI cut |
| **Lane 9** | |
| **Band size (kb)** | **Assignment** |
| 6.7/6.8 | XhoI cut + BsaI cut or BsaI cut + BbvCI cut or XhoI cut + BbvCI cut |
| 3.4/3.3 | XhoI cut + BsaI cut or BsaI cut + BbvCI cut or XhoI cut + BbvCI cut |

*Figure S3.1* Full assignment of bands observed in Fig. 3.1. **a** Diagram of plasmid pDRM1 together with an agarose gel showing all the test digestions of pDRM1 (gel is the same gel shown in **Fig. 3.1c**, shown here for clarity). **b** Assignment of every DNA band observed in the gel in **a**. (*Continues on the next page*)

**c**




**d**

| Lane 1 | |
|---|---|
| **Band size (kb)** | **Assignment** |
| >10 | Nicked plasmid |
| 7 | Supercoiled plasmid |
| **Lane 2** | |
| **Band size (kb)** | **Assignment** |
| 10.1 | XhoI cut |
| **Lane 3** | |
| **Band size (kb)** | **Assignment** |
| 10.1 | AscI cut |
| **Lane 4** | |
| **Band size (kb)** | **Assignment** |
| >10 | Nicked plasmid |
| 10.1 | BbvCI single cut |
| 7 | Supercoiled plasmid |
| 5.0/5.1 | BbvCI double cut |
| **Lane 5** | |
| **Band size (kb)** | **Assignment** |
| >10 | Nb. BbvCI nicked plasmid |
| 10.1 | Nb. BbvCI single cut |
| 5.0/5.1 | Nb. BbvCI double cut |
| **Lane 6** | |
| **Band size (kb)** | **Assignment** |
| 10.1 | XhoI cut or AscI cut |
| 7.1 | XhoI cut + AscI cut |
| 3 | XhoI cut + AscI cut |
| **Lane 7** | |
| **Band size (kb)** | **Assignment** |
| 10.1 | XhoI cut or BbvCI single cut |
| 7.5/7.6 | XhoI cut + BbvCI single cut |
| 5.0/5.1 | BbvCI double cut |
| 2.5/2.6 | XhoI cut + BbvCI single cut |
| **Lane 8** | |
| **Band size (kb)** | **Assignment** |
| 10.1 | AscI cut or BbvCI single cut |
| 5.6 | AscI cut + BbvCI single cut |
| 4.5 | AscI cut + BbvCI single cut |
| **Lane 9** | |
| **Band size (kb)** | **Assignment** |
| 10.1 | XhoI cut or AscI cut or BbvCI single cut |
| 7.1 | XhoI cut + AscI cut |
| 4.5 | AscI cut + BbvCI single cut |
| 3 | XhoI cut + AscI cut |
| 2.5/2.6 | XhoI cut + BbvCI single cut |

*Figure S3.1 (Continued).* **c** Diagram of plasmid pDRM2 together with an agarose gel showing all the test digestions of pDRM2 (gel is the same gel shown in **Fig. 3.1d**, shown here for clarity). **d** Assignment of every DNA band observed in the gel in **c**.

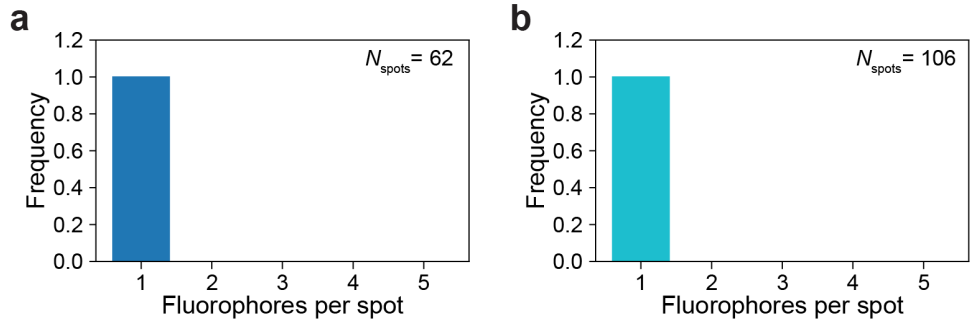



*Figure S3.2* Distribution of numbers of Cy3 fluorophores within each diffraction-limited spot of DNA constructs C-DRM1 and C-DRM2. **a** Distribution of numbers of Cy3 fluorophores within each diffraction-limited spot found in DNA construct C-DRM1 (**Fig. 3.3b**). **b** Distribution of numbers of Cy3 fluorophores within each diffraction-limited spot found in DNA construct C-DRM2 (**Fig. 3.3i**).

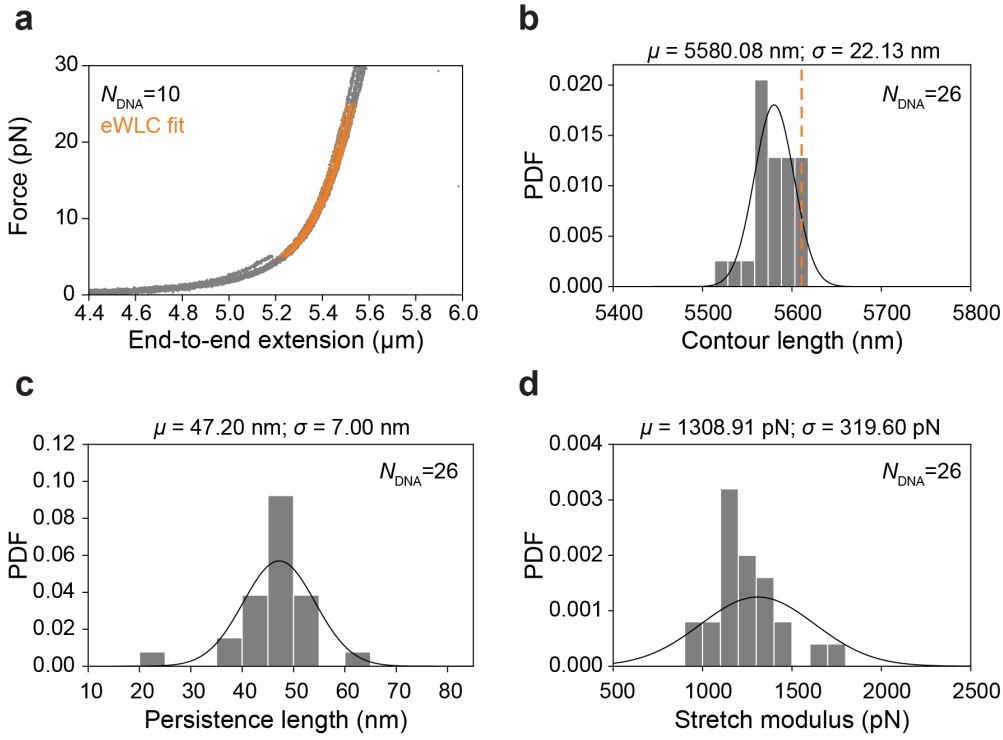



*Figure S3.3* Single-molecule characterization of DNA construct CZL7. **a** Example force-extension curves of construct C-ZL7 (gray) together with the fitted extensible worm-like chain (eWLC) model plotted in the force range used for fitting (orange). **b-d** Probability density function of **b** contour lengths, **c** persistence lengths, and **d** stretch moduli obtained from the extensible worm-like chain model fits of force-extension curves of C-ZL7; black lines show a Gaussian fit of the data, and μ and σ are the mean and the standard deviation of the data, respectively. Dotted orange line in panel B shows the expected contour length of C-ZL7.

## 3.14 Supplementary tables

*Supplementary table 3.1* gBlocks used in this study

| gBlock1: |
| --- |
| CAGCTCATTTGTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGA<br>GTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTAT<br>CAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATC<br>GGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAG<br>AAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGC<br>GCTTAATGCGCCGCTACAGGGCGCGTACTATGGTTGCTTTGACGTATGCGGTGTGAAATACCGCACAGATGCGTAA<br>GGAGAAAATACCGCATCAGGCGCCGCCTCAGCTCATGTCATCCTCAGCACACTTGACCCTCAGCTCAGCTAGCCTC<br>AGCCTACAATCACCTCAGCGGAGCTCGGTACCCGATCCAGATCCCGAACGCCTATCTTAAAGTTTAAACATAAAGAC<br>CAGACCTATAGACCAGACCTAAAGACACTACATAAAGAGCAGACCTAAAGACGCCTTGTTGTTAGCCATAAAGTGAT<br>AACCTTTAATCATTGTCTTTATTAATACAACTCACTATAAGGAGAGACAACTTAAAGAGACTTAAAAGATTAATTTAAAA<br>TTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCATCGAGAGGGACACGGGGAAACACC<br>ACCATCATCACCATCATCCTGACTAGAGTGCTTGGCGAACCGGTGTTTGACGTCCAGGAATGTCAAATCCGTGGCG<br>TGACCTATTCCGCACCGCTGCGCGTTAAACTGCGTCTGGTGATCTATGAGCGCGAAGCGCCGGAAGGCACCGTAA<br>AAGACATTAAAGAACAAGAAGTCTACATGGGCGAAATTCCGCTCATGACAGACAACGGTACCTTTGTTATCAACGGT<br>ACTGAGCGTGTTATCGTTTCCCAGCTGCACCGTAGTCCGGGCGTCTTCTTTGACTCCGACAAAGGTAAAACCCACTC<br>TTCGGGTAAAGTGCTGTATAACGCGCGTATCATCCCTTACCGTGGTTCCTGGCTGGACTTCGAATTCGATCCGAAG<br>GACAACCTGTTCGTACGTATCGACCGTCGCCGTAAACTGCCTGCGACCATCATTCTGCGCGCCCTGAACTACACCA<br>CAGAGCAGATCCTCGACCTGTTCTTTGAAAAAGTTATCTTTGAAATCCGTGATAACAAGCTGCAGATGGAACTGGTG<br>CCGGAACGCCTGCGTGGTGAAACCGCATCTTTTGACATCGAAGCTAACGGTAAAGTGTACGTAGAAAAAGGCCGCC<br>GTATCACTGCGCGCCACATTCGCCAGCTGGAAAAAGACGACGTCAAACTGATCGAAGTCCCGGTTGAGTACATCGC<br>AGGTAAAGTGGTTGCTAAAGACTATATTGATGAGTCTACCGGCGAGCTGATCTGCGCAGCGAACATGGAGCTGAGC<br>CTGGATCTGCTGGCTAAGCTGAGCCAGTCTGGTCACAAGCGTATCGAAACGCTGTTCACCAACGATCTGGATCACG<br>GCCCATATATCTCTGAAACCTTACGTGTCGACCCAACTAACGACCGTCTGAGCGCACTGGTAGAAATCTACCGCATG<br>ATGCGCCCTGGCGAGCCGCCGACTCGTGAAGCAGCTGAAAGCCTGTTCGAGAACCTGTTCTTCTCCGAAGACCGTT<br>ATGACTTGTCTGCGGTTGGTCGTATGAAGTTCAACCGTTCTCTGCTGCGCGAAGAAATCGAAGGTTCCGGTATCCTG<br>AGCAAAGACGACATCATTGATGTTATGAAAAAGCTCATGGATATCCGTAACGGTAAAGGCGAAGTCGATGATATCGA<br>CCACCTCGGCAACCGTCGTATCCGTTCCGTTGGCGAAATGGCGGAAAACCAGTTCCGCGTTGGCCTGGTACGTGTA<br>GAGCGTGCGGTGAAAGAGCGTCTGTCTCTGGGCGATCTGGATACCCTGATGCCACAGGATATGATCAACGCCAAG<br>CCGATTTCCGCAGCAGTGAAAGAGTTCTTCGGTTCCAGCCAGCTGTCTCAGTTTATGGACCAGAACAACCCGCTGT<br>CTGAGATTACGCACAAACGTCGTATCTCCGCACTCGGCCCAGGCGGTCTGACCCGTGAACGTGCAGGCTTCGAAG<br>TTCGAGACGTACACCCGACTCACTACGGTCGCGTATGTCCAATCGAAACCCCTGAAGGTCCGAACATCGGTCTGAT<br>CAACTCTCTGTCCGTGTACGCACAGACTAACGAATACGGCTTCCTTGAGACTCCGTATCGTAAAGTGACCGACGGT<br>GTTGTAACTGACGAAATTCACTACCTGTCTGCTATCGAAGAAGGCAACTACGTTATCGCCCAGGCGAACTCCAACTT<br>GGATGAAGAAGGCCACTTCGTAGAAGACCTGG |
| gBlock2: |
| GAAGGCCACTTCGTAGAAGACCTGGTAACTTGCCGTAGCAAAGGCGAATCCAGCTTGTTCAGCCGCGACCAGGTTG<br>ACTACATGGACGTATCCACCCAGCAGGTGGTATCCGTCGGTGCGTCCCTGATCCCGTTCCTGGAACACGATGACGC<br>CAACCGTGCATTGATGGGTGCGAACATGCAACGTCAGGCCGTTCCGACTCTGCGCGCTGATAAGCCGCTGGTTGG<br>TACTGGTATGGAACGTGCTGTTGCCGTTGACTCCGGTGTAACTGCGGTAGCTAAACGTGGTGGTGTCGTTCAGTAC<br>GTGGATGCTTCCCGTATCGTTATCAAAGTTAACGAAGACGAGATGTATCCGGGTGAAGCAGGTATCGACATCTACAA<br>CCTGACCAAATACACCCGTTCTAACCAGAACACCTGTATCAACCAGATGCCGTGTGTGTCTCTGGGTGAACCGGTT<br>GAACGTGGCGACGTGCTGGCAGACGGTCCGTCCACCGACCTCGGTGAACTGGCGCTTGGTCAGAACATGCGCGTA<br>GCGTTCATGCCGTGGAATGGTTACAACTTCGAAGACTCCATCCTCGTATCCGAGCGTGTTGTTCAGGAAGACCGTTT<br>CACCACCATCCACATTCAGGAACTGGCGTGTGTGTCCCGTGACACCAAGCTGGGTCCGGAAGAGATCACCGCTGA<br>CATCCCGAACGTGGGTGAAGCTGCGCTCTCCAAACTGGATGAATCCGGTATCGTTTACATTGGTGCGGAAGTGACC<br>GGTGGCGACATTCTGGTTGGTAAGGTAACGCCGAAAGGTGAAACTCAGCTGACCCCAGAAGAAAAACTGCTGCGTG<br>CGATCTTCGGTGAGAAAGCCTCTGACGTTAAAGACTCTTCTCTGCGCGTACCAAACGGTGTATCCGGTACGGTTATC<br>GACGTTCAGGTCTTTACTCGCGATGGCGTAGAAAAAGACAAACGTGCGCTGGAAATCGAAGAAATGCAGCTCAAAC<br>AGGCGAAGAAAGACCTGTCTGAAGAACTGCAGATTGATATGATTTTTATTGTTAAGGATGACCATAATGAAGAACGT<br>GATATTTCAATAGCTAACCACGTTATTAATATTCATACAGGAAATGCTAATGCTATGCTGGGCCTGACAGACGAAGAG<br>AAACAAAATCAGCTGGAACAGCTGGCTGAGCAGTATGACGAACTGAAACACGAGTTCGAGAAGAAACTCGAAGCGA<br>AACGCCGCAAAATCACCCAGGGCGACGATCTGGCACCGGGCGTGCTGAAGATTGTTAAGGTATATCTGGCGGTTAA<br>ACGCCGTATCCAGCCTGGTGACAAGATGGCAGGTCGTCACGGTAACAAGGGTGTAATTTCTccctGGAGACCCCTCG<br>AAGCGGGTCTGTTCAGCCGTATCCGTGCTGTGCTGGTAGCCGGTGGCGTTGAAGCTGAGAAGCTCGACAAACTGC<br>CGCGCGATCGCTGGCTGGAGGGTCTCCgtgcAAGATCAACCCGATCGAAGATATGCCTTACGATGAAAACGGTACGC<br>CGGTAGACATCGTACTGAACCCGCTGGGCGTACCGTCTCGTATGAACATCGGTCAGATCCTCGAAACCCACCTGGG<br>TATGGCTGCGAAAGGTATCGGCGACAAGATCAACGCCATGCTGAAACAGCAGCAAGAAGTCGCGAAACTGCGCGA<br>ATTCATCCAGCGTGCGTACGATCTGGGCGCTGACGTTCGTCAGAAAGTTGACCTGAGTACCTTCAGCGATGAAGAA<br>GTTATGCGTCTGGCTGAAAACATGTCATTTGATAGACCGGAAATATACAGTGCTCCTGTTTTACAAGGAGAATCTCCT<br>AACGACGATGATAATACTGAAATCATAAAGTCCTTTAAGAATTTCATTTTGGAGTTCAGACTTGACTCGCAATTTATTT<br>ACAGAGATCAGTTAAGGAACAACATCCTTGTGAAGAATTATTCTTTAACGGTTAACATGGAGCATTTGATCGGATATA<br>ACGAAGACATATATAAGAAACTATCAGACGAACCTTCAGATATCATTCCATTATTCGAAACCGCGATCACACAAGTGG |

CTAAAAGGATAAGTATTCTAAGCAGAGCTCAATCTGCTAATAACAATGACAAAGATCCAGAAAATACTAGTATGGATA
CTGATTCTCTCTTATTGAACTCTTTACCAACATTTCAATTAATTTTAAACTCCAATGCAAATCAGATTCCATTGAGAGAT
TTGGATTCTGAACACGTCTCCAAGATTGTCCGTTTATCAGGTATTATAATATCCACGTCAGTTTTATCTTCCCGTGCC
ACGTACCTTTCTATAATGTGCAGAAATTGCAGACACACAACATCAATAACAATCAACAATTTCAATTCTATCACAGGC
AATACCGTCAGTTTACCACGTTCTTGCTTATCTACGATTGAGAGTGAATCTTCTATGGCAAACGAGTCGAATATTGGT
GATGAATCGACCAAGAAAAATTGTGGACCTGATCCATATATTATTATCCATGAGTCTTCAAAGTTTATTGATCAACAGT
TTTTAAAATTACAGGAAATCCCAGAACTGGTTCCAG

## gBlock3:

CAGGAAATCCCAGAACTGGTTCCAGTAGGTGAGATGCCTAGAAACTTAACAATGACTTGTGACCGATACCTAACAAA
CAAAGTTATTCCTGGTACGAGAGTCACTATAGTAGGTATTTATTCCATCTATAATTCTAAAAATGGTGCCGGATCCGG
AAGGAGCGGGGGTGGAAATGGAGGAAGTGGTGTTGCTATTAGAACACCTTATATCAAAATATTAGGTATTCAGTCCG
ATGTAGAAACCTCCTCTATTTGGAATTCAGTAACTATGTTTACTGAGGAGGAAGAAGAGGAATTTCTACAGCTAAGTA
GAAACCCGAAGCTTTATGAAATTTTGACCAACTCTATTGCCCCCTCTATTTTTGGTAATGAAGATATAAAAAAAGCCAT
TGTATGTTTATTGATGGGTGGTTCCAAGAAGATATTACCCGATGGTATGAGGTTAAGAGGTGATATCAATGTACTATT
ATTAGGTGATCCAGGTACCGCCAAATCTCAACTATTGAAATTTGTGGAGAAAGTGTCACCTATTGCGGTATATACATC
TGGTAAGGGATCTTCTGCAGCTGGGTTAACTGCCAGTGTACAAAGAGATCCGATGACAAGAGAATTTTATTTGGAAG
GTGGTGCTATGGTGCTTGCCGATGGTGGTGTTGTATGCATGGATGAATTCGATAAAATGAGAGATGAAGATAGAGT
GGCCATTCATGAAGCTATGGAGCAGCAAACAATCTCCATCGCAAAAGCTGGTATCACTACAGTGCTAAATTCTAGAA
CTAGTGTTTTAGCGGCTGCTAATCCGATATACGGCCGGTATGATGATTTGAAGTCTCCTGGTGACAACATTGATTTC
CAAACTACTATTTTATCCCGTTTTGATATGATTTTTATTGTTAAGGATGACCATAATGAAGAACGTGATATTTCAATAG
CTAACCACGTTATTAATATTCATACAGGAAATGCTAATGCTATGCAAAACCAACAAGAGGAAAATGGCAGTGAAATTA
GTATTGAAAAGATGAAACGTTACATTACGTATTGTAGATTGAAATGTGCACCAAGACTTTCACCGCAGGCCGCTGAA
AAACTGTCATCGAACTTCGTCACCATTAGGAAGCAATTATTAATCAACGAATTAGAGTCAACGGAAAGGTCGTCTATT
CCAATTACCATTCGTCAATTAGAAGCTATTATTAGAATAACAGAATCATTAGCCAAGTTAGAATTAAGTCCTATTGCAC
AGGAGAGACATGTTGACGAAGCTATTAGATTGTTTCAAGCTTCCACAATGGACGCAGCGTCTCAGGATCCAATTGGC
GGCTTAAATCAAGCAAGCGGAACATCTTTGTCAGAAATCCGTCGTTTTGAACAAGAACTAAAAAGAAGATTACCTATA
GGCTGGTCTACTTCTTATCAAACTTTGAGGAGAGAATTTGTAGATACACATAGATTTTCTCAATTAGCACTGGATAAG
GCCTTATATGCCCTAGAGAAGCATGAAACAATTCAATTGAGACACCAGGGACAGAATATTTACAGAAGTGGTGTATG
AAGCCCCGGCCTGCAGGAATTCGATATCAAGCTTATGGATACCGTCGACCGCGAGGGGGGGCCCGGTACCCAGCT
TTTGTTCCCTTTAGTGAGGGTTAATTCCGAGCTTGGCGTAATCATGGTCATAGCTGAAACCTGTGTGATATTGAAATC
CGCTCACAATTCCACACAACATAGGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGGTAACT
CACATTAATTGCGTTGCGCTCACTCTGCGCAAAGGTATGCCAATCGCAACGCCGGTGTTCGACGGTGCGAAAGAAG
CAGAAATTAAAGAGCTGCTGAAACTTGGCGACCTGCCGACTTCCGGTCAGATCCGCCTGTACGATGGTCGCACTGG
TGAACAGTTCGAGCGTCCGGTAACCGTTGGTTACATGTACATGCTGAAACTGAACCACCTGGTCGACGACAAGATG
CACGCGCGTTCCACCGGTTCTTACAGCCTGGTTACTCAGCAGCCGCTGGGTGGTAAGGCACAGTTCGGTGGTCAG
CGTTTCGGGGAGATGGAAGTGTGGGCGCTGGAAGCATACGGCGCAGCATACACCCTGCAGGAAATGCTCACCGTT
AAGTCTGATGACGTGAACGGTCGTACCAAGATGTATAAAAACATCGTGGACGGCAACCATCAGATGGAGCCGGGCA
TGCGAGAGTAGGGAACCTCGAGTGCCAGGCATCAAAGAAAAGGGAAGGCACAGTCGAAAGACTGGGCCTTGGGTT
TTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCA
ACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGA
CGGATGGCCTTTTTGCGTTTCTACAAACTCTTCCTGTCGTCATATCTACAAGCCATCCCCCCACAGATACGGTAAACT
AGCCTCGTTTTTGCATCAGGAAAGCAGAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATC
CG

## gBlock4:

GCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT
GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTC
GTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTC
GCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTT
ATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA
GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGG
TGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGA
CCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAG
GTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTG
CGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGT
AACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGG
CAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAG
AAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGA
TTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAAC
TTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGC
CTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGA
GATCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCT
GCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTC
CCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTT
GTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATC
CGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCT
CTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCT

TCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGAT
CTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATA
AGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCA
TGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCA
CCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCG
CGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATG
CCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCA
TCAGAGCAGATTGTACTGAGAGTGCACCATAAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTT
AAATCAGCTCATTTGTTAACCAATAGGCCG

gBlock5:

GCTGTTTCCTGTGTGAAATTGTTATCCGTCTACAATTCCACACAACATACGAGCCGCAAGCATAAAGTGTAAAGCCT
GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTC
GTGCCTGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCGGCTTCCTC
GCTCACTGACTCGCTGCGCTCGGTCGTCTGGCTGCGGCGAGCGGTATCATCTCACTCAAAGGCGGTAATACGGTTA
TCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAG
GCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
GGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGAC
CCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGG
TATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGC
GCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAG
AAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCA
AACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAA
GATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAGCAAAACTCACGTTAAGGGATTTTGGTCATGAGATT
ATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTT
GGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCT
GACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA
TCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGC
AACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCG
CAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCC
AACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTC
AGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGT
AAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTT
GCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCG
GGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTC
AGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGG
GCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGA
GCGCATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT
GACGTCTAAGAAACCATCCTCAGCAGAGCTAGCCTCAGCCTACAATCACCTCAGCCGAGGCCCTTTCGTCTCGCGC
GTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGCAGACGGTCAGAGCTTGTCTGTAAGCGGATGC
CGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCAT
CAGAGCAGATTGTACTGAGAGTGCACCATAAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTA
AATCAGCTCATTTGTTAACCAATAGGCCG

gBlock6:

CAGCTCATTTGTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGA
GTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTAT
CAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATC
GGAACCCTAAAGGGAGCCCCCGAGGCGCGCCTTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGG
AAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACA
CCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTACTATGGTTGCTTTGACGTATGCGGTGTGAAATACCGCACA
GATGCGTAAGGAGAAAATACCGCATCAGGCGCCGGGAGCTCGGTACCCGATCCAGATCCCGAACGCCTATCTTAAA
GTTTAAACATAAAGACCAGACCTATAGACCAGACCTAAAGACACTACATAAAGAGCAGACCTAAAGACGCCTTGTTG
TTAGCCATAAAGTGATAACCTTTAATCATTGTCTTTATTAATACAACTCACTATAAGGAGAGACAACTTAAAGAGACTT
AAAAGATTAATTTAAAATTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCATCGAGAGGG
ACACGGGGAAACACCACCATCATCACCATCATCCTGACTAGAGTGCTTGGCGAACCGGTGTTTGACGTCCAGGAAT
GTCAAATCCGTGGCGTGACCTATTCCGCACCGCTGCGCGTTAAACTGCGTCTGGTGATCTATGAGCGCGAAGCGCC
GCAAGGCACCGTAAAAGACATTAAAGAACAAGAAGTCTACATGGGCGAAATTCCGCTCATGACAGACAACGGTACC
TTTGTTATCAACGGTACTGAGCGTGTTATCGTTTCCCAGCTGCACCGTAGTCCGGGCGTCTTCTTTGACTCCGACAA
AGGTAAAACCCACTCTTCGGGTAAAGTGCTGTATAACGCGCGTATCATCCCTTACCGTGGTTCCTGGCTGGACTTCG
AATTCGATCCGAAGGACAACCTGTTCGTACGTATCGACCGTCGCCGTAAACTGCCTGCGACCATCATTCTGCGCGC
CCTGAACTACACCACAGAGCAGATCCTCGACCTGAATTCCTCGATTTTTTTATGTTTAGTTTCGCGGACGACGGTTTC
GAGGTGGCGGTCTGGACCACGCATGAGAGCGTCGAAGCGGAGGCGGTGTTCGCCGAGATCGGCTCGCGCATGGC
CGAGTTGAGCGGTTCCAGGCTGGCCAAACAGCATCAGATGGAACTAAACATAAAAATACACCGGTCCCAGCATCCG
GTATCGCAGCTGGCCACCGTCGGCGTCTCGCACGACCACCAGTGCAAGTATCTGAGCAGCGCTATCGTGCTCCTC
GGAGTGGAGGCAGCCGAATGCGACGGTGTGCTTGCCTTCCTGGAGACCTCCGCGCTCCGCAACCTCCACTTCTAC
GAGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAGGTGCCTTAAGGACATCGCACCTGGTGCATGACCCGCAAG
CTTGGTGCCTGACGCTCGCCACACGACCCGCAGCGCCCGACCGAAAGGAGCGCACGACCCGGGGTTCGAGAACC
TGTTCTTCTCCGAAGACCGTTATGACTTGTCTGCGGTTGGTCGTATGAAGTTCAACCGTTCTCTGCTGCGCGAAGAA

ATCGAAGGTTCCGGTATCCTGAGCAAAGACGACATCATTGATGTTATGAAAAAGCTCATGGATATCCGTAACGGTGA
TATTTCAATAGCTAACCACGTTATTAATATTCATACAGGAAATGCTAATGCTATGCAAAACCAACAAGAGGAAAATGG
CAGTGAAATTAGTATTGAAAAGATGAAACGTTACATTACGTATTGTAGATTGAAATGTGCACCAAGACTTTCACCGCA
GGCCGCTGAAAAACTGTCATCGAACTTCGTCACCATTAGGAAGCAATTGGGAATCAACGAATTAGAGTCAACGGAAA
GGTCGTCTATTCCAATTACCATTCGTCAATTAGAAGCTATTATTAGAATAACGATTTCATTAGCCAAGTTAGATATAAG
TCCTATTGCACAGGAGAGACATGTTGACGAAGCTATTAGATTGTTTCAAGCTTCCACAATGGACGAAGGTCCGAACA
TCGGTCTGATCAACTCTCTGTCCGTGTACGCACAGACTAACGAATACGGCTTCCTTGAGACTCCGTATCGTAAAGTG
ACCGACGGTGTTGTAACTGACGAAATTCACTACCTGTCTGCTATCGAAGAAGGCAACTACGTTATCGCCCAGGCGA
ACTCCAACTTGGATGAAGAAGGCCACTTCGTAGAAGACCTGG

gBlock7:

GAAGGCCACTTCGTAGAAGACCTGGTAACTTGCCGTAGCAAAGGCGAATCCAGCTTGTTCAGCCGCGACCAGGTTG
ACTACATGGACGTATCCACCCAGCAGGTGGTATCCGTCGGTGCGTCCCTGATCCCGTTCCTGGAACACGATGACGC
CAACCGTGCATTGATGGGTGCGAACATGCAACGTCAGGCCGTTCCGACTCTGCGCGCTGATAAGCCGCTGGTTGG
TACTGGTATGGAACGTGCTGTTGCCGTTGACTCCGGTGTAACTGCGGTAGCTAAACCCGGGTCGTGCGCTCCTTTC
GGTCGGGCGCTGCGGGTCGTGTGGCGAGCGTCAGGCACCAAGCTTGCGGGTCATGCACCAGGTGCGATGTCCTT
AAGGCACCTCGACGTCGGCGGTGACGGTGAAGCCGAGCCGCTCGTAGAAGTGGAGGTTGCGGAGCGCGGAGGTC
TCCAGGAAGGCAAGCACACCGTCGCATTCGGCTGCCTCCACTCCGAGGAGCACGATAGCGCTGCTCAGATACTTG
CACTGGTGGTCGTGCGAGACGCCGACGGTGGCCAGCTGCGATACCGGATGCTGGGACCGGTGTATTTTTATGTTT
AGTTCCATCTGATGCTGTTTGGCCAGCCTGGAACCGCTCAACTCGGCCATGCGCGAGCCGATCTCGGCGAACACC
GCCTCCGCTTCGACGCTCTCATGCGTGGTCCAGACCGCCACCTCGAAACCGTCGTCCGCGAAACTAAACATAAAAA
AATCGAGGAATTCGACATTCTGGTTGGTAAGGTAACGCCGAAAGGTGAAACTCAGCTGACCCCAGAAGAAAAACTG
CTGCGTGCGATCTTCGGTGAGAAAGCCTCTGACGTTAAAGACTCTTCTCTGCGCGTACCAAACGGTGTATCCGGTA
CGGTTATCGACGTTCAGGTCTTTACTCGCGATGGCGTAGAAAAAGACAAACGTGCGCTGGAAATCGAAGAAATGCA
GCTCAAACAGGCGAAGAAAGACCTGTCTGAAGAACTGCAGATTGATATGATTTTTATTGTTAAGGATGACCATAATGA
AGAACGTGATATTTCAATAGCTAACCACGTTATTAATATTCATACAGGAAATGCTAATGCTATGCTGGGCCTGACAGA
CGAAGAGAAACAAAATCAGCTGGAACAGCTGGCTGAGCAGTATGACGAACTGAAACACGAGTTCGAGAAGAAACTC
GAAGCGAAACGCCGCAAAATCACCCAGGGCGACGATCTGGCACCGGGCGTGCTGAAGATTGTTAAGGTATATCTG
GCGGTTAAACGCCGTATCCAGCCTGGTGACAAGATGGCAGGTCGTCACGGTAACAAGGGTGTAATTTCTCCCTGGA
GACCCCTCGAAGCGGGTCTGTTCAGCCGTATCCGTGCTGTGCTGGTAGCCGGTGGCGTTGAAGCTGAGAAGCTCG
ACAAACTGCCGCGCGATCGCTGGCTGGAGGGTCTCCGTGCAAGATCAACCCGATCGAAGATATGCCTTACGATGAA
AACGGTACGCCGGTAGACATCGTACTGAACCCGCTGGGCGTACCGTCTCGTATGAACATCGGTCAGATCCTCGAAA
CCCACCTGGGTATGGCTGCGAAAGGTATCGGCGACAAGATCAACGCCATGCTGAAACAGCAGCAAGAAGTCGCGA
AACTGCGCGAATTCATCCAGCGTGCGTACGATCTGGGCGCTGACGTTCGTCAGAAAGTTGACCTGAGTACCTTCAG
CGATGAAGAAGTTATGCGTCTGGCTGAAAACATGTCATTTGATAGACCGGAAATATACAGTGCTCCTGTTTTACAAG
GAGAATCTCCTAACGACGATGATAATACTGAAATCATAAAGTCCTTTAAGAATTTCATTTTGGAGTTCAGACTTGACT
CGCAATTTATTTACAGAGATCAGTTAAGGAACAACATCCTTGTGAAGAATTATTCTTTAACGGTTAACATGGAGCATTT
GATCGGATATAACGAAGACATATATAAGAAACTATCAGACGAACCTTCAGATATCATTCCATTATTCGAAACCGCGAT
CACACAAGTGGCTAAAAGGATAAGTATTCTAAGCAGAGCTCAATCTGCTAATAACAATGACAAAGATCCAGAAAATAC
TAGTATGGATACTGATTCTCTCTTATTGAACTCTTTACCAACATTTCAATTAATTTTAAACTCCAATGCAAATCAGATTC
CATTGAGAGATTTGGATTCTGAACACGTCTCCAAGATTGTCCGTTTATCAGGTAGCTGAGGTGATTGTAGGCTGAGG
CTAGCTCTGCTGAGGTACCTTTCTATAATGTGCAGAAATTGCAGACACACAACATCAATAACAATCAACAATTTCAAT
TCTATCACAGGCAATACCGTCAGTTTACCACGTTCTTGCTTATCTACGATTGAGAGTGAATCTTCTATGGCAAACGAG
TCGAATATTGGTGATGAATCGACCAAGAAAAATTGTGGACCTGATCCATATATTATTATCCATGAGTCTTCAAAGTTTA
TTGATCAACAGTTTTTAAAATTACAGGAAATCCCAGAACTGGTTCCAG

*Supplementary table 3.2* Oligos used in this study

| **Oligo** | **5' to 3' sequence** |
|---|---|
| DRM_140 | AGATTGTACTGAGAGTGCACCATAAAATTG |
| DRM_141 | CACTTTATGGCTAACAACAAGGCG |
| DRM_142 | CCGAACATCGGTCTGATCAACTC |
| DRM_143 | GAGTGAATCTTCTATGGCAAACGAGTC |
| DRM_144 | GATGACGTGAACGGTCGTACC |
| DRM_145 | GCAGTATGACGAACTGAAACACGAG |
| DRM_159 | [5’Phos]TGAGGTGATTGTAGGCTGAGGCTAGCTCTGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT[3’Cy3] |
| DRM_177 | CGGCCTATTGGTTAACAAATGAGCTG |
| DRM_180 | GAAGGCCACTTCGTAGAAGACCTG |
| DRM_181 | CTGGAACCAGTTCTGGGATTTCC |
| DRM_182 | GAAGGCCACTTCGTAGAAGACCTG |
| DRM_186 | [5'Phos]TGAGGTGATTGTAGGCTGAGGCTAGCTGAGCTGAGGGTCAAGTGTGCTGAGGATGACATGAGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT[3’Cy3] |
| Vector.FOR | GATATATACATCCTGTATTTGACCCCAGAAGAAAAACTGCT |
| Vector.REV | CTATCCACGCAGCGGCAATCAGCTGAGTTTCACCTTTCGGCGT |
| ZL1 | GGAGCTCGGTACCCGATCC |
| ZL2 | CGGCGCCTGATGCGG |
| ZL3 | GTTTAAACATATGACCAGACCTATAGACCAGACC |
| ZL4 | GGTCTGGTCATATGTTTAAACTTTAAGATAGGCG |
| ZL5 | [5'Phos]GTATTTGACCCCAGGTAAAGACGCCTTGTTGTTAGCCATAAAGTG |

*Indicates phosphorothioate backbone linkages

# 4

# A biophysics toolbox for reliable data acquisition and processing in integrated force-confocal fluorescence microscopy

*This chapter is published as: Liu Z.*, van Veen E.*, Sánchez H., Solano B., Palmero Moya F., McCluskey K. A., Ramírez Montero D., van Laar T., and Dekker, N.H. (2024). A biophysics toolbox for reliable data acquisition and processing in integrated force-confocal fluorescence microscopy. ACS Photonics. *Equal contribution. doi:10.1021/acsphotonics.3c01739.*

## 4.1 Abstract

Integrated single-molecule force-fluorescence spectroscopy setups allow for simultaneous fluorescence imaging and mechanical force manipulation/measurements on individual molecules, providing comprehensive dynamic and spatiotemporal information. Dual-beam optical tweezers (OT) combined with a confocal scanning microscope form a force-fluorescence spectroscopy apparatus broadly used to investigate various biological processes, in particular protein:DNA interactions. Such experiments typically involve imaging of fluorescently labeled proteins bound to DNA, and force spectroscopy measurements on trapped individual DNA molecules. Here, we present a versatile state-of-the-art toolbox including the preparation of protein:DNA complex samples, design of a microfluidic flow cell incorporated with OT, automation of OT-confocal scanning measurements, and the development and implementation of a streamlined data analysis package for force and fluorescence spectroscopy data processing. Its components can be adapted to any commercialized or home-built dual-beam OT setup equipped with a confocal scanning microscope, which will facilitate single-molecule force-fluorescence spectroscopy studies on a large variety of biological systems.

## 4.2 Introduction

Over the past three decades, single-molecule techniques have evolved into a versatile approach for probing the fundamental mechanisms of various biological processes. Single-molecule force spectroscopy techniques, including optical tweezers (OT), magnetic tweezers (MT), and atomic force microscopy (AFM), are widely used to quantify the mechanical properties of biomolecules and monitor their dynamics involving force or contour length changes[1–5]. Single-molecule fluorescence imaging techniques, including confocal scanning microscopy and total internal reflection fluorescence (TIRF) microscopy, are able to directly visualize molecules of interest, and provide information including their stoichiometry and location[6]. They are uniquely suited for investigating complex systems with multiple molecules involved, as different molecules can be visualized using spectroscopically distinct fluorophores.

It is highly desirable for researchers to combine force and fluorescence spectroscopy techniques to gain a comprehensive insight into complex biological systems. Such integrated force-fluorescence microscopy setups have been realized using AFM[7,8], MT[9–12], single-beam OT[13], and dual-beam OT[14,15], in combination with confocal scanning or TIRF microscopy. Among these techniques, dual-beam OT have unique advantages when implemented with fluorescence imaging. In a dual-beam OT setup, two beads are trapped by laser beams that propagate along the (optical) *z*-axis (**Fig. 4.1**). The beads then tether an individual biomolecule between them, in the *x*-*y* plane. Typically, the fluorescence excitation laser beam also propagates along the *z*-axis (**Fig. 4.1b**). Therefore, the tethered molecule is perpendicular to both the trapping and excitation laser beams. In contrast, the orientations of molecules immobilized by MT or single-beam OT typically align with the *z*-axis or have a substantial component along it. For the most commonly used imaging techniques (confocal scanning and TIRF), it is more convenient to image in the entire *x*-*y* plane, rather than to scan along the *z*-axis; furthermore, the increased size of the point-spread function along the *z*-axis relative to the *x*-*y* plane reduces resolution in this dimension[16]. Therefore, use of dual-beam OT integrated with fluorescence imaging allows for the convenient extraction of highly resolved information over the entire biomolecule. An additional advantage of dual-beam OT over MT and single-beam OT is that instead of being tethered on the surface at one end, the molecules of interest are separated from the surface by several microns, effectively inhibiting undesirable surface adhesion. Taken together, dual-beam OT force-fluorescence spectroscopy has become a powerful single-molecule biophysics technique, broadly used to investigate a number of biological processes including, but not limited to, DNA replication, DNA repair, and protein folding

dynamics[3,17–22]. The focus of this article is the application of dual-beam OT force-fluorescence spectroscopy in studying protein:DNA interactions, one of the fields that has benefited most from the development of this approach.

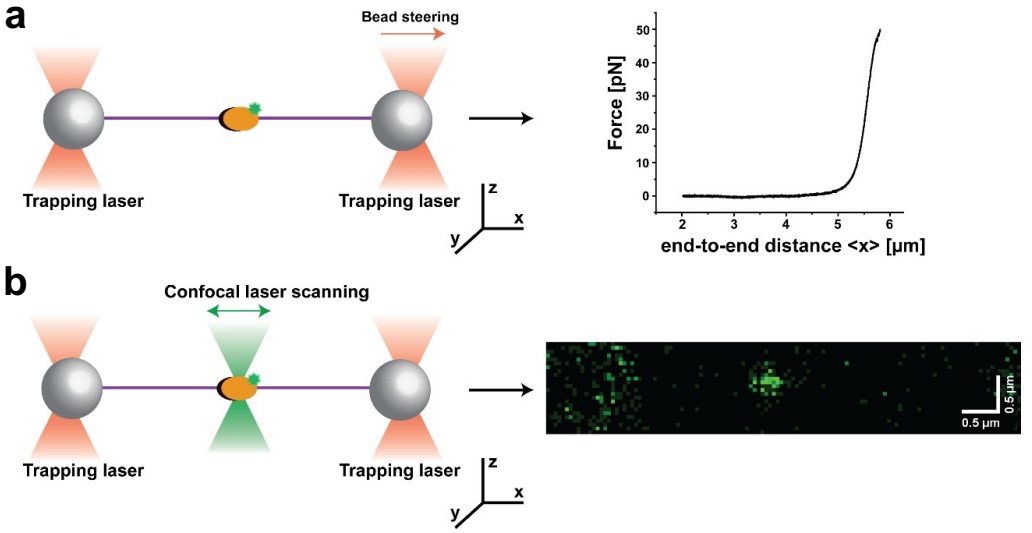



*Figure 4.4.2-1* Figure 1 Dual-beam OT setup for force-fluorescence spectroscopy measurements. An individual DNA molecule (purple) is tethered between two micron-sized beads (grey spheres) that are held in optical traps (red) to act as a substrate for a fluorescently labeled molecule of interest (orange oval with green dot). **a** (left) One of the beads is moved away from the other along the x-axis in order to exert a stretching force $F_x$ on the tethered DNA molecule. (right) One then records force-distance data by plotting this force against the end-to-end distance of the DNA measured along the x-coordinate. **b** (left) A confocal scanning laser (green) integrated with a dual-beam OT performs a line scan along the x-axis. (right) A series of offset line scans imaging the x-y plane to form a 2D image, where a fluorescently labeled CMG holo-helicase[23] is shown as a fluorescent spot. This allows one to capture the diffraction-limited spots of fluorescently labeled molecules bound to the tethered DNA molecule and monitor their dynamics.

As shown in **Fig. 4.1a**, a dual-beam OT setup typically generates force spectroscopy data by moving one bead in the *x*-direction away from the other one to apply stretching forces to a DNA molecule immobilized between the beads. The forces are recorded and plotted against the end-to-end distance of the DNA measured along the *x*-coordinate[3,17]. Such force-distance curves can be subsequently fitted with elasticity models or transformed into contour length space to extract mechanical information about the DNA molecule, and/or the molecules bound to DNA (orange oval in **Fig. 4.1**).

In studies of protein:DNA interactions, dual-beam OT can be integrated with confocal scanning microscopy (**Fig. 4.1b**) to image fluorescently labeled proteins bound to the trapped DNA and monitor their dynamics[15,24,25]. A confocal scanning

microscope benefits from a very small detection volume (on the order of femtoliters[26,27]) and is able to reject the fluorescence signal outside of the detection volume. This allows for a relatively high concentration of labeled proteins (on the order of ~10 nM[25,28]) in the measurement solution to monitor their dynamic interactions with DNA.

In this article, we will discuss several experimental and data analysis aspects required to obtain reliable biophysical data using dual-beam OT integrated with confocal scanning microscopy. First, we will discuss sample preparation strategies suited for investigating complex biological systems, in particular protein:DNA interactions, using dual-beam OT integrated with confocal scanning microscopy and a multichannel microfluidic system that enables well-controlled and flexible sample handling on the microscope. We will then focus on the theoretical background, development, and implementation of a carefully designed analytical tool to extract reliable spatiotemporal, stoichiometric, and mechanical information from dual-beam OT-confocal scanning microscopy data. In doing so, we will refer to scripts that we have developed to enable automatic data acquisition and storage on an integrated dual-beam OT-confocal scanning microscope that significantly enhance experimental throughput and reproducibility.

It is worth noting that, while the aforementioned analysis tools were developed based on investigations of protein:DNA interactions carried out using the commercially available Lumicks C-trap instrument (https://lumicks.com/products/c-trap-optical-tweezers-fluorescence-label-free-microscopy/), they can be easily adapted to a large variety of experiments carried out with any dual-beam OT-confocal scanning microscopy systems.

## 4.3 Data acquisition

### 4.3.1 Sample preparation for investigating protein:DNA interactions

One of the prerequisites for the DNA substrate in dual-beam OT-confocal scanning microscope measurements is that they should be relatively long. This ensures a sufficient distance between the fluorescently labeled proteins of interest and the edges of the beads, thus rejecting the noise signal from the beads generated by additional DNA molecules and proteins bound to the beads. The relatively long trap-trap distance also helps avoid interference between optical traps. Different strategies have been developed to prepare such long DNA substrates, including the

use of λ phage DNA[29,30] or plasmid DNA[31,32], synthesis via PCR amplification[33,34], and *de novo* synthesis of plasmid DNA[35]. Prior to single-molecule experiments, the DNA substrates are linearized if necessary, and their ends are functionalized to allow them to be tethered to the beads. Widely used functional groups for nucleic acid immobilization include small molecule-protein pairs such as biotin:streptavidin, digoxigenin:antidigoxigenin, and to a lesser extent, fluorescein:antifluorescein. These binding groups are highly specific and allow for mechanically stable non-covalent interactions with high affinity[36–39]. The next step is to bind the proteins of interest to the functionalized DNA molecules. This can be achieved either in bulk or at the single-molecule level, or using a combination of both (**Fig. 4.2**).

The primary advantage of pre-binding the proteins of interest on the DNA through bulk incubation prior to single-molecule measurements (**Fig. 4.2a**) is specific to proteins that require a complex reaction or long incubation with the DNA

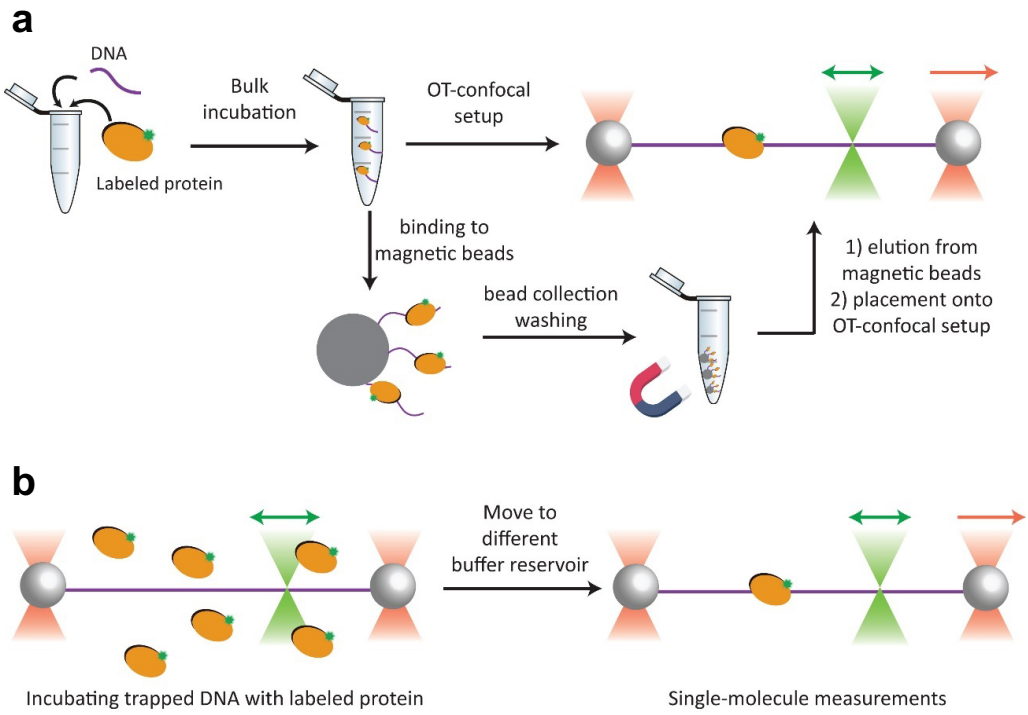



*Figure 4.2* Strategies to prepare protein:DNA substrates. **a** A protein of interest is bound to a DNA substrate through bulk incubation. The assembled protein:DNA complex can be either directly loaded into the dual-beam OT-confocal scanning microscope setup for single-molecule force-fluorescence measurements, or purified in bulk prior to single-molecule measurements. **b** An individual DNA substrate molecule is trapped in the dual-beam OT and incubated with a fluorescently labeled protein of interest to form a protein:DNA complex. The complex can be subsequently steered into a separate buffer reservoir for single-molecule measurements in the absence of background fluorescence.

substrate to function. This is more practically achieved in the bulk context than in the context of a dual-beam OT setup that is limited in terms of trapping duration and throughput. Another advantage of pre-binding is that the resulting protein:DNA complex can be purified, if necessary, once the proteins are bound, thereby removing free protein that could otherwise aggregate onto the DNA and complicate its manipulation (**Fig. 4.2a**). This is illustrated by our recent work[19], in which we pre-bound the DNA substrate onto streptavidin-coated magnetic beads via desthiobiotin:streptavidin interactions and reconstituted the origin-based assembly and activation of the replicative helicase CMG in bulk[40]. Different washing steps were used to remove unbound protein and aggregates from the magnetic bead-bound protein:DNA complexes, which were subsequently eluted from the magnetic beads using an excess of biotin. We took advantage of the orthogonality between biotin:streptavidin and digoxigenin:anti-digoxigenin interactions to then trap the protein:DNA complexes on OT using anti-digoxigenin-coated polystyrene beads.

While pre-binding the protein in bulk provides flexibility in forming a protein:DNA complex of interest, it is difficult to monitor the intermediate steps in its assembly. Another strategy is to incubate the proteins of interest with individual DNA substrates at the single-molecule level (**Fig. 4.2b**). This approach provides excellent control of incubation timing by steering the DNA substrate first into and then out of the buffer containing the protein of interest, enabling measurements of protein association/dissociation kinetics and the observation of early intermediates generated in the reactions involving protein:DNA interactions. However, should protein aggregates form, it is more challenging to remove them by washing the protein:DNA complex trapped between the beads, because the beads prevent dissociation of proteins from the DNA ends.

These two approaches can also be combined for complex systems with various proteins involved. For example, in our recent publication[22], we pre-bound histones on double-stranded DNA in bulk to assemble nucleosomes, which requires overnight salt gradient dialysis. Another protein complex of interest, the eukaryotic DNA replication initiator, origin recognition complex (ORC), was subsequently bound onto the DNA by incubating such a trapped single DNA in a buffer containing ORC.

To implement the aforementioned two approaches, we designed a microfluidic chip system (**Fig. 4.3a**) that is able to handle very small volume (~100 μL) of samples with low concentrations (down to ~10 pM), and provide compartmentalization to separate beads, nucleic acids, proteins, and different

buffers. This flow cell is reusable and can be cleaned, followed by surface passivation using BSA and Pluronics, within ~4 hours. Alternatively, a better passivation could be achieved by overnight incubation with passivation reagents. The details of the microfluidic chip design are described in **Supplementary Method 4.1**.

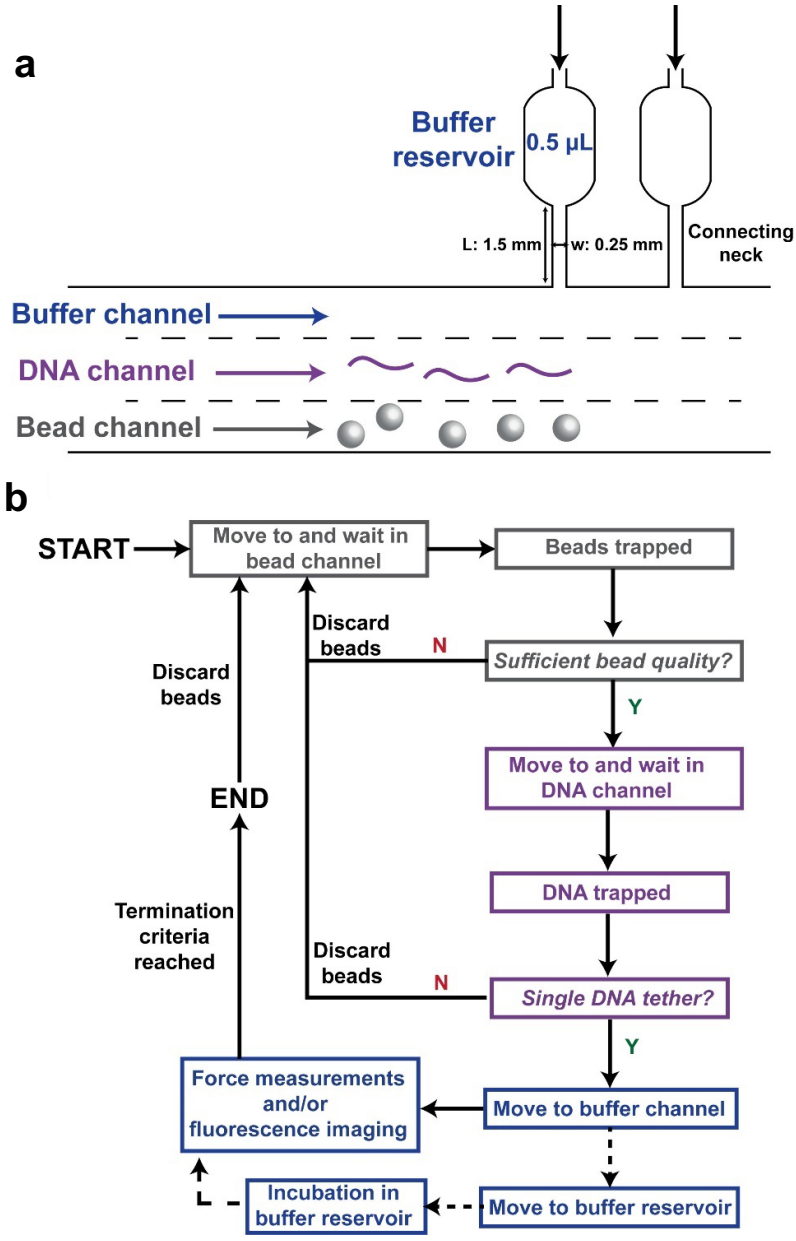



*Figure 4.3* Microfluidic system design and procedure of automated dual-beam OT force-fluorescence microscopy measurements. *(Continues on the next page).*

*Figure 4.3 (Continued).* **a** A microfluidic chip designed to carry out dual-beam OT force-fluorescence microscopy measurements. The bead, DNA, and buffer channels are separated by laminar flow, without a physical barrier. Multiple large buffer reservoirs (0.5 µL) are connected to the buffer channel by thin (0.25 mm wide) and long (1.5 mm long) necks. These necks ensure separation between the reservoir and the main channel in the absence of flow. **b** A flow chart showing the procedure for performing automated force and/or fluorescence spectroscopy measurements on an individual DNA molecule. The procedure starts by moving the two trapping laser foci to the bead channel in order to trap beads. Once beads of sufficient quality are trapped, they are steered to the DNA channel to tether a single DNA molecule. The qualities of beads and tethered DNA are monitored as indicated, and not passing the checkpoints leads to a restart of the protocol. Dashed lines show that the movement to buffer reservoir and subsequent incubation are optional, depending on the experiment. Force spectroscopy and/or fluorescence imaging are carried out in the buffer channel or a separate buffer reservoir. When measurement termination criteria, e.g. fluorescence intensity, or stretching force, are reached, the procedure is restarted to trap and measure another DNA molecule. This procedure is repeated until the desired replicate number of DNAs is reached, where the experiment is terminated.

In the next sections, we will describe a fully automated data acquisition pipeline suited for the aforementioned sample preparation methods and microfluidic system, which significantly facilitates dual-beam OT-confocal scanning microscopy measurements.

### 4.3.2 Motivation and code structure for automated data acquisition

A dual-beam OT setup can only trap DNA molecules one at a time. The procedure of trapping beads, trapping DNA, and conducting force spectroscopy and/or fluorescence imaging measurements is repeated for every molecule measured. Conventionally, the measurements are carried out by the experimenter using external input devices such as mouse, keyboard, or joystick. The repetitive and extensive human input limits both the experimental throughput and the reproducibility of measurements. Automation of the data acquisition helps address these difficulties by minimizing human input, and additionally frees the researchers from repetitive operations on the instruments. In our group, automated data acquisition can increase experimental throughput of certain force-fluorescence spectroscopy measurements by up to 2 fold.

A few scripts are already available on public repositories (https://harbor.lumicks.com/scripts) to automate basic operations of OT, including catching beads, trapping DNA, and carrying out force spectroscopy measurements. We integrated these existing functions with custom-written scripts that enable automatic collection of confocal scanning data, thereby forming a complete and automated data acquisition pipeline (**Fig. 4.3b**) that can be directly used on the Lumicks C-Trap system with the aforementioned microfluidic chip (**Fig. 4.3a**). The code repository consists of three parts that we describe in turn: beads

and DNA trapping, confocal scanning, and force spectroscopy. A separate *parameters.yml* file is used to input user-specified parameters.

### 4.3.3 Automated beads and DNA trapping

The measurements always start by moving the trapping laser to the bead channel and turning on the flow. On the C-Trap system, the trapped beads are assigned a matching score based on the similarity between the bright field image of the bead and a template image preset by the user. Beads with a matching score below the threshold (*bead_match_threshold,* typically set to 90 out of 100) are discarded, as they are likely to be multiple beads trapped in a single trapping laser focus.

Once two beads with matching scores exceeding the threshold are trapped, they are steered to the DNA channel with the flow kept on. DNA molecules with both ends functionalized are brought in contact with the bead surface by the flow, and initially only one end of the DNA is attached to the beads. As shown in **Fig. S4.1a**, the flow direction is from the lefthand side to the righthand side. Here we focus on the DNA molecules with one end attached to the lefthand bead (Bead 1), which are stretched by the flow. The unrestricted end of the DNA is brought close to the righthand bead (Bead 2), which is repeatedly approached to and moved away from Bead 1 along the *x*-axis. Once a DNA molecule is successfully tethered between the beads (**Fig. S4.1b**), rightward movement of Bead 2 generates tension on the DNA, resulting in a restoring force on the beads that is detected once it exceeds a preset threshold (*force_threshold*). The flow is then turned off, leaving the remaining DNA molecules with only one end attached to the beads in a collapsed, coiled conformation (**Fig. S4.1c**).

The trapped DNA is subsequently steered to the buffer channel or a buffer reservoir (**Fig. 4.3**). Incubation in the buffer reservoir may be carried out depending on the experiment, e.g. for binding proteins to trapped DNA (**Fig. 4.2b**). Prior to fluorescence imaging and/or force spectroscopy measurements, it is necessary to check whether a single DNA molecule is tethered, which is accomplished by comparing the end-to-end extension of the DNA at a specific force with the expected value predicted by the extensible worm-like chain (eWLC) model[41]. While the eWLC model is limited in its applicable force range (5-30 pN), it is convenient to implement in the code and suitable for most experimental conditions. The final step prior to measurements is to ensure that the DNA is maintained at constant values of the *y*- and z-coordinates in the lab frame over its entire length (**Fig. 4.1**). This alignment ensures that both the forces applied to the

DNA and the movement of fluorescently labeled proteins along it have only *x*-components, which significantly simplifies subsequent data analysis.

### 4.3.4 Automated confocal scanning imaging

In automated continuous confocal scanning measurements, it is helpful to monitor the fluorescence intensity of the fluorophores and stop the measurement once the fluorophores of interest are bleached, as this avoids collecting unusable data. For this purpose, the fluorescence intensities are constantly extracted and projected against the *x*-axis of the image (**Fig. S4.2**). Once the mean intensities of the relevant fluorescence signals fall below the respective threshold values set by the user (*intensity_threshold*), the confocal scanning measurement is terminated and the image data is saved. It is worth noting that the images typically contain the beads at the ends of the tethered DNA molecule. This spurious signal from the beads should be excluded from the analysis through the specification of a bead margin (upper panels in **Figure S4.2a, b**).

### 4.3.5 Automated force spectroscopy measurement

Typical constant pulling speed force spectroscopy measurements are carried out by steering one bead away from the other along the *x*-axis at a constant speed, which results in an increasing stretching force on the tethered DNA molecule. The pulling speed, as well as the initial and final end-to-end extensions of the DNA can be specified by the user. In addition, it is also possible to collect repeated forward-reverse force-distance traces on a single DNA molecule until either a preset replicate number is reached, or the DNA tether is broken. Such forward-reverse measurements are particularly useful in studying protein folding dynamics[17].

## 4.4 Quantification of confocal scanning images

### 4.4.1 Data visualization methods

Kymography and full confocal scans are two commonly used approaches to visualize the dynamics of biomolecules monitored by confocal scanning microscopy. Kymography creates a single image with which to visualize a dynamic process by making a stack of the line scans acquired in the confocal scanning area at consecutive time intervals (**Figs. 4.4a** and **S4.3**). This provides an overview of the fluorescent intensities along a line collected over time that benefits from the fast imaging of line scans and attendant high time resolution. Therefore, kymographs are particularly useful for visualizing rapid motion dynamics of

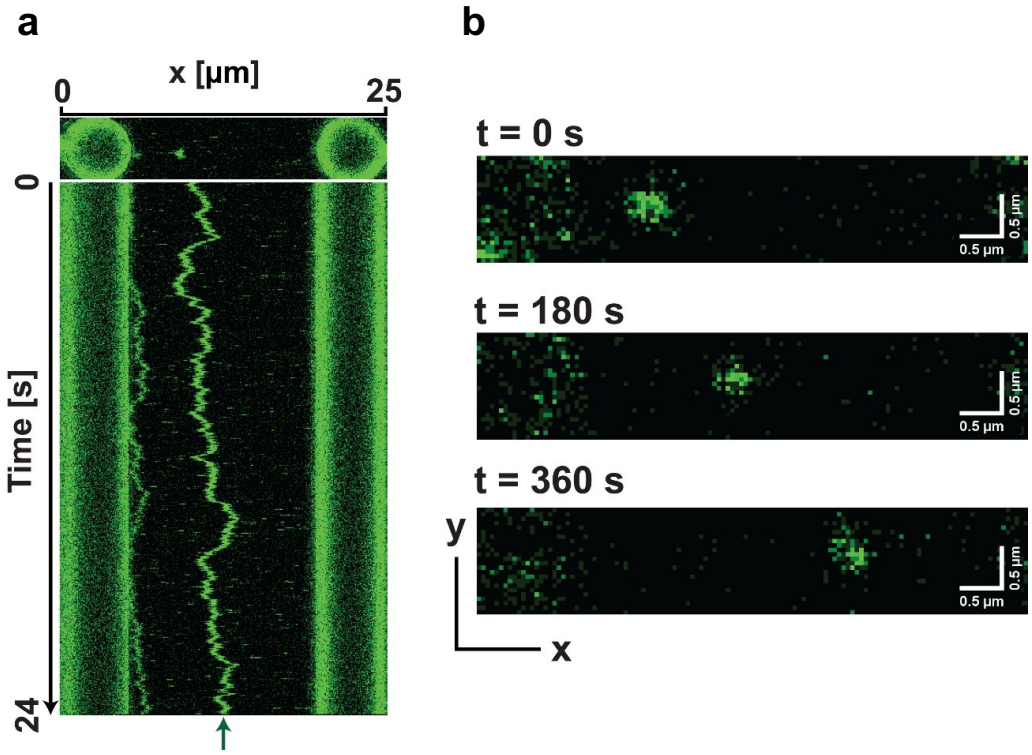



*Figure 4.4* Examples of kymograph and 2D full confocal scanning images collected using dual-beam OT-confocal scanning microscope. **a** A kymograph illustrating the diffusive motion of a fluorescently labeled Mcm2-7 helicase on DNA (green trace indicated by the green arrow). The upper panel shows a full 2D scan at the start of the measurement that includes the beads and the fluorescent Mcm2-7 helicase. The kymograph is constructed by repeating 1D scans along a given line on the x-axis and stacking the scans (**Fig. S4.3**). **b** Three 2D confocal scanning images that sample the x-y plane at the time stamps indicated. The green spot illustrates the unidirectional translocation of a fluorescently labeled CMG holo-helicase[23] on DNA oriented along the x-axis.

proteins on DNA[42–44]. However, for in-depth fluorescent spot motion analysis (discussed in **Motion analysis** section), it is important to localize the fluorophore with high precision. The localization precision of fluorescent spots can be measured by calculating the standard deviation of localizations of a static fluorophore (**Fig. S4.4**). As kymographs consist of only 1D line scans, their localization precision is limited as a result of the loss of information relative to the 2D space. In our dual-beam OT force-fluorescence spectroscopy setup, full 2D confocal scans (**Fig. 4.4b**) yield a ~2 fold higher localization precision compared to 1D scans (**Fig. S4.4**), at the expense of reduced time resolution and increased motion blur. Due to these limitations, 2D confocal scans are most suitable to monitor processes at sub-second or slower time scales. A simulation method to quantify the effect of spot velocity on motion blur is discussed in **Supplementary Method 4.2**. Because one of the main

interests of our confocal scanning data analysis pipeline is the motion analysis of fluorescently labeled proteins bound to DNA, whose accuracy benefits from increased localization precision, the data and analyses presented in the following sections are all based on full 2D confocal scans.

To maximize the insight achievable from tracking the positions of labeled proteins bound to DNA in the instrument, it is necessary to determine the exact correspondence between the fitted pixel position of a fluorophore in a confocal scanning image and its genomic coordinate along the DNA. To achieve this, we must first locate the ends of the DNA in the confocal scanning image and then determine the appropriate conversion of pixels in this image to microns and subsequently to kilobases (kbs).

Starting with the first point, *a priori* it would seem possible to determine the locations of the DNA ends directly from the bead images visible in the 2D confocal scans (**Fig. 4.4b**); however, the spurious fluorescence signal from the edges of the beads makes it challenging to pinpoint their edges precisely and introduces uncertainty into the location of the ends of the DNA molecule. Fortunately, this difficulty can be overcome by determining the bead positions in the brightfield images, where fluorescence signal is not detected. To then map these positions onto the confocal images, we need to quantify the shift between the two sets of images, which will be discussed in the next section.

### 4.4.2 Quantification of brightfield to confocal offset and confocal scanning image pixel size

To quantify the brightfield to confocal offset, we start by collecting a confocal scanning image dataset on a DNA molecule that includes a statically bound fluorescent protein (green dots in **Fig. 4.5a**) located at a known and unique sequence distinct from the DNA center. This can for example be achieved using a fluorescently labeled dCas9[19,45] bound to a specific and unique sequence on the DNA. Because the DNA molecule can be randomly trapped in two opposite orientations in a dual-beam OT-confocal scanning microscope, the collected confocal images will display fluorescent spots in two locations, which we designate $x_{dye,confocal}$ and $x_{dye,confocal,mirrored}$ (**Fig. 4.5a**). Consequently, the center of the tethered DNA molecule in the confocal image, $x_{DNA\ center,confocal}$, is given by $x_{DNA\ center,confocal} = \frac{1}{2}(x_{dye,confocal} + x_{dye,confocal,mirrored})$. Conversely, under the assumption that the two optical traps have identical stiffnesses and the beads trapped are equal in size, the center of the DNA molecule in the brightfield images, $x_{DNA\ center,BF}$, is given by $x_{DNA\ center,BF} = \frac{1}{2}(x_{Bead1,BF} + x_{Bead2,BF})$ (**Fig.**

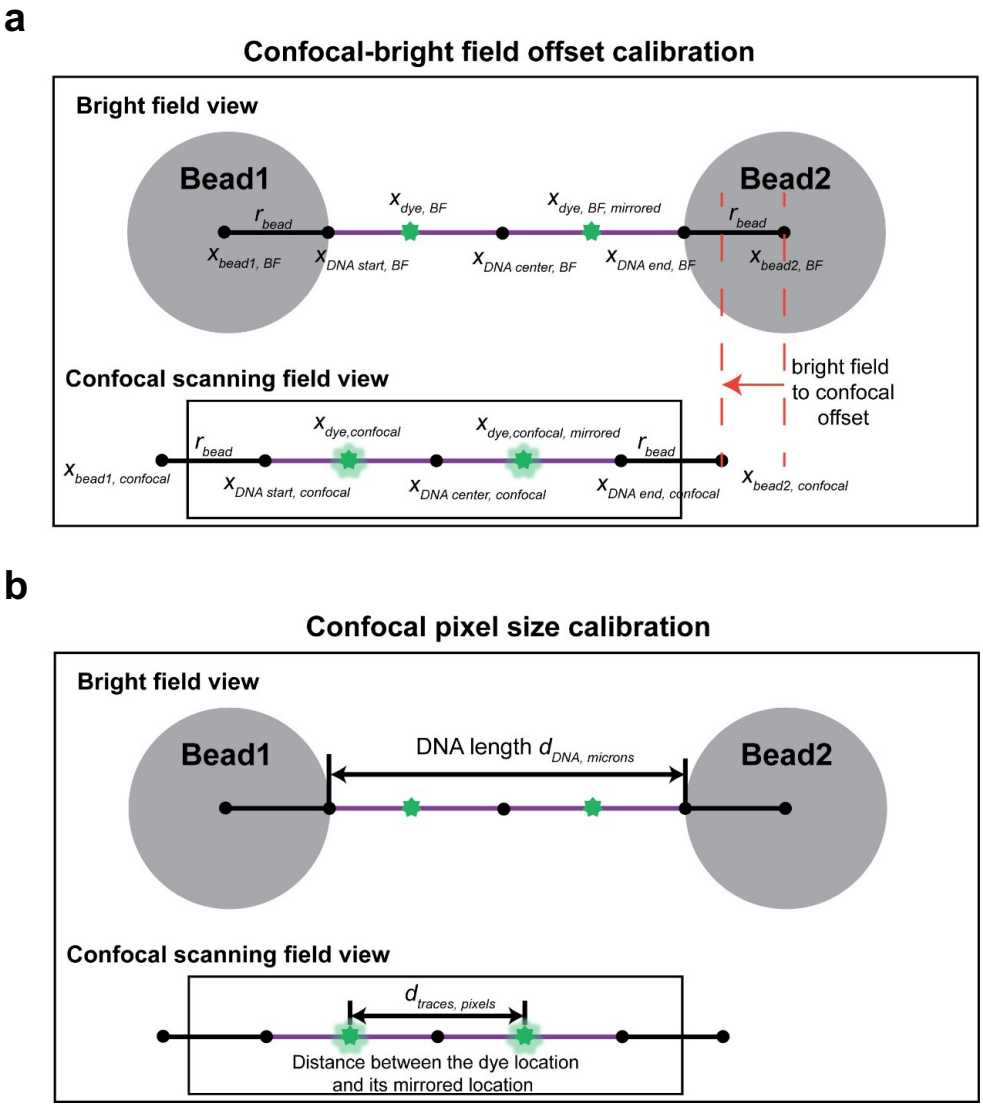



***Figure 5* Use of site-specific fluorophores to determine the genomic locations of fluorescent spots.** Determination of the bead extremity and pixel-to-nm conversion in the confocal scanning images is required for determining the locations of DNA ends. **a** The location of the bead extremity in the brightfield image is mapped onto the corresponding location in the confocal scanning image by calibration of the offset between these two sets of images (indicated by the red dashed lines and arrow). To do so, a static fluorophore is bound to the DNA at a known distance away from the DNA center. This fluorophore will appear at two x-locations in the confocal scanning image (green dots, $x_{dye,confocal}$ and $x_{dye,confocal,mirrored}$). The arithmetic mean between these two locations yields the center of the DNA in this image ($x_{DNA\ center,confocal}$). Note that the fluorophores are not visible in the bright field image. Therefore, the center of the DNA in the brightfield image is given by arithmetic mean between the bead locations $x_{DNA\ center,BF} = \frac{1}{2}(x_{Bead1,BF} + x_{Bead2,BF})$. The offset equals the difference between the DNA center locations in bright field and confocal scanning images. **b** Dividing the distance between the locations of the two fluorophores in the population of traces measured in the confocal scanning images (in pixels) by the length of the DNA measured in brightfield image (in microns) yields the confocal scanning image pixel size.

**4.5a**). In consequence, the offset in the *x*-coordinate between the brightfield and confocal images in microns is given by:

$$x_{shift,micron} = x_{DNA\ center,confocal} - x_{DNA\ center,BF} \qquad \textbf{(Eq. 4.3.1)}$$

We note that the direct readout of the confocal scanning image is in pixels, which nominally have a size (in nanometers) that is set when carrying out the confocal scanning measurements. However, the actual pixel size may differ from this nominal size due to image distortion, and therefore it must be calibrated. To do so, we first determine the distance in pixels for a given stretching force applied on the DNA, $d_{traces,px}(F)$, between the oppositely oriented fluorescent spots described in the previous section. As the genomic coordinates of the underlying fluorescently labeled DNA-bound proteins are known, this distance is also known in kbp and designated $d_{traces,kbp}$ (**Fig. 4.5b**). Dividing these two quantities by each other yields a force-dependent pixel-kbp conversion factor:

$$c_{confocal,kbp/px}(F) = \frac{d_{traces,kbp}}{d_{traces,px}(F)} \qquad \textbf{(Eq. 4.3.2)}$$

Additionally, we know the total length of the DNA in kbp, $d_{DNA,kbp}$, and, from the brightfield image, its end-to-end extension in microns, $d_{DNA,micron}$, for a given applied force on the DNA (**Fig. 4.5b**). This yields a force-dependent micron-kbp conversion factor:

$$c_{BF,micron/kbp}(F) = d_{DNA,micron}(F) / d_{DNA,kbp} \qquad \textbf{(Eq. 4.3.3)}$$

Combining these two quantities, we obtain an overall micron-pixel conversion factor, i.e. the corrected pixel size in microns, which is not force-dependent, as the pixel-kbp and micron-kbp conversion factors are determined at the same force applied on the DNA:

$$c_{confocal,micron/px} = c_{confocal,kbp/px}(F) \times c_{BF,micron/kbp}(F) \qquad \textbf{(Eq. 4.3.4)}$$

With the image offset and pixel size conversion factor $c_{confocal,micron/px}$ known, we can calculate the *x*-coordinates of the edges of the DNA in the confocal image. The bead locations $x_{bead1,BF}$ and $x_{bead2,BF}$ in microns are known from the bright field image (**Fig. 4.5a**), and the bead radius is taken to be a constant, which

is a reasonable assumption given typical vendor specifications, for example, polystyrene particles from Spherotech Inc. (https://www.spherotech.com/pol_par.htm). The DNA start and end locations in the confocal image can be calculated as follows:

$$x_{DNA\ start,confocal,px} = \frac{x_{Bead1,BF} + x_{shift,micron} + r_{bead,micron}}{c_{confocal,micron/px}} \quad \textbf{(Eq. 4.3.5)}$$

$$x_{DNA\ end,confocal,px} = x_{DNA\ start,confocal,px} + \frac{d_{DNA,microns}}{c_{confocal,micron/px}} \quad \textbf{(Eq. 4.3.6)}$$

So finally, for a pixel location measured in the confocal image $x_{confocal,px}$, the corresponding location on the DNA in base pairs can be calculated using:

$$x_{kbp} = c_{confocal,kbp/px} \times (x_{confocal,px} - x_{DNA\ start,confocal,px}) \quad \textbf{(Eq. 4.3.7)}$$

The aforementioned calibration parameters are stored in a file named `config.yml` and an offset correction/tracking parameter file named `params_offset_tracking.yml`, and listed in **Tables S4.1, S4.2**. A full list of optional parameters is provided in the code documentation.

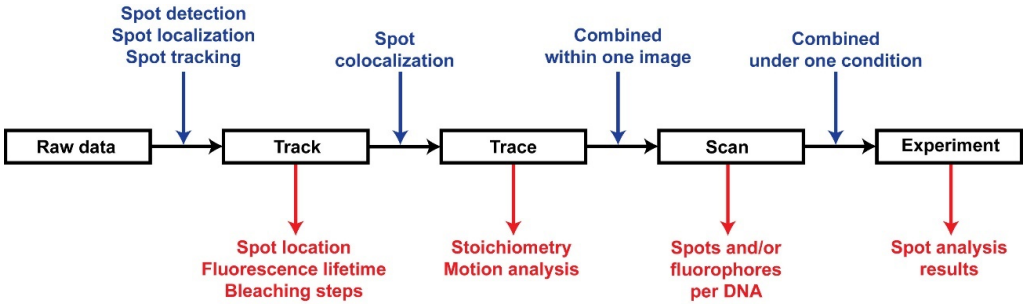



***Figure 4.6*** Flow chart summarizing the data hierarchy of confocal scanning data analysis pipeline and the output at each level. The analysis performed to move to the next level are shown in blue. The information extracted at each level is shown in red. The input data include the raw image data and the associated metadata. Spot detection, localization, and tracking algorithms are used to identify fluorescent spots in the image (Track) of a certain color, providing information on locations, lifetime, and number of bleaching steps of the fluorophores. Colocalized spots (Tracks) of different colors are combined within one Trace, which provide information on stoichiometry and motion properties. Different Traces within one image are stored in the same Scan object, providing information on the number of spots and/or fluorophores on one DNA molecule. All the Scans collected under one experimental condition are stored in one Experiment object, where the spots analysis results are reported in a table.

## 4.5 Confocal scanning data analysis

### 4.5.1 Analysis input and internal data hierarchy

The input data of our confocal scanning data analysis pipeline include: 1) multi-frame .tiff image data that consists of intensities (ADU, integer values) for three colors at every pixel location, and 2) metadata associated with the images. The contents and format of metadata file are specified in the code documentation.

The analysis pipeline extracts the following information from the input data: 1) detection of fluorescent spots, 2) tracking of the motion of fluorescent spots in multi-frame images, 3) colocalization of spots in different colors, and 4) stoichiometries of different colors in fluorescent spots. The aforementioned information is stored in a hierarchical set of classes to keep track of spots in different colors. This set of classes consists of four levels that are designated Track, Trace, Scan, and Experiment, and summarized in **Fig. 4.6**.

The most low-level class is the Track class, which stores information about a tracked spot of a single color, such as a numpy array of the $x$-locations in kbp (`Track.x_kbp`) at each timepoint in seconds (`Track.time_s`). Colocalized Tracks are stored in an upper-level class called Trace. In other words, a Trace might contain multiple Tracks of different colors which are at the same location (within the colocalization distance) on the DNA. All the Traces within one confocal scan image are stored in a Scan object. With the Scan object, we can calculate the number of fluorescent spots as well as the number of fluorophores on the DNA. Finally, the highest-level class, Experiment, contains all Scans that were collected under the same experimental conditions. The technical details of how to detect, track, and analyze spots will be explained below.

### 4.5.2 Detection, localization, and tracking of fluorescent spots

For detecting spots in 2D confocal scanning images, we use the Laplacian of Gaussian (LoG) 'blob detector' implementation from **skimage** (https://scikit-image.org/docs/stable/api/skimage.feature.html#skimage.feature.blob_log). A LoG detector has two user inputs, $\sigma$ and $T_{LoG}$. The detector convolves the image with a Gaussian kernel with standard deviation $\sigma$, after which a Laplacian operator is applied. This results in a strong response for blobs with a radius $r = \sqrt{2}\sigma$, where $r$ is the expected point spread function (PSF) radius. This radius can be set for each color separately and should be approximately half the wavelength. We then look for such responses in the transformed image, selecting any local maxima above the threshold parameter $T_{LoG}$.

After detecting the fluorescent spots using LoG, we next determine the subpixel location of each detected spot by fitting its intensity profile to a 2D Gaussian function with a background noise term (**Supplementary Method 4.3**). The overall fluorescence intensity of a spot is calculated by summing the intensity values for all the pixels that lie within the expected PSF radius $r$, minus the background noise.

For the tracking of fitted spots, the user inputs two values: a maximum frame-to-frame linking distance, and a maximum frame skip value for connecting track segments. In general, these parameters are picked by considering expected diffusion coefficients and fluorophore blinking rates. The Linear Assignment Problem (LAP) framework[46] is used to find out the links between fluorescent spots in different frames. First, track segments are found by listing all possible frame-to-frame connections between fitted spots as elements in a matrix, after which a LAP solver (in our case **scipy**'s linear sum assignment optimizer) is used to find the combination of connections with the lowest cost. Then, another such matrix is made for connections between track segments, using the same solver to find connected segments, resulting in full tracks. All spots that are found to be in the same track are given the same *track_id* in the output table. At this stage, we do not allow for track splitting and/or merging.

The spot detection, localization, and tracking results can be displayed in two ways:

*Full location plots*, showing all spot detections over time, with connections between spots to indicate tracks.

*Histograms of initial locations*, showing spot locations fitted with 2D Gaussian on the *x*-axis and the corresponding counts on the *y*-axis. Here we take the mean location of each trace in the first 5 frames. Bin sizes are usually chosen taking into account the uncertainty in fluorescent spot localization (this section and **Fig. S4.4**) as well as the uncertainty in the localization of the DNA ends (**Quantification of confocal scanning images** section). The trapped DNA has two possible orientations, and in some experiments, it is not possible to know this orientation *a priori*. In this case, a fluorescent molecule bound to a specific sequence on the DNA has two possible locations on the confocal scanning image symmetric with respect to the DNA center (**Quantification of confocal scanning images** section). For such experiments, it makes more sense to plot the distance from the DNA center on the *x*-axis.

### 4.5.3 Fluorescence lifetime and bleaching traces analysis

Subsequent analysis of tracks include fluorescence lifetime analysis and determination of stoichiometry.

*Lifetime analysis*: for each frame, we count the number of fluorophores that are still present in all the tracks in an experiment. This time-*vs*-fluorophore count data can be fitted with a single exponential decay function, yielding a fitted value for the mean lifetime, or with multiple exponential functions (whereby the Bayesian Information Criterion (BIC) can be used to determine the number of components), yielding distinct mean lifetimes per component.

*Stoichiometry determination*: a well-established fluorophore stoichiometry measurement method is to illuminate the fluorophores until photobleaching, plot the fluorescent intensity over time, and count the number bleaching steps[47–49]. To detect bleaching steps, we use change-point analysis (CPA) as implemented in the python library **ruptures**. In CPA we minimize a cost function that includes a penalty term for introducing a step (a 'change-point'). There are a number of cost functions that one can choose from; we use the least squared deviation (CostL2) to detect the shifts in the mean value of the fluorescence intensity in these bleaching traces.

This approach requires two user parameters: a minimum plateau length (in frames) that sets the minimum number of data points that should be in a plateau between two steps, and a minimum step size $\Delta I_{min}$, which can be determined using fluorescently labeled dCas9 (**Supplementary Method 4.4**). We set the CPA penalty to $\Delta I_{min}^{2} \times \log(N_{sig})$, where $N_{sig}$ is the number of data points in the signal. This formula comes from the Bayesian Information Criterion (BIC) (see also Eq. 30 in Ref.[50]).

After CPA with the given minimum plateau length and penalty term, it is still possible to find steps with values below the minimum step size if these steps are deemed significant according to the BIC. This could occur for example because a spot moved slightly out of focus, a fluorophore in the background bleached, or a neighboring spot bleached, causing the signal to decrease slightly but permanently. These incorrectly identified steps are filtered out by *pruning*: we eliminate any steps smaller than the minimum step size, and set the local fit (between the preceding and following step) to the weighted (by length) average of the preceding and following plateaus.

### 4.5.4 Colocalization of tracks and crosstalk correction

Distances between tracks are defined as their average separation over the first 5 frames of the scan. For track colocalization, we again use the LAP framework, where the cost of colocalizing two spots of different colors is equal to the absolute value of their separation up to a user-specified distance threshold $T_{col}$. For distances greater than the threshold, two spots are not considered to be colocalized. If two tracks are colocalized, we call them a trace, and they receive the same *trace_id* index in the output table. Quantities like stoichiometry are stored at the trace level. For example: a trace containing a green track with two bleaching steps, and a red track with one bleaching step, has a stoichiometry of (2 green + 1 red).

To process colocalized spots (fluorophores $f_1$ and $f_2$ with main signals in detection channels $c_1$ and $c_2$, respectively), CTrapPy automatically corrects their intensities by subtracting the noise caused by crosstalk between different channels of fluorescence signal. For example, to correct the intensity in channel 2 for crosstalk from channel 1, we use the relationship $I_{f2,c2,corr} = I_{f2.c2} - I_{f1,c1} \cdot c_{f1,c1\to c2}$, where $c_{f1,c1\to c2}$ is the crosstalk correction factor (i.e. the relative amount of leakage from channel 1 into channel 2 for fluorophore $f_1$). This correction can be performed for any combinations of channels.

### 4.5.5 Motion analysis

In measurements requiring long imaging time (typically >10 min), drift in the spot location over the measurement might occur. To correct the drift, a control experiment is conducted by imaging a static fluorescent spot on the DNA (**Fig. 4.5**), and computing the average velocity of this static spot $v_{drift}$. This average velocity can be used to correct the spot locations in measurements with moving fluorescent spots: $x_{corr}(t) = x_{raw}(t) - v_{drift} \cdot t$. After the drift correction, CTrapPy performs different types of motion analysis.

*Processivity* is simply the trace end location minus the trace start location; a processivity plot is a histogram with processivity on the *x*-axis and trace counts on the *y*-axis.

*Diffusion calculations* are done following previously reported approach[51]. For each trace, the mean squared displacement (MSD) is computed, and the diffusion coefficient is calculated by fitting an optimized number of MSD data points. The baseline for classifying a spot as diffusive can be established by running a static spot dataset (e.g. dCas9) through diffusion analysis; the average diffusion

coefficient over all static traces gives a baseline value for comparison to other tracks.

*Instantaneous velocity analysis* yields a fitted value of the instantaneous velocity at each time point of every trace of interest. Calculating it directly using $v = \Delta x / \Delta t$ would result in very wide velocity distributions as a result of the noise present in location measurements. Instead, we perform a CPA fit to detect linear segments in the velocity (again using ruptures, with the "CostLinear" cost function). This is a way to remove noise and estimate the instantaneous velocity; it also shows how often a trace exhibits velocity changes. Technical details of instantaneous velocity calculation are explained in **Supplementary Method 4.5**.

*Anomalous diffusion analysis* can be performed on spots that exhibit significant motion, i.e., spots with a maximum instantaneous velocity above a certain threshold, usually a multiple of a baseline velocity spread in calibration measurements on static proteins. We followed a previously reported method[52] to perform the anomalous diffusion fit; the anomalous diffusion exponent α is an indicator for motion type ($\alpha \ll 1$: sub-diffusive or constrained diffusion, $\alpha \approx 1$: diffusive, $\alpha \gg 1$, super-diffusion or unidirectional motion). Technical details of anomalous diffusion analysis are explained in **Supplementary Method 4.6.**

### 4.5.6 Spot analysis output

The full Experiment object can be exported as a table. Each row contains a detected spot (with an *x*-location, intensity, color, etc.) at a certain frame, with a `scan_id`, `trace_id`, and `track_id`. These three indices form a unique identifier for each Track in the Experiment, and show which Tracks are colocalized.

Spot tables can be filtered based on stoichiometry, location, and starting frame.

Fluorescent spots (Traces) with large stoichiometries (usually >5 or >10 fluorophore counts, depending on the experiment) are indicative of protein aggregation. These traces are usually not interesting for further analysis and can be filtered out.

If the initial location of a trace is too close to one of the ends of the DNA, this usually indicates that the signals originated from protein bound to the bead, and are of no interest for further analysis. Hence, we filter out traces starting too close to the beads.

Finally, in most experiments we do not expect any proteins to bind to the DNA during the confocal scanning process itself. Hence, traces that start at a later time

during this process are indicative of tracking errors. To prevent false spot counts, traces starting at later time frames (usually >5 frames) are filtered out.

The most important columns in the output table are provided in **Table S4.3**. The full output specification can be found in the code documentation.

## 4.6 Force spectroscopy data analysis

As shown in **Figure 4.1a** , dual-beam OT force spectroscopy measurements are typically performed by steering one bead away from the other one on the *x*-axis at a constant velocity, thus applying stretching forces to the elastic biopolymer (e.g. DNA) trapped between the beads. Such measurements are mostly used for two purposes: 1) to extract the elasticity parameters, such as contour length and persistence length, of biomolecules[53]; 2) to detect protein or DNA conformational changes, such as protein folding/unfolding or DNA wrapping/unwrapping in protein-DNA complexes[54,55].

For the first purpose, we fit the force-distance curves generated by stretching the elastic molecule of interest with elasticity models. Our analysis toolbox provides three elasticity models: the worm-like chain model (WLC)[56], the WLC model with enthalpic corrections[57], and the extensible worm-like chain (eWLC) model[41]. The resulting fitting parameters are: persistence length $L_p$, contour length $L_c$, and stretch modulus $S$ (for the eWLC model). The parameters of each force-distance curve as well as a statistical summary are recorded and can be exported.

For the second purpose, we have developed a contour length increment analysis tool, as protein and DNA conformational changes are typically associated with contour length changes. The force-distance data measured using force spectroscopy is transformed into a contour length space using the eWLC model. For each data point on the force-distance curve, the contour length is calculated and plotted against force. CPA (**Fluorescence lifetime and bleaching traces analysis** section) is used to fit the contour length-force plot and detect contour length changes. The contour length increment, i.e. the difference between the mean contour lengths of the plateaus before and after the change, and the force at which the change happens, are recorded and exported in a summary table. Currently only the eWLC model is available for contour length increment analysis.

## 4.7 Conclusions and outlook

The power of integrated dual-beam OT-confocal scanning force-fluorescence spectroscopy measurements has been demonstrated in the investigation of various biological systems[58–61]. To make the best use of this powerful tool, careful preparation of biological samples, proper experimental procedures, and rigorous data analysis are all crucial. Here, we have presented a versatile data acquisition and analysis pipeline designed for investigating protein:DNA interactions using OT-confocal scanning microscope. We discussed different sample preparation strategies for assembling protein:DNA complexes in bulk and/or at the single-molecule level, and the biological samples for which these approaches are most suitable. In addition, we designed a microfluidic chip that allows long incubation of the trapped nucleic acid substrate in a buffer reservoir well-separated from the other channels without a physical barrier (**Sample preparation for investigating protein:DNA interactions** section and **Supplementary Method 4.1**). The aforementioned sample preparation and handling methods form the basis of high-quality data collection. The data acquisition pipeline is further complemented by an experiment automation protocol, which improves experimental reproducibility and throughput.

The data processing pipeline that we developed provides a number of functions to extract spatiotemporal information from confocal scanning images, where we focus on the analysis of fluorescent spots in 2D full confocal scanning images. A number of calibration and filtering steps are taken to remove spurious data, and various analysis tools are used to extract information from the collected data and inform about the dynamics of measured biomolecules. Different uses and approaches to the analysis of force spectroscopy data were briefly discussed.

The methods discussed in this article have been demonstrated in a number of studies investigating the dynamics of various protein:DNA interaction systems[19,22,62]. However, it is worth noting that these tools can be adapted in many other research topics using the same technique, including but not limited to protein folding and dynamics of nucleic acid secondary structures. The broad single-molecule biophysics community could also benefit from the data processing tools presented here, as confocal scanning microscopy and force spectroscopy techniques are also widely used in a number of other single-molecule measurements, e.g. AFM or MT single-molecule force spectroscopy.

## 4.8 Author contributions

Z.L., E.v.V., and N.D. conceived the article. Z.L., E.v.V., H.S., B.S., D.R.M., and N.D drafted the article. All authors reviewed the manuscript and provided input. E.v.V. and F.P.M. developed and maintained the associated Python codes with input from Z.L., H.S., B.S., K.M., D.R.M, and N.D. H.S. and B.S. designed the microfluidic flow cell. Z.L, H.S., B.S., K.M., D.R.M, and T.v.L. performed the biochemical and biophysical experiments, supervised by N.D.

## 4.9 Acknowledgments

N.D. acknowledges funding from the Netherlands Organization for Scientific Research (NWO) through Top grant 714.017.002 and a Spinoza Prize, from the Dutch Foundation on Fundamental Research on Matter (part of NWO) through grant 16PR1047, from 'BaSyC—Building a Synthetic Cell' Gravitation grant (024.003.019) of the Netherlands Ministry of Education, Culture and Science (OCW), and from the European Research Council through an Advanced Grant (REPLICHROMA; grant number 789267). Z.L. acknowledges funding from EMBO Postdoctoral Fellowship (grant number ALTF 484-2022). D.R.M. acknowledges funding from a Boehringer Ingelheim Fonds PhD fellowship. K.M. acknowledges funding from the Marie Curie-Skłodowska Actions program of the European Union (TOPOREF; grant number REP-800488). The authors thank Andrea Candelli for help in designing the microfluidic flow cell, Aafke van den Berg for help in experiment automation, John Diffley for discussions of biochemical experiments, Carlas Smith for discussions of localization precision, and Pang Yen Wang for discussions and assistance in dual-beam OT-confocal scanning microscope experiments.

## 4.10 Supplementary methods

### Supplementary Method 4.1: microfluidic chip design

In a typical dual-beam OT-confocal fluorescence imaging measurement, the trapping lasers are used to trap two beads, after which one traps an individual DNA molecule between them. The resulting bead-DNA complex is then steered towards a separate compartment to load proteins onto the DNA, and/or to carry out force spectroscopy or fluorescence imaging measurements. To avoid intermixing between the bead, DNA, and buffer solutions, it is highly preferable to maintain

them in distinct compartments. A widely used compartmentalization strategy is to create distinct channels using laminar flow[24,63], as this obviates the introduction of physical barriers (see the bead, nucleic acid, and buffer channels in **Figure 3A**). While highly convenient, this approach does come with the limitation that the flow has to be maintained to ensure separation between the channels during long measurements or incubation. Furthermore, constant flow increases the shear stress on the DNA substrate, which facilitates dissociation of protein:DNA complexes, and can complicate force spectroscopy measurements.

To overcome these issues, we have designed a flow cell with relatively large buffer reservoirs (~0.5 µL) separated from the main channel by a long thin connecting neck (1.5 mm long, 0.25 mm wide, **Fig. 4.3a**) to reduce the rate of diffusion between them. This allows the user to turn off the flow in all channels and carry out more precise imaging and force spectroscopy measurements there. The inclusion of several such reservoirs provides the flexibility to include different storage reservoirs for reagents used in different incubation and/or imaging steps. For example, a protein can be loaded onto the tethered DNA molecule in one reservoir at the single-molecule level (**Sample preparation for investigating protein:DNA interactions** section), and subsequent measurements can be carried out in a different reservoir, if necessary. In our instrument, steering the DNA molecule from one reservoir to an adjacent reservoir by moving the sample stage relative to the trapped beads takes ~30 s (moving at a speed of 0.2 mm/s). At this movement speed, the drag force applied by the surrounding fluid on the beads (1.7 µm diameter) is estimated to be around 3 pN, which is sufficiently low to present minimal influence on the stability of the beads in the OT and the DNA molecules trapped between them.

## Supplementary Method 4.2: simulation of blurred point spread function

To quantify the motion blurring occurring in 2D confocal scan images, we have simulated the scan process based on the following parameters:

| Scan length $L$ [px] | Scan width $H$ [px] | Pixel size [nm] | Pixel time $t$ [ms] |
|---|---|---|---|
| 124 | 24 | 50 | 0.2 |

Our simulations show that the blurring produces the elongation (in the moving direction of the fluorescent spot) and rotation of the PSF function. These two quantities are our metrics to determine the motion blurring. We use a rotated Gaussian fitting function to measure them:

$$g\,(x,y) = Aexp[-a(x\,-x_0)^2\,+2b(x\,-x_0)(y\,-y_0)+c(y\,-\,y_0)^2]+B$$

$$a = \frac{1}{2}\left(\frac{cos^2\theta}{2\sigma_x^2}+\frac{sin^2\theta}{2\sigma_y^2}\right)$$

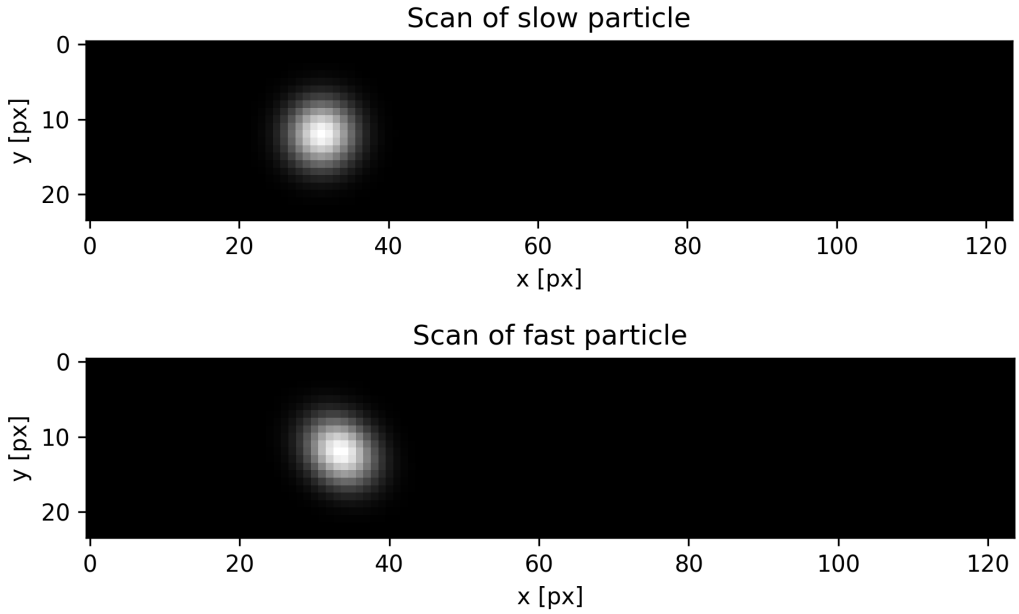
$$b = \frac{1}{2}\left(\frac{sin2\theta}{4\sigma_y^2}+\frac{cos2\theta}{4\sigma_x^2}\right)$$

$$c = \frac{1}{2}\left(\frac{sin^2\theta}{2\sigma_x^2}+\frac{cos^2\theta}{2\sigma_y^2}\right)$$

Here, $A$ represents the amplitude, $B$ the noise background, ($x_0$, $y_0$) the center coordinates, ($\sigma_x$ ,$\sigma_y$ ) the standard deviations, and $\theta$ the rotation angle.

For example, we simulated the unidirectional movement of a fluorescent spot at 1.7 nm/s (corresponding to a eukaryotic replicative DNA helicase)[64], which gives rise to negligible motion blur. In contrast, a fast moving fluorescent spot (200 fold faster, 340 nm/s), gives rise to a clearly rotated Gaussian point spread function, as shown in the table and figures below.

| $v$ [nm/s] | $\sigma$ [px] | $\sigma_y$ [px] | $\sigma_x$ [px] | $\theta$ [deg] |
|---|---|---|---|---|
| 1. 7 | 3 | 3.0015 | 2.9986 | - |
| 340 | 3 | 3.3195 | 2.7149 | 41.93 |



## Supplementary Method 4.3: fitting of spot intensity profile

The intensity profiles of detected fluorescence spots are fitted by 2D Gaussian function with a background term:

$$g\,(x,y) = Aexp\left[-\frac{(x\ -x_0)^2}{2\sigma_x^2}-\frac{(y\ -y_0)^2}{2\sigma_y^2}\right]+B$$

where the fitted parameters are:

$A$: The amplitude (intensity value corresponding to the mean position) of the detected spot.

($x_0$, $y_0$): The $x$ and $y$ coordinates of the detected spot, respectively.

($\sigma_x$ ,$\sigma_y$ ): The standard deviation in $x$ and $y$ of the detected spot, respectively.

$B$: The intensity correction due to background noise.

The estimated intensity value of any spot corresponds then to the sum of photons in a circular mask (given a detection radius), taking into account the background noise $B$.

## Supplementary Method 4.4: fluorophore bleaching step size calibration

To calibrate the minimum and average bleaching step size ($\Delta I_{min}$ and $\Delta I_{avg}$, respectively) of a fluorophore, we typically singly label dCas9 with this fluorophore, attach the labeled dCas9 to a unique location on the DNA molecule through the use of an appropriately selected guide RNA, and continuously image the dCas9-bound fluorophore until it bleaches. In this case, provided that the fluorophore is well-behaved, the overwhelming majority of fluorescent spots should exhibit one bleaching step. Quantification of its associated mean intensity provides an excellent basis for starting the calibration process. The user inputs an initial guess for the minimal step size $\Delta I_{min}$ for the dCas9 dataset, after which CTrapPy generates fitted intensity plots and a step size distribution graph. The step size distribution has a mean $\Delta I_{avg}$ and a standard deviation $\sigma_{\Delta I}$. The user updates the initial guess until two conditions are met:

- A large majority of the fitted intensity plots look good upon visual inspection;

- $\Delta I_{min} \leq \Delta I_{avg} - 2\sigma_{\Delta I}$, meaning that at least 95% of bleaching steps will be larger than the minimum step size.

In the case of multi-color measurements, it is important to check the crosstalk between different channels, and correct the fluorescence intensity taking the crosstalk into consideration. The method to correct for crosstalk is described in **Colocalization of tracks and intensity correction** section.

## Supplementary Method 4.5: piecewise linear motion fit

First, a Kalman filter is applied to the spot locations of each trace using the pykalman Python library, in order to reduce noise. The resulting locations are fit to linear segments using CPA, with a linear cost function, to fit multiple linear segments. The minimum segment size and penalty term can be set by the user After the fit, every detected spot has an associated fitted velocity $v_{CPA}$, which has lower noise than an instantaneous velocity calculated by simply subtracting subsequent spot locations. Of course, the piecewise linear fit only makes sense for static traces (with or without drift), and traces exhibiting piecewise linear motion.

The mean velocity $\mu_v$ of $v_{CPA}$ for all traces in the static calibration dataset can be used for drift correction, using $x_{corrected}(t) = x(t) - t \cdot \mu_v$. This drift correction is not necessary if $t_{image} \cdot \mu_v \ll \sigma_x$, where $\sigma_x$ is the location error. The standard deviation $\sigma_v$ of the velocity distribution can be used as a cutoff to determine if a spot is static or is undergoing motion.

## Supplementary Method 4.6: anomalous diffusion analysis

After MSD analysis on the drift-corrected traces, we can fit an anomalous diffusion exponent α. This exponent gives us a characterization of motion type for each trace. We start with:

$$MSD(\tau) = D_\alpha \tau^\alpha + 2\sigma_x^2$$

where $D_\alpha$ is the anomalous diffusion constant, $\tau$ is the lag time and $\sigma_x$ is the location error. $\alpha \ll 1$ corresponds to spots exhibiting confined diffusion; for diffusing particles we expect $\alpha \approx 1$; $\alpha \gg 1$ indicates particles undergoing unidirectional motion. The actual fit of α is done using the formula

$$log(MSD(\tau) - 2\sigma_x^2) = log(D_\alpha) + \alpha log(\tau^\alpha),$$

so that we can use linear regression. We use least squares to fit up to a maximum lag time corresponding to one third of the length of the actual fit. After

that, the MSD generally exhibits fluctuations that are too large to perform reliable fitting.

## 4.11 References


(1) Neuman, K. C.; Nagy, A. Single-Molecule Force Spectroscopy: Optical Tweezers, Magnetic Tweezers and Atomic Force Microscopy. *Nat. Methods* **2008**, *5* (6), 491–505. https://doi.org/10.1038/nmeth.1218.

(2) Sumbul, F.; Rico, F. Single-Molecule Force Spectroscopy: Experiments, Analysis, and Simulations. *Methods Mol. Biol.* **2019**, *1886*, 163–189. https://doi.org/10.1007/978-1-4939-8894-5_9.

(3) Bustamante, C. J.; Chemla, Y. R.; Liu, S.; Wang, M. D. Optical Tweezers in Single-Molecule Biophysics. *Nat. Rev. Methods Primers* **2021**, *1* (1). https://doi.org/10.1038/s43586-021-00021-6.

(4) Choi, H.-K.; Kim, H. G.; Shon, M. J.; Yoon, T.-Y. High-Resolution Single-Molecule Magnetic Tweezers. *Annu. Rev. Biochem.* **2022**, *91* (1), 33–59. https://doi.org/10.1146/annurev-biochem-032620-104637.

(5) Yang, B.; Liu, Z.; Liu, H.; Nash, M. A. Next Generation Methods for Single-Molecule Force Spectroscopy on Polyproteins and Receptor-Ligand Complexes. *Front Mol Biosci* **2020**, *7*, 85. https://doi.org/10.3389/fmolb.2020.00085.

(6) Shashkova, S.; Leake, M. C. Single-Molecule Fluorescence Microscopy Review: Shedding New Light on Old Problems. *Biosci. Rep.* **2017**, *37* (4), BSR20170031. https://doi.org/10.1042/bsr20170031.

(7) Plochberger, B.; Röhrl, C.; Preiner, J.; Rankl, C.; Brameshuber, M.; Madl, J.; Bittman, R.; Ros, R.; Sezgin, E.; Eggeling, C.; Hinterdorfer, P.; Stangl, H.; Schütz, G. J. HDL Particles Incorporate into Lipid Bilayers – a Combined AFM and Single Molecule Fluorescence Microscopy Study. *Sci. Rep.* **2017**, *7* (1). https://doi.org/10.1038/s41598-017-15949-7.

(8) Fukuda, S.; Uchihashi, T.; Iino, R.; Okazaki, Y.; Yoshida, M.; Igarashi, K.; Ando, T. High-Speed Atomic Force Microscope Combined with Single-Molecule Fluorescence Microscope. *Rev. Sci. Instrum.* **2013**, *84* (7), 073706. https://doi.org/10.1063/1.4813280.

(9) van Loenhout, M. T. J.; de Grunt, M. V.; Dekker, C. Dynamics of DNA Supercoils. *Science* **2012**, *338* (6103), 94–97. https://doi.org/10.1126/science.1225810.

(10) Gunn, K. H.; Marko, J. F.; Mondragón, A. An Orthogonal Single-Molecule Experiment Reveals Multiple-Attempt Dynamics of Type IA Topoisomerases. *Nat. Struct. Mol. Biol.* **2017**, *24* (5), 484–490. https://doi.org/10.1038/nsmb.3401.

(11) Kemmerich, F. E.; Swoboda, M.; Kauert, D. J.; Grieb, M. S.; Hahn, S.; Schwarz, F. W.; Seidel, R.; Schlierf, M. Simultaneous Single-Molecule Force and Fluorescence Sampling of DNA Nanostructure Conformations Using Magnetic

Tweezers. *Nano Lett.* **2016**, *16* (1), 381–386. https://doi.org/10.1021/acs.nanolett.5b03956.

(12) Aldag, P.; Rutkauskas, M.; Madariaga-Marcos, J.; Songailiene, I.; Sinkunas, T.; Kemmerich, F. E.; Kauert, D. J.; Siksnys, V.; Seidel, R. Dynamic Interplay between Target Search and Recognition for the Cascade Surveillance Complex of Type I-E CRISPR-Cas Systems. *bioRxiv*, 2022. https://doi.org/10.1101/2022.12.18.520913.

(13) Hohng, S.; Zhou, R.; Nahas, M. K.; Yu, J.; Schulten, K.; Lilley, D. M. J.; Ha, T. Fluorescence-Force Spectroscopy Maps Two-Dimensional Reaction Landscape of the Holliday Junction. *Science* **2007**, *318* (5848), 279–283. https://doi.org/10.1126/science.1146113.

(14) Ishijima, A.; Kojima, H.; Funatsu, T.; Tokunaga, M.; Higuchi, H.; Tanaka, H.; Yanagida, T. Simultaneous Observation of Individual ATPase and Mechanical Events by a Single Myosin Molecule during Interaction with Actin. *Cell* **1998**, *92* (2), 161–171. https://doi.org/10.1016/s0092-8674(00)80911-3.

(15) Comstock, M. J.; Ha, T.; Chemla, Y. R. Ultrahigh-Resolution Optical Trap with Single-Fluorophore Sensitivity. *Nat. Methods* **2011**, *8* (4), 335–340. https://doi.org/10.1038/nmeth.1574.

(16) Cole, R. W.; Jinadasa, T.; Brown, C. M. Measuring and Interpreting Point Spread Functions to Determine Confocal Microscope Resolution and Ensure Quality Control. *Nat. Protoc.* **2011**, *6* (12), 1929–1941. https://doi.org/10.1038/nprot.2011.407.

(17) Bustamante, C.; Alexander, L.; Maciuba, K.; Kaiser, C. M. Single-Molecule Studies of Protein Folding with Optical Tweezers. *Annu. Rev. Biochem.* **2020**, *89* (1), 443–470. https://doi.org/10.1146/annurev-biochem-013118-111442.

(18) Comstock, M. J.; Whitley, K. D.; Jia, H.; Sokoloski, J.; Lohman, T. M.; Ha, T.; Chemla, Y. R. Protein Structure. Direct Observation of Structure-Function Relationship in a Nucleic Acid-Processing Enzyme. *Science* **2015**, *348* (6232), 352–354. https://doi.org/10.1126/science.aaa0130.

(19) Ramírez Montero, D.; Sánchez, H.; van Veen, E.; van Laar, T.; Solano, B.; Diffley, J. F. X.; Dekker, N. H. Nucleotide Binding Halts Diffusion of the Eukaryotic Replicative Helicase during Activation. *Nat. Commun.* **2023**, *14* (1), 2082. https://doi.org/10.1038/s41467-023-37093-9.

(20) Anand, R.; Buechelmaier, E.; Belan, O.; Newton, M.; Vancevska, A.; Kaczmarczyk, A.; Takaki, T.; Rueda, D. S.; Powell, S. N.; Boulton, S. J. HELQ Is a Dual-Function DSB Repair Enzyme Modulated by RPA and RAD51. *Nature* **2022**, *601* (7892), 268–273. https://doi.org/10.1038/s41586-021-04261-0.

(21) Kaczmarczyk, A. P.; Déclais, A.-C.; Newton, M. D.; Boulton, S. J.; Lilley, D. M. J.; Rueda, D. S. Search and Processing of Holliday Junctions within Long DNA by Junction-Resolving Enzymes. *Nat. Commun.* **2022**, *13* (1), 5921. https://doi.org/10.1038/s41467-022-33503-6.

(22) Sánchez, H.; Liu, Z.; van Veen, E.; van Laar, T.; Diffley, J. F. X.; Dekker, N. H. A Chromatinized Origin Reduces the Mobility of ORC and MCM through

Interactions and Spatial Constraint. *Nat. Commun.* **2023**, *14* (1), 6735. https://doi.org/10.1038/s41467-023-42524-8.

(23) Langston, L. D.; Zhang, D.; Yurieva, O.; Georgescu, R. E.; Finkelstein, J.; Yao, N. Y.; Indiani, C.; O'Donnell, M. E. CMG Helicase and DNA Polymerase ε Form a Functional 15-Subunit Holoenzyme for Eukaryotic Leading-Strand DNA Replication. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (43), 15390–15395. https://doi.org/10.1073/pnas.1418334111.

(24) Candelli, A.; Wuite, G. J. L.; Peterman, E. J. G. Combining Optical Trapping, Fluorescence Microscopy and Micro-Fluidics for Single Molecule Studies of DNA-Protein Interactions. *Phys. Chem. Chem. Phys.* **2011**, *13* (16), 7263–7272. https://doi.org/10.1039/c0cp02844d.

(25) Sirinakis, G.; Ren, Y.; Gao, Y.; Xi, Z.; Zhang, Y. Combined Versatile High-Resolution Optical Tweezers and Single-Molecule Fluorescence Microscopy. *Rev. Sci. Instrum.* **2012**, *83* (9), 093708. https://doi.org/10.1063/1.4752190.

(26) Wilson, T. Resolution and Optical Sectioning in the Confocal Microscope. *J. Microsc.* **2011**, *244* (2), 113–121. https://doi.org/10.1111/j.1365-2818.2011.03549.x.

(27) Jonkman, J.; Brown, C. M.; Wright, G. D.; Anderson, K. I.; North, A. J. Tutorial: Guidance for Quantitative Confocal Microscopy. *Nat. Protoc.* **2020**, *15* (5), 1585–1611. https://doi.org/10.1038/s41596-020-0313-9.

(28) Heller, I.; Hoekstra, T. P.; King, G. A.; Peterman, E. J. G.; Wuite, G. J. L. Optical Tweezers Analysis of DNA-Protein Complexes. *Chem. Rev.* **2014**, *114* (6), 3087–3119. https://doi.org/10.1021/cr4003006.

(29) Newton, M. D.; Taylor, B. J.; Driessen, R. P. C.; Roos, L.; Cvetesic, N.; Allyjaun, S.; Lenhard, B.; Cuomo, M. E.; Rueda, D. S. DNA Stretching Induces Cas9 Off-Target Activity. *Nat. Struct. Mol. Biol.* **2019**, *26* (3), 185–192. https://doi.org/10.1038/s41594-019-0188-z.

(30) Gruszka, D. T.; Xie, S.; Kimura, H.; Yardimci, H. Single-Molecule Imaging Reveals Control of Parental Histone Recycling by Free Histones during DNA Replication. *Science Advances* **2021**, *6*, eabc0330. https://doi.org/10.1126/sciadv.abc0330.

(31) Schauer, G. D.; Spenkelink, L. M.; Lewis, J. S.; Yurieva, O.; Mueller, S. H.; van Oijen, A. M.; O'Donnell, M. E. Replisome Bypass of a Protein-Based R-Loop Block by Pif1. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (48), 30354–30361. https://doi.org/10.1073/pnas.2020189117.

(32) 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. Tunability of DNA Polymerase Stability during Eukaryotic DNA Replication. *Mol. Cell* **2020**, *77* (1), 17-25.e5. https://doi.org/10.1016/j.molcel.2019.10.005.

(33) Bell, N. A. W.; Molloy, J. E. Efficient Golden Gate Assembly of DNA Constructs for Single Molecule Force Spectroscopy and Imaging. *Nucleic Acids Res.* **2022**, *50* (13), e77. https://doi.org/10.1093/nar/gkac300.

(34) Burnham, D. R.; Kose, H. B.; Hoyle, R. B.; Yardimci, H. The Mechanism of DNA Unwinding by the Eukaryotic Replicative Helicase. *Nat. Commun.* **2019**, *10* (1). https://doi.org/10.1038/s41467-019-09896-2.

(35) Daniel Ramírez Montero, Zhaowei Liu, and Nynke H. Dekker. De Novo Fabrication of Custom-Sequence Plasmids for the Synthesis of Long DNA Constructs with Extrahelical Features. *Biophysical Journal*. **2024**, 123 (1), 31-41. http://doi.org/10.1016/j.bpj.2023.11.008

(36) Rico, F.; Russek, A.; González, L.; Grubmüller, H.; Scheuring, S. Heterogeneous and Rate-Dependent Streptavidin-Biotin Unbinding Revealed by High-Speed Force Spectroscopy and Atomistic Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (14), 6594–6601. https://doi.org/10.1073/pnas.1816909116.

(37) Merkel, R.; Nassoy, P.; Leung, A.; Ritchie, K.; Evans, E. Energy Landscapes of Receptor–Ligand Bonds Explored with Dynamic Force Spectroscopy. *Nature* **1999**, *397* (6714), 50–53. https://doi.org/10.1038/16219.

(38) Neuert, G.; Albrecht, C.; Pamir, E.; Gaub, H. E. Dynamic Force Spectroscopy of the Digoxigenin-Antibody Complex. *FEBS Lett.* **2006**, *580* (2), 505–509. https://doi.org/10.1016/j.febslet.2005.12.052.

(39) Ros, R.; Schwesinger, F.; Anselmetti, D.; Kubon, M.; Schäfer, R.; Plückthun, A.; Tiefenauer, L. Antigen Binding Forces of Individually Addressed Single-Chain Fv Antibody Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (13), 7402–7405. https://doi.org/10.1073/pnas.95.13.7402.

(40) Douglas, M. E.; Ali, F. A.; Costa, A.; Diffley, J. F. X. The Mechanism of Eukaryotic CMG Helicase Activation. *Nature* **2018**, *555* (7695), 265–268. https://doi.org/10.1038/nature25787.

(41) Odijk, T. Stiff Chains and Filaments under Tension. *Macromolecules* **1995**, *28* (20), 7016–7018. https://doi.org/10.1021/ma00124a044.

(42) Kono, S.; van den Berg, A.; Simonetta, M.; Mukhortava, A.; Garman, E. F.; Tessmer, I. Resolving the Subtle Details of Human DNA Alkyltransferase Lesion Search and Repair Mechanism by Single-Molecule Studies. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119* (11), e2116218119. https://doi.org/10.1073/pnas.2116218119.

(43) Heller, I.; Sitters, G.; Broekmans, O. D.; Farge, G.; Menges, C.; Wende, W.; Hell, S. W.; Peterman, E. J. G.; Wuite, G. J. L. STED Nanoscopy Combined with Optical Tweezers Reveals Protein Dynamics on Densely Covered DNA. *Nat. Methods* **2013**, *10* (9), 910–916. https://doi.org/10.1038/nmeth.2599.

(44) Avellaneda, M. J.; Franke, K. B.; Sunderlikova, V.; Bukau, B.; Mogk, A.; Tans, S. J. Processive Extrusion of Polypeptide Loops by a Hsp100 Disaggregase. *Nature* **2020**, *578* (7794), 317–320. https://doi.org/10.1038/s41586-020-1964-y.

(45) Deng, W.; Shi, X.; Tjian, R.; Lionnet, T.; Singer, R. H. CASFISH: CRISPR/Cas9-Mediated in Situ Labeling of Genomic Loci in Fixed Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (38), 11870–11875. https://doi.org/10.1073/pnas.1515692112.

(46) Jaqaman, K.; Loerke, D.; Mettlen, M.; Kuwata, H.; Grinstein, S.; Schmid, S. L.; Danuser, G. Robust Single-Particle Tracking in Live-Cell Time-Lapse

Sequences. *Nat. Methods* **2008**, *5* (8), 695–702. https://doi.org/10.1038/nmeth.1237.
(47) McGuire, H.; Aurousseau, M. R. P.; Bowie, D.; Blunck, R. Automating Single Subunit Counting of Membrane Proteins in Mammalian Cells. *J. Biol. Chem.* **2012**, *287* (43), 35912–35921. https://doi.org/10.1074/jbc.M112.402057.
(48) Das, S. K.; Darshi, M.; Cheley, S.; Wallace, M. I.; Bayley, H. Membrane Protein Stoichiometry Determined from the Step-Wise Photobleaching of Dye-Labelled Subunits. *Chembiochem* **2007**, *8* (9), 994–999. https://doi.org/10.1002/cbic.200600474.
(49) Tsekouras, K.; Custer, T. C.; Jashnsaz, H.; Walter, N. G.; Steve Pressé. A Novel Method to Accurately Locate and Count Large Numbers of Steps by Photobleaching. *Mol. Biol. Cell* **2016**, *27* (22), 3601–3615. https://doi.org/10.1091/mbc.e16-06-0404.
(50) Truong, C.; Oudre, L.; Vayatis, N. Selective Review of Offline Change Point Detection Methods. *Signal Processing* **2020**, *167* (107299), 107299. https://doi.org/10.1016/j.sigpro.2019.107299.
(51) Michalet, X. Mean Square Displacement Analysis of Single-Particle Trajectories with Localization Error: Brownian Motion in an Isotropic Medium. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2010**, *82* (4 Pt 1), 041914. https://doi.org/10.1103/PhysRevE.82.041914.
(52) Kepten, E.; Weron, A.; Sikora, G.; Burnecki, K.; Garini, Y. Guidelines for the Fitting of Anomalous Diffusion Mean Square Displacement Graphs from Single Particle Tracking Experiments. *PLoS One* **2015**, *10* (2), e0117722. https://doi.org/10.1371/journal.pone.0117722.
(53) Broekmans, O. D.; King, G. A.; Stephens, G. J.; Wuite, G. J. L. DNA Twist Stability Changes with Magnesium(2+) Concentration. *Phys. Rev. Lett.* **2016**, *116* (25). https://doi.org/10.1103/physrevlett.116.258102.
(54) Díaz-Celis, C.; Cañari-Chumpitaz, C.; Sosa, R. P.; Castillo, J. P.; Zhang, M.; Cheng, E.; Chen, A. Q.; Vien, M.; Kim, J.; Onoa, B.; Bustamante, C. Assignment of Structural Transitions during Mechanical Unwrapping of Nucleosomes and Their Disassembly Products. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119* (33), e2206513119. https://doi.org/10.1073/pnas.2206513119.
(55) Jahn, M.; Buchner, J.; Hugel, T.; Rief, M. Folding and Assembly of the Large Molecular Machine Hsp90 Studied in Single-Molecule Experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (5), 1232–1237. https://doi.org/10.1073/pnas.1518827113.
(56) Bustamante, C.; Marko, J. F.; Siggia, E. D.; Smith, S. Entropic Elasticity of Lambda-Phage DNA. *Science* **1994**, *265* (5178), 1599–1600. https://doi.org/10.1126/science.8079175.
(57) Bouchiat, C.; Wang, M. D.; Allemand, J.-F.; Strick, T.; Block, S. M.; Croquette, V. Estimating the Persistence Length of a Worm-like Chain Molecule from Force-Extension Measurements. *Biophys. J.* **1999**, *76* (1), 409–413. https://doi.org/10.1016/s0006-3495(99)77207-3.
(58) Spakman, D.; Clement, T. V. M.; Biebricher, A. S.; King, G. A.; Singh, M. I.; Hickson, I. D.; Peterman, E. J. G.; Wuite, G. J. L. PICH Acts as a Force-Dependent

Nucleosome Remodeler. *Nat. Commun.* **2022**, *13* (1), 7277. https://doi.org/10.1038/s41467-022-35040-8.

(59) Tanasie, N.-L.; Gutiérrez-Escribano, P.; Jaklin, S.; Aragon, L.; Stigler, J. Stabilization of DNA Fork Junctions by Smc5/6 Complexes Revealed by Single-Molecule Imaging. *Cell Rep.* **2022**, *41* (10), 111778. https://doi.org/10.1016/j.celrep.2022.111778.

(60) Schaich, M. A.; Schnable, B. L.; Kumar, N.; Roginskaya, V.; Jakielski, R. C.; Urban, R.; Zhong, Z.; Kad, N. M.; Van Houten, B. Single-Molecule Analysis of DNA-Binding Proteins from Nuclear Extracts (SMADNE). *Nucleic Acids Res.* **2023**, *51* (7), e39. https://doi.org/10.1093/nar/gkad095.

(61) Galvanetto, N.; Ivanović, M. T.; Chowdhury, A.; Sottini, A.; Nüesch, M. F.; Nettels, D.; Best, R. B.; Schuler, B. Extreme Dynamics in a Biomolecular Condensate. *Nature* **2023**, *619* (7971), 876–883. https://doi.org/10.1038/s41586-023-06329-5.

(62) Sánchez, H.; McCluskey, K.; van Laar, T.; van Veen, E.; Asscher, F. M.; Solano, B.; Diffley, J. F. X.; Dekker, N. H. DNA Replication Origins Retain Mobile Licensing Proteins. *Nat. Commun.* **2021**, *12* (1), 1908. https://doi.org/10.1038/s41467-021-22216-x.

(63) van Mameren, J.; Modesti, M.; Kanaar, R.; Wyman, C.; Wuite, G. J. L.; Peterman, E. J. G. Dissecting Elastic Heterogeneity along DNA Molecules Coated Partly with Rad51 Using Concurrent Fluorescence Microscopy and Optical Tweezers. *Biophys. J.* **2006**, *91* (8), L78-80.

(64) Kose, H. B.; Xie, S.; Cameron, G.; Strycharska, M. S.; Yardimci, H. Duplex DNA Engagement and RPA Oppositely Regulate the DNA-Unwinding Rate of CMG Helicase. *Nat. Commun.* **2020**, *11* (1), 3713.

## 4.12 Supplementary figures

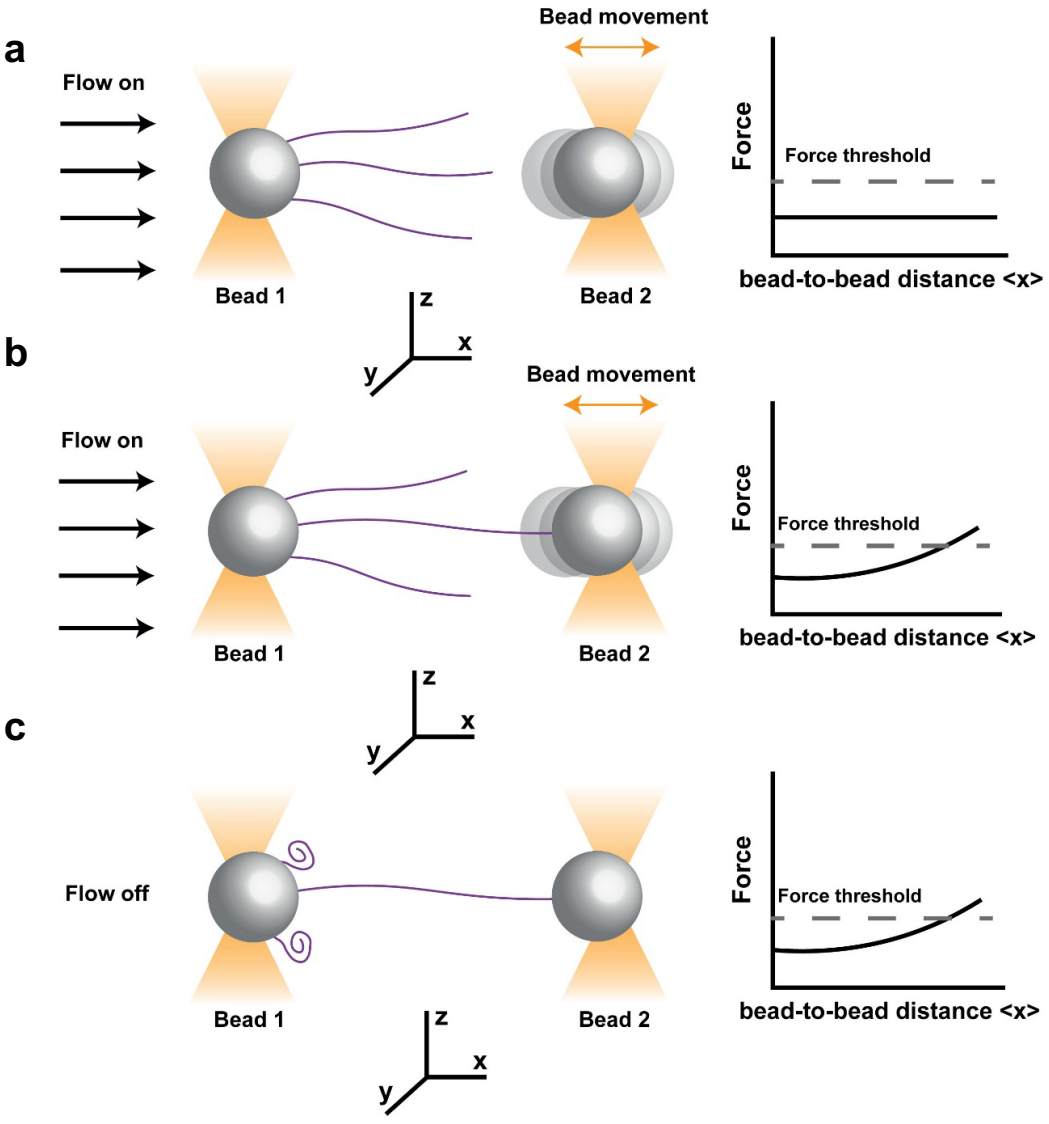



*Figure S4.1* Automated trapping of DNA in dual-beam OT. **a** Functionalized DNA molecules are brought in contact with the beads by the flow to have one end of the DNA attached to the beads. Those DNA molecules attached to Bead 1 are stretched by the flow and the other end of the DNA is brought close to Bead 2 by the flow. Bead 2 is constantly moved close to and away from Bead 1 on the x-axis. The force exerted on the beads is low and independent of the bead-to-bead distance when the DNA is not tethered between the beads. **b** Once the other end of the DNA is attached to Bead 2, rightward movement of Bead 2 generates a force that exceeds a force threshold set by the user, indicating that one or more DNA molecules are tethered between the beads. **c** The flow is then turned off, leaving the remaining DNA molecules attached to the beads in a coiled and collapsed conformation.

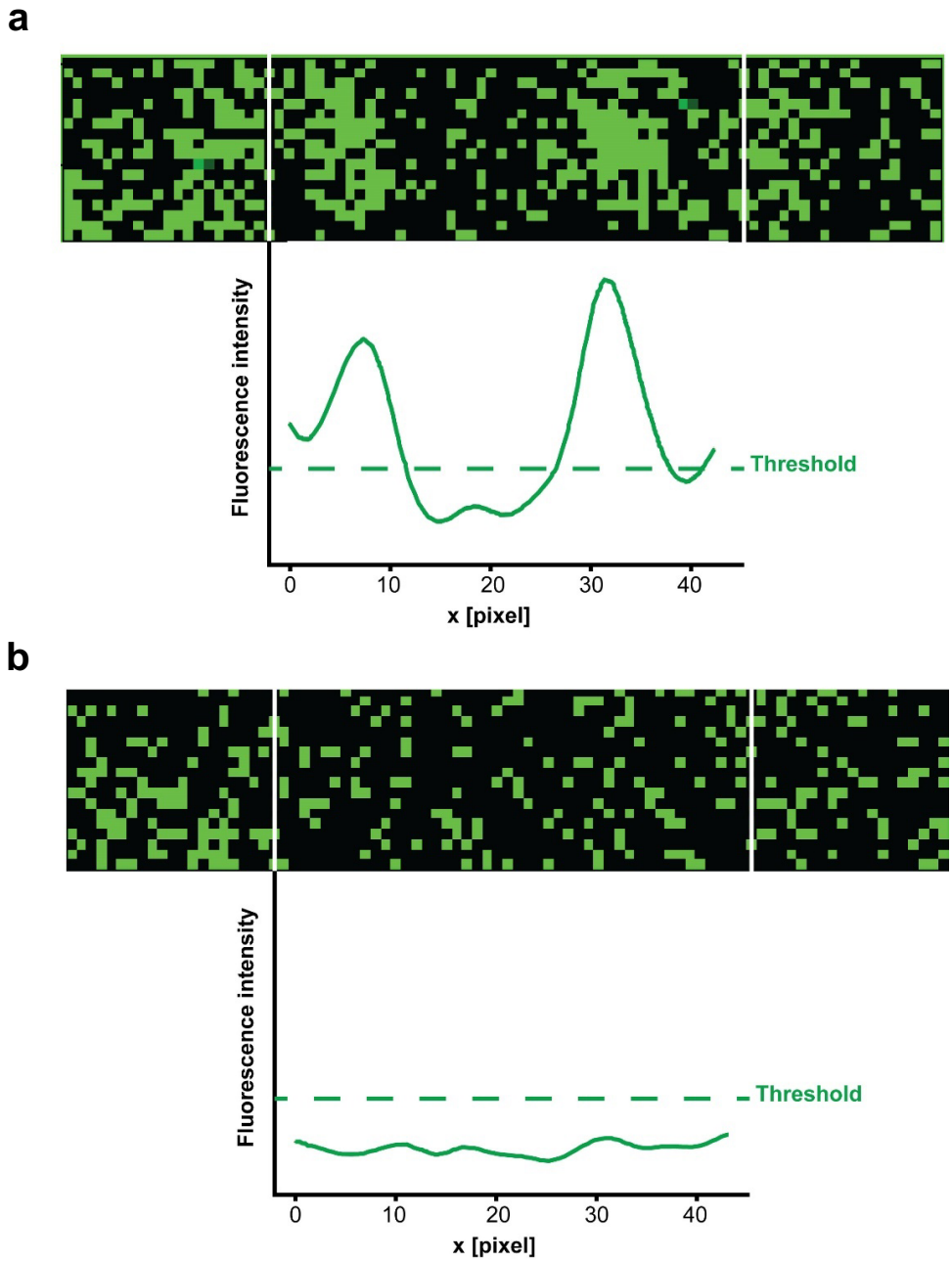



*Figure S4.2* Projection of fluorescence intensities on the x-axis to monitor fluorophore bleaching. **a** Two fluorescent spots are present in the 2D confocal scanning image (upper panel), giving rise to two peaks on the fluorescence intensity projection plot (lower panel) exceeding the preset intensity threshold. The signals close to the beads (beyond the white lines on the upper panel) are excluded from the analysis by specifying a bead margin. **b** After the fluorophores bleach, the background fluorescence intensity falls below the threshold. The continuous confocal scan is subsequently terminated.

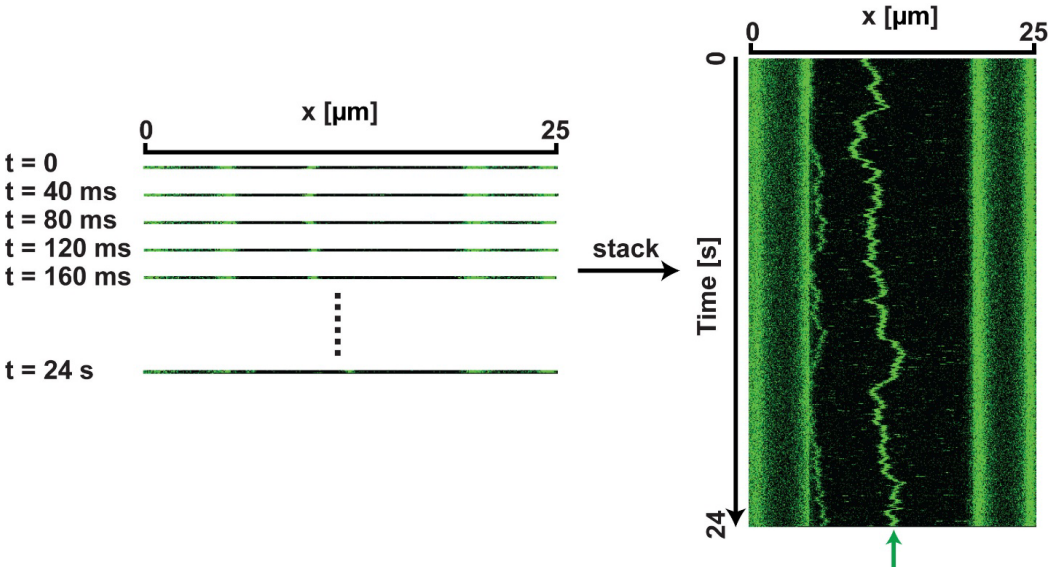



*Figure S4.3* Construction of a kymograph. (left panel) 1D scans are repeated along a selected line along the x-axis. A kymograph is constructed by stacking these 1D scans to form a visual representation (right panel) in which one axis is the x-axis and the other axis represents time. The motion of the fluorescent molecule over time is shown by a green line (indicated by the green arrow in the right panel).

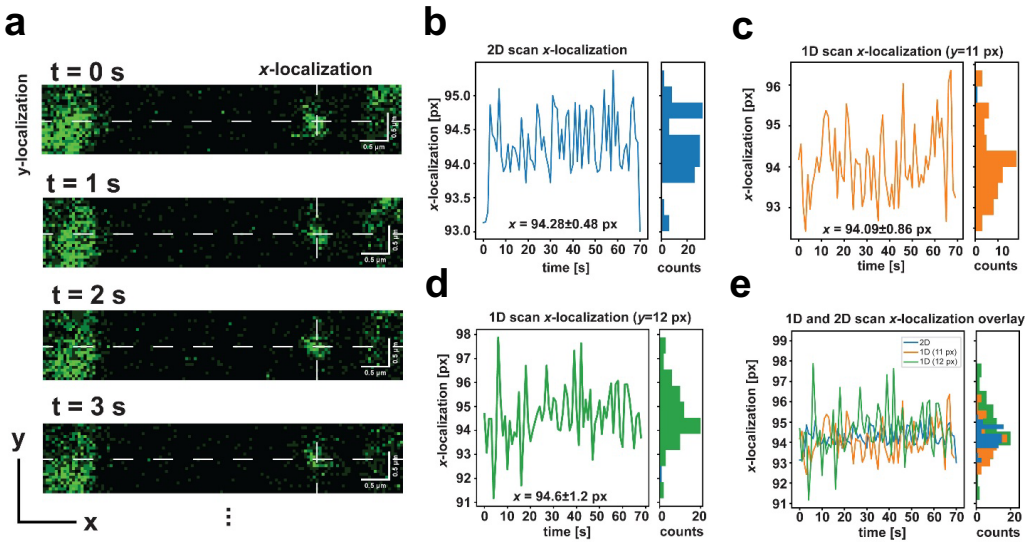



*Figure S4.4* Comparison of the localization precision of 1D and 2D scans. **a** A fluorescently labeled DNA oligo is hybridized to a single-stranded flap on a double-stranded DNA trapped between two beads to form a static fluorescent spot in the confocal scanning image. The fluorescent spot is detected by continuous 2D confocal scans with a pixel size of 50 nm and a frame rate of 1 s/frame until the fluorophore bleaches. The x- and y- localizations of the spot are indicated by the white dashed lines, respectively. **b** The x-localization of the fluorescent spot in each 2D image is determined by 2D Gaussian fitting and plotted against time. Mean and standard deviation of the localization are indicated in the figures. The distribution of the x-localization is shown in the histograms at right. **c, d** The x-localization of the fluorescent spot in each frame is determined by 1D Gaussian fitting to a selected line in the 2D confocal scan image (white dashed line in panel **a**) to mimic the line scans used to build kymographs. Two different y-coordinates close to the center of the fluorescent spot are selected to carry out this 1D scan analysis. The average locations and standard deviations are shown in the figures. The distribution of the x-localizations is shown in histogram. **e** The time traces and distributions of x-localizations of the fluorescent spot determined by 2D images versus 1D scans are overlayed. The standard deviations of x-localizations determined by 1D scan analysis are approximately 2 fold higher compared to 2D scan analysis. Therefore, 2D scan analysis yields a ~2 fold higher fluorophore localization precision than 1D scans.

## 4.13 Supplementary tables

*Table S4.1:* Description of configuration file parameters.

| Field | Type | Description |
|---|---|---|
| `colocalization_distance` | float | Colocalization distance (micron) |
| `step_counting_divide` | boolean | Do we divide steps by the expected step size? |
| `confocal_pixel_size` | float | Width of one pixel in confocal image (micron) |
| `min_step_sizes_{lasers}` | 3 integers | Minimum step sizes for r, g, b, for each possible combination of active confocal lasers |
| `avg_step_sizes_{lasers}` | 3 integers | Average step sizes for r, g, b, for each possible combination of active confocal lasers |
| `bleed_through` | 3x3x3 floats | Bleed through matrix. For axes (*i, j, k*) and value *x*: laser *i* causes the *j* fluorophore to leak *x*% of their signal into channel *k*. |

*Table S4.2:* Description of tracking parameter file parameters.

| Field | Type | Description |
|---|---|---|
| `offset` | float | Brightfield-confocal shift (micron) |
| `laser colors` | boolean | Active confocal lasers; 'r', 'g', 'b', 'rg', 'rb', 'gb', or 'rgb' |
| `thresholds` | 3 floats | LoG detection threshold; numbers for RGB separately |
| `radius (pixels)` | 3 floats | LoG detection radius; numbers for RGB separately |
| `max frame skip` | 3 floats | Max time skip for tracking (frames) |

| | | |
|---|---|---|
| max dist between spots (pixels) | 3 floats | Max distance skip for tracking (pixels). |

*Table S4.3:* Description of scan table columns.

| Column name | Type | Level | Description |
|---|---|---|---|
| scan_id | integer | scan | identifier for scan |
| trace_id | integer | trace | identifier for colocalized tracks of different colors |
| track_id | integer | track | identifier for a spot of a single color, tracked over time |
| color_id | integer | track | identifier for spot color: (0, 1, 2) for ('r', 'g', 'b'), respectively |
| frame | integer | spot | frame number |
| x_pixel | float | spot | x-location of spot in pixels |
| time_s | float | spot | time in seconds |
| x_kbp | float | spot | x-location of spot in kbp from left bead edge |
| intensity | float | spot | sum of pixel values in detected spot |
| corrected_intensity | float | spot | intensity corrected for crosstalk |
| dna_start_pixel | float | scan | starting pixel coordinate of DNA |
| dna_end_pixel | float | scan | ending pixel coordinate of DNA |
| dna_length_kbp | float | experiment | DNA length in kbp |
| file_name | string | scan | path to tiff file associated with scan |
| lifetime | integer | track | lifetime of the track in frames |
| step_count | integer | track | number of steps in track, from step fitting the bleaching trace |

# 5

# High-throughput imaging of fully reconstituted CMG

*This chapter is in preparation for submission for peer review publication as: Ramírez Montero D., Palmero Moya F., Talele S., Liu Z., Sánchez H., van Laar T., and Dekker N.H. (**2024**). A high-throughput single-molecule assay to image the motion of fully reconstituted CMG.*

## 5.1 Abstract

Given the biochemical complexity in the study of the dynamics of fully reconstituted CMG as well as those of entire replisomes, these studies would greatly benefit from having a high-throughput manner to image CMG/replisome motion. Here, we describe a hybrid ensemble and single-molecule assay that allowed us to observe and quantify the motion of fully reconstituted CMG along surface-tethered flow-stretched DNA molecules. To develop this assay, we designed a microfluidic flow cell and developed a novel surface functionalization strategy that allowed us to adapt the hybrid ensemble and single-molecule assay developed in Chapter 2 to make it compatible with flow-stretched DNA TIRF microscopy. To validate our assay, we performed controls on DNA:Mcm2-7 and DNA:CMG complexes to show the feasibility of flow-stretching pre-formed DNA:protein complexes eluted from magnetic beads. Furthermore, our preliminary positional and motion analysis of Mcm2-7 and CMG complexes bound to flow-stretched DNA molecules shows consistency with our previous optical-tweezers work. Although further work is still required for a fully functional high-throughput single-molecule assay to image and quantify the motion of fully reconstituted CMG, our preliminary data suggests that our high-throughput assay is likely pave the way for high-throughput single-molecule imaging of fully reconstituted replisomes.

## 5.2 Introduction

In **Chapter 2**, we described a novel hybrid ensemble and single-molecule assay that allowed us to fully reconstitute the assembly and activation of CMG onto magnetic-bead-bound DNA, elute intact DNA:CMG complexes, and image the motion of CMG at the single-molecule level using correlative optical tweezers and confocal microscopy[1]. While this assay allowed us to image and quantify the motion of fully reconstituted CMG at the single-molecule level for the first time, the inherently low throughout of optical tweezers has its limitations: for example, the acquisition of enough data points to draw statistically significant conclusions can take weeks of work for each condition tested. Furthermore, although the reconstitution of CMG activation is one of the most complex biochemical reactions studied at the single-molecule level so far[1,2], it only represents the initial steps in DNA and chromatin replication[3–5]. As mentioned in the introduction, our long-term goal is to study the dynamics of a full replisome at the single-molecule level, and furthermore, to study these dynamics in the context of a full replisome cruising through chromatin. Reaching this goal will unavoidably increase the complexity of the reaction, as the full reconstitution of DNA and chromatin replication involves several extra reaction steps and protein components[4,6–9]. What is more, because no step is not 100% efficient, with every additional step one lowers the overall yield of the reaction. These observations highlight the need for high-throughput ways to image CMG and replisome motion at the single-molecule level.

As described in detail in **Chapter 1**, three high-throughput single-molecule techniques have been previously used to study *in vitro* reconstituted CMG/replisome dynamics using pre-formed CMG, namely, magnetic tweezers[10], flow tweezers[11], and TIRF microscopy[8,12–14]. As discussed in Chapter 1, in the first two techniques, motion of the CMG/replisome is deduced indirectly by monitoring the length change as dsDNA is converted into ssDNA (during unwinding)[10], or vice versa (as happens to the lagging strand during leading-strand replication)[11], in other words, without directly imaging the replisome. While very powerful, as these techniques allow one to control the force applied to the DNA tethers, the lack of direct imaging complicates the analysis of protein dynamics during replisome progression, often relying on conclusions from drop-out experiments[11]. On the other hand, TIRF microscopy is a high-throughput technique that allows for the direct observation of protein exchange dynamics during replisome progression[14]. For this reason, we set out to study at the single-molecule level the motion of fully reconstituted CMG in high-throughput employing TIRF microscopy.

In this chapter, we describe our pioneering work using TIRF microscopy to image the motion of fully reconstituted CMG in high-throughput along individual flow-stretched surface-tethered DNA molecules. To that end, we adapted the hybrid ensemble and single-molecule assay developed in **Chapter 2** to make it compatible with TIRF microscopy. This development required the design of a microfluidics flow cell, as well as the development of a novel surface functionalization strategy, the combination of which allowed us to flow-stretch and image CMG:DNA complexes eluted from magnetic beads, as in our prior optical tweezers work[1].

As described in **Chapter 2**, imaging fully reconstituted CMG motion at the single-molecule level was made possible thanks to the fact that we assembled and activated CMG onto magnetic-bead-bound DNA[1,5], which allowed us to: 1) remove all the excess protein that would otherwise aggregate on the DNA; and 2) wash the DNA with to remove any intermediates and aggregates from the DNA[1]. This purification procedure yielded a clean and virtually aggregate-free DNA:protein substrate for our single-molecule studies using correlative optical tweezer and confocal microscopy[1]. This highlights an important difference that should be noticed between our work studying the motion of fully reconstituted CMG and previous work studying pre-formed CMG motion with flow-stretched DNA TIRF: in previous flow-stretched DNA TIRF work, naked DNA molecules were first flow-stretched and bound to a surface, and then pre-formed CMG and other protein components were added to the surface-bound DNA[13,14]. In the past, we have tried a similar approach with fully reconstituted CMG. That is, flow-stretching fully duplex naked DNA containing an origin of replication onto a glass surface, and then adding

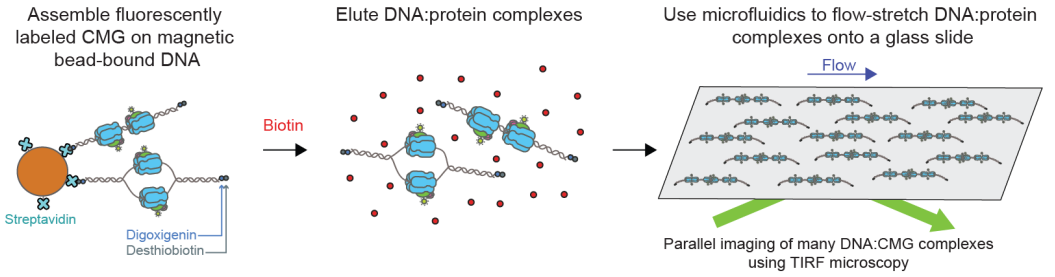



*Figure 5.1* Pictorial description of high-throughput assay to image the motion of fully reconstituted CMG at the single-molecule level. A linear DNA constructs labeled at both ends with digoxigenin and desthiobiotin moieties is bound to streptavidin-coated magnetic beads. CMG complexes are then assembled and activated on the magnetic-bead bound DNA, which allows for the purification of the DNA:CMG complexes away from unbound excess proteins and aggregates. The complexes can then be eluted by the addition of free biotin, and the eluted intact complexes can then be flow-stretched onto a glass surface, so that the motion of CMG can be imaged in high-throughput using TIRF microscopy.

the loading and firing factors at their required concentrations (10-200 nM) at different stages[4,5]. Nonetheless, this approach resulted in a lot of protein aggregation and non-specific adhesion to the surface, complicating the analysis and affecting the motion of CMG (Humberto Sánchez and Belén Solano, personal communication). For this reason, we decided to implement a new approach: fully reconstituting CMG onto origin-containing DNA, purifying the resulting DNA:protein complexes away from the excess protein and protein aggregates, and then flow-stretching the purified DNA:CMG complexes onto a glass surface (**Fig. 5.1**). In the following sections, we will first describe in detail different components of this assay that we engineered, as well as show some preliminary CMG positional and motion analysis that validates our novel approach.

## 5.3 Results

### 5.3.1 Design of a microfluidic flow cell to image flow-stretched CMG:DNA complexes

A key component of flow-stretched TIRF microscopy is the use of microfluidics to bind and flow-stretch long (> 20 kb) linear DNA molecules onto a functionalized glass surface[11,13–20]. Furthermore, in flow-stretched DNA experiments, the number of DNA molecules bound to the surface is a function of the cross-sectional area of the flow cell, the concentration of the DNA in the flown solution, the flow rate and the volume of DNA solution that is flown[11,14,15,21–24]. Therefore, as we aimed to adapt the hybrid ensemble and single-molecule assay developed in **Chapter 2** to make it compatible with TIRF microscopy, we decided to design a new flow cell that would adapt to our specific experimental needs.

Several aspects needed to be considered when designing our custom-made flow cell. First of all, as described in **Chapter 2**, when we elute the DNA:CMG complexes from the magnetic beads, we end up with volume of solution containing the DNA:CMG complexes on the scale of hundredths of µl[1], which cannot be reduced by a large factor, as we need to add enough buffer to easily resuspend milligrams of magnetic beads[1]. Therefore, an ideal flow cell would allow us to flow such large volumes at a fast enough flow rate to be compatible with the required 3-3.4 hours of biochemical preparation prior to the single-molecule imaging[1] (without requiring additional hours of flow-stretching). In second place, the functionalization of the glass surface is time-consuming, and yet, only a small fraction of the glass surface is imaged[25]. Thus, an ideal flow cell would allow us to create several channels of controlled geometry on a given glass slide, to conduct

more than one experiment using the same slide. Finally, an ideal flow-cell would be easy and fast to assemble, to easily combine it with the ensemble biochemical part of the experiments.

After some attempts, we ended up with a flow cell design that fulfilled all these requirements, which is illustrated in **Fig. 5.2a**. Briefly, five channels are cut out of a transparent adhesive optical spacer using a custom-made stencil. The spacer is then sandwiched between a 1-mm thick quartz slide with inlet and outlet holes drilled with laser cutter, and a passivated and functionalized coverslip. As we will describe below, this flow cell allowed us to readily and reproducibly flow stretched purified DNA:CMG complexes, and image CMG motion along the DNA in high-throughput.

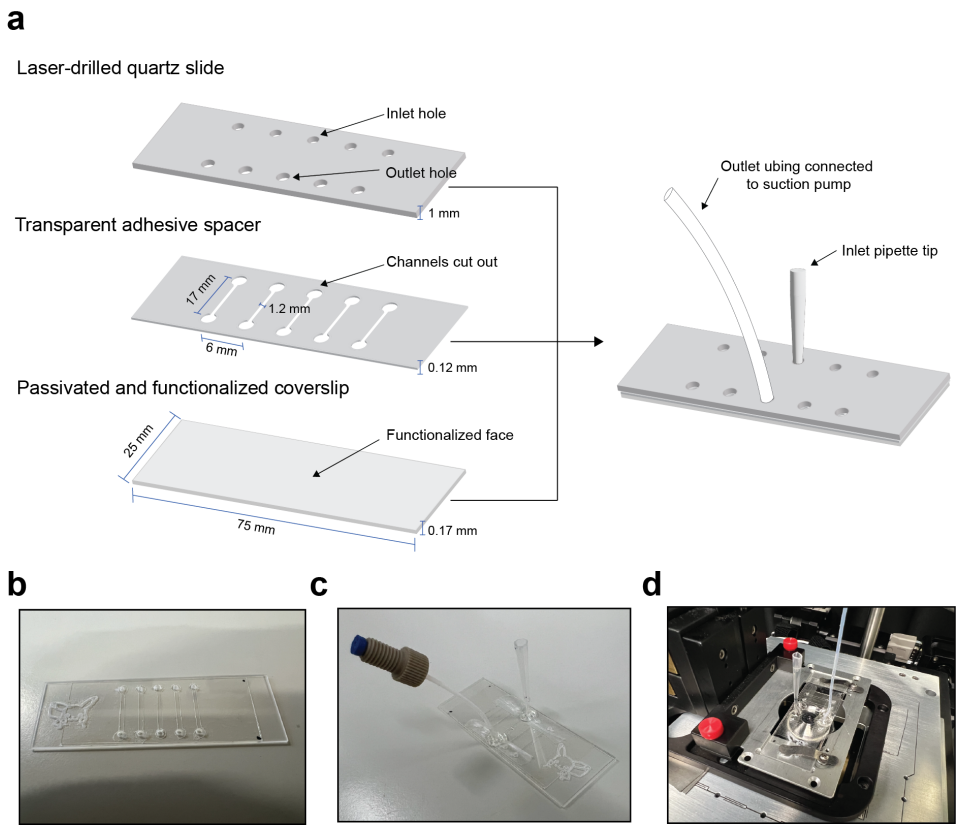



*Figure 5.2* Design of a custom micro fluidic flow cell to image the motion of fully reconstituted CMG in high-throughput using TIRF microscopy. **a** Pictorial description of the flow cell assembly. *Left*: a transparent double-sided adhesive optical spacer with the channels cut out is sandwiched between a quartz microscope slide with inlet and outlet holes drilled and a functionalized coverslip. *Right*: a pipette tip is then attached to the inlet hole, and a plastic tubing connected to a suction pump to the outlet hole. **b** Photograph of flow cell. **c** Photograph of a flow cell with an inlet pipette tip and outlet tubing connected to one channel (all other channels are covered with parafilm). **d** Photograph of a full flow cell mounted on a TIRF microscope and ready for imaging.

### 5.3.2 A novel surface functionalization strategy to image fully reconstituted CMG on DNA in high-throughput.

A critical aspect in TIRF microscopy experiments is the surface used. This is because in such experiments, the sample to be imaged is in close proximity to a glass surface[26–28]. An ideal surface for TIRF microscopy studies must meet two criteria: first, the surface must be functionalized for the specific attachment of biological molecules; and second, the surface must be otherwise passivated to prevent non-specific interactions of biological molecules with the surface. Having a well-passivated surface is particularly important for our experiments, as CMG and several other loading, firing and replication factors easily bind surfaces.

One of the most commonly used passivation strategies is to covalently attach PEG chains to the glass surface via primary amine-NHS ester interactions[11,14,16,20,22–24]. In a typical passivation procedure, a clean and aminosilanized glass slide covalently bound to a long (typically 3.5-5 kDa) PEG chain[14,23,29,30], followed by an optional second and third rounds of additional PEGylation with either the same PEG chain or a shorter chain that can more easily access holes in the PEG later[31–33]. What is more, more recent studies have shown that in marginal PEG solubility conditions, PEG adopts a brush-like conformation in which the PEG chains are closer together, allowing for a better surface passivation[30]. One additional advantage of using PEG as a passivation reagent is that it can easily be covalently conjugated to biotin, which allows for the covalent attachment of biotin to the surface in the form of biotinylated PEG (typically added as a mixture of PEG and biotinylated PEG[11,13,14,16,22–24,32–36]. The surface-bound biotin can then be bound to streptavidin, which has four binding sites for biotin, therefore exposing additional biotin binding sites that can be used to bind biotin-conjugated biological molecules to the surface[13,14,16,17]. Thus, PEG/biotinylated PEG can both passivate and functionalize the surface.

As described in **Chapter 2**, in our previous optical-tweezers work, CMG was assembled and activated onto a linear DNA construct with both ends functionalized with two orthogonal moieties: desthiobiotin and digoxigenin[1]. The desthiobiotin moieties allowed us to bind the DNA to streptavidin-coated magnetic beads, and then assemble and activate CMG onto magnetic-bead-bound DNA (including important washing steps to purify the DNA:CMG complexes away from excess protein and aggregates) [1]. The purified DNA:CMG complexes were then eluted from the magnetic beads with the addition of an excess of free biotin, which outcompetes the desthiobiotin-streptavidin interaction[37] and releases intact DNA:protein complexes[1,38,39]. Therefore, a streptavidin-functionalized surface would not allow

us to bind the DNA to a streptavidin-functionalized surface, as the excess free biotin would outcompete the DNA. In **Chapter 2**, we overcame this issue by adding an additional desthiobiotin moiety to each end of the linear DNA substrate (which tightly binds an antibody raised against it in an orthogonal fashion to biotin:streptavidin interactions), and binding the DNA:CMG complexes to polystyrene beads coated with anti-digoxigenin (anti-Dig) antibody[1]. Thus, an ideal surface for our purposes would be well passivated as well as functionalized with anti-Dig.

Thus far, some groups have worked with anti-Dig-coated glass surfaces have either purposely made the passivation suboptimal so that the anti-Dig can bind the surface in a non-specific manner[40–42], or pre-coat their glass slides with nitrocellulose to make them very sticky so that the anti-Dig easily attaches to the

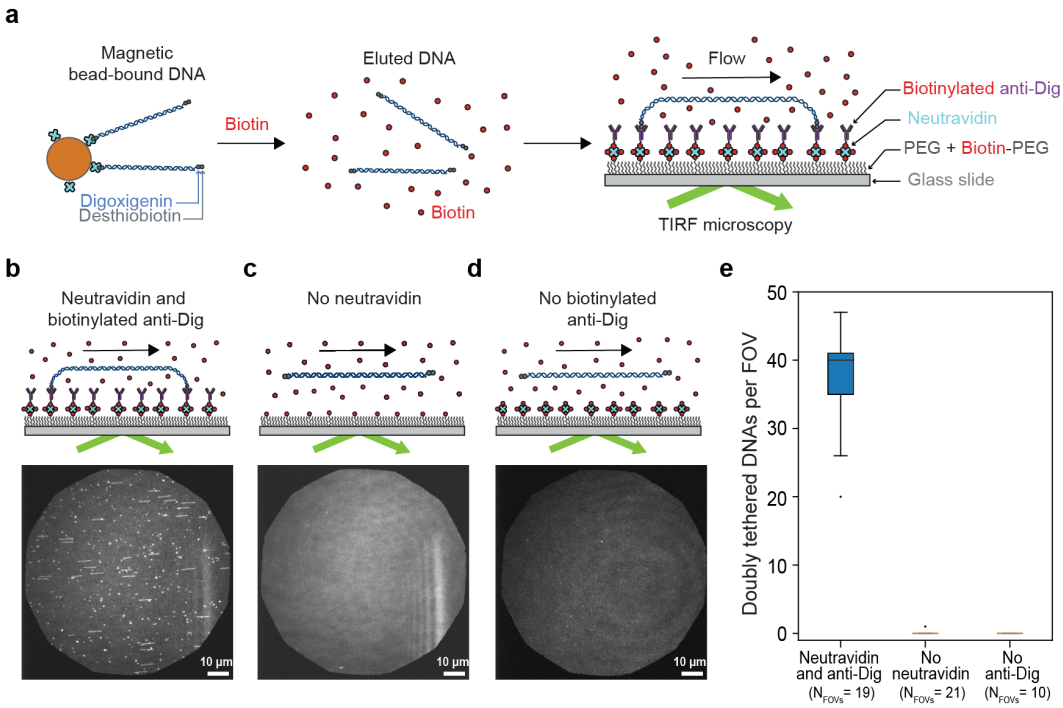



*Figure 5.3* A novel surface functionalization strategy to image flow-stretched doubly tethered DNA molecules eluted from magnetic beads. **a** Linear DNA molecules doubly functionalized at both ends with desthiobiotin and digoxigenin moieties were eluted from streptavidin-coated magnetic beads with an excess of biotin. The eluted DNA molecules, in a buffer solution containing excess biotin, were doubly tethered onto a glass surface through a novel surface functionalization strategy: glass coverslips were covalently passivated with a mixture of PEG and biotinylated PEG. Subsequently, the surface was incubated with neutravidin, and then with biotinylated anti-Dig antibody. **b-d** Representative fields of view showing the relative amounts of doubly tethered DNA molecules when the surface was incubated with **b** neutravidin and then biotinylated anti-Dig, **c** only with biotinylated anti-Dig, or **d** only with neutravidin. Images were acquired in the absence of flow. **e** Boxplots of number of doubly tethered DNAs per field of view for the three conditions described in **b-d**.

surface[43]. Neither approach would be ideal for us, as we need as passivated a surface as possible to prevent CMG from interacting with the surface as it moves on the DNA. Furthermore, such approaches lack control of the orientation of the antibody, reducing the number of accessible antigen binding sites.

To overcome these limitations, we developed a novel surface passivation and functionalization strategy that 1) passivates the glass surface to minimize non-specific interactions, and 2) exposes properly oriented anti-Dig molecules on the surface to bind DNA molecules eluted from magnetic beads. This surface functionalization strategy is outlined in **Fig. 5.3a** and consists of the following steps: first, neutravidin was bound to a coverslip passivated under cloud-point conditions with a mixture of PEG and biotinylated PEG (**Methods**); the neutravidin binds to the surface-bound biotin via one of its binding sites, but exposes at least one of its three additional biotin-binding sites. Then, a biotin-conjugated anti-Dig antibody was bound to the surface-bound neutravidin through its conjugated biotin. As seen in **Fig. 5.3 b**, this surface functionalization strategy allowed us to flow-stretch linear DNA molecules (doubly functionalized at both ends with desthiobiotin and digoxigenin moieties) that were previously bound to streptavidin-coated magnetic beads and eluted with excess free biotin[1]. Importantly, we also generated control surfaces in which we omitted either the neutravidin or the biotinylated anti-Dig. If the attachment of the doubly tethered DNAs to the surface is indeed through biotinylated anti-Dig:neutravidin:surface interactions, the omission of either component should reduce the number of DNA double tethers. As seen in **Fig. 5.3b-f**, we saw a drastic decrease in the number of double tethers when either component was omitted, compared with the surface in which both components were added (**Fig. 5.3c-f**). This is to be expected because, if neutravidin is omitted, the biotinylated anti-Dig could not bind to the surface. On the other hand, if the biotinylated anti-Dig is omitted, the free biotin in the eluted DNA solution would saturate the available biotin-binding sites in the surface-bound neutravidin, preventing the DNA from binding to it (**Fig. 5.3c-f** upper diagrams). Taken together, these results validate our development of a novel surface functionalization assay that allowed us to flow stretch DNA molecules eluted from streptavidin-coated magnetic beads.

### 5.3.3 A high-throughput assay to image the motion of fully reconstituted CMG at the single-molecule level

#### 5.3.3.1 Assay description and biophysical tests of doubly tethered DNA

Having developed and validated the surface passivation and functionalization strategy described **Section 5.3.2** and the microfluidic flow cell described in **Section 5.3.1**, we tested whether we could combine these two developments to image fully reconstituted CMG at the single-molecule level in high-throughput. To this end, we modified the hybrid ensemble and single-molecule assay developed in **Chapter 2**. Briefly, we synthesized a 23.6 kb linear DNA substrate containing a naturally occurring ARS1 origin of replication, functionalized both of its ends with desthiobiotin and digoxigenin moieties, and bound it to streptavidin-coated magnetic beads (**Methods**). We then assembled and activated CMG onto the magnetic bead-bound DNA employing red fluorescently labeled Mcm2-7$^{JF646}$ and green fluorescently labeled Cdc45$^{LD555}$. We then eluted the intact DNA:CMG complexes with an excess of biotin, and flowed-stretched them onto a PEGylated surface functionalized with anti-Dig (**Fig. 5.4a, Section 5.3.2**). After flow stretching the DNA, we flowed buffer solution supplemented with SYTOX green to stain the DNA, which then shows in the blue channel (**Methods**). As seen in **Fig. 5.4b-d**, our newly developed protocol allowed us to image Mcm2-7 (**Fig. 4.4b**) and Cdc45 (**Fig. 4.4c**) colocalized with flow-stretched DNA tethers. Furthermore, the triple colocalization of Mcm2-7, Cdc45, and DNA shows our ability to observe CMG on flow-stretched DNA (**Fig. 5.4d**).

Of note, the percentage of DNA tethers colocalized with Mcm2-7 and Cdc45 was consistent across field of views (**Fig. 5.4e**), and reproducible across biological replicates (**Fig. 5.4f**), attesting to the robustness of our assay. In addition, we observed more DNA double tethers colocalized with Mcm2-7 spots (**Fig. 5.4e**) than with Cdc45 (**Fig. 5.4f**), which is in agreement with previous observations that CMG assembly is a less efficient reaction than Mcm2-7 loading[1,5,44]. Of note, while the percentage of double tethers colocalized with Cdc45 (**Fig. 5.4f**, bottom) was only slightly lower than the ~33% that we observed in our optical tweezers work with fully reconstituted CMG[1] (**Chapter 2**), the average percentage of Mcm2-7 spots per field of view that was colocalized with Cdc45 (**Fig. 5.4g**) was 4-fold lower than what we previously observed in our optical tweezers work (**Chapter 2**). Future work will be required to understand the reason(s) behind this discrepancy.

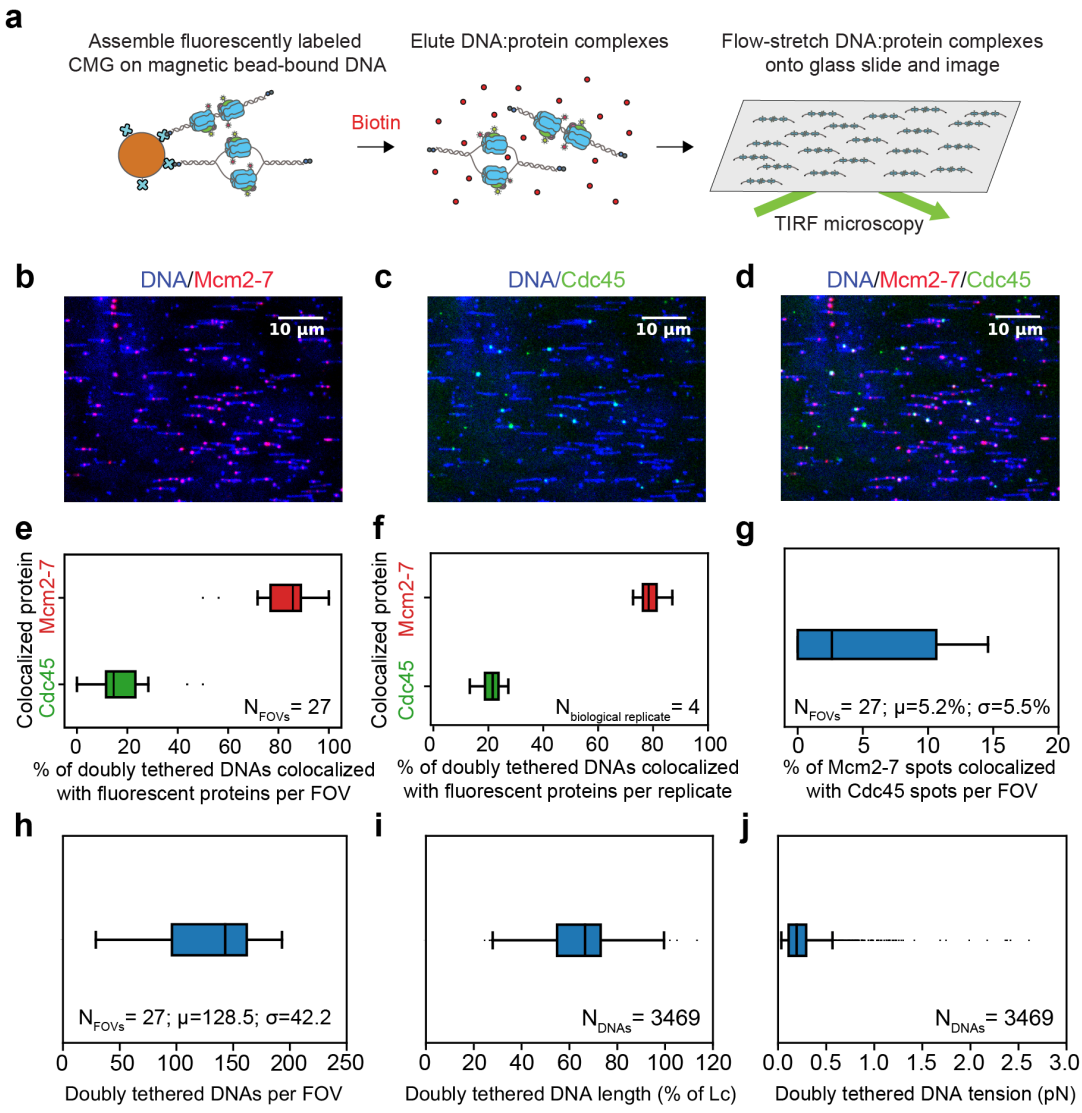



*Figure 5.4 A hybrid ensemble and single-molecule assay to image the motion of fully reconstituted CMG in high-throughput.* **a,** Pictorial description of the assay. Fluorescently labeled CMG complexes were assembled and activated onto magnetic-bead bound linear DNA molecules using fluorescently labeled Mcm2-7 (red) and Cdc45 (green). Intact CMG:DNA complexes were then eluted from the magnetic beads with excess biotin, flow-stretched onto a passivated glass slide, and imaged in the absence of flow with TIRF microscopy. **b-d** Representative images of flow-stretched DNA:CMG complexes. Overlays of **b** DNA (blue) and Mcm2-7 (red), **c** DNA (blue) and Cdc45 (green), and **d** DNA (blue), Mcm2-7 (red) and Cdc45 (green). **e-f** Percentage of doubly tethered DNA molecules colocalized with either Mcm2-7 fluorescent spots or Cdc45 fluorescent spots per **e** field of view, or **f** biological replicate**. g** % of Mcm2-7 spots colocalized with Cdc45 spots. **h** Number of flow-stretched doubly tethered DNAs per field of view. **i** End-to-end extensions of doubly tethered DNAs as a percentage of the contour length of the DNA. **j** Estimated tension in the doubly tethered DNA molecules, deduced by applying the worm-like chain model to the end-to-end extensions shown in **i (Methods)**.

As discussed in the introduction, our flow-stretching protocol differs from previous work in that we flow-stretch pre-formed DNA:CMG complexes, as opposed to bare DNA. Since the proteins bound to the eluted DNA could bend and compact the DNA[44–46], we wanted to monitor whether we could generate doubly tethered DNA molecules of lengths comparable with those obtained via more conventional flow-stretching of naked DNA[13,14,16]. As seen in **Fig. 5.4h**, after some optimization of the initial amounts of magnetic bead-bound DNA as well as the final dilution factor of the eluted DNA:CMG complexes (**Methods**), we were able to achieve an average of 129 doubly tethered DNA molecules per field of view. Furthermore, under our flow-stretching conditions (**Methods**), the length distribution of doubly tethered DNA molecules (**Fig. 5.4i**) was similar to that obtained with conventional ways of producing doubly tethered DNA molecules on PEGylated surfaces via biotin-avidin linkages[16]. From the mean of this length distribution, we estimated the mean DNA tensions to lie below 1 pN by applying a worm-like chain model (Eq. 10 in Ref.[47]) (**Fig. 5.4j**, **Methods**). These tensions are below those used in our optical tweezers experiments, suggesting that tension-related artifacts are not a concern in our assay. This analysis of the tethers obtained using our novel experimental protocol confirm that it is possible to assemble CMG on origin-containing DNA bound to magnetic bead, elute DNA:CMG complexes, and image them in high-throughput after flow-stretching them onto a functionalized glass surface.

#### 5.3.3.2 Mcm2-7 and CMG stoichiometry and positional analysis

To further assess the functionality of our high-throughput assay, we wanted to compare how the distributions of Mcm2-7 and CMG stoichiometries compared between our previous optical tweezers work (**Chapter 2**) and our new flow-stretched DNA TIRF work. For this, for each diffraction-limited spot colocalized with DNA, we counted the number of associated bleaching steps and measured its initial position on the DNA (**Fig. 5.5**, **Methods**). Of note, because in these experiments we cannot differentiate between the two possible orientations that the DNA can have, we visualize the initial positions as distances from the DNA center, as we have done in previous optical tweezers work[1,48,49].

Firstly, by counting the number of Mcm2-7 and CMG spots per DNA and the number of bleaching steps per spot, we could count the total number of Mcm2-7 and CMG complexes per DNA. With this analysis, we found that, in general, DNAs in our TIRF assay have similar numbers of Mcm2-7 and CMG spots to those in our optical tweezers work (**Figs. 5.5a,d; 2.1b;** and **S2.2b**), but as these spots generally have reduced stoichiometries (**Figs. 5.5b,e; 2.1c**; and

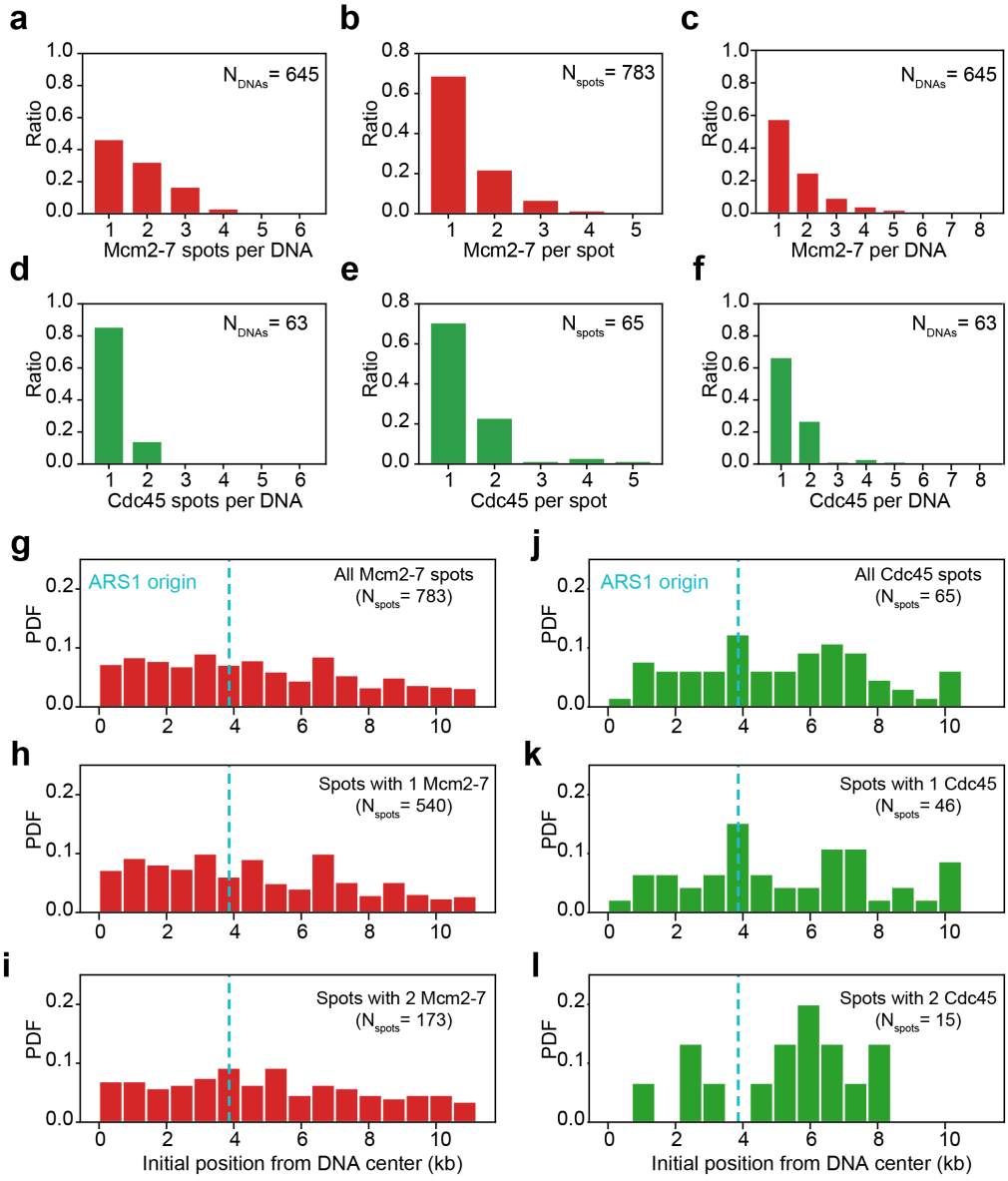



*Figure 5.5* Stoichiometry and positional analysis of Mcm2-7 and CMG. **a** Distribution of numbers of Mcm2-7 diffraction-limited spots per DNA. **b** Distribution of numbers of Mcm2-7 complexes within each diffraction-limited spot. **c** Distribution of number of Mcm2-7 complexes per DNA, calculated by combining **a** and **b**. **d** Distribution of numbers of Cdc45 diffraction-limited spots per DNA. **e** Distribution of numbers of Cdc45 complexes within each diffraction-limited spot. **f** Distribution of number of Cdc45 complexes per DNA, calculated by combining **d** and **e**. **g-i** Distribution of initial positions on the DNA of **g** all Mcm2-7 diffraction-limited spots, **h** diffraction-limited spots containing 1 Mcm2-7 or **i**, diffraction-limited spots containing 2 Mcm2-7; the ARS1 origin of replication is indicated by the dashed cyan line. **j-l** Distribution of initial positions on the DNA of **j** all Cd45 diffraction-limited spots, **k** diffraction-limited spots containing 1 Cdc45, or **l**, diffraction-limited spots containing 2 Cdc45; the ARS1 origin of replication is indicated by the dashed cyan line.

**S2.2c**), the total numbers of Mcm2-7 and CMG molecules are lower than in our optical tweezers assay (**Figs. 5.5c,f; S2.1c;** and **S2.2d**). Further studies will be required to understand the reasons behind these differences, which will be an important step in increasing the throughput of our assay. Secondly, by looking at the distributions of initial positions of Mcm2-7 and CMG spots, we observed a similarly broad positional distribution of Mcm2-7 and CMG spots on the DNA as in our previous optical tweezers work (**Figs. 5.5g-l** and **2.1d-f, h**). Furthermore, in the case of CMG, we observe a similar preference for the bin containing the origin of replication (**Fig. 5.5j-k**) as we observed in our optical tweezers work (**Fig. 2.1d,f**), suggesting that the flow-stretching does not greatly affect the positions of the spots.

Taken together, our results show that it is possible to image Mcm2-7 as well as CMG on flow-stretched DNA molecules in high-throughput, increasing the number of data points that can be acquired in a single movie compared to our correlative optical tweezers and confocal microscopy work developed in **Chapter 2**.

#### 5.3.3.3 DNA-bound Mcm2-7 and CMG exhibit more motion than surface-adhered Mcm2-7 and CMG

As mentioned in the introduction, CMG and some of the loading and firing factors easily adhere to surfaces. In our previous optical tweezers work, we imaged CMG away from the surface employing dual-beam optical tweezers[1]. However, in our high-throughput assay, CMG is closer to a surface coated with neutravidin and anti-Dig antibody. It is thus a possibility that Mcm2-7 or CMG could interact with the surface in a non-specific manner even if bound to a doubly tethered DNA. To monitor whether this is the case, we monitored the root mean squared displacements (RMSD) of Mcm2-7 and CMG spots that are colocalized with DNA molecules, and compare this motion with that of spots that are not colocalized with DNA and are therefore non-specifically bound to the surface (**Fig. 5.6a**). To this end, we image the flow-stretched DNA:CMG complexes in imaging buffer supplemented with ATPγS, which we previously showed to halt CMG motion along the DNA. We then calculated the root mean squared displacement of Mcm2-7 (red) and CMG (green) spots colocalized and not-colocalized with DNA. Importantly, we separately analyzed root mean squared displacements along the x- and y-axes, because the x axis of the image mostly aligns with the direction of the flow applied to flow-stretch the DNA molecules, and therefore most of the DNAs are aligned with this axis. By separating out the motion analysis between the two axes, we can also differentiate between motion along the DNA axis and perpendicular to it (**Fig. 5.6a**). Of note, in this analysis, we did not project the spot positions onto their respective DNAs, so

that we could fairly compare DNA-colocalized spots with non-DNA-colocalized spots.

As observed in **Fig. 5.6b-i**, in all cases, we observed a population centered at around 0.04 µm RMSD. As the same population appears in both axes, in red and green spots, and in DNA-colocalized and non-DNA-colocalized spots (**Fig. 5.6b-i**), we propose that this population likely represents the lower bound in our motion detection. Such motion may result from e.g. drift in the system. Nevertheless, in both red and green spots, we detected a second population centered at around 0.09 µm RMSD, which is most prevalent along the x-axis (i.e. transverse to the orietations of most DNA molecules) and in the case of DNA-colocalized spots (**Fig. 5.6d, h**). Of note, we expect the DNA double tethers to oscillate transversely to their long axis due to thermal fluctuations[50]. Therefore,

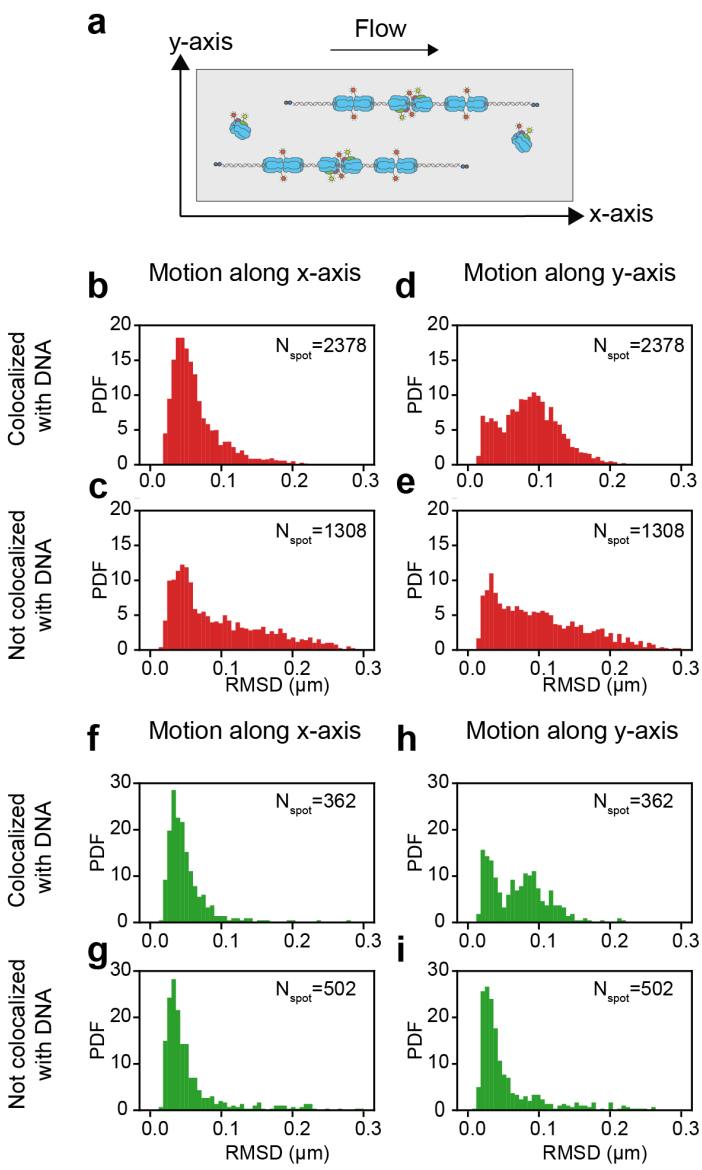



*Figure 5.6* Comparing motion of DNA- and surface-bound proteins. **a** Diagram of top view of the flow cell showing the axis along which motion was assessed. **b** Root mean squared displacement along the x-axis of Mcm2-7 spots colocalized with DNA. **c** Root mean squared displacement along the x-axis of Mcm2-7 spots not colocalized with DNA. **d** Root mean squared displacement along the y-axis of Mcm2-7 spots colocalized with DNA. **e** Root mean squared displacement along the y-axis of Mcm2-7 spots not colocalized with DNA. **F** Root mean squared displacement along the x-axis of CMG spots colocalized with DNA. **c** Root mean squared displacement along the x-axis of CMG spots not colocalized with DNA. **d** Root mean squared displacement along the y-axis of CMG spots colocalized with DNA. **c** Root mean squared displacement along the y-axis of CMG spots not colocalized with DNA.

the fact that this additional population is most prevalent in DNA-colocalized spots suggests that there is a population of DNA-bound Mcm2-7 and CMC that fluctuates jointly with the DNA, supporting the notion that these spots are not strongly interacting with the surface. This observation is particularly important given the fact that we are employing a new anti-Dig functionalized surface, which has not been employed to study protein motion before (**Section 5.3.2**). Of note, the range of these transverse fluctuations falls within the estimated range of DNA fluctuations (**Fig. S5.2**), suggesting that they represent bona-fide oscillations of DNA-bound proteins.

#### 5.3.3.4 Mcm2-7 and CMG diffuse on flow-stretched DNA and nucleotide binding halts CMG diffusion

In previous ensemble and single-molecule, it was shown that Mcm2-7 hexamers exhibit diffusive motion on the DNA[48,51]. We therefore wanted to test whether we could detect the same diffusive motion in our TIRF experiments. To this end, we loaded red fluorescently labeled Mcm2-7 onto origin-containing magnetic bead-bound DNA, eluted and flow-stretched the DNA:Mcm2-7 complexes, and imaged in buffer supplemented with SYTOX green and ATP. As observed in **Fig. 5.7a,c**, this approach allowed us to observe diffusive motion of Mcm2-7 spots colocalized with DNA. We also measured diffusion constants along the DNA for every Mcm2-7 spot, which yielded the distribution shown in **Fig. 5.7f**. From this analysis, we noticed that the mean diffusion constant of $10^{-4.68}$ $kb^2/s$ measured for Mcm2-7 (**Fig. 5.7f**) is about 10-fold lower than what we measured in previous correlative dual-beam optical tweezers and confocal microscopy studies (Ref.[48]). Further studies will be required to understand this difference.

In addition, we showed in **Chapter 2** that CMG also exhibits diffusive motion in the absence of nucleotide, but that the addition of ATPγS halts this diffusive motion[1]. We therefore also tested whether we could detect diffusive motion of CMG motion in the absence of nucleotide. To this end, we assembled and activated CMG onto magnetic bead-bound DNA, eluted and flow-stretched the DNA:CMG complexes, and imagined in buffer supplemented with SYTOX green and without any nucleotide. As observed in **Fig. 5.7b,d**, this approach allowed us to observe diffusive motion of CMG spots colocalized with DNA, the quantification of which is shown in **Fig. 5.7f**. Furthermore, when we imaged CMG in buffer supplemented with ATPγS, and analyzed the motion of the DNA-colocalized CMG spots (**Fig. 5.7e,h**), we obtained a distribution of diffusion constants that was on average 10-fold slower than those of Mcm2-7 (**Fig. 5.7f**) and CMG in the absence of nucleotide

(**Fig. 5.7g**). This observation is in agreement with our previous observation that the presence of ATPγS halts the diffusive motion of CMG[1].

That being said, we noticed that the mean diffusion constants of $10^{-4.88}$ kb$^2$/s measured for CMG in the absence of nucleotide (**Fig. 5.7g**) or the mean diffusion constant of $10^{-5.93}$ kb$^2$/s measured for CMG in the presence of ATPγS (**Fig. 5.7h**) are about 10-fold lower than those previously measured (Ref.[48] and **Fig. S2.5i**). Further work will be required to understand these discrepancies, and well as to calibrate the motion noise in our TIRF setup to be able to separate static from mobile molecules in a quantitative manner, as we have previously done in our optical tweezers work[1].

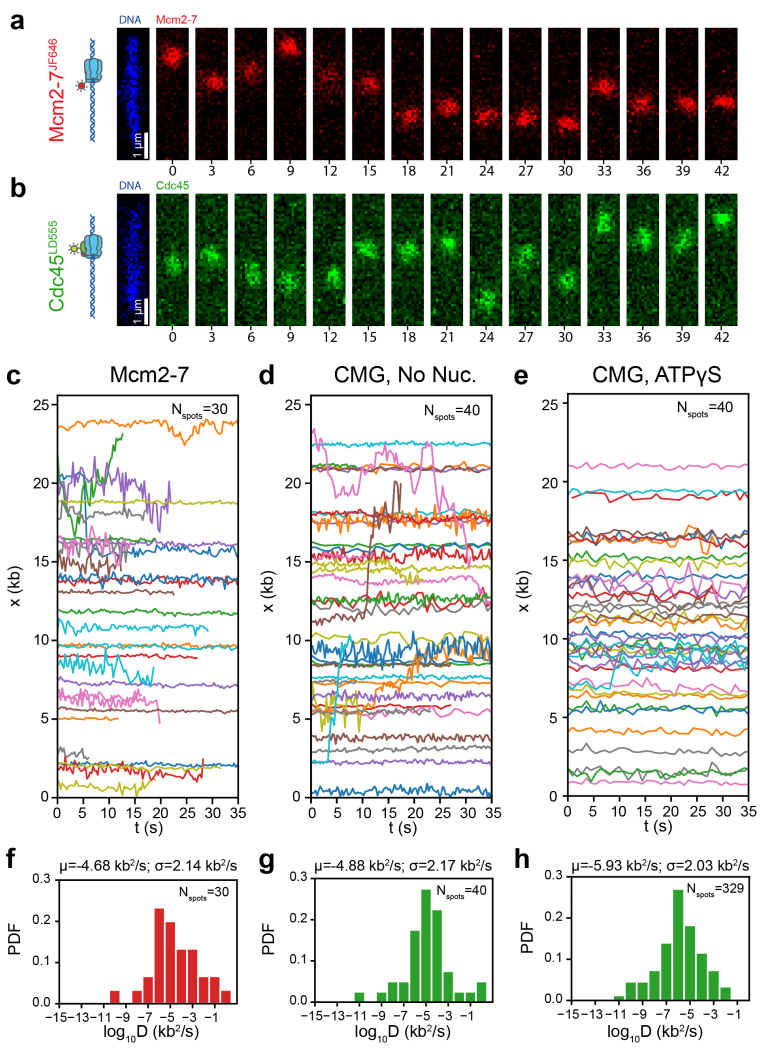



*Figure 5.7* Mcm2-7 and CMG diffuse on flow-stretched DNA. **a** Example kymograph of diffusive Mcm2-7 complexes on flow-stretched DNA. **b** Example kymograph of diffusive CMG on flow-stretched DNA. **c** Position along flow-stretched DNA vs. time plots of Mcm2-7 spots in the presence of ATP. **d** Position along flow-stretched DNA vs. time plots of CMG spots in the absence of nucleotide. **e** Position along flow-stretched DNA vs. time plots of CMG spots in the presence of ATPγS. **f** Distribution of diffusion constants of Mcm2-7 spots in the presence of ATP. **g** Distribution of diffusion constants of CMG spots in the absence of nucleotide. **h** Distribution of diffusion constants of CMG spots in the presence of ATPγS.

### 5.3.3.5 Analyzing unidirectional motion of CMG on flow-stretched DNA

As a final validation of our high-throughput single-molecule assay, we set out to image the unidirectional motion of fully reconstituted CMG that we previously observed in the presence of ATP[1]. We therefore flow-stretched DNA:CMG complexes as previously described, but this time imaged then in buffer supplemented with ATP. Unfortunately, as shown in **Fig. 5.8a**, we have not been able to detect the long-range motion of CMG observed in our optical tweezers work in our flow-stretched DNA TIRF assay (compare **Fig. 5.8a** with **Fig. 2.2b**). Furthermore, we applied the same change-point analysis employed in **Chapter 2** to obtain a denoised distribution of instantaneous velocities of CMG (**Fig. 5.8b**), and observed overall less motion than what we previously observed in our optical tweezers work (compare **Fig. 5.8b** with **Fig. S2.5a**). We are currently conducting experiments to understand why this is the case, which will allow us to tackle any experimental limitations preventing us from observing the unidirectional motion of CMG in our high-throughput experiments.

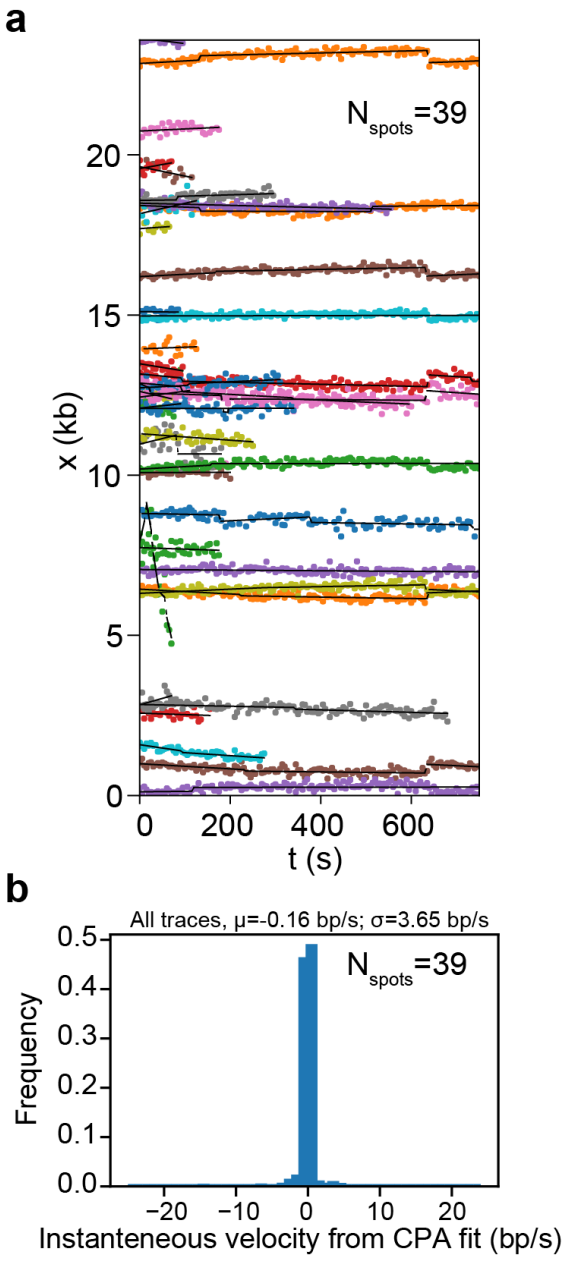



*Figure 5.8* Analysis of unidirectional CMG motion. **a** Position along flow-stretched DNA vs. time plots of CMG spots in the presence of ATP. **b** Distribution of instantaneous velocities coming from the CPA fits of CMG spots bound to flow-stretched DNA in the presence of ATP.

## 5.4 Discussion and outlook

We report the first high-throughput single-molecule imaging of fully reconstituted CMG motion on flow-stretched DNA molecules. This was achieved by adapting our previously developed hybrid ensemble and single-molecule assay from correlative dual-beam optical tweezers and confocal microscopy to flow-stretched DNA TIRF microscopy. To develop this assay, we developed a custom-made microfluidic flow cell (**Fig. 5.2**) that can be easily assembled and allows us to image several channels per glass slide used. Further, we also developed a novel surface functionalization strategy to coat a glass surface with anti-Dig antibody while keeping the surface otherwise passivated with PEG (**Fig. 5.3**).

Our initial data with CMG has demonstrated the feasibility of flow-stretching pre-formed DNA:CMG complexes onto a glass surface, showing that it is possible to achieve stretched-enough DNA molecules containing Mcm2-7 and Cdc45 (**Fig. 5.4**). We note that the amount of Mcm2-7 and Cdc45 per flow-stretched DNA that we observe (**Fig 5.5a-f**), as well as the overall CMG assembly efficiency (**Fig. 5.4g**), are lower than what we observed in our optical tweezers experiments (**Chapter 2**). Further work will be required to understand the reasons behind this discrepancy; nonetheless, though lower than initially expected, the overall number of CMG complexes that we observe by TIRF in a single day of biochemical reconstitution in one field of view (~20-25 CMG complexes per FOV (**Fig. 5.4 f,h**)) is already larger than what we observed with our optical tweezers assay in a day. Further biochemical optimization to increase the number of fluorescent molecules per DNA will be required to increase the overall yield. In addition, experimental advancements in the imaging, such as quick field of view alternation, or the employment of photoactivatable caged ATP[52] will also greatly increase the yield.

In agreement with previous work, including our optical tweezers work with fully reconstituted CMG, we have been able to detect diffusive motion of both Mcm2-7 (**Fig. 5.7a,c**) as well as CMG in the absence of nucleotide (**Fig. 5.7b,d**). Furthermore, in agreement with our optical tweezers work, CMG diffusive motion is halted in the presence of ATPγS (**Fig. 5.7a,c**). Our measured diffusion constants for the three conditions we tested (**Fig. 5.7f-h**) are one order of magnitude slower than what we previously measured in our optical tweezer assay (Ref.[48] and **Fig. S2.5i**). This, together with the fact that we have not been able to detect long-range unidirectional motion of CMG when ATP is added (**Fig. 5.8**), suggest that there may be a systematic problem with our high-throughput assay, or that we have not properly separated static and mobile molecules. We initially anticipated that this may have been due to non-specific interactions of Mcm2-7 and CMG with the anti-

Dig-coated surface. Nevertheless, two lines of evidence suggest that this may not be the case. First, DNA-bound Mcm2-7 and CMG move transversely on the DNA, suggesting that DNA-bound proteins are more mobile than their surface-bound counterparts. Furthermore, we are currently conducting experiments comparing the motion of pre-formed CMG on DNA molecules held in place in an optical trap with that of pre-formed CMG on DNA molecules flow-stretched onto a standard neutravidin-coated surface (both conditions being tested in the presence of ATP). In these preliminary experiments, we have observed substantially less motion of CMG on flow-stretched DNA when compared with optical tweezers (Zhaowei Liu, personal communication), suggesting that the anti-Dig-coated surface may not be the (only) reason behind the lack of motion observed. Work employing different surfaces PEGylated in different ways is currently being carried out to further discard this possibility.

Another difference between our optical tweezers experiments and the TIRF work described here is that in the former we are able to increase the temperature of the sample to 30 °C, whereas we did not have this control in place on our TIRF setup at the time of these experiments. It is therefore a possibility that the activation energy barrier for DNA unwinding by CMG may be not surpassed at the current temperature at which we conduct our experiments. Current work addressing this question is currently being carried out.

Once these limitations have been tackled, our high-throughput single-molecule assay to image CMG will pave the way for the study of this important helicase in the presence of different elongation factors, to obtain an unparalleled mechanist insight of how the motion of CMG is modulated. Furthermore, once the yield is increased, our assay will facilitate a smooth progression in our single-molecule studies of CMG motion to studying a full replisome, and a full replisome in the context of chromatinized DNA.

## 5.5 Methods

### 5.5.1 Biological Materials and Preparation

#### 5.5.1.1 Protein purification

ORC, Cdc6, Mcm2-7$^{Halo\text{-}Mcm3}$/Cdt1, DDK, S-CDK, Sld3/7, Cdc45-iS6, Sld2, GINS, Pol ε, Dbp11, Mcm10 and Sfp phosphopantetheinyl transferase were expressed and purified as described in Chapter 2.

#### 5.5.1.2 Protein labelling

**Mcm2-7$^{JF646\text{-}Mcm3}$**: labeling of Mcm2-7$^{Halo\text{-}Mcm3}$ with fluorescent dye JF646-HaloTag ligand was carried out as described in Chapter 2. Labelling efficiency was measured to be 69.5 ± 3 % (measured value ± instrumental error) by measuring the absorption at 280 nm and 555 nm.

**Cdc45$^{LD555}$:** Cdc45-iS6 was fluorescently labelled by incubating Cdc45-iS6 with Sfp phosphopantetheinyl transferase and LD555-CoA (Lumidyne Technologies, custom synthesis) in a 1:1:4.5 molar ratio in Cdc45 gel filtration buffer (250 K-phosphate, pH 7.6, 150 mM KOAc, 10 % glycerol, and 0.5 mM DTT) supplemented with 10 mM $MgCl_2$ at RT for 1 h. Cdc45$^{LD555}$ was then separated from unincorporated dye and Sfp phosphopantetheinyl transferase by gel filtration in a Superdex 200 Increase 10/300 GL column (Cytiva #15182085). After gel filtration, positive fractions were pooled and concentrated in an Amicon Ultra-4 centrifugal filter Ultracel 30 k (Millipore # UFC803024). Labelling efficiency was measured to be 80.1 ± 3 % (measured value ± instrumental error) by measuring the absorption at 280 nm and 555 nm.

### 5.5.2 Surface preparation for TIRF experiments

#### 5.5.2.1 Slide cleaning

25 x 75 mm # 1.5 glass coverslips were cleaned sonicated in a mixture of 70 % acetone and 30 % methanol at room temperature for 20 min. The coverslips were then washed three times with water and then sonicated in water at room temperature for 5 min. Subsequently, the coverslips were sonicated at room temperature for 1 h in 1 M KOH, washed 3 times with water and sonicated in water at room temperature for 5 min. After the last sonication, the coverslips were stored in water overnight.

#### 5.5.2.2 Piranha etching

Clean coverslips were placed in a Teflon holder and incubated in piranha solution (3:1 mixture of $H_2SO_4$ and $H_2O_2$) until the temperature of the piranha solution reaches room temperature (~ 2 h). After the piranha etching, coverslips were washed 3 times with water.

#### 5.5.2.3 Aminosilanization

Etched coverslips were placed in a Teflon holder previously washed 10 times with methanol. Once in the holder, the coverslips were immediately washed with methanol, and sonicated in fresh methanol at room temperature for 5 min. The coverslips were then incubated for 1 h with shaking with a 5 % (v/v) acetic acid and 10 % (v/v) (3-aminopropyl)triethoxysilane (Sigma Aldrich # 440140) solution in methanol. Subsequently, the coverslips were washed 3 times with methanol, washed 2 times with water and rinsed in fresh water at room temperature for 5 min. After the washes, the coverslips were dried with $N_2$ and immediately PEGylated.

#### 5.5.2.4 PEGylation

The PEGylation protocol used is an adaptation of a previously published protocol[30].

Buffers:

**SBC buffer**: 100 mM $NaHCO_3$ in water.

**SBPS buffer**: SBC buffer supplemented with 550 mM $K_2SO_4$.

Aminosilanized coverslips were covered with 100 µl of buffer SBPS containing 0.02 % (m/v) Biotin-PEG-NHS (5000 Da) (Iris Biotech GmbH # PEG1057.0001) and 0.1 % (w/v) mPEG-NHS (5000 Da) (Tebu Bio # 228PG1-SPA-5k) overnight at 4 °C in a dark and humid chamber. The following day, coverslips were thoroughly washed with water and dried with $N_2$. Subsequently, coverslips were PEGylated overnight with a 100 µl of buffer SBC containing 0.01 % MS(PEG)4 (Thermo Fisher # 10210404) as described above. The following day, the coverslips were thoroughly washed with water, dried with $N_2$ and stored at – 20 °C in $N_2$.

One day prior to each experiment, individual coverslips were PEGylated overnight with 0.01 % MS(PEG)4 as described above. Flow cells were assembled immediately after this third PEGylation step.

### 5.5.3 Single-molecule Instrumentation and imaging

#### 5.5.3.1 Buffers

**Buffer A**: 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, and 1 M NaCl.

**Buffer B**: 10 mM HEPES-KOH pH 7.6, 1 mM EDTA, and 1 M KOAc.

**Buffer C**: 10 mM HEPES-KOH pH 7.6, and 1 mM EDTA.

***Loading Buffer*:** 25 mM HEPES-KOH pH 7.6, 100 mM K glutamate, 10 mM MgOAc, 0.02 % NP40 substitute, 10 % glycerol, 2 mM DTT, 100 µg/ml BSA, and 5 mM ATP.

***HSW Buffer*:** 25 mM HEPES-KOH pH 7.6, 300 mM KCl, 10 mM MgOAc, 0.02 % NP40 substitute, 10 % glycerol, 1 mM DTT, and 400 µg/ml BSA.

***CMG Buffer*:** 25 mM HEPES-KOH pH 7.6, 250 mM K glutamate, 10 mM MgOAc, 0.02 % NP40 substitute, 10 % glycerol, 1 mM DTT, and 400 µg/ml BSA.

***Elution Buffer*:** CMG buffer supplemented with 100 µM biotin and 1 mM ATPγS (Roche # 11162306001).

***Imaging Buffer*:** CMG buffer supplemented with 2 mM 1,3,5,7 cyclooctatetraene, 2 mM 4-nitrobenzylalchohol, and 2 mM Trolox.

#### 5.5.3.2 DNA functionalization and binding to magnetic beads

30 µg of 23.6 kb plasmid pGL50-ARS1 (synthesized as described in **Chapter 2**) containing a natural ARS1 origin were linearized overnight with AflII (NEB # R0520L). The resulting 4-nt overhangs TTAA at both ends of the linear DNA were functionalized by incorporating desthiobiotinylated dATP (Jena Bioscience # NU-835-Desthiobio) and digoxigenylated dUTP (Jena Bioscience #NU-803-DIGXL) with Klenow Fragment (3'→5' exo-) (NEB # M0212L) at 30 °C for 1 h; unincorporated nucleotides were removed with Microspin™ S-400 HR spin columns (GE Healthcare # GE27-5140-01). The functionalized DNA was bound overnight at 4 °C to 4 mg of Dynabeads™ M-280 Streptavidin magnetic beads (Invitrogen # 11205D) in Buffer A. After binding, beads were washed twice with Buffer B, twice with Buffer C, and stored at 4 °C in Buffer C. The amount of bound DNA was measured by comparing the concentration of DNA in the supernatant before and after binding, yielding a binding efficiency of 2.40-2.71 µg DNA (~165-190 fmol)/mg beads.

#### 5.5.3.4 Hybrid ensemble and high-throughput single-molecule assay

##### *5.5.3.4.1 Ensemble CMG assembly and activation*

CMG assembly and activation reactions were carried out in two stages: Mcm2-7 loading and phosphorylation, and CMG assembly and activation. Unless otherwise specified, each step of the reaction was conducted at 30 °C with 800 rpm shaking:

***Mcm2-7 loading and phosphorylation***: 2 mg of magnetic DNA-bound magnetic beads were washed with 300 µl of Loading Buffer, and resuspended in 100 µl of

Loading Buffer. To load Mcm2-7 hexamers onto the origin-containing DNA, 50 nM ORC, 50 nM Cdc6, and 100 nM Mcm2-7$^{JF646\text{-}Halo\text{-}Mcm3}$/Cdt1 were incubated with the beads for a total of 30 min, but added to the reaction at 0 min, 5 min and 10 min, respectively. Subsequently, 100 nM DDK was added and the reaction incubated for 30 min. The supernatant was then removed, and the beads were washed once with 300 µl of HSW buffer and once with 300 µl of CMG Buffer.

***CMG assembly and activation***: after washing, beads were resuspended in 50 µl of CMG Buffer supplemented with 5 mM ATP (Thermo Fisher # 10304340). 200 nM S-CDK was then added, and the reaction was incubated for 5 min. Subsequently, 30 nM Dpb11, 100 nM GINS, 20 nM Polε, 80 nM Cdc45-iS6$^{LD555}$, 30 nM Sld3/7, 50 nM Sld2, and 10 nM Mcm10 were added to the reaction and incubated for 15 min. For this step, a master mix of all the proteins was made immediately before and incubated on ice. After CMG assembly and activation, the supernatant was removed, and the beads were washed three times with 300 µl of HSW Buffer and once with 300 µl of CMG Buffer supplemented with 1 mM ATPγS (Roche # 11162306001). After washing, the assembled DNA-protein complexes were eluted by resuspending the magnetic beads in 200 µl of Elution Buffer, and incubated at RT for 30 min with 800 rpm shaking. The supernatant was then removed and diluted by the addition of 300 µl of CMG Buffer supplemented with 1 mM ATPγS.

#### *5.5.3.4.2 Microfluidic flow cell assembly*

Custom microfluidic flow cells were assembled as follow: five channels with the dimensions specified in **Fig. 5.2** were cut out of a 75 mm X 25 mm X 0.12 mm SecureSeal™ Double Sided Adhesive optical spacer (Grace Bio-Labs # 620001) using a custom made metal stencil and a scalpel. The spacer was then sandwiched between a 75 mm X 25 mm X 1mm quartz slide (Finkenbeiner # 1X3X1MM) with laser-drilled inlet and outlet holes, and a PEGylated coverslip 75 mm X 25 mm # 1.5 coverslip. A pipette tip was used as an inlet and plastic tubing connected to a suction peristaltic pump was used as an outlet.

#### *5.5.3.4.3 Surface functionalization and DNA/DNA:CMG complex flow-stretching*

DNA:CMG complexes eluted from magnetic beads were flow-stretched employing the flow cell described in **Section 5. 3**. Prior to each experiment, the flow cell channel was prepared as follows. All solutions were flowed into each channel at a rate of 50 µl/min.

Prior to flow-stretching, 200 µl of PBS was flown into the channel to hydrate the PEG layer and incubated for 10 minutes. Then, 50 µl of a solution containing 0.2 mg/ml of neutravidin (Thermo Fisher # 10682654) in PBS were flown into the channel and incubated for 20 minutes. After the incubation, excess unbound neutravidin was washed away with 200 µl of PBS. Then, 50 µl of a solution containing 0.2 mg/ml of biotinylated anti-Digoxigenin antibody (Abam # ab419) in PBS were flown into the channel, and incubated for 20 minutes. After the incubation, excess unbound antibody was washed away with 200 µl of PBS.

Following the surface functionalization, 200 µl of CMG Buffer (containing 400 µg/ml of BSA) were flown into the channel and incubated for 10 minutes to further passivate the channel. Then, 500 µl of solution containing the DNA/DNA:CMG eluted from magnetic beads were flown into the channel with constant flow to flow-stretch the DNA molecules/DNA:CMG complexes. Then, 200 µl of CMG buffer were flown to wash any unbound DNA/DNA:CMG. After flow stretching, 50 µl of Imaging Buffer supplemented with 150 nM SYTOX green (Thermo Fisher # S7020), or Imaging Buffer supplemented with 150 nM SYTOX green (Thermo Fisher # S7020) and supplemented either 1 mM ATPγS or 5 mM ATP were flown into the channel, and the sample was imaged right away.

#### *5.5.3.4.4 Single-molecule imaging*

Single-molecule experiments were performed on a commercial micromirror TIRF microscope (Mad City Labs) equipped with 730 nm, 638 nm, 561 nm, and 488 nm lasers. One pixel in the acquired images is equivalent to an area of 96 nm x 96 nm in the sample. All exposure times were set to 300 ms. In multi-color experiments, lasers were alternated every 300 ms. As specified in each figure, the sample was imaged with 1 mW 488 nm laser, 10 mW 561 nm laser and/or 15 mW 638 nm laser, as measured at back focal plane of the objective. In addition, the sample was continuously illuminated with 1 mW 730 nm laser, which does not directly excite any of the fluorophores employed, but is used to correct for drift in Z and keep the sample in focus. All single molecule experiments were conducted at room temperature (22-23 °C).

## 5.5.4 Data analysis

### 5.5.4.1 Software and code

We employed Python 3.10 with the following libraries: numpy==1.24.3; matplotlib==3.7.3; streamlit==1.27.0; scipy==1.11.2; scikit-image==0.22.0; scikit-learn==1.3.1; pyyaml==6.0.1; pandas==2.1.1; pillow==9.5.0; tifffile== 2023.9.26;

jupyterlab==4.0.9; notebook==6.0.3; ruptures==1.1.9; label-studio==1.11.0; picassosr==0.6.5; lmfit==1.2.2; numba==0.58.0.

#### 5.5.4.2 Estimation of DNA tensions from lengths

To estimate DNA tensions, we acquired force-extension curves of a 14.2 kb DNA substrate in the same Imaging Buffer employed in our flow-stretched DNA TIRF assays. Force-extension curves were acquired in an automated manner with commercial dual-beam optical tweezers (C-Trap, LUMICKS) as previously described[49]. To extract the value of the persistence length of DNA in this buffer, force-extension curves were fitted in a force range of 0.5-40 pN to an extensible worm-like chain model (Eq. 13 in Ref.[47]):

$$F = \frac{k_B T}{L_p}\left[\frac{1}{4(1-l)^2} - \frac{1}{4} + l + \sum_{i=2}^{i\leq 7} a_i (l)^i\right], \text{with } l = \frac{d}{l_c} - \frac{F}{S} \quad \textbf{(Eq. 5.1)}$$

where $F$= force, $d$=end-to-end distance of DNA, $L_p$=persistence length, $L_c$=contour length, $S$=stretch modulus, $k_B$=Boltzmann's constant, $T$=temperature, $a_2 = -0.5164228$, $a_3 = -2.737418$, $a_4 = -16.07497$, $a_5 = -38.87607$, $a_6 = -39.49944$, and $a_7 = -14.17718$

From this fit, we extracted persistence lengths, contour lengths and stretch moduli of the DNA (**Fig. S5.1**). The average persistence length value (48.9 nm) (**Fig. S5.1b**) was then used to estimate the tension of the flow-stretched 23.6 kb DNA molecules by applying an the following inextensible worm-like chain model (Eq. 10 in Ref.[47]) to the measured DNA lengths:

$$F = \frac{k_B T}{L_p}\left[\frac{1}{4(1-l)^2} - \frac{1}{4} + l + \sum_{i=2}^{i\leq 7} a_i (l)^i\right], \text{with } l = \frac{d}{l_c} \quad \textbf{(Eq. 5.2)}$$

where $F$= force, $d$=end-to-end distance of DNA, $L_p$=persistence length, $L_c$=contour length, $S$=stretch modulus, $k_B$=Boltzmann's constant, $T$=temperature, $a_2 = -0.5164228$, $a_3 = -2.737418$, $a_4 = -16.07497$, $a_5 = -38.87607$, $a_6 = -39.49944$, and $a_7 = -14.17718$.

#### 5.5.4.3 Overview of fluorescence data analysis

After acquiring TIRF microscopy images, the raw images were processed to generate a table containing the spot detections in each frame. These detected spots are then connected between frames to generate traces containing location and intensity over time.

#### 5.5.4.4 DNA identification

DNA double tethers were manually identified in each field of view and labelled using Python Label-Studio. A tight rectangular box was drawn around each double tether. The mid-points of the two shorter sides of the rectangle were taken as the DNA top and bottom coordinate, respectively, and the DNA was modelled as a line segment connecting these two dots. Each DNA was given a unique identified called "idSegment".

#### 5.5.4.5 Spot detection and tracking

For spot detection and localization we used Picasso Localize[53], which localizes identified spots with sub-pixel resolution by fitting a 2-dimensiotnal Gaussian curve to each detected spot. The localized spots in each frame (in the green and/or red channels) were then mapped onto the DNA (blue) channel using S/T polynomial decomposition[54]. All subsequent analysis was conducted using the coordinates of the spots mapped onto the blue channel.

After mapping fluorescent spots onto the blue channel, we tracked the spots frame-by-frame using our own implementation of the Linear Assignment Problem (LAP) framework[55] as previously described[1,48,49].

#### 5.5.4.6 Colocalization determination

##### *5.5.4.6.1 Colocalization of fluorescent proteins with DNA segments*

To determine whether a fluorescent spot mapped to the DNA channel is colocalized with a DNA segment, we defined a colocalization distance as follows: we generated a time-projection of the raw data image of three DNA molecules (projected through 20 frames or 6 s) to obtain a projected image capturing the oscillatory motion of the DNA along its long axis. For clarity, we defined the axis parallel to the DNA long axis as the x-axis, and the perpendicular axis as the y-axis (**Fig. S5.2a**). We then measured the intensity profile along the y-axis for each row of pixels along the x-axis, fitted each intensity profile to a Gaussian function and extracted the standard deviation $\sigma_y$ for every pixel column along the x-axis. Then, to define a distance threshold to determine colocalization of fluorescent spots with DNA molecules, we modelled the distribution of $\sigma_y$ as a continuous distribution using a kernel density estimate (KDE), and used the mode of the resulting distribution (10.29 pixels or 1.076 μm) as our DNA-colocalization distance threshold (**Fig. S5.2b**). Thus, a fluorescent spot is considered colocalized with a DNA segment if the distance between the spot (once mapped to the blue channel) and a DNA (modelled as a line

segment) is less than 10.29 pixels. For all red and green spots, colocalization with the DNA was determined at frame number 3.

#### *5.5.4.6.2 Colocalization of green and red fluorescent proteins*

To determine whether a red fluorescent spot is colocalized with a green fluorescent spot, we calculated a colocalization distance specific to each pair of spots as follows: we first calculated the average standard deviation along the x-axis of each spot in a track (of note, the fitted spots were symmetric, so $\sigma_x = \sigma_y$). We then defined a colocalization distance between a red and a green spot as:

$$D_{coloc} = 3 * \left( \frac{\hat{\sigma}_{x,red} + \hat{\sigma}_{x,green}}{2} \right)$$

Then, a red spot and a green spot are considered colocalized if the distance between their initial position (average of the first 3 frames) is less or equal to $D_{coloc}$.

### 5.5.4.7 Determination of number of fluorophores per diffraction-limited spot

To count the number of bleaching steps of each diffraction-limited spot, we fitted the intensity vs. time traces to a piecewise constant function using Change-Point Analysis (CPA) (we use a Python implementation called 'ruptures'[56]). This analysis was done as previously described[1], using a minimum step size of $2 \times 10^7$ photons, which we determined empirically by looking at several single bleaching step curves.

### 5.5.4.8 Positional analysis

To analyze positions along the DNA, we projected each DNA-colocalized spot onto its respective DNA axis in every frame, and measured the distance between each projected spot and one end of the DNA (in pixels). We then converted these distances from pixels to kb by calculating a pixel-to-kb conversion factor for each DNA molecule. In the initial position plots, we plot the average position of the first three frames of each trace. The bin size of the initial position histograms was set to 700 bp to be close to the diffraction limit while having the ARS1 origin positioned near the center of its corresponding bin[1].

#### 5.5.4.9 Motion analysis

##### *5.5.4.9.1 Calculation of MSD and diffusion coefficients*

For RMSD plots along x- and y- axes (**Fig. 5.6**), we calculated the RMSD for each trace (in microns) with a delay time of 1 frame (900 ms). In this analysis, we did not project the spot positions onto their respective DNA to be able to fairly compare the motion of spots colocalized with DNA with that of spots not colocalized with DNA.

To calculate diffusion coefficients along the DNA long axis (**Fig. 5.7**), we first projected each DNA-colocalized spot onto its respective DNA, and calculated the diffusion coefficient along the DNA axis as previously described[1,48], using a previously published appropriate range of delay times[57].

##### *5.5.4.9.2 CPA analysis of CMG*

CPA analysis of all CMG traces in buffer supplemented with ATP was conducted almost exactly as described in Chapter 2, to be able to compare our data with that of Chapter 2. The only difference was that we did not perform a drift correction of our traces, as we currently do not have a measurement of the drift.

#### 5.5.4.10 Data filtering

The resulting data table of traces including with number of fluorophores per spot was to reduce noise, outliers, and tracking errors, as follows:

1. Diffraction-limited spots containing more than 5 fluorescent proteins, are filtered out to prevent aggregates from entering our analysis.
2. Traces starting after frame 3 are filtered out because we do not expect fluorescent proteins to land on the DNA during the scan.
3. Traces with fewer than 10 frames are filtered out because we expect these to represent incorrectly detected spots.
4. Spots with widths ($\sigma_x$) less than 1.4 pixels are filtered out to remove incorrectly detected noise.
5. Finally, for our diffusion analysis, we only used tracks with at least 20 frames to obtain better fits.

#### 5.5.4.11 Boxplot parameters

All boxplots in this study show quartiles Q1, Q2 and Q3, and the whiskers show the minimum and maximum non-outlier values, where outliers are defined as data

points smaller than Q1 − 1.5 × interquartile range or larger than Q3 + 1.5 × interquartile range.

## 5.6 Data availability

Raw and processed data have been deposited in the 4TU.ResearchData repository and can be found at https://doi.org/10.4121/851a4f32-af27-4e63-ab8d-f86f4566f93f. This repository contains:

- A table with an overview of experiments.
- Spot position and intensity tables sorted by experimental condition
- Force-distance curves
- Motion analysis summary tables, containing motion information for each trace. The scan_id, trace_id fields link this table to the table containing the full trace information.

## 5.7 Code availability

All the code used in this manuscript is available at:

https://gitlab.tudelft.nl/nynke-dekker-lab/public/fsDNA

## 5.8 References

1. Ramírez Montero, D. *et al.* Nucleotide binding halts diffusion of the eukaryotic replicative helicase during activation. *Nat. Commun.* **14**, (2023).

2. de Jesús-Kim, L., Friedman, L. J., Ramsoomair, C., Gelles, J. & Bell, S. P. DDK regulates replication initiation by controlling the multiplicity of Cdc45-GINS binding to Mcm2-7. *Elife* **10:e65471**, 1–30 (2021).

3. Bell, S. P. & Labib, K. Chromosome duplication in Saccharomyces cerevisiae. *Genetics* **203**, 1027–1067 (2016).

4. Yeeles, J. T. P., Deegan, T. D., Janska, A., Early, A. & Diffley, J. F. X. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**, 431–435 (2015).

5. Douglas, M. E., Ali, F. A., Costa, A. & Diffley, J. F. X. The mechanism of eukaryotic CMG helicase activation. *Nature* **555**, 265–268 (2018).

6. Kurat, C. F., Yeeles, J. T. P., Patel, H., Early, A. & Diffley, J. F. X. Chromatin

Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates. *Mol. Cell* **65**, 117–130 (2017).

7. Yeeles, J. T. P., Janska, A., Early, A. & Diffley, J. F. X. How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication. *Mol. Cell* **65**, 105–116 (2017).

8. Lewis, J. S. *et al.* 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**, 10630–10635 (2017).

9. Lewis, J. S. *et al.* Tunability of DNA polymerase stability during eukaryotic DNA replication. *Suppl. Inf.* **77**, 602086 (2019).

10. Burnham, D. R., Kose, H. B., Hoyle, R. B. & Yardimci, H. The mechanism of DNA unwinding by the eukaryotic replicative helicase. *Nat. Commun.* **10**, 1–14 (2019).

11. Lewis, J. S. *et al.* 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**, 10630–10635 (2017).

12. Kose, H. B., Larsen, N. B., Duxin, J. P. & Yardimci, H. Dynamics of the Eukaryotic Replicative Helicase at Lagging-Strand Protein Barriers Support the Steric Exclusion Model. *Cell Rep.* **26**, 2113-2125.e6 (2019).

13. Kose, H. B., Xie, S., Cameron, G., Strycharska, M. S. & Yardimci, H. Duplex DNA engagement and RPA oppositely regulate the DNA-unwinding rate of CMG helicase. *Nat. Commun.* **11**, 1–15 (2020).

14. Lewis, J. S. *et al.* Tunability of DNA polymerase stability during eukaryotic DNA replication. *Mol. Cell* **77**, 1–9 (2020).

15. Yardimci, H. *et al.* Bypass of a protein barrier by a replicative DNA helicase. *Nature* **492**, 205–209 (2012).

16. Yardimci, H., Loveland, A. B., van Oijen, A. M. & Walter, J. C. Single-molecule analysis of DNA replication in Xenopus egg extracts. *Methods* **57**, 179–186 (2012).

17. Sparks, J. L. *et al.* The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell* **176**, 167-181.e21 (2019).

18. Low, E., Chistol, G., Zaher, M. S., Kochenova, O. V. & Walter, J. C. The DNA replication fork suppresses CMG unloading from chromatin before termination. *Genes Dev.* **34**, 1534–1545 (2020).

19. Vrtis, K. B. *et al.* Single-strand DNA breaks cause replisome disassembly. *Mol. Cell* **81**, 1309-1318.e6 (2021).

20. Gruszka, D. T., Xie, S., Kimura, H. & Yardimci, H. Single-molecule imaging reveals control of parental histone recycling by free histones during DNA replication. *Sci. Adv.* **6**, (2020).

21. Geertsema, H. J., Kulczyk, A. W., Richardson, C. C. & Van Oijen, A. M. Single-molecule studies of polymerase dynamics and stoichiometry at the bacteriophage T7 replication machinery. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4073–4078 (2014).

22. Scherr, M. J., Wahab, S. A., Remus, D. & Duderstadt, K. E. Mobile origin-licensing factors confer resistance to conflicts with RNA polymerase ll ll Mobile origin-licensing factors confer resistance to conflicts with RNA polymerase. *CellReports* **38**, (2022).

23. Dequeker, B. J. H. *et al.* MCM complexes are barriers that restrict cohesin-mediated loop extrusion. *Nature* **606**, 197–203 (2022).

24. Kose, H. B., Larsen, N. B., Duxin, J. P. & Yardimci, H. Dynamics of the Eukaryotic Replicative Helicase at Lagging-Strand Protein Barriers Support the Steric Exclusion Model. *Cell Rep.* **26**, 2113-2125.e6 (2019).

25. McCluskey, K. & Dekker, N. H. Principles and best practices of optimizing a micromirror-based multicolor TIRF microscopy system. *Opt. Commun.* **538**, (2023).

26. Axelrod, D., Thompson, N. L. & Burghardt, T. P. Total internal reflection fluorescent microscopy. *J. Microsc.* **129**, 19–28 (1983).

27. Martin-Fernandez, M. L., Tynan, C. J. & Webb, S. E. D. A "pocket guide" to total internal reflection fluorescence. *J. Microsc.* **252**, 16–22 (2013).

28. White, D. S., Smith, M. A., Chanda, B. & Goldsmith, R. H. Strategies for Overcoming the Single-Molecule Concentration Barrier. *ACS Meas. Sci. Au* **3**, 239–257 (2023).

29. Pradhan, B. *et al.* The Smc5/6 complex is a DNA loop-extruding motor. *Nature* **616**, 843–848 (2023).

30. Park, S. R. *et al.* A Single-Molecule Surface-Based Platform to Detect the Assembly and Function of the Human RNA Polymerase II Transcription Machinery. *Structure* **28**, 1337–1343 (2020).

31. Chandradoss, S. D. *et al.* Surface passivation for single-molecule protein studies. *J. Vis. Exp.* 1–8 (2014) doi:10.3791/50549.

32. Ticau, S., Friedman, L. J., Ivica, N. A., Gelles, J. & Bell, S. P. Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell* **161**, 513–525 (2015).

33. Ticau, S. *et al.* Mechanism and timing of Mcm2-7 ring closure during DNA

replication origin licensing. *Nat. Struct. Mol. Biol.* **24**, 309–315 (2017).

34. Zhang, A., Friedman, L. J., Gelles, J. & Bell, S. P. Changing protein–DNA interactions promote ORC binding- site exchange during replication origin licensing. *Proc. Natl. Acad. Sci.* **120**, 1–12 (2023).

35. Champasa, K., Blank, C., Friedman, L. J., Gelles, J. & Bell, S. P. A conserved Mcm4 motif is required for Mcm2-7 double-hexamer formation and origin DNA unwinding. *Elife* **8**, 1–23 (2019).

36. Gupta, S., Friedman, L. J., Gelles, J. & Bell, S. P. A helicase- - tethered ORC flip enables bidirectional helicase loading. *Elife* **10:e74282**, 1–31 (2021).

37. Hirsch, J. D. *et al.* Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: Uses for protein labeling, detection, and isolation. *Anal. Biochem.* **308**, 343–357 (2002).

38. Goswami, P. *et al.* Structure of DNA-CMG-Pol epsilon elucidates the roles of the non-catalytic polymerase modules in the eukaryotic replisome. *Nat. Commun.* **9**, (2018).

39. Eickhoff, P. *et al.* Molecular Basis for ATP-Hydrolysis-Driven DNA Translocation by the CMG Helicase of the Eukaryotic Replisome. *Cell Rep.* **28**, 2673-2688.e8 (2019).

40. Seifert, M. *et al.* Inhibition of sars-cov-2 polymerase by nucleotide analogs from a single-molecule perspective. *Elife* **10**, 1–26 (2021).

41. Janissen, R. *et al.* Global DNA Compaction in Stationary-Phase Bacteria Does Not Affect Transcription. *Cell* **174**, 1188-1199.e14 (2018).

42. Janissen, R. *et al.* Induced intra- and intermolecular template switching as a therapeutic mechanism against RNA viruses. *Mol. Cell* **81**, (2021).

43. Bell, N. A. W. & Molloy, J. E. Efficient golden gate assembly of DNA constructs for single molecule force spectroscopy and imaging. *Nucleic Acids Res.* **50**, E77 (2022).

44. Lewis, J. S. *et al.* Mechanism of replication origin melting nucleated by CMG helicase assembly. *Nature* (2022) doi:10.1038/s41586-022-04829-4.

45. Abid Ali, F. *et al.* Cryo-EM structure of a licensed DNA replication origin. *Nat. Commun.* **8**, 1–10 (2017).

46. Parker, M. W. *et al.* A new class of disordered elements controls DNA replication through initiator self-assembly. *Elife* **8**, 1–35 (2019).

47. Bouchiat, C. *et al.* Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.* **76**, 409–413 (1999).

48. Sánchez, H. *et al.* DNA replication origins retain mobile licensing proteins. *Nat. Commun.* **12**, (2021).

49. Sánchez, H. *et al.* A chromatinized origin reduces the mobility of ORC and MCM through interactions and spatial constraint. *Nat. Commun.* **14**, (2023).

50. Crut, A., Lasne, D., Allemand, J. F., Dahan, M. & Desbiolles, P. Transverse fluctuations of single DNA molecules attached at both extremities to a surface. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **67**, 051910/1-051910/6 (2003).

51. Remus, D. *et al.* Concerted Loading of Mcm2-7 Double Hexamers around DNA during DNA Replication Origin Licensing. *Cell* **139**, 719–730 (2009).

52. Sabantsev, A. *et al.* Spatiotemporally controlled generation of NTPs for single-molecule studies. *Nat. Chem. Biol.* **18**, 1144–1151 (2022).

53. Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12**, 1198–1228 (2017).

54. McCluskey, K. A. *et al.* Global correction of optical distortions in multicolor single-molecule microscopy using Zernike polynomial gradients. *Opt. Express* **29**, (2021).

55. Jaqaman, K. *et al.* Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* **5**, 695–702 (2008).

56. Truong, C., Oudre, L. & Vayatis, N. Selective review of offline change point detection methods. *Signal Processing* **167**, 107299 (2020).

57. Michalet, X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **82**, 1–13 (2010).

## 5.9 Acknowledgements

We thank Anne Early, Lucy Drury, and Max Douglas for providing the yeast strains for the overexpression of unlabeled proteins, Mike Filius, Dimitri de Roos and Koushik Sreenivasa for assistance in making the custom-made microfluidic flow cell; N.D. lab members Julien Gros for assistance in purifying DDK, and Kaley McCluskey for assistance in microscope alignment and training; and Edo van Veen for assistance in developing initial data analysis routines. We also thank Zhaowei Liu, Humberto Sánchez, Pang-Yen Wang, Jacob Lewis, Alessandro Costa, Hasan Yardimci, and Anita Meier for useful scientific discussions. D.R.M. acknowledges funding from a Boehringer Ingelheim Fonds PhD Fellowship, and N.D.

acknowledges funding from the Netherlands Organisation for Scientific Research (NWO) through Top grant 714.017.002, and from the European Research Council through an Advanced Grant (REPLICHROMA; grant number 789267).

## 5.10 Author contributions statement

D.R.M. and N.D. conceived the study. D.R.M. and T.v.L. purified the proteins. D.R.M. and H.S. labeled the proteins. D.R.M. developed the ensemble hybrid and single-molecule assay with input from N.D. D.R.M. engineered and developed the custom-made microfluidic flow cell with input from S.T and N.D.. D.R.M. developed and validated the novel surface functionalization strategy. D.R.M. acquired the single-molecule TIRF data with assistance from S.T. Z.L. acquired the force-distance curves. F.P.M. designed and wrote data analysis routines with input from all authors. D.R.M. and F.P.M. performed the data analysis. S.T. maintained and aligned the microscope. All authors were involved in the discussion of the data. D.R.M. wrote the manuscript with input from F.P.M. and N.D.

## 5.11 Competing interests statement

The authors declare no competing financial interests.

## 5.12 Materials & correspondence

Correspondence and requests for materials should be addressed to N.D. (n.h.dekker@tudelft.nl)

## 5.13 Supplementary figures

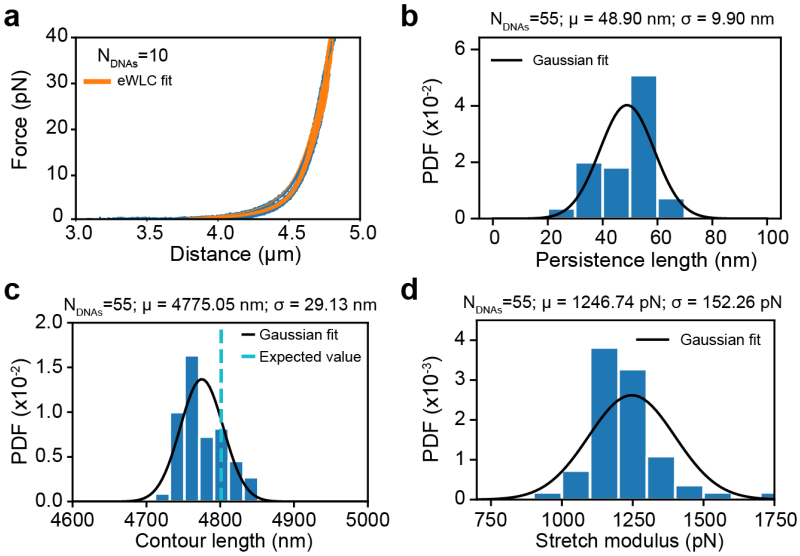



*Figure S5.1* Measuring the persistence length of DNA in Imaging Buffer **a** Example force-extension curves of 14.2 DNA construct (blue) measured in dual-beam optical tweezers, together with the fitted extensible worm-like chain (eWLC) model (Eq. 13 in Ref.[47]) plotted in the force range used for fitting (orange) (**Methods**). Probability density function of **b** persistence lengths, **c** contour lengths, and **d** stretch moduli obtained from the extensible worm-like chain model fits of force-extension curves of a 14.2 kb DNA construct acquired in the Imaging Buffer used in TIRF experiments; black lines show a Gaussian fit of the data, and μ and σ are the mean and the standard deviation of the data, respectively. Dotted cyan line in panel **c** shows the expected contour length of the DNA construct used. The average persistence length value was used to estimate the tension of flow-stretched DNA molecules (**Methods**).

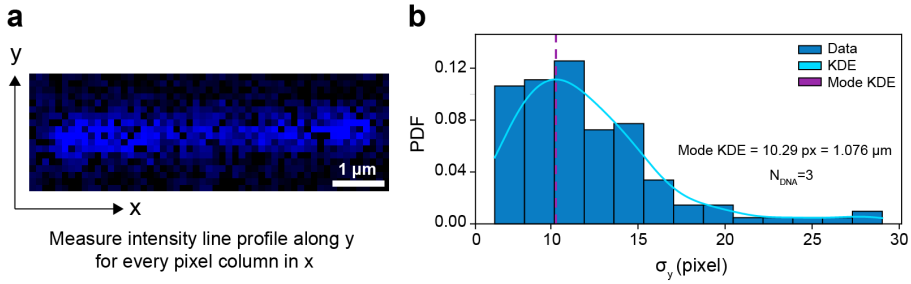



*Figure S5.2* Determination of distance threshold for the colocalization of fluorescent spots with DNA molecules. **a** Example Z-projection of a flow-stretched DNA molecule through time (20 frames or 6 s). The width of the projected image captures the range of motion of the DNA molecule about its tethering axis. We then measured intensity line profiles along the y-axis for every pixel column along the x-axis, fitted each intensity profile to a Gaussian function and extracted the standard deviation $\sigma_y$ for every pixel column. **a** Distribution of $\sigma_y$ values (measured as described in **a**) of three different DNA molecules. To define a distance threshold to determine colocalization of fluorescent spots with DNA molecules, we modeled the distribution of $\sigma_y$ as a continuous distribution using a kernel density estimate (KDE), and used the mode of the resulting distribution as our DNA-colocalization distance from a DNA molecule (which we model as a line segment (**Methods**)).

# 6

# A hybrid ensemble and single-molecule assay to image the motion of fully reconstituted CMG

*This chapter is in preparation for submission for peer review publication as: Ramírez Montero D., Sánchez H., van Veen E., van Laar T., and Dekker, N.H. (2024). A hybrid ensemble and single-molecule assay to image the motion of fully reconstituted CMG.*

## 6.1 Abstract

Eukaryotes have one replicative helicase known as CMG, which centrally organizes and drives the replisome, and leads the way at the front of replication forks. Obtaining a deep mechanistic understanding of the dynamics of CMG is critical to understanding how cells achieve the enormous task of efficiently and accurately replicating their entire genome once per cell cycle. Single-molecule techniques are uniquely suited to quantify the dynamics of CMG due to their unparalleled temporal and spatial resolution. Nevertheless, single-molecule studies of CMG motion have thus far relied on pre-formed CMG purified from cells as a complex, which precludes the study of the steps leading up to its activation. Here, we describe a hybrid ensemble and single-molecule assay that allowed us to image at the single-molecule level the motion of fluorescently labeled CMG after fully reconstituting its assembly and activation from 36 different purified *S. cerevisiae* polypeptides. This assay relies on the double functionalization of the ends of a linear DNA substrate with two orthogonal attachment moieties and can be adapted to study similarly complex DNA-processing mechanisms at the single-molecule level.

## 6.2 Introduction

DNA replication in eukaryotes is carried out by a dynamic protein complex known as the replisome[1]. A key component of this complex is the eukaryotic replicative helicase Cdc45/Mcm2-7/GINS (CMG) helicase, which is the main molecular motor driving and centrally organizing the replisome, and leads the way at the front of replication forks[1,2]. Obtaining a deep quantitative understanding of the dynamics of CMG is therefore critical to understanding the dynamics of the replisome. Such an understanding could be acquired with single-molecule techniques, which are uniquely suited to study molecular motors, such as CMG, due to their unmatched spatial and temporal resolution, and provide us with an unparalleled quantitative understanding of their function, stochasticity and dynamics[2–9].

*In vivo*, CMG is loaded and activated in temporally separated fashion to ensure that replication occurs only once per cell cycle[1,10,11]. First, in the G1-phase of the cell cycle, a set of proteins known as loading factors loads the first component of CMG, the Mcm2-7 hexameric complex, onto dsDNA[12–16] in the form of double hexamers located in a head-to-head conformation[15,17,18]. In the specific case of yeast, this initial process occur at specific DNA sequences known as origins of replication[1]. Although Mcm2-7 is the motor core of the replicative helicase, it is by itself unable to unwind DNA[19] without the two helicase-activating factors Cdc45 and GINS, which need to be recruited to the loaded Mcm2-7 to give rise to fully active CMG[11,19–21]. The process of helicase activation takes place in the G1-phase of the cell cycle, and starts by the selective phosphorylation of Mcm2-7 double hexamers by the cell cycle-regulated kinase DDK[22–24]. These phosphorylation events facilitate the recruitment of Cdc45 and GINS to the Mcm2-7 double hexamers[10,22–26] by a second set of proteins known as firing factors[10,11,26]. The binding of Cdc45 and GINS gives rise to two sister CMG helicases, which are initially enclosing both strands of the parental DNA and located in a head-to-head configuration[11,27]. In a final activation step, the firing factor Mcm10 catalyzes the ATP hydrolysis-dependent extrusion of one DNA strand from each sister CMG[11]. After strand extrusion, sister CMG helicases bypass and separate from one another by translocating along ssDNA in an ATP hydrolysis dependent manner[11,20,21,28], unwinding DNA by sterically excluding the non-translocation strand[29]. This entire process has been fully reconstituted *in vitro* from a minimal set of 36 purified *S. cerevisiae* polypeptides[10,11].

Despite the exquisite *in vivo* regulation of CMG assembly and activation described above, *in vitro* reconstituted single-molecule motion studies of CMG[2,30–34] have thus far relied on pre-activated CMG purified as a complex from cells[20,21], missing all the steps prior to its activation and the bidirectional nature of its motion. This pre-activated CMG approach has been the gold standard in the single-molecule field partly due to the biochemical complexity of the fully reconstituted CMG assembly reaction[10,11]. This biochemical reaction has been challenging to translate from the bulk biochemical level to the single-molecule level for several reasons. First, to maximize reaction efficiencies, the loading and firing factors needed for CMG assembly and activation are required at concentrations in the range of 10-200 nM[10,11,27]. These ranges of concentration correspond to the high end of what most single-molecule techniques can tolerate, especially when using fluorescently labeled components[35]. Finally, CMG has evolved to cruise through thousands of base pairs in a cell[36–39]. Therefore, to study its motion at a biologically relevant spatial scale, one requires long DNA substrates (typically of lengths in the order of tens of kilobases)[30,31,34,40–42]. Employing such long DNA substrates poses the additional challenge that, the longer the DNA substrate, the more potential non-specific binding sites for proteins and protein aggregates. In the case of CMG, this latter point is particularly important, as several of the loading and firing factors involved in CMG assembly and activation contain intrinsically disordered regions[43] and are aggregation-prone.

Here, we report a hybrid ensemble and single-molecule assay that allowed us to observe and quantify the motion of CMG after fully reconstituting its assembly and activation from 36 purified *S. cerevisiae* polypeptides[28]. This assay relies on the double functionalization of both ends of a DNA substrate with two orthogonal attachment moieties: desthiobiotin and digoxigenin[2]. The first moiety, desthiobiotin, is used to reversibly bind the DNA substrate to streptavidin-coated magnetic beads[44]. We then assemble and activate CMG onto the bead-bound DNA, and employ a magnetic rack to purify and wash the resulting magnetic bead-bound DNA:CMG complexes. In doing so, we remove excess protein that would otherwise aggregate on the DNA substrate; this provides us with virtually aggregation-free DNA:CMG complexes. Intact complexes are then eluted from the magnetic beads by the addition of a molar excess of free biotin which can outcompete the desthiobiotin-streptavidin interaction. We then bind individual DNA:CMG complexes between two optically trapped polystyrene beads coated with anti-digoxigenin antibody (anti-Dig); for this step, we take advantage of the fact that the second moiety on the DNA, digoxigenin, can bind to anti-Dig even in a buffer solution containing an excess of free biotin. We anticipate that this assay can be

easily adapted for the study at the single-molecule level of similarly complex DNA:protein interactions.

## 6.3 Protocol

### 6.3.1 Synthesis of doubly functionalized linear DNA substrate and binding to magnetic beads.

#### 6.3.1.1 Dual functionalization of DNA substrate with desthiobiotin and digoxigenin moieties (**Fig. 6.1a**)

1. Linearize 20 µg of 23.6 kb plasmid pGL50-ARS1 (containing a natural ARS1 origin of replication) with 200 units of restriction enzyme AflII for 16 h at 37 °C in a final volume of 200 µL of 1X CutSmart Buffer. This step can be reduced to 4 hours without reducing the yield, if more convenient.
2. Inactivate AflII by incubating the reaction at 65 °C for 20 min.
3. Blunt the resulting 4-nucleotide TTAA overhangs by supplementing the 200-µL linearization reaction with 60 units of Klenow Fragment (3'→5' exo-) polymerase, 17 µL of 10X NEBuffer 2, 33 µM D-Desthiobiotin-7-dATP, 33 µM Digoxigenin-11-dUTP, 33 µM dCTP, 33 µM dGTP, and MilliQ water to a final volume of 370 µM. Incubate the reaction at 37 °C for 30 min.
4. Supplement the reaction with 10 mM EDTA and inactivate the Klenow Fragment by incubating the reaction at 75 °C for 20 min.
5. Bring the volume of the reaction up to 400 µL with MilliQ water.
6. Take 4 x S-400 spin columns, vortex them for at least 30 s to resuspend the resin, and then centrifuge them for 1 min at 0.8 x g to remove the storage buffer. Transfer the columns to clean 1.5 mL tubes.
7. Immediately, add 100 µL of the DNA solution to each column, and centrifuge them for 2 min at 0.8 x g. The DNA is now in the flow-through and the columns may be discarded.
8. Pool together the flowthrough of the four columns and measure the exact volume and DNA concentration, which will be used to calculate the amount of DNA bound to the beads.

#### 6.3.1.2 Binding doubly functionalized linear DNA to streptavidin-coated magnetic beads (**Fig. 6.1b**)

1) Vortex M-280 streptavidin-coated magnetic beads for 30 s to resuspend them.

2) Transfer 4 mg of resuspended M-280 streptavidin-coated magnetic beads to a clean 1.5 mL tube. Place the tube in a magnetic rack, wait 1 min for the beads to be collected by the magnet, and remove the storage buffer.

3) Add 1 mL of Buffer A (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, and 1 M NaCl) to the beads and resuspend by vortexing for 5 s. Incubate the beads at room temperature for 5 min.

4) Place the tube in a magnetic holder, wait 1 min for the beads to be collected by the magnet, and remove Buffer A.

    a. Important: do not discard the supernatant before measuring its volume and the concentration of unbound DNA.

    b. Calculate the total amount of DNA bound to the magnetic beads by comparing the total amount of DNA added to the beads with the amount of DNA left in the supernatant. The yield should be in the range of 2.3-2.9 mg of DNA (~150-190 fmol) per mg of magnetic beads.

5) Resuspend the beads in 400 µL of 2X Buffer A (10 mM Tris-HCl pH 7.5, 1 mM EDTA, and 2 M NaCl) by pipetting, and then add the 400 µL of functionalized DNA solution. Mix gently by pipetting.

6) Incubate the bead/DNA mixture overnight at 4 °C with end-over-end rotation to allow the functionalized DNA to bind to the beads.

7) With the help of the magnetic rack, remove the buffer and wash the beads twice with 500 µL of Buffer B (10 mM HEPES-KOH pH 7.6, 1 mM EDTA, and 1 M KOAc), and twice with 500 µL of Buffer C (10 mM HEPES-KOH pH 7.6, and 1 mM EDTA).

8) Finally, resuspend the beads in 300 µL of Buffer C and store at 4 °C. We recommend making 4 single-use aliquots of 1 mg of magnetic beads, and not storing the DNA for longer than two weeks to prevent nicking.

## 6.3.2 Hybrid ensemble and single-molecule assay to image and quantify the motion of fully reconstituted CMG with correlative dual-beam optical tweezers and confocal microscopy

### 6.3.2.1 Ensemble assembly and activation of fluorescently labeled CMG onto magnetic bead-bound *DNA* (**Fig. 6.1c**).

CMG assembly and activation reactions were carried out in two stages: Mcm2-7 loading and phosphorylation, and CMG assembly and activation. Unless otherwise specified, all buffer exchange steps were conducted with the help of a magnetic rack, by allowing the beads to be collected by the magnet for 1 min and then removing the supernatant. All incubations were conducted in a temperature-controlled heat block with a lid to prevent photobleaching of fluorescent proteins. If a lid is now available, the tubes should be covered in tin foil. The purification and fluorescent labeling of all the proteins employed in this protocol has been carried out as previously described[10,11,28].

#### *6.3.2.1.1 Mcm2-7 loading and phosphorylation*

1) Take 1 mg of DNA-bound streptavidin-coated magnetic beads and remove the storage buffer (Buffer C).
2) Wash the beads with 200 µL of Loading Buffer (25 mM HEPES-KOH pH 7.6, 100 mM K glutamate, 10 mM MgOAc, 0.02 % NP40 substitute, 10 % glycerol, 2 mM DTT, 100 µg/ml BSA, and 5 mM ATP).
3) Remove the Loading Buffer and resuspend the beads in 75 µl of Loading Buffer. Mix gently by pipetting.
4) Add ORC at a final concentration of 35.7 nM to the bead-bound DNA and incubate the reaction for 5 min at 30 °C with 800 rpm agitation.
5) Add Cdc6 at a final concentration of 50 nM Cdc6 and incubate the reaction for 5 min at 30 °C with 800 rpm agitation.
6) Add Mcm2-7/Cdt1 (or fluorescently labeled Mcm2-7$^{JF646\text{-}Halo\text{-}Mcm3}$/Cdt1) at a final concentration of 100 nM and incubate the reaction for 20 min at 30 °C with 800 rpm agitation.

7) Add DDK at a final concentration of 100 nM and incubate the reaction for 30 min at 30 °C with 800 rpm agitation.

8) Remove the supernatant and wash the bead-bound DNA (which now contains phosphorylated Mcm2-7 hexamers) with 200 µL of high-salt wash (HSW) Buffer (25 mM HEPES-KOH pH 7.6, 300 mM KCl, 10 mM MgOAc, 0.02 % NP40 substitute, 10 % glycerol, 1 mM DTT, and 400 µg/ml BSA). Mix by pipetting, ensuring that the beads are fully resuspended in the buffer and no clumps are visible.

9) Remove the HSW Buffer and wash the beads once with 200 µL of CMG buffer (25 mM HEPES-KOH pH 7.6, 250 mM K glutamate, 10 mM MgOAc, 0.02 % NP40 substitute, 10 % glycerol, 1 mM DTT, and 400 µg/ml BSA).

#### *6.3.2.1.2 Assembly and activation of fluorescently labelled CMG onto magnetic bead bound origin-containing DNA and elution of intact DNA:CMG complexes from the magnetic beads.*

1) Remove the CMG buffer and resuspend the DNA-bound beads in 50 µL of CMG Buffer supplemented with 5 mM ATP.

2) Add 50 nM Dpb11, 200 nM GINS, 30 nM Polε, 20 nM S-CDK, 50 nM Cdc45LD555, 30 nM Sld3/7, 55 nM Sld2, and 10 nM Mcm10 to the bead-bound DNA. For this step, mix all the proteins in one tube immediately before adding them to the DNA, and place it on ice. Add the resuspended bead-bound DNA to the protein mix. Incubate the reaction for 15 min at 30 °C with 800 rpm agitation.

3) Wash the beads three times with 200 µL of HSW Buffer.

4) Wash the beads once with 200 µL of CMG Buffer.

### *6.3.1.3 Elution of intact DNA:CMG complexes from magnetic beads* (**Fig. 6.1d**).

1) Remove CMG buffer and elute DNA:CMG complexes from the magnetic beads by resuspending the CMG-containing DNA-bound magnetic beads in 200 µL of Elution Buffer (CMG buffer supplemented with 10 mM biotin). Incubate at room temperature for 1 h with 800 rpm agitation.

2) Place the tube in a magnetic rack, and allow the beads to be collected for 5 minutes. Carefully collect the supernatant (which now contains the eluted DNA:CMG complexes) without disrupting the settled beads, and transfer it to a new tube.

3) To ensure that there are no beads left in the solution, place again the collected supernatant in a magnetic rack and allow any remaining beads to be collected for another 5 minutes. Carefully collect the supernatant and transfer it to a new tube.

4) Add 1400 µL of CMG Buffer to the 200 µL-supernatant. The sample is now ready for single-molecule imaging.

#### 6.3.2.2 Single-molecule imaging of fully reconstituted CMG with correlative dual-beam optical tweezers and confocal microscopy (**Fig. 6.1e-g**).

The single-molecule part of the assay is conducted in a commercial setup that combines dual-beam optical tweezers with confocal microscopy and microfluidics[45,46] (**Fig. 6.1e-g**), but may also be conducted in a home-built setup. The commercial setup we employed is equipped with a microfluidic flow cell with five inlets (referred to as Channels 1-5, respectively) and one outlet (**Fig. 6.1e**). Channels 1-3 are injected from the left and used for bead trapping (Channel 1), DNA-protein complex binding (Channel 2), and to check for the presence of CMG (Channel 3). Channels 4 and 5 are used as protein reservoirs and buffer exchange locations. Before each experiment, the microfluidic flow cell needs to be passivated for at least 30 min with 1 mg/mL bovine serum albumin, followed by 0.5% Pluronic F-127 dissolved in ultrapure water.

The content of each channel in each experiment is as follows (**Fig. 6.1e**):

- ***Channel 1*:** Anti-digoxigenin coated polystyrene beads (2.06 µm diameter) diluted 1:50 in PBS.
- ***Channel 2*:** DNA:CMG complexes eluted from magnetic beads.
- ***Channel 3*:** Imaging Buffer (CMG buffer supplemented with 2 mM 1,3,5,7 cyclooctatetraene, 2 mM 4-nitrobenzylalchohol, and 2 mM Trolox)
- ***Channels 4 and 5:*** Imaging Buffer supplemented with 25 nM RPA, 10 nM Mcm10, and either 5 mM ATP, 5 mM ATPγS, or no nucleotide.

1) Before the experiment, adjust the trapping laser power was adjusted to achieve a stiffness of 0.3 pN/nm in both traps[33,46]. Set confocal pixel size was set to 50 x 50 nm, illumination time per pixel to 0.2 ms, and tframe rate to 5 s.

2) Flow all solutions into the flow cell at a constant pressure in the injection port of 0.5 bar. Then, turn off the flow in channels 4 and 5. After initially flowing all solutions, reduce the pressure to 0.2 bar.

3) Move trapping lasers to Channel 1 until one bead is caught in each optical trap.

4) Move trapped beads to Channel 2 and fish a DNA:CMG complex by increasing and decreasing the distance between the beads until a DNA tether is trapped (which is known by monitoring the force-extension curve of the trapped DNA[47]).

5) Move the beads to Channel 3 and immediately stop the flow in all channels. Take a one-frame test scan in Channel 3 to confirm the presence of CMG.

6) If CMG is present, move the DNA tether to either channel 4 or 5, for imaging.

7) In channel 4 or 5, adjust the distance between the beads to achieve a tension of 2 pN in the DNA tether.

8) Image $CMG^{Cdc45\text{-}LD555}$ with a 561 nm laser at a power of 4 μW as measured at the objective, and by taking 2D confocal scans over an area of 160 x 18 pixels (**Fig. 6.1f**). In such scans, CMG will show as a two-dimensional diffraction-limited spot, such as the example shown in **Fig. 6.1g**.

We note that, although this publication focuses on the hybrid ensemble and single-molecule assay, we have previously published a comprehensive description of the analysis of the data that we have generated with this assay[48].

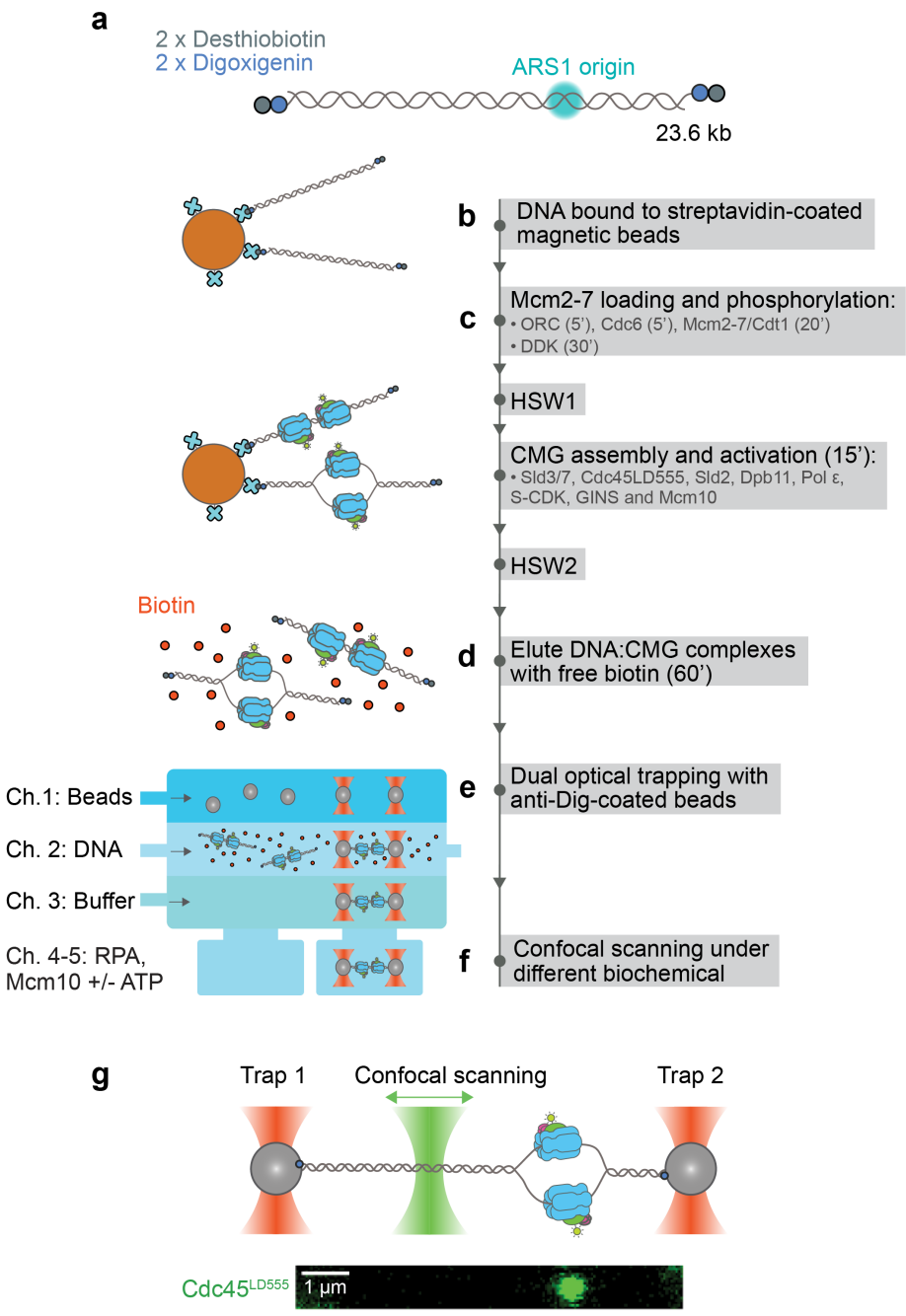



***Figure 6.1*** Pictorial description of hybrid ensemble and single-molecule assay to image and quantify the motion of fully reconstituted CMG. **a** 23.6 kb linear DNA containing a naturally occurring ARS1 origin of replication is doubly functionalized at both ends with desthiobiotin and digoxigenin moieties. **b** Doubly functionalized DNA is bound to streptavidin-coated magnetic beads through its desthiobiotin moieties. **c** CMG is stepwise assembled and activated onto the magnetic bead-bound DNA with different washing steps included to remove excess unbound protein and protein aggregates. (*Continues on the next page*).

***Figure 6.1** (Continued).* **d** Intact DNA:CMG complexes are then eluted from the magnetic beads by the addition of an excess of free biotin, which outcompetes the desthiobiotin-streptavidin interaction. **e** Individual DNA:CMG complexes are bound between two anti-digoxigenin-coated optically trapped polystyrene beads with the help of a microfluidic flow cell. Note that the Dig-anti-Dig interaction is orthogonal to biotin-avidin interactions, so it is not affected by the presence of free biotin. **f** Once held in place by the optical tweezers, DNA:CMG complexes are transferred into different buffer conditions, where the DNA plane is then scanned with a confocal scanning laser to image the motion of CMG along the DNA over time. **g** (top) Diagram of a DNA:CMG complex held in place between two optically trapped anti-Dig-coated polystyrene beads being scanned by a confocal scanning laser. (bottom) Example 2D scan of CMG bound to a DNA held in place with an optical trap. The DNA is unlabeled in these experiments, but it can be thought of as a horizontal line running through the middle of the image.

## 6.4 Representative results

When carried out correctly, the protocol described here should yield virtually aggregate-free DNA:CMG complexes. An aggregation-free reaction should not clog any of the channels in the microfluidic flow cell, and it should be possible to stretch the trapped DNA molecule to an end-to-end extension within 10% of its contour length without breaking the DNA. On the other hand, if there is aggregation in the reaction, DNA molecules may sometimes become compacted by the aggregates, often causing DNA breakage if stretched. A good way to recognize protein aggregates is by looking at the 2D scans of the DNA. In an aggregation-free reaction, CMG will appear as discrete, symmetrical diffraction-limited spots sparsely crowding the DNA, such as the ones in the scans shown in **Fig. 6.2a**. On the contrary, aggregates will show as less discrete, sometimes asymmetrical blobs crowding a larger length on the DNA, like those in the scans shown in **Fig. 6.2b**. Furthermore, if the assay is successfully executed, and high purify of the purified proteins was achieved, one should detect long-range motion of CMG in the presence of ATP, as in the kymograph shown in **Fig. 6.2c**.

## 6.5 Discussion

### 6.5.1 Critical steps and important reagent quality checks

There are several critical steps and biological reagent quality checks in the assay we describe here that should be highlighted. First, the purity of the proteins used is highly important because DNA degradation caused by even small nuclease contaminants in the protein samples will adversely affect the data. This is because only intact DNA molecules can be trapped in the dual-beam optical tweezers, so any cleaved DNA molecules cannot be imaged. More importantly, nicks on the DNA will

cause CMG to dissociate[41], complicating the observation of CMG long range motion. We strongly recommend testing each purified protein for nuclease activity, as well as constantly monitoring the integrity of the starting plasmid substrate to ensure that nicking is reduced to a minimum. A second important step is the careful removal of the magnetic beads following the elution of DNA:CMG complexes. Supernatant removal in this step should be conducted slowly so as not to perturb the collected beads. If magnetic beads are left in the sample flown into the optical tweezers, they will often hit the optically trapped polystyrene beads, causing them

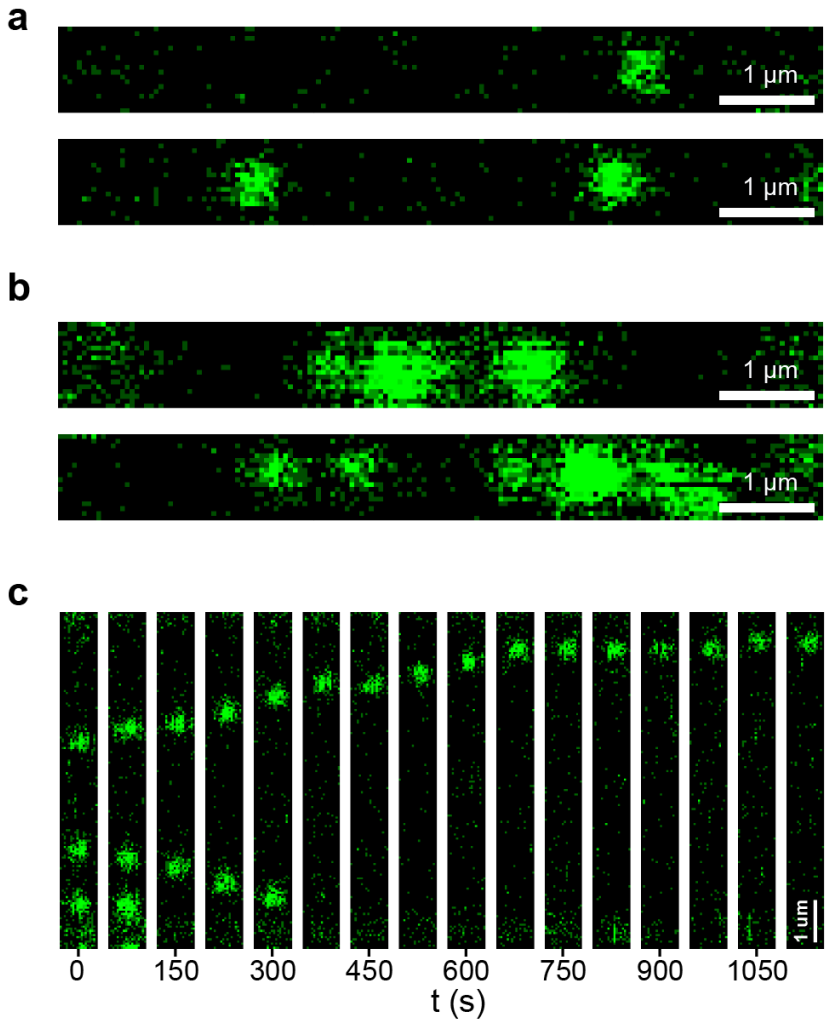



*Figure 6.2* Examples of data from a successful and an unsuccessful experiment. **a** Example 2D scans of a CMG-containing DNA in an aggregation-free sample. In both scans, CMG shows as symmetrical and discrete diffraction-limited spots sparsely distributed along the DNA. **b** Example 2D scans of a CMG-containing DNA in a sample containing aggregates. In both scans, CMG shows as less symmetrical blobs crowding the DNA. **c** Kymograph showing the position on the DNA of CMG diffraction-limited spots over time in the presence of ATP, showing the long-range motion of CMG complexes.

to escape the optical trap and complicating the data acquisition. Finally, DNA:CMG complexes should be handled carefully in the optical tweezers. To this end, we recommend not increasing the tension of the DNA above 10 pN, as the application of force may dissociate CMG from the DNA. Furthermore, moving between channels in the microfluidic flow cell, should be done as slowly as possible (~ 0.2 mm/s) to prevent the resulting drag forces from dissociating CMG from the DNA.

### 6.5.2 Modifications of the method

There are several steps of the assay that could be modified. For instance, we have shown that the elution time can be reduced form 60 min to 30 min without significantly affecting the elution yield. In addition, we recommend supplementing the elution buffer with either a low (below 1 mM) concentration of ATP or ATPγS to prevent CMG from diffusing off the DNA ends as well as to stabilize CMG[28]. Further, although the buffer compositions and protein concentrations we report here are greatly based on those employed in prior ensemble biochemical and single-molecule work[11,18], the assay we describe should be fully compatible with other protocols to assemble CMG[26,27]. Therefore, any biochemical advancement reported to increase the efficiency of CMG assembly or activation could and should be implemented in the bulk part of the assay to increase the yield. Finally, increasing the time between frames increases the total time in which CMG can be imaged, facilitating the observation of long-range CMG motion before fluorophore bleaching.

### 6.5.3 Limitations of the method

The hybrid method we describe is limited in that one can only image CMG following its activation in bulk. Further work will be required to be able to be able observe the activation of CMG in real time. Another important limitation is that, while we expect CMG to be assembled in pairs[17,26,27], the total number of CMG complexes per DNA that we observe is mostly one[28], suggesting that CMG, or at least that Cdc45 is dissociating from the DNA during the handling of the sensitive DNA:CMG complexes. Reducing the number of handling steps prior to the single-molecule imaging, as well as a developing a better passivation of the plastic tubing and glass of the microfluidic flow cell are poised to increase this yield.

### 6.5.4 Significance of the method

Single-molecule motion studies of CMG have thus far employed pre-activated CMG purified as a complex from cells. While relatively simpler, this pre-activated CMG approach is limited in that it misses any the steps leading up to CMG activation, as well as the bidirectional nature of CMG and replisome motion. On the other hand,

the full reconstitution of CMG assembly and activation has the potential to study any pre-activation steps, as well as to study CMG motion in a bidirectional manner. Nevertheless, this approach is harder to translate from the bulk biochemical level to the single-molecule level, as it involves a lot more purified protein factors and steps. The assay we describe here has helped us to overcome these challenges, by allowing us to image the motion of fully reconstituted CMG at the single-molecule level, allowing us to access some previously missed pre-activation dynamics[28].

Another advantage of this assay compared to previous CMG motion lies in the fully double-stranded nature of the DNA substrate we employ (**Fig. 6.1a**). In previous pre-activated CMG work, the most common way of binding CMG to the DNA substrate is through a 3' ssDNA flap. This results in a DNA construct that cannot be easily torsionally constrained and thus prohibits study of the role of supercoiling in replisome progression. Conversely, the new approach we describe here could has the potential to be adapted to study the role of torque in this process, as the DNA substrate used is fully double-stranded.

### 6.5.5 Broader applications of the method

We hope that the hybrid assay we describe will pave the way towards the full reconstitution of a complete eukaryotic replisome, allowing us and others to observe and quantify the important dynamics that allow the replisome to success at all its different tasks. DNA replication aside, the assay that we report represents an important advancement in translating a complicated biochemical reaction from the bulk biochemical to the single-molecule level. We anticipate that this assay can be easily modified to study similarly complex DNA:protein interactions involved in different DNA processing mechanisms.

## 6.6 Acknowledgements

The authors thank Anne Early, Lucy Drury, and Max Douglas for providing yeast strains for the overexpression of unlabeled proteins, and Jacob Lewis for providing M.HpaII methyltransferase., as well as N.D. lab members Anuj Kumar, Katinka Ligthart, and Julien Gros for their help purifying loading factors and DDK. The authors also thank Kaley McCluskey, Dorian Mikolajczak, Joseph Yeeles, Jacob Lewis, Alessandro Costa, Hasan Yardimci, and Taekjip Ha for useful scientific discussions. This work was supported by a Boehringer Ingelheim Fonds PhD Fellowship awarded to D.R.M., and by Netherlands Organisation for Scientific

Research (NWO) Top grant 714.017.002, and European Research Council Advanced Grant (REPLICHROMA; grant number 789267) awarded to N.D.

## 6.7 Disclosures

The authors have nothing to disclose.

## 6.8 References

1. Bell, S. P. & Labib, K. Chromosome duplication in Saccharomyces cerevisiae. *Genetics* **203**, 1027–1067 (2016).

2. Burnham, D. R., Kose, H. B., Hoyle, R. B. & Yardimci, H. The mechanism of DNA unwinding by the eukaryotic replicative helicase. *Nat. Commun.* **10**, 1–14 (2019).

3. Ticau, S., Friedman, L. J., Ivica, N. A., Gelles, J. & Bell, S. P. Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell* **161**, 513–525 (2015).

4. Ticau, S. *et al.* Mechanism and timing of Mcm2-7 ring closure during DNA replication origin licensing. *Nat. Struct. Mol. Biol.* **24**, 309–315 (2017).

5. Gupta, S., Friedman, L. J., Gelles, J. & Bell, S. P. A helicase- - tethered ORC flip enables bidirectional helicase loading. *Elife* **10:e74282**, 1–31 (2021).

6. Lewis, J. S. *et al.* 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**, 10630–10635 (2017).

7. Dulin, D., Berghuis, B. A., Depken, M. & Dekker, N. H. Untangling reaction pathways through modern approaches to high-throughput single-molecule force-spectroscopy experiments. *Curr. Opin. Struct. Biol.* **34**, 116–122 (2015).

8. Lewis, J. S., van Oijen, A. M. & Spenkelink, L. M. Embracing Heterogeneity: Challenging the Paradigm of Replisomes as Deterministic Machines. *Chem. Rev.* **123**, 13419–13440 (2023).

9. Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460–465 (2005).

10. Yeeles, J. T. P., Deegan, T. D., Janska, A., Early, A. & Diffley, J. F. X. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**,

431–435 (2015).

11. Douglas, M. E., Ali, F. A., Costa, A. & Diffley, J. F. X. The mechanism of eukaryotic CMG helicase activation. *Nature* **555**, 265–268 (2018).

12. Remus, D. *et al.* Concerted Loading of Mcm2-7 Double Hexamers around DNA during DNA Replication Origin Licensing. *Cell* **139**, 719–730 (2009).

13. Frigola, J., Remus, D., Mehanna, A. & Diffley, J. F. X. ATPase-dependent quality control of DNA replication origin licensing. *Nature* **495**, 339–343 (2013).

14. Coster, G., Frigola, J., Beuron, F., Morris, E. P. & Diffley, J. F. X. Origin Licensing Requires ATP Binding and Hydrolysis by the MCM Replicative Helicase. *Mol. Cell* **55**, 666–677 (2014).

15. Coster, G. & Diffley, J. F. X. Bidirectional eukaryotic DNA replication is established by quasi-symmetrical helicase loading. *Science (80-. ).* **357**, 314–318 (2017).

16. Miller, T. C. R., Locke, J., Greiwe, J. F., Diffley, J. F. X. & Costa, A. Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM. *Nature* **575**, 704–710 (2019).

17. Abid Ali, F. *et al.* Cryo-EM structure of a licensed DNA replication origin. *Nat. Commun.* **8**, 1–10 (2017).

18. Sánchez, H. *et al.* DNA replication origins retain mobile licensing proteins. *Nat. Commun.* **12**, (2021).

19. Ilves, I., Petojevic, T., Pesavento, J. J. & Botchan, M. R. Activation of the MCM2-7 Helicase by Association with Cdc45 and GINS Proteins. *Mol. Cell* **37**, 247–258 (2010).

20. Moyer, S. E., Lewis, P. W. & Botchan, M. R. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10236–10241 (2006).

21. Langston, L. D. *et al.* CMG helicase and DNA polymerase ε form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15390–15395 (2014).

22. Sheu, Y. J. & Stillman, B. Cdc7-Dbf4 Phosphorylates MCM Proteins via a Docking Site-Mediated Mechanism to Promote S Phase Progression. *Mol. Cell* **24**, 101–113 (2006).

23. Sheu, Y. J. & Stillman, B. The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature* **463**, 113–117 (2010).

24. Randell, J. C. W. *et al.* Mec1 Is One of Multiple Kinases that Prime the Mcm2-7 Helicase for Phosphorylation by Cdc7. *Mol. Cell* **40**, 353–363 (2010).

25. Greiwe, J. F. *et al.* Structural mechanism for the selective phosphorylation of DNA-loaded MCM double hexamers by the Dbf4-dependent kinase. *Nat. Struct. Mol. Biol.* **29**, (2021).

26. de Jesús-Kim, L., Friedman, L. J., Ramsoomair, C., Gelles, J. & Bell, S. P. DDK regulates replication initiation by controlling the multiplicity of Cdc45-GINS binding to Mcm2-7. *Elife* **10:e65471**, 1–30 (2021).

27. Lewis, J. S. *et al.* Mechanism of replication origin melting nucleated by CMG helicase assembly. *Nature* (2022) doi:10.1038/s41586-022-04829-4.

28. Ramírez Montero, D. *et al.* Nucleotide binding halts diffusion of the eukaryotic replicative helicase during activation. *Nat. Commun.* **14**, (2023).

29. Kose, H. B., Larsen, N. B., Duxin, J. P. & Yardimci, H. Dynamics of the Eukaryotic Replicative Helicase at Lagging-Strand Protein Barriers Support the Steric Exclusion Model. *Cell Rep.* **26**, 2113-2125.e6 (2019).

30. Lewis, J. S. *et al.* 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**, 10630–10635 (2017).

31. Lewis, J. S. *et al.* Tunability of DNA polymerase stability during eukaryotic DNA replication. *Mol. Cell* **77**, 1–9 (2020).

32. Schauer, G. D. *et al.* Replisome bypass of a protein-based R-loop block by Pif1. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 30354–30361 (2020).

33. Wasserman, M. R., Schauer, G. D., O'Donnell, M. E. & Liu, S. Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase. *Cell* **178**, 600-611.e16 (2019).

34. Kose, H. B., Xie, S., Cameron, G., Strycharska, M. S. & Yardimci, H. Duplex DNA engagement and RPA oppositely regulate the DNA-unwinding rate of CMG helicase. *Nat. Commun.* **11**, 1–15 (2020).

35. White, D. S., Smith, M. A., Chanda, B. & Goldsmith, R. H. Strategies for Overcoming the Single-Molecule Concentration Barrier. *ACS Meas. Sci. Au* **3**, 239–257 (2023).

36. Liachko, I. *et al.* A comprehensive genome-wide map of autonomously replicating sequences in a naive genome. *PLoS Genet.* **6**, 22 (2010).

37. Kapadia, N. *et al.* Processive Activity of Replicative DNA Polymerases in the Replisome of Live Eukaryotic Cells. *Mol. Cell* **80**, 114-126.e8 (2020).

38. Claussin, C., Vazquez, J. & Whitehouse, I. Single-molecule mapping of replisome progression. *Mol. Cell* **82**, 1372-1382.e4 (2022).

39. Polo Rivera, C. & Deegan, T. D. Replicon-seq: seeing is believing. *Trends*

*Genet.* **38**, 987–988 (2022).

40. Sparks, J. L. *et al.* The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell* **176**, 167-181.e21 (2019).

41. Vrtis, K. B. *et al.* Single-strand DNA breaks cause replisome disassembly. *Mol. Cell* **81**, 1309-1318.e6 (2021).

42. Low, E., Chistol, G., Zaher, M. S., Kochenova, O. V. & Walter, J. C. The DNA replication fork suppresses CMG unloading from chromatin before termination. *Genes Dev.* **34**, 1534–1545 (2020).

43. Parker, M. W. *et al.* A new class of disordered elements controls DNA replication through initiator self-assembly. *Elife* **8**, 1–35 (2019).

44. Hirsch, J. D. *et al.* Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: Uses for protein labeling, detection, and isolation. *Anal. Biochem.* **308**, 343–357 (2002).

45. Candelli, A., Wuite, G. J. L. & Peterman, E. J. G. Combining optical trapping, fluorescence microscopy and micro-fluidics for single molecule studies of DNA-protein interactions. *Phys. Chem. Chem. Phys.* **13**, 7263–7272 (2011).

46. Candelli, A. *et al.* Visualization and quantification of nascent RAD51 filament formation at single-monomer resolution. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15090–15095 (2014).

47. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic Elasticity of lambda-Phage DNA. *Science (80-. ).* **265**, 1599–1600 (1994).

48. Liu, Z. *et al.* A biophysics toolbox for reliable data acquisition and processing in integrated force-confocal fluorescence microscopy. *ACS Photonics* (2024).

## 6.9 Table of materials

| Name of Material/ Equipment | Company | Catalog Number |
|---|---|---|
| AflII | NEB | R0520L |
| CutSmart Buffer | NEB | B6004S |
| Klenow Fragment (3'→5' exo-) | NEB | M0212L |
| NEBuffer2 | NEB | B7002S |
| D-Desthiobiotin-7-dATP | Jena Bioscience | NU-835-Desthiobio |
| Digoxigenin-11-dUTP | Jena Bioscience | NU-803-DIGXL |
| dCTP | Promega | U122B |
| dGTP | Thermo Fisher | 10218014 |
| Microspin™ S-400 HR spin columns | GE Healthcare | GE27-5140-01 |
| Dynabeads™ M-280 Streptavidin magnetic beads | Invitrogen | 11205D |
| ATP solution | Thermo Fisher | R0441 |
| C-Trap | Lumicks | |
| BSA | NEB | B9000S |
| Pluronic F-127 | Sigma | P2443 |
| Anti-digoxigenin coated polystyrene beads | Spheroteck | DIGP-20-2 |
| ATPγS | Roche | 11162306001 |

# 7

# Concluding remarks

As stated in the introduction, this thesis is my contribution to the long-term goal of imaging and quantifying the dynamics of a full eukaryotic replisome in a chromatinized context. Achieving this goal will of undoubtedly take more than the 4.5 years of work included in this thesis, and it will likely involve many more researchers in different laboratories. Nonetheless, it is my hope that the work presented in the previous chapters will lay the foundation for future work in our lab and in others. In this chapter, I will discuss potential directions in which the research presented in the previous chapters could be continued to get us closer to our long-term goal.

## 7.1 Single-molecule imaging of fully reconstituted CMG

The reconstitution of regulated eukaryotic DNA and chromatin replication from purified *S. cerevisiae* proteins[1–3] has been a breakthrough in the field of ensemble biochemistry, allowing many researchers worldwide to study and understand the different intermediate steps in eukaryotic DNA initiation, elongation, and termination[3–12]. Conversely, *in vitro* reconstituted single-molecule studies of the motion dynamics of CMG and full replisomes[13–18] have thus far employed pre-formed CMG purified as a complex from cells[19,20], which misses all the steps leading up to activation. To fill this gap, we set out to study the motion of CMG after fully reconstituting its assembly and activation.

In **Chapter 2**, we report the first imaging and quantification of the motion of fully reconstitute CMG at the single-molecule level using correlative dual-beam optical tweezers and confocal microscopy[21], providing the field with a different way of studying the dynamics of involved in CMG activation. This was achieved by the development of a novel hybrid ensemble and single-molecule assay relying on the double functionalization of DNA with two orthogonal attachment types[21]. This assay, initially described in **Chapter 2** and described in depth in **Chapter 6**, allowed us to translate the complex biochemical reaction of CMG activation from the

ensemble biochemical level[3] to the single-molecule level. Furthermore, we illustrate the adaptability of this powerful assay in **Chapter 5** by showing that it can be made compatible with high throughput TIRF microscopy imaging.

Although a growing body of structural[22–25] and biochemical[1,20,26,27] work suggests that there are no major differences between the pre-formed CMG and the fully reconstituted CMG systems, only the fully reconstituted system allows us to interrogate intermediate steps prior to helicase activation[3,25,28–37]. Furthermore, it is yet to be ascertained whether the phosphorylation states of pre-formed and fully reconstituted CMG are fully identical. If differences exist, they might affect the protein-protein interactions that CMG can support, as CMG phosphorylation has been important to modulate some of its protein-protein interactions[4,38].

The work studying the motion of fully reconstituted CMG presented in this thesis constitutes only an initial approach that should be expanded upon and further optimized in future work. In the following sections, we will discuss some of the potential ways in which this can be achieved.

## 7.2 Real-time imaging of CMG activation

One of the limitations of the hybrid and single-molecule approaches described in **Chapters 2** is that it only allows us to image the motion of fully reconstituted CMG following its activation in bulk, rendering the observation of CMG activation in real time highly unlikely. In the future, it would be of interest to modify these assays to be able to visualize CMG activation *in situ*. To that end, one could consider omitting the firing factor Mcm10 from the ensemble part of the assay, to allow for the assembly of head-to-head sister CMGs but without strand extrusion[3,25], and then add Mcm10 during the single-molecule imaging (**Fig. 7.1a**). However, it is yet to be determined whether Mcm10 is by itself able to catalyze strand extrusion from the central channel of CMG, or whether other firing factors need to be bound to the head-to-head sister CMG helicases in order for Mcm10-catalyzed strand extrusion to occur[25,39]. If the latter is true, the preservation of intact DNA:CMG:firing factors complex prior to the single-molecule imaging may be difficult to achieve due to the washing steps required to remove intermediates. Future work testing different washing conditions may provide us with gentle enough conditions to preserve such sensitive complexes. Alternatively, one could test the addition of Mcm10 in addition to other firing factors during the single-molecule imaging, to find the minimal set of proteins required for strand extrusion. Of note, in work not presented in this thesis, we have tried to omit Mcm10 from the CMG activation part of the reaction in our optical tweezers work presented in **Chapter 2**. Nonetheless, under such

conditions, the 23.6 kb DNA molecules we employ were compacted by what appeared to be either protein aggregates or a large protein complex. This compaction prevented us from acquiring meaningful data in the optical tweezers, as the compacted DNA:CMG complexes would often break as we tried to stretch them.

Another possible approach could be to load and phosphorylate Mcm2-7 double hexamers in bulk, and add all the firing factors[1,3] at the single-molecule level (**Fig. 7.1b**). This approach may nonetheless require further optimization of the CMG assembly conditions, as our experience tells us that firing factors easily aggregate onto the optically trapped polystyrene beads, greatly complicating the single-molecule imaging (we tried this approach in work not presented in this thesis. In those experiments, we observed that the optically trapped beads were

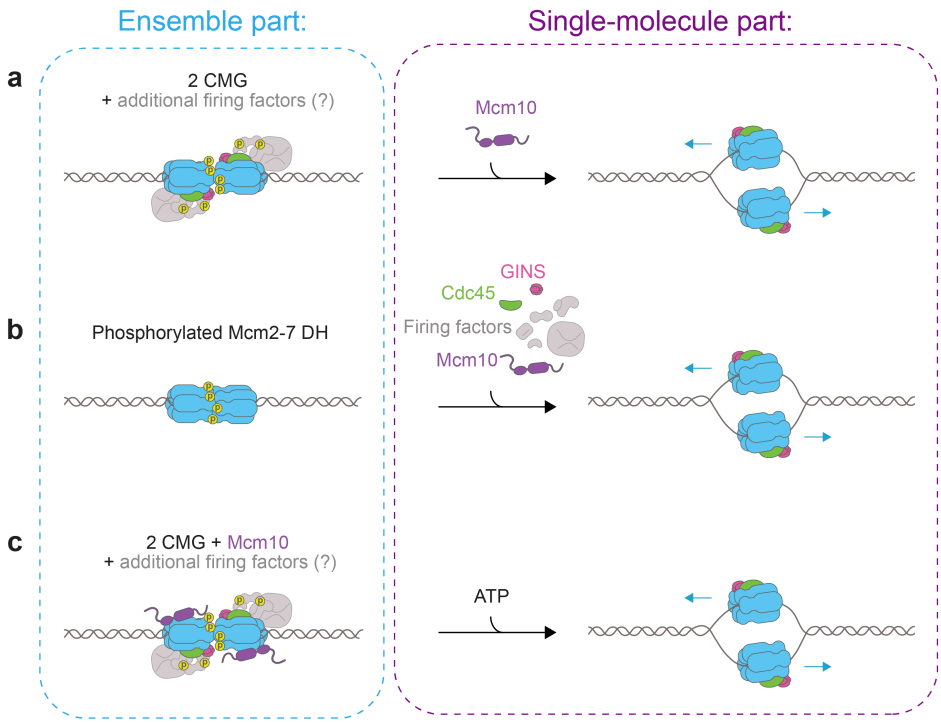



*Figure 7.1* Alternative experimental approaches to image CMG activation *in situ*. **a** Mcm10 could be omitted from the ensemble part of the reaction to allow for the formation of sister CMGs without strand extrusion (left); other firing factors may form a complex with CMGE. Addition of Mcm10 during the single-molecule imaging could lead to CMG activation (right). **b** Loading and phosphorylation can be carried during the ensemble part of the reaction (left). Addition of all the firing factors (including Mcm10) during the single-molecule imaging could lead to CMG assembly and activation (right). **c** Employing pre-phosphorylated Sld2 and Sld3 during the ensemble part of the reaction could allow for the formation of sister CMGs in complex with Mcm10 and possibly other firing factors in the presence of ATPγS (left). The addition of ATP during the single-molecule imaging could lead to CMG activation (right). Orange circles represent the phosphorylated state of Mcm2-7, and/or Sld2 and Sld3.

quickly covered with fluorescent protein aggregates that covered our entire field of view and prevented us from imaging the DNA). Furthermore, the fact that in optical tweezers, DNA:CMG complexes are studied one at a time adds extra complications to this approach, as the firing factors would have to be incubated in the flow cell for long periods of time (up to ~5 minutes per DNA molecule to allow for CMG assembly alone[3], plus the time required to image CMG translocation), likely aggregating over time. For this reason, this specific alternative approach might be simpler to implement employing the flow-stretched DNA TIRF high-throughput assay described in **Chapter 5**, as one could simultaneously observe hundreds of CMG assembly and activation steps, greatly reducing variability between individual DNA molecules.

A third approach that would be interesting to investigate would be to pre-phosphorylate Sld3 and Sld2 with S-phase CDK[40] and purify the phosphorylated proteins, as previously done[3]. This would allow for CMG assembly to take place in the presence of ATPγS[3], but prevent strand extrusion, previously shown to require ATP hydrolysis[3]. Under such conditions, one may be able to trap an intermediate complex composed of CMG bound to Mcm10 (and potentially to other firing factors). Then, during the single-molecule imaging, one could add ATP, to hopefully observe sister CMG separation in real time (Fig. 7.1c).

## 7.3 Direct imaging of Mcm10

The mechanism of Mcm10-mediated strand extrusion during CMG activation is a critical step in DNA replication initiation that remains poorly understood[3,18]. Therefore, in all the alternative approaches described in the preceding section, it will be of great interest to image fluorescently labeled Mcm10 in addition to CMG. Visualizing Mcm10 would be especially informative if the real-time observation of CMG activation is achieved, as one could then quantify how many Mcm10 molecules are required for sister CMG separation to occur, as well as to monitor how many of these Mcm10 molecules, if any, remain bound to the CMG following strand extrusion.

Even without any of the additional modifications described in the preceding section, the hybrid ensemble and single-molecule assay we developed in **Chapter 2** could benefit from a direct visualization of Mcm10. For instance, one could carry out the ensemble activation of CMG in the presence of fluorescently labeled Mcm10, and then monitor the colocalization of Mcm10 with CMG during the single-molecule imaging part of the assay. Of note, Mcm10 has been reported to form a stable

complex with CMG[5,41], suggesting that it might remain bound to CMG during the handling prior to the single-molecule imaging. The presence of colocalized Mcm10 with CMG may then be used as a proxy to monitor which CMG complexes might have undergone strand extrusion (those bound to Mcm10), and which ones have not (those not bound to Mcm10). In turn, this analysis could help to test the final model proposed in **Chapter 2**, as one could correlate the presence of colocalized Mcm10 with the different motion types we observed. It should be mentioned that this approach may require further optimization of the concentration of Mcm10 employed, as Mcm10 has been shown to bind Mcm2-7, albeit weakly, prior to CMG activation[42].

## 7.4 Differentiating between CMG motion double- or single-stranded DNA

Another limitation of the hybrid ensemble and bulk assay reported in **Chapters 2, 5 and 6** is that one cannot determine which CMG molecules are enclosing ssDNA and which ones are enclosing dsDNA. This is particularly important as growing body of evidence has suggested that CMG can translocate on dsDNA in addition to ssDNA[43–45].

A potential way to fill this gap would be the addition of fluorescently labeled RPA during the single-molecule imaging. Work from our lab employing pre-formed CMG has shown that RPA binding to long stretches of unwound ssDNA behind a unidirectionally moving CMG can only be observed at very low salt concentrations and only when employing human RPA, as previously reported[15] (Pang Yen Wang and Zhaowei Liu, personal communication). Further work will be required to assess whether one can identify conditions that allow for the observation of long *S. cerevisiae* RPA streaks while maintaining physiological salt concentrations.

An alternative RPA-independent way to ascertain whether CMG molecules move along ssDNA or dsDNA would be the addition of a nick on the translocation strand, or the use of a hybrid ssDNA/dsDNA subtrate[46], both of which could be achieved employing nicking Cas9[46] or other commercially available nicking enzymes. CMG complexes moving on ssDNA should dissociate from the DNA at the nicks or gaps[47], whereas CMG complexes moving on dsDNA should not, allowing us to differentiate between them. Using this approach, it would be interesting to investigate whether the two velocity populations of unidirectionally moving CMG that we observed in **Chapter 2** correspond to CMG molecules translocating on ssDNA and dsDNA. Furthermore, it has been previously suggested that pre-formed

CMG is able to transition between surrounding ssDNA and dsDNA though the action of Mcm10[18]. It will therefore be of interest to investigate whether any other elongation factors and/or DNA features[45] can affect these transitions.

## 7.5 Expanding our understanding of fully reconstituted CMG motion

The work presented in this thesis has focused on CMG motion in isolation. Nevertheless, the hybrid assay developed in **Chapter 2** and described in detail in **Chapter 7** could be easily modified by the addition of different auxiliary factors during the single-molecule imaging. These modifications are relatively small but have great potential to increase our understanding of how CMG motion is regulated. Such an understanding will in turn provide us with important insights into how the speed of the replisome is modulated, which is important for its function[48].

An important set of auxiliary factors that modulates the speed of CMG is the Mrc1–Tof1–Csm3 (MTC) complex, previously shown to allow the replisome to achieve rates comparable to those observed *in vivo*[49]. Further, ensemble biochemical experiments showed that the presence of Mrc1 directly increases the unwinding rate of CMG in experiments where Csm3/Tof1 was present[6]. Finally, in previous single-molecule work, the MTC complex was reported to interact dynamically with a pre-formed CMG-based replisome, causing temporary accelerations in its motion[50]. In these experiments, neither Mrc1 nor Csm3/Tof1 was directly imaged, so the separate binding dynamics of these protein factors could not be studied. It would therefore be of interest to use the CMG motion assay we developed in **Chapter 2** to image the binding dynamics of fluorescently labeled Mrc1 and Csm3/Tof1 (labeled in different colors). Such experiments would allow us to understand the individual exchange dynamics of these important auxiliary factors, and to correlate them with changes in CMG velocity. Of note, interactions between the Tof1/Csm3 complex and CMG were previously shown to be modulated by phosphorylation of CMG by DDK[38]. Therefore, if there are any differences of the phosphorylation state of fully reconstituted CMG and pre-formed CMG, the exchange dynamics of the MTC complex may differ between the two approaches to study CMG.

Another factor that has been shown to stimulate the unwinding rate of CMG is the firing factor Mcm10[5,41,50], which crosslinking mass spectrometry has shown to have a binding site on CMG close to that of Mrc1[27,51], suggesting a form of binding competition. Furthermore, it was also shown that the checkpoint kinase RAD53

regulates the unwinding rate of CMG[6,8] through the phosphorylation of Mrc1 and Mcm10[6]. Nevertheless, phosphorylated Mcm10 only slows down CMG unwinding in the absence of unphosphorylated Mrc1[6]. It would thus be of interest to study this competition with and without RAD53 through the simultaneous imaging of Mcm10 and Mrc1 fluorescently labeled in different colors. Such experiments would allow us to correlate the binding of either protein to differences in CMG velocity.

In **Chapter 2** the velocity distribution of unidirectionally moving CMG was analyzed with a change-point analysis similar to the one employed in previous studies[50,52]. Nevertheless, a more detailed dwell-time analysis of CMG motion, such as those employed in magnetic tweezers studies of pre-formed CMG[16] and other molecular motors[53–56], will provide the field with a deeper understanding of fully reconstituted CMG, potentially uncovering different translocation and pausing modes. This type of analysis will be particularly important when comparing CMG motion in the presence of different elongation factors, as some of these factors may affect CMG in subtle ways (for instance, by decreasing the probability of entering a pausing or backtracking state[16]) that may only be noticed with such analysis.

In **Chapter 3**, we developed a way of synthesizing fully custom-sequence DNA constructs, providing us with full control of the DNA sequence context in which we can study the motion of pre-formed CMG. In future work, if would be of interest to implement this approach to fully reconstituted CMG. Of note, previous magnetic tweezers work on pre-formed CMG showed that it often backtracks and enters different pause states[16]. Further analysis of this work could unveil DNA sequences in which CMG is more likely to enter these different states. In turn, the method to synthesize custom-sequence constructs for single-molecule experiments described in **Chapter 3** could then be used to generate DNA substrates that bias CMG motion towards these different states. These constructs could in turn be used to understand how such CMG motion states can be affected by auxiliary protein factors in a more controlled manner. In addition, we employed the method described in **Chapter 3** to introduce extrahelical/non-canonical structures into DNA constructs used in single-molecule studies. In future work, it would be interesting to employ this approach to introduce different obstacles in the course of CMG, complementing previous work done on replisome-obstacle encounters[57–59].

## 7.6 Beyond CMG motion: single-molecule motion studies of full replisomes

In **Chapter 5**, we have shown some promising yet in-progress work towards the high-throughput single-molecule imaging of CMG motion. One of the main limitations of high-throughput approach is our current lack of observation of CMG unidirectional motion. At the end of **Chapter 5**, we have identified and listed specific experimental changes that could be tested to observe CMG motion, as well as potential avenues to increase the overall yield.

Optimizing this high-throughput assay to image CMG motion should be a priority as we move approach our goal of imaging a fully reconstituted replisome in a naked[1,49] and chromatinized DNA context[2,60]. This is because every added step is not 100 % efficient and thus decreases the overall yield of the reaction, eventually yielding our low-throughput optical tweezers approaches[21,29,61] unfeasible. Further optimization of the high-throughput assay developed in **Chapter 5**, together with the ability to image CMG activation *in situ*, may allow for a smooth transition towards the full reconstitution of bidirectional DNA replication initiation and elongation, which one would hope would imply the simple addition of additional replication factors into the flow cell[1,12,49].

## 7.7 Applications of this work beyond DNA replication

DNA replication aside, the work presented in this thesis may me adapted to other fields of biochemistry and biophysics. First, the translation of the complex biochemistry required to assemble study the motion of fully reconstituted CMG from the ensemble biochemical level[3] to the single-molecule level represents a major advancement in the single-molecule field. We anticipate that the hybrid ensemble and single-molecule approaches developed in **Chapter 2** and **Chapter 5** will enable the interrogation at the single-molecule level of similarly complex nucleic-acid processing mechanisms, paving the way for a better integration of biochemistry and single-molecule biophysics in the future. In addition, the data analysis suite, and the user-friendly graphical user interface that we report in **Chapter 4**[62], will allow the single-molecule biophysics and biochemistry community alike to analyze the data of correlative dual-beam optical tweezers and confocal microscopy experiments in a more streamlined and rigorous manner, conducting all the necessary controls to draw more quantitatively sound conclusions from such experiments. Finally, the method to synthesize fully custom-sequence DNA constructs for single-molecule studies described in **Chapter 3**

provides the single-molecule biophysics community with unprecedented control of the sequence context in which to study other molecular motors, allowing us to expand the sequence space in their study.

## 7.8 References


1. Yeeles, J. T. P., Deegan, T. D., Janska, A., Early, A. & Diffley, J. F. X. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**, 431–435 (2015).

2. Kurat, C. F., Yeeles, J. T. P., Patel, H., Early, A. & Diffley, J. F. X. Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates. *Mol. Cell* **65**, 117–130 (2017).

3. Douglas, M. E., Ali, F. A., Costa, A. & Diffley, J. F. X. The mechanism of eukaryotic CMG helicase activation. *Nature* **555**, 265–268 (2018).

4. de Jesús-Kim, L., Friedman, L. J., Ramsoomair, C., Gelles, J. & Bell, S. P. DDK regulates replication initiation by controlling the multiplicity of Cdc45-GINS binding to Mcm2-7. *Elife* **10:e65471**, 1–30 (2021).

5. Lõoke, M., Maloney, M. F. & Bell, S. P. Mcm10 regulates DNA replication elongation by stimulating the CMG replicative helicase. *Genes Dev.* **31**, 291–305 (2017).

6. McClure, A. W. & Diffley, J. F. X. Rad53 checkpoint kinase regulation of dna replication fork rate via mrc1 phosphorylation. *Elife* **10**, 1–24 (2021).

7. Hill, J., Eickhoff, P., Drury, L. S., Costa, A. & Diffley, J. F. X. The eukaryotic replisome requires an additional helicase to disarm dormant replication origins. *bioRxiv* 2020.09.17.301366 (2020).

8. Devbhandari, S. & Remus, D. Rad53 limits CMG helicase uncoupling from DNA synthesis at replication forks. *Nat. Struct. Mol. Biol.* **27**, 461–471 (2020).

9. Devbhandari, S., Jiang, J., Kumar, C., Whitehouse, I. & Remus, D. Chromatin Constrains the Initiation and Elongation of DNA Replication. *Mol. Cell* **65**, 131–141 (2017).

10. Deegan, T. D., Baxter, J., Ortiz Bazán, M. Á., Yeeles, J. T. P. & Labib, K. P. M. Pif1-Family Helicases Support Fork Convergence during DNA Replication Termination in Eukaryotes. *Mol. Cell* **74**, 231-244.e9 (2019).

11. Deegan, T. D., Mukherjee, P. P., Fujisawa, R., Rivera, C. P. & Labib, K. CMG helicase disassembly is controlled by replication fork dna, replisome components and a ubiquitin threshold. *Elife* **9**, 1–33 (2020).

12. Aria, V. & Yeeles, J. T. P. Mechanism of Bidirectional Leading-Strand Synthesis Establishment at Eukaryotic DNA Replication Origins. *Mol. Cell* **73**, 199-211.e10 (2019).

13. Lewis, J. S. *et al.* 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**, 10630–10635 (2017).

14. Lewis, J. S. *et al.* Tunability of DNA polymerase stability during eukaryotic DNA replication. *Mol. Cell* **77**, 1–9 (2020).

15. Kose, H. B., Xie, S., Cameron, G., Strycharska, M. S. & Yardimci, H. Duplex DNA engagement and RPA oppositely regulate the DNA-unwinding rate of CMG helicase. *Nat. Commun.* **11**, 1–15 (2020).

16. Burnham, D. R., Kose, H. B., Hoyle, R. B. & Yardimci, H. The mechanism of DNA unwinding by the eukaryotic replicative helicase. *Nat. Commun.* **10**, 1–14 (2019).

17. Schauer, G. D. *et al.* Replisome bypass of a protein-based R-loop block by Pif1. *Proc. Natl. Acad. Sci.* (2020) doi:10.1073/PNAS.2020189117.

18. Wasserman, M. R., Schauer, G. D., O'Donnell, M. E. & Liu, S. Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase. *Cell* **178**, 600-611.e16 (2019).

19. Moyer, S. E., Lewis, P. W. & Botchan, M. R. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10236–10241 (2006).

20. Langston, L. D. *et al.* CMG helicase and DNA polymerase ε form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15390–15395 (2014).

21. Ramírez Montero, D. *et al.* Nucleotide binding halts diffusion of the eukaryotic replicative helicase during activation. *Nat. Commun.* **14**, (2023).

22. Yuan, Z. *et al.* Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. *Nat. Struct. Mol. Biol.* **23**, 217–224 (2016).

23. Georgescu, R. *et al.* 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**, E697–E706 (2017).

24. Eickhoff, P. *et al.* Molecular Basis for ATP-Hydrolysis-Driven DNA Translocation by the CMG Helicase of the Eukaryotic Replisome. *Cell Rep.* **28**, 2673-2688.e8 (2019).

25. Lewis, J. S. *et al.* Mechanism of replication origin melting nucleated by CMG helicase assembly. *Nature* (2022) doi:10.1038/s41586-022-04829-4.

26. Georgescu, R. E. *et al.* Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat. Struct. Mol. Biol.* **21**, 664–670 (2014).

27. Baretić, D. *et al.* Cryo-EM Structure of the Fork Protection Complex Bound to CMG at a Replication Fork. *Mol. Cell* **78**, 926-940.e13 (2020).

28. Gupta, S., Friedman, L. J., Gelles, J. & Bell, S. P. A helicase- - tethered ORC flip enables bidirectional helicase loading. *Elife* **10:e74282**, 1–31 (2021).

29. Sánchez, H. *et al.* DNA replication origins retain mobile licensing proteins. *Nat. Commun.* **12**, (2021).

30. Frigola, J., Remus, D., Mehanna, A. & Diffley, J. F. X. ATPase-dependent quality control of DNA replication origin licensing. *Nature* **495**, 339–343 (2013).

31. Remus, D. *et al.* Concerted Loading of Mcm2-7 Double Hexamers around DNA during DNA Replication Origin Licensing. *Cell* **139**, 719–730 (2009).

32. Coster, G., Frigola, J., Beuron, F., Morris, E. P. & Diffley, J. F. X. Origin Licensing Requires ATP Binding and Hydrolysis by the MCM Replicative Helicase. *Mol. Cell* **55**, 666–677 (2014).

33. Coster, G. & Diffley, J. F. X. Bidirectional eukaryotic DNA replication is established by quasi-symmetrical helicase loading. *Science (80-. ).* **357**, 314–318 (2017).

34. Ticau, S., Friedman, L. J., Ivica, N. A., Gelles, J. & Bell, S. P. Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell* **161**, 513–525 (2015).

35. Ticau, S. *et al.* Mechanism and timing of Mcm2-7 ring closure during DNA replication origin licensing. *Nat. Struct. Mol. Biol.* **24**, 309–315 (2017).

36. Greiwe, J. F. *et al.* Structural mechanism for the selective phosphorylation of DNA-loaded MCM double hexamers by the Dbf4-dependent kinase. *Nat. Struct. Mol. Biol.* **29**, (2021).

37. Miller, T. C. R., Locke, J., Greiwe, J. F., Diffley, J. F. X. & Costa, A. Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM. *Nature* **575**, 704–710 (2019).

38. Bastia, D. *et al.* Phosphorylation of CMG helicase and Tof1 is required for programmed fork arrest. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3639–E3648 (2016).

39. Kanke, M., Kodama, Y., Takahashi, T. S., Nakagawa, T. & Masukata, H. Mcm10 plays an essential role in origin DNA unwinding after loading of the CMG

components. *EMBO J.* **31**, 2182–2194 (2012).

40. Zegerman, P. & Diffley, J. F. X. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* **445**, 281–285 (2007).

41. Langston, L. D. *et al.* Mcm10 promotes rapid isomerization of CMG-DNA for replisome bypass of lagging strand DNA blocks. *Elife* **6**, 1–21 (2017).

42. Douglas, M. E. & Diffley, J. F. X. Recruitment of Mcm10 to Sites of Replication Initiation Requires Direct Binding to the Minichromosome Maintenance ( MCM ) Complex. *J. Biol. Chem.* **291**, 5879–5888 (2016).

43. Langston, L. & O'Donnell, M. Action of CMG with strand-specific DNA blocks supports an internal unwinding mode for the eukaryotic replicative helicase. *Elife* **6**, 1–23 (2017).

44. Jenkyn-Bedford, M. *et al.* A conserved mechanism for regulating replisome disassembly in eukaryotes. *Nature* **600**, 743–747 (2021).

45. Low, E., Chistol, G., Zaher, M. S., Kochenova, O. V. & Walter, J. C. The DNA replication fork suppresses CMG unloading from chromatin before termination. *Genes Dev.* **34**, 1534–1545 (2020).

46. Belan, O. *et al.* Generation of versatile ss-dsDNA hybrid substrates for single-molecule analysis. *STAR Protoc.* **2**, 100588 (2021).

47. Vrtis, K. B. *et al.* Single-strand DNA breaks cause replisome disassembly. *Mol. Cell* **81**, 1309-1318.e6 (2021).

48. Saxena, S. & Zou, L. Hallmarks of DNA replication stress. *Mol. Cell* **82**, 2298–2314 (2022).

49. Yeeles, J. T. P., Janska, A., Early, A. & Diffley, J. F. X. How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication. *Mol. Cell* **65**, 105–116 (2017).

50. Lewis, J. S. *et al.* 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**, 10630–10635 (2017).

51. Mayle, R. *et al.* Mcm10 has potent strand-annealing activity and limits translocase-mediated fork regression. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 798–803 (2019).

52. Duderstadt, K. E. *et al.* Simultaneous Real-Time Imaging of Leading and Lagging Strand Synthesis Reveals the Coordination Dynamics of Single Replisomes. *Mol. Cell* **64**, 1035–1047 (2016).

53. Dulin, D., Berghuis, B. A., Depken, M. & Dekker, N. H. Untangling reaction

pathways through modern approaches to high-throughput single-molecule force-spectroscopy experiments. *Curr. Opin. Struct. Biol.* **34**, 116–122 (2015).

54. Dulin, D. *et al.* Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase. *Cell Rep.* **10**, (2015).

55. Dulin, D. *et al.* Signatures of Nucleotide Analog Incorporation by an RNA-Dependent RNA Polymerase Revealed Using High-Throughput Magnetic Tweezers. *Cell Rep.* **21**, 1063–1076 (2017).

56. Janissen, R. *et al.* Global DNA Compaction in Stationary-Phase Bacteria Does Not Affect Transcription. *Cell* **174**, 1188-1199.e14 (2018).

57. Sparks, J. L. *et al.* The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell* **176**, 167-181.e21 (2019).

58. Schauer, G. D. *et al.* Replisome bypass of a protein-based R-loop block by Pif1. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 30354–30361 (2020).

59. Sato, K., Martin-Pintado, N., Post, H., Altelaar, M. & Knipscheer, P. Multistep mechanism of G-quadruplex resolution during DNA replication. *Sci. Adv.* **7**, (2021).

60. Chacin, E. *et al.* Establishment and function of chromatin organization at replication origins. *Nature* (2023) doi:10.1038/s41586-023-05926-8.

61. Sánchez, H. *et al.* A chromatinized origin reduces the mobility of ORC and MCM through interactions and spatial constraint. *Nat. Commun.* **14**, (2023).

62. Liu, Z. *et al.* A biophysics toolbox for reliable data acquisition and processing in integrated force-confocal fluorescence microscopy. *ACS Photonics* (2024).

# Summary

Prior to cell division, a cell must generate an exact copy of its entire genome to ensure that each of the daughter cells obtains a full copy. This process, known as DNA replication, is an essential process of life and is vital for the health and survival of all cellular organisms.

In eukaryotes, DNA replication is catalyzed by a megadalton-sized dynamic protein complex known as the replisome, which generally carries out its important function in a remarkably efficient and accurate manner. This Herculean task becomes even more impressive when one considers the thousands of roadblocks along the way around which the replisome must navigate, such as tightly bound DNA-binding proteins, covalent DNA-protein crosslinks, highly stable DNA secondary structures, and several forms of DNA damage. In addition to replicating DNA, the replisome is also responsible for a critical process in epigenetic inheritance: the disassembly of nucleosomes on the parental DNA ahead of the replication fork, and their reassembly onto the newly synthesized DNA behind it. Finally, adding to the list of complex tasks it must achieve, the replisome is also able to sense some forms of DNA damage and facilitate their repair.

The core replisome consists of the Cdc45/Mcm2-7/GINS (CMG) replicative helicase, which separates the two strands of the parental DNA, and three polymerases that use the unwound parental strands as templates for DNA synthesis. This core organizes the exquisitely coordinated recruitment and release of accessory proteins that are required at different stages of DNA replication. These specialized accessory factors speed up the replisome and allow it to conduct its additional functions, such as facilitating the repair of DNA lesions, disassembling nucleosomes ahead of the replication fork, and reassembling parental and new histones into nucleosomes behind the replication fork. The dynamic nature of the replisome provides it with a high level of plasticity and is key to the successful achievement of its many functions.

A key component of the core replisome is the replicative helicase CMG, which is the molecular motor driving the replisome and acts as an organizing center to

recruit and release accessory proteins. Obtaining a deep quantitative understanding of the dynamics of CMG is therefore crucial for understanding how cells efficiently and faithfully copy their entire genome once per cell cycle. An excellent avenue to obtain such an understanding of the dynamics of CMG are single-molecule techniques, which can study molecular motors with unparalleled temporal and spatial resolution. Furthermore, single-molecule techniques are uniquely suited to observe low-probability events and access kinetic constants that would be otherwise hidden in the asynchronous average output of ensemble biochemical techniques. This thesis focuses on studying the motion dynamics of CMG at the single-molecule level, and it specifically focuses on the CMG ortholog from the budding yeast *Saccharomyces cerevisiae*.

In a cell, DNA replication starts by the cell-cycle regulated piecewise assembly and activation of CMG, a process that has been fully reconstituted *in vitro* from a minimal set of 36 *S. cerevisiae* polypeptides in bulk biochemical studies. This fully reconstituted approach successfully recapitulates the *in vivo* cell cycle regulation of CMG assembly and allows for the study of the intermediate steps in this process. Conversely, single-molecule studies of CMG motion have thus far employed pre-activated CMG purified as a complex from cells. Previous single-molecule studies have therefore missed all the steps in the assembly and activation of this important helicase. In this thesis, we report the first single-molecule studies of the motion of CMG after fully reconstituting its assembly and activation from 36 purified *S. cerevisiae* proteins.

In **Chapter 1**, we provide an introduction to eukaryotic DNA replication, CMG assembly and activation, as well as to the single-molecule techniques employed in this study. We also provide a comprehensive survey of previous *in vitro* reconstituted single-molecule studies of CMG to contextualize the pioneering single-molecule work developed in this work to study the motion of fully reconstituted CMG. We also highlight the challenges involved in translating the full reconstitution of CMG motion to the single-molecule level.

In **Chapter 2**, we report the first single-molecule imaging and quantification of the motion of fully reconstituted CMG. For this, we developed a hybrid ensemble and single-molecule assay relying on the functionalization of both ends of a linear DNA substrate with two orthogonal surface attachment moieties. This hybrid assay allowed us to: 1) fully reconstitute the origin-based assembly and activation of fluorescently labeled CMG from 36 purified *S. cerevisiae* polypeptides; and 2) image and quantify the motion of fluorescently labeled CMG along individual DNA molecules held in place in an optical trap. In these studies, we observed that CMG

exhibits two different motion types according to the nucleotide present in solution. We observe that CMG exhibits the expected unidirectional motion when ATP is present in solution, but that it exhibits diffusive motion in the absence of nucleotide. Furthermore, we show that ATP binding, and not hydrolysis, halts this diffusive motion. Taken together, these observations led us to propose a model whereby ATP keeps newly assembled CMG near origins of replication and facilitates the initial DNA melting required to initiate DNA replication.

Single-molecule studies of CMG as well as other molecular motors as they move on DNA often require DNA constructs with specific sequences relevant to the specific biological process being studied. Furthermore, these substrates require extra-helical/non-canonical structural features to interrogate different DNA transactions, which often require the presence of specific restriction enzyme sites at specific locations within the initial DNA substrate. These types of DNA constructs are commonly synthesized by modifying pre-existing plasmids, or by amplifying pre-existing regions within a plasmid. These synthesis methods rely on time-consuming molecular cloning and limit the sequence space that can be studied. In **Chapter 3**, we report a novel method to assemble fully custom-sequence long (> 10 kb) plasmids *de novo* from synthetic fragments. Using these plasmids, we developed a fast and efficient method to assemble > 10 kb linear DNA constructs with modified ends for surface attachment, as well as internal extrahelical structures. As a proof-of-principle, we synthesized two different DNA constructs with biotinylated ends and internal 3' ssDNA flaps, which we validate using a combination of ensemble molecular biology and single-molecule correlative dual-beam optical tweezers and confocal microscopy. Finally, we validate our ssDNA flap incorporation method by showing that pre-formed fluorescently labeled CMG can bind to the flap and then translocate in the expected direction. This new synthesis method provides us and others with the ability to fully control the DNA sequence context in which CMG and other molecular motors can be studied.

During our correlative dual-beam optical tweezers and confocal microscopy studies of CMG motion, we have developed new ways of more efficiently acquiring data, as well as developed different calibration and data analysis routines to draw sound quantitative conclusions from such data. In **Chapter 4**, we provide a comprehensive survey of all the sample preparation, data acquisition, experimental automation, and data analysis routines that we have developed to study molecular motors employing correlative dual-beam optical tweezers and confocal microscopy. We also report the development of a publicly available user-friendly graphical user interface to allow the community to implement our data analysis pipeline to their own data.

Correlative dual-beam optical tweezers and confocal microscopy are very powerful technique as they allow for the direct observation of fluorescently labeled molecules along a DNA molecule, while providing exquisite control of the force applied to the DNA substrate. Nonetheless, this technique is limited by its inherently low throughput, a particularly important consideration in our fully reconstituted studies of CMG, and in our future goal of extending these studies to full replication. This is because these processes consist of a series of biochemical steps that are not 100% efficient, reducing the overall yield of the reaction with every added step. In **Chapter 5**, we set out to overcome this limitation by adapting the hybrid ensemble and single-molecule assay developed in Chapter 2 to make it compatible with total internal reflection fluorescence (TIRF) microscopy. This modified assay allowed us to image the motion of fully reconstituted CMG along surface-tethered flow-stretched DNA molecules in high-throughput. We report several developments required for the development of this assay, including the design and development of a tailor-made microfluidic flow-cell, as well as the development of a novel surface passivation and functionalization strategy. Finally, we report the preliminary analysis of fully reconstituted CMG motion along flow-stretched surface-tethered DNA molecules, attesting to the potential of our new assay.

In **Chapter 6**, we report an in-depth step-by-step description of the hybrid ensemble and single-molecule assay developed in Chapter 2 that allowed us to image the motion of fully reconstituted CMG at the single-molecule level. We believe that modification of this powerful assay will allow the single-molecule biophysics community to interrogate at the single-molecule level similarly complex DNA-processing reactions.

Finally, in **Chapter 7**, we summarize the main conclusions of these chapters and provide potential directions in which the research that started with this thesis can continue. In this final chapter, we contextualize the pioneering work developed in previous chapters, and discuss the implications that such work could have for the fields of DNA replication and single-molecule biophysical studies of other DNA-protein interactions.

# Samenvatting

Voorafgaand aan de celdeling kan plaatsvinden, dupliceert een cel het gehele genoom zodat dat elke dochtercel een complete kopie bevat. Dit kopieerproces, bekend als DNA-replicatie, is een essentieel proces in het leven zoals wij het kennen en voor de gezondheid en overleving van alle cellulaire organismen.

In eukaryotische cellen wordt DNA replicatie gekatalyseerd door een MegaDalton-groot dynamisch eiwitcomplex, genaamd het replisoom. Dit replisoom voert deze belangrijke functie uit op een onwaarschijnlijk efficiënte en nauwkeurige manier. Deze complexe taak wordt nog indrukwekkender wanneer men in acht neemt dat er duizenden obstakels op de DNA aanwezig zijn waar het replisoom zich omheen moet navigeren, zoals bijvoorbeeld sterk gebonden eiwitten, covalente DNA-eiwit bindingen, stabiele secundaire DNA structuren en verschillende vormen van DNA schade. Naast DNA replicatie is het replisoom ook verantwoordelijk voor een kritisch epigenetisch proces, namelijk het ontkoppelen van nucleosomen op het oorspronkelijke DNA zodat deze overgezet kunnen worden op het nieuw gekopieerde DNA. Naast deze complexe taken heeft het replisoom ook het vermogen om vormen van DNA schade te detecteren en reparatie hiervan in gang te zetten.

De basis van het replisoom bestaat uit de Cdc45/Mcm2-7/GINS (CMG) replicatieve helicase, die de twee strengen van het oorsponkelijke DNA van elkaar scheidt, alsmede drie DNA polymerases die deze DNA strengen benutten als mal voor het kopieer proces. Deze kerneiwitten organiseren zo optimaal mogelijk de gecoördineerde associatie en dissociatie van andere eiwitten die vereist zijn tijdens de verschillende fases van DNA replicatie. Deze gespecialiseerde eiwitten versnellen het replisoom en laten additionele functies toe, waaronder het herstel van DNA schade, decondensatie van nucleosomen op het DNA in het pad van de replicatievork en hun condensatie op het DNA achter de replicatievork met behulp van zowel gerecycelde als nieuw gesynthetiseerde histonen. Het dynamische karakter van het replisoom voorziet in een hoge mate van aanpassingsvermogen en is de sleutel tot het succesvol vervullen van deze vele functies.

De replicatieve helicase CMG is de moleculaire motor die het replisoom aandrijft en acteert als een organisatiecentrum voor de associatie en dissociatie van andere benodigde eiwitten aan en van het replisoom. Een uitstekende methodiek om inzicht in de dynamiek van CMG te verkrijgen zijn de zogeheten enkel-molecuul technieken, die het mogelijk maken om de werking van moleculaire motoren met hoge temporele en spatiale resolutie te bestuderen. Deze technieken zijn ook bij uitstek geschikt om gebeurtenissen met een lage waarschijnlijkheid waar te nemen en daarmee toegang te krijgen tot kinetisch constanten die anders verborgen zouden blijven in de gemiddelde output van asynchrone biochemische technieken. Dit proefschrift richt zich op het bestuderen van de bewegingsdynamiek van CMG op het niveau van een enkel molecuul, en richt zich specifiek op de CMG-ortholoog van *Saccharomyces cerevisiae*.

In de cel begint de DNA replicatie met een gereguleerde stapsgewijze assemblage en activering van CMG, een proces dat *in vitro* biochemische onderzoek nagebootst kan worden door gebruik van een minimale set van 36 *S. cerevisiae*-polypeptiden. Omdat deze op volledige reconstitutie gebaseerde aanpak een vergelijkbare CMG assemblage teweegbrengt als in de *in vivo* celcyclus, maakt de reconstitutie de studie van de tussenstappen in dit proces mogelijk. Tot op heden is bij onderzoeken naar CMG beweging met enkel-molecuul technieken vooral gebruik gemaakt van een vooraf geactiveerd CMG complex dat direct in zijn geheel gezuiverd wordt uit cellen waarin de verschillende subunits tot overexpressie gebracht zijn. Deze enkel-molecuul studies hebben daarom niet alle stappen in de assemblage en activering van deze belangrijke helicase kunnen beschrijven. In dit proefschrift rapporteren wij het eerste enkel-molecuul onderzoek naar de beweging van CMG nadat wij de assemblage en activering ervan volledig hebben gereconstitueerd uit 36 gezuiverde *S. cerevisiae* eiwitten.

**Hoofstuk 1** omvat een inleiding tot DNA replicatie in eukaryoten, CMG-assemblage en -activering, evenals enkel-molecuul technieken die in deze studie worden gebruikt. We bieden ook een uitgebreid overzicht van eerdere *in vitro* onderzoeken naar CMG met enkel-molecuul technieken zodat het onderzoek in dit proefschrift naar de beweging van CMG in context geplaatst kan worden. We benadrukken ook de uitdagingen die gepaard zijn gegaan met het vertalen van de volledige reconstructie van CMG beweging naar het enkel-molecuul niveau.

In **Hoofdstuk 2** rapporteren we de eerste beeldopnames en kwantificering van de beweging van volledig gereconstitueerd CMG op enkel-molecuul niveau. Hiervoor hebben we een hybride bulk en enkel-molecuul methodiek ontwikkeld, gebaseerd op de functionaliteit van beide uiteinden van een lineair DNA substraat

met twee orthogonale oppervlakte-bindende bestandsdelen. Deze hybride methodiek stelde ons in staat om: 1) de op oorsprong-gebaseerde assemblage en activering van fluorescent gelabeld CMG uit 36 gezuiverde *S. cerevisiae*-polypeptiden volledig te reconstrueren; en 2) de beweging van fluorescent gelabeld CMG langs individuele DNA moleculen (die op hun plaats worden gehouden in een optische pincet) zichtbaar te maken en te kwantificeren. In dit onderzoek hebben we gezien dat CMG twee verschillende bewegingstypen vertoont, afhankelijk van het soort nucleotide dat in oplossing aanwezig is. Zo zien we dat CMG een verwachte voorwaartse beweging (translocatie) vertoont wanneer ATP in oplossing aanwezig is, maar dat het diffuse beweging vertoont in afwezigheid hiervan. Bovendien laten we zien dat louter de binding van ATP, en niet de hydrolyse ervan, in staat is deze diffusiebeweging te stoppen. Bij elkaar genomen leiden deze observaties tot een model waarin we voorstellen dat de aanwezigheid van ATP nieuw geassembleerd CMG dicht bij de replicatie-oorsprong houdt om vervolgens het ontwinden van DNA te faciliteren dat nodig is voor de initiatie van DNA replicatie.

Voor enkel-molecuul onderzoeken naar CMG en zo ook naar andere moleculaire motoren die zich op DNA voortbewegen, zijn vaak DNA-constructen gewenst die bepaalde sequenties omvatten die relevant zijn voor de bestudering van specifieke biologische processen. Bovendien vereisen deze substraten extra-helix-achtige/niet-kanonieke structurele kenmerken om onderzoek aan specifieke wisselwerkingen op het DNA mogelijk te maken, waardoor vaak de aanwezigheid van specifieke restrictie-enzym herkennings-sequenties op specifieke locaties op het initiële DNA-substraat gewenst is. Dergelijke DNA constructen worden doorgaans gesynthetiseerd door reeds bestaande plasmiden te modificeren, of door reeds bestaande gebieden binnen een plasmide te amplificeren. Deze synthesemethoden zijn afhankelijk van tijdrovend moleculair klonen en beperken de sequentieruimte die kan worden bestudeerd. In **Hoofdstuk 3** rapporteren we daarom een nieuwe methode om volledig op maat gemaakte lange (>10 kb) plasmiden *de novo* samen te stellen uit synthetische fragment. Met behulp van deze plasmiden hebben we een snelle en efficiënte methode ontwikkeld om lineaire DNA-constructen van > 10 kb samen te stellen die zowel gemodificeerde uiteinden voor oppervlaktebinding omvatten alsmede interne extra-helix structuren. Als toepassing hiervan hebben we twee verschillende DNA constructen gesynthetiseerd met gebiotinyleerde uiteinden en interne 3'-ssDNA-flappen, die we valideren met behulp van een combinatie van moleculaire biologie en optische pincetten uitgerust met confocale microscopie. Ten slotte valideren we onze methode voor het inbouwen van de ssDNA-flap door aan te tonen dat

voorgevormde fluorescent gelabelde CMG aan de interne flap kan binden en vervolgens in de verwachte richting kan verplaatsen. Deze nieuwe synthese methode biedt ons en anderen de mogelijkheid om volledige controle uit te oefenen op de DNA-sequentie context voor CMG en andere moleculaire motoren.

Tijdens onze experimenten met optische pincetten uitgerust met confocale microscopie naar CMG-beweging hebben we ook nieuwe manieren ontwikkeld om efficiënter data te verzamelen, en hebben we verschillende kalibratie- en gegevensanalyse routines ontwikkeld om goede kwantitatieve conclusies uit dergelijke gegevens te trekken. In **Hoofdstuk 4** geven we een uitgebreid overzicht van de monstervoorbereiding, data-acquisitie, experimentele automatisering en data-analyseroutines die we hebben ontwikkeld om moleculaire motoren te bestuderen met behulp van bovengenoemde optische pincetten uitgerust met confocale microscopie. We beschrijven ook de ontwikkeling van een publiekelijk beschikbare, gebruiksvriendelijke grafische gebruikersinterface waarmee de overige wetenschappers onze data-analysepijplijn op hun eigen gegevens kunnen toepassen.

Correlatieve krachts- en confocale microscopie op basis van optische pincetten zijn zeer krachtige technieken omdat ze de directe observatie van gelabelde eiwitten op een DNA molecuul via fluorescentie microcopie mogelijk maken, terwijl ze tegelijkertijd controle bieden over de kracht die op het DNA substraat wordt uitgeoefend. Desalniettemin wordt deze techniek beperkt door het inherent lage aantal moleculen dat per dag bestudeerd kan worden, wat impact heeft op ons onderzoek naar gereconstitueerd CMG, maar ook op ons toekomstige doel om dit onderzoek uit te breiden tot volledige DNA replicatie. De opbouw van deze processen bestaat uit een reeks biochemische stappen waarvan er geen enkel 100% efficiënt is, met als gevolg dat de totale opbrengst van de reactie met elke toegevoegde stap afneemt. In **Hoofdstuk 5** proberen we deze beperking te omzeilen door de hybride bulk en enkel-molecuul methodiek, ontwikkeld in Hoofdstuk 2, aan te passen om deze compatibel te maken met totale interne reflectie fluorescentie (TIRF) microscopie. Met dit aangepaste protocol konden we de beweging van volledig gereconstitueerd CMG moleculen op een veeltal van DNA moleculen (aan weerszijden gebonden aan het oppervlak te zijn uitgerekt door middel van stroming) tegelijkertijd in beeld brengen. We beschrijven de verschillende technische ontwikkelingen die nodig zijn voor de realisatie van deze nieuwe meetmethodiek, waaronder het ontwerp en de ontwikkeling van een op maat gemaakte microfluïdische “vloeistofcel”, evenals de ontwikkeling van een nieuwe strategie voor oppervlaktepassivering en functionaliteit. Ten slotte rapporteren we onze eerste analyse van de beweging van volledig gereconstitueerd

CMG in deze nieuwe context, wat het potentieel van deze nieuwe meetmethodiek aantoont.

In **Hoofdstuk 6** rapporteren we een diepgaande stapsgewijze beschrijving van de hybride bulk en enkel-molecuul methodiek ontwikkeld in Hoofdstuk 2, waarmee we de beweging van volledig gereconstitueerd CMG in beeld konden brengen op het enkel-molecuul niveau. Wij geloven dat modificatie van deze krachtige methodiek de enkel-molecuul biofysica gemeenschap in staat zal stellen om op het niveau van één molecuul vergelijkbare complexe DNA verwerkingsreacties te onderzoeken.

Ten slotte vatten we in **Hoofdstuk 7** de belangrijkste conclusies van de andere hoofdstukken samen en geven we mogelijke richtingen aan waarin het onderzoek dat met dit proefschrift is gestart, kan worden voortgezet. In dit laatste hoofdstuk zetten we de implicaties van het in eerdere hoofdstukken gepresenteerde werk uiteen en plaatsen we dit in het kader van DNA replicatie en enkel-molecuul onderzoek naar andere DNA-eiwit interacties.

# Curriculum vitae

## Daniel Ramírez Montero

12/01/1993 Born in Mexico City, Mexico

### Education

2010-2012 **International Baccalaureate Bilingual Diploma**
United World College Atlantic College, Wales, U.K.

2012-2016 **B.Sc. Biochemistry and Molecular Biology, *magna cum laude* with honors**
Brown University, Providence, U.S.A.
Bachelor's thesis supervisor: Dr. Nicolas L. Fawzi

2018-2019 **M.Sc. Biology**
Massachusetts Institute of Technology, Cambridge, U.S.A.
Master's thesis supervisor: Dr. Joseph H. Davis

2019-2024 **Ph.D. Biophysics**
Delft University of Technology, Delft, The Netherlands
Ph.D. supervisor: Dr. Nynke H. Dekker

### Additional research experience

2014-2016 **Undergraduate Researcher**
Fawzi Laboratory
Brown University, Providence, U.S.A.

2016-2018 **Graduate Scientist**
AstraZeneca Innovative Medicine and Early Development (IMED) Graduate Programme
AstraZeneca Pharmaceuticals, Waltham, U.S.A.

2024 **Research stay**
University of Oxford, Oxford, United Kingdom

## Fellowships, scholarships, and awards

| | |
|---|---|
| 2023 | **Block Scholarship** (award for a promising young physicist at the Aspen Winter Conferences)<br>Single Molecule Biophysics Aspen Winter Conference |
| 2022 | **Poster Prize**<br>Single Molecule Approaches to Biology Gordon Research Conference |
| 2020-2023 | **Boehringer Ingelheim Fonds PhD Fellowship** (personal fellowship to conduct doctoral research)<br>Boehringer Ingelheim Fonds |
| 2018-2019 | **Presidential Graduate Fellowship** (personal fellowship to conduct graduate training)<br>Massachusetts Institute of Technology |
| 2016 | **Leallyn Clapp Prize for an Outstanding Undergraduate Thesis in Biochemistry**<br>Brown University Chemistry Department |
| 2015 | **Junior Prize in Biochemistry**<br>Brown University Chemistry Department |
| 2014 | **Karen T. Romer Undergraduate Teaching and Research Award for Summer Research**<br>Brown University Dean of the College |
| 2013 | **Foster Freshman Prize in French**<br>Brown University French Studies Department |
| 2012-2016 | **Brown University Scholarship** (awarded to complete undergraduate studies)<br>Brown University |
| 2012-2016 | **Davis United World College Scholarship** (awarded to study at Brown University)<br>Davis United World College Scholars Program |
| 2010-2012 | **United World Colleges Mexico Scholarship** (full scholarship to represent Mexico at UWC Atlantic College)<br>UWC Mexico |

# List of publications

**Ramírez Montero D.**, Palmero Moya F., Talele S., Liu Z., Sánchez H., van Laar T., and Dekker N.H. (**2024**). A high-throughput single-molecule assay to image the motion of fully reconstituted CMG. (*Manuscript in preparation)*

**Ramírez Montero D.**, Sánchez H., van Veen E., van Laar T., and Dekker, N.H. (**2024**). A hybrid ensemble and single-molecule assay to image the motion of fully reconstituted CMG. (*Manuscript under review*)

Liu Z.*, van Veen E.*, Palmero Moya F., Sánchez H., Solano B., McCluskey K.A., **Ramírez Montero D.**, van Laar T., and Dekker N.H. (**2024**). A biophysics toolbox for reliable data acquisition and processing in integrated force-confocal fluorescence microscopy. *ACS Photonics*. DOI: 10.1021/acsphotonics.3c01739. *Denotes equal contribution.

**Ramírez Montero D.**, Liu Z., and Dekker, N.H. (**2024**). De novo fabrication of custom-sequence plasmids for the synthesis of long DNA constructs with extrahelical features. *Biophysical Journal* 123(1):31-41.

**Ramírez Montero D.**, Sánchez H., van Veen E., van Laar T., Solano B., Diffley J.F.X., and Dekker, N.H. (**2023**). Nucleotide binding halts diffusion of the eukaryotic replicative helicase during activation. *Nature Communications* 14:2082.

Wang A.*, Conicella A.*, Martin E., Broder Schmidt H., Reeb A, Nourse A., **Ramirez Montero D.**, Ryan V., Rohatgi R., Naik M., Ayala Y., Mittag T., and Fawzi N. (**2018**). A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation and RNA splicing. *EMBO Journal* 37: e97452. *Denotes equal contribution.

# Acknowledgements/Agradecimientos

Many people agree that doing a PhD can be hard. Indeed, during my PhD, there were some challenging moments when experiments did not work and when it was hard to see the light at the end of the tunnel. What is more, the COVID-19 pandemic hit a few months after I started my PhD, which further complicated things. Luckily, I had the fortune of having many people by my side who helped and encouraged me throughout this journey. I would like to dedicate this section to the many, many people without the help of whom this PhD would not have been possible.

First, I would like to thank **Nynke** for the opportunity to work in her lab and on such a fascinating topic. Thank you, Nynke, for all your kindness and mentoring during the past four and a half years, for all the stimulating scientific discussions, for your constant encouragement even when experiments were not going as expected, and for always motivating me to pursue excellence in science. Although there were some stressful periods during my PhD, meeting with you always calmed me down motivated me. In particular, I want to thank you for having dared to take on a biochemist transitioning to the field of single-molecule biophysics, like me, and for investing in my education as a quantitative biologist and biophysicist. I can confidently say that I have made the field of single-molecule biophysics my own and for this I will be forever grateful to you.

I would also like to thank **Christophe** for his mentoring during the past four and a half years. Thank you, Christophe, for your valuable feedback during our progress meetings, and for continuing to mentor and support me all the way from Toulouse. Above all, thank you for your kindness and for your words of encouragement. Merci pour tout, Christophe. J'espère qu'on se reverra bientôt !

I would also like to thank **Dr. John Diffley** not only for taking time to read my thesis and for agreeing to be part of my defense committee, but for all the invaluable biological and technical input in several of the chapters presented in this thesis. Thank you, John, for all those stimulating discussions and for never hesitating to help me and the team in any way you could.

I would also like to extend my gratitude to all the members of my committee, **Prof. dr. Chirlmin Joo**, **Prof. dr. Christian Kaiser**, **Prof. dr. Karl Duderstadt**, **Dr. Meindert Lamers**, and **Prof. dr. Bernd Rieger** for taking their time to read my thesis and for challenging its content during my defense.

Next, I would like to thank all the people from the laboratory who trained me and helped me on countless occasions during my PhD. I would like to start this section by thanking the person who spent the most time training me in the lab:

**Humberto**. Thank you so much, Humberto, for sharing so much of your knowledge with me, and for all your patience and encouragement during my PhD. Thank you for always pushing me to do the best science I can do. You and I spent a lot of time working together during the COVID-19 pandemic. In fact, you were probably the person with whom I interacted the most during those months (as my accent in Spanish can attest, *tío*!). We all know that life was hard and chaotic during all those months of confinement, but seeing you every day was one of the things that kept me going and helped me get through it all. Thank you for all the chats and laughs we had in the office and outside the office. I will always be thankful to you for all your mentoring and training, but more than that, for constantly taking care of me and for your friendship. ¡Hasta pronto, amigo!

Next, I would like to thank **Belén**, whose friendly presence in the lab always brought a smile to my face and calmed me down. Thank you very much, Belén, for all the nice conversations that we had over the past four and a half years. Running into you in the office and having a short (or sometimes not so short) chat with you always left me feeling relaxed and at peace, even during the most stressful parts of my PhD. Thank you for always making me feel understood, at home and welcomed; your friendship means a lot to me and I will always cherish it. Thank you as well, for introducing me to your lovely family and for all the music sessions together. ¡Nos vemos en México!

I also want to thank **Theo**, not only for spending hours and hours generating many of the biological reagents employed in this work, but also for teaching me a lot of tricks and techniques in the lab. Thank you for always dropping whatever you were doing to help me whenever I needed help. Thank you as well for your sense of humor, which always provided comic relief to stressful moments. Ook bedankt dat je mij hebt geholpen met het verbeteren van mijn Nederlands!

I would also like to thank my colleague and friend, **Zhaowei**, for his help and friendship. Thank you, Zhaowei, for all that you taught me about force spectroscopy and for helping me with the automation of the C-Trap, and for all the stimulating discussions about CMG and about DNA replication that we had over the years. I always loved how in-depth our scientific conversations were and I sincerely hope to always have colleagues with whom I can reach such depth. I cannot wait to see the great science you will do in the future!

To the *Master of the CoSMoS*, **Saurabh**, I would like to thank for all his help with the TIRF/CoSMoS microscope. I always admired your ability to understand and tweak microscopes, but above all, your kindness and patience explaining things to others. Although we only overlapped in the lab for a year and a half, your contribution to the TIRF work in this thesis has been extremely important and I have learned a lot from you. I truly enjoyed working closely with you on those long and complicated TIRF experiments; most of all, though, I have enjoyed becoming your friend. Thanks for all the laughs and the nice conversations. I very much look forward to seeing the cool science that you will do in the future!

To my friend and colleague, **Francisco**, I would like to thank for all his help with programming, especially with the analysis of the TIRF data. I particularly enjoyed a lot all the conversations about math and biology that we had. Though we were only together in the lab for a few months, it has been great getting to know you and working with you.

To my fellow PhD students/soon to be students, **Pang Yen**, and **Jackson**, thank you very much for the nice working environment and for all the energy that you brought to the lab. To **Vincent**, my once bachelor student and now colleague, I thank for his diligence and engagement during his initial training. It has been a pleasure to work with you during the past few months, and to witness your scientific growth from bachelor student to PhD student. I look forward to the work that you will do during your PhD.

I would also like to thank former members of the lab who are now scattered around the world, but who helped and supported me during my PhD: **Kaley**, **Edo**, **Julien**, **Dorian** and **Louis**.

Thank you, **Kaley**, for teaching me an indescribable amount of physics and biophysics. You are a great teacher and I will always be thankful for all you taught me. Thanks a lot for all the nice conversations we had in the office, ranging from fractals to the pets of Aquaman. Above all, thank you for always encouraging me with your kind words.

Thank you, **Edo** (aka *the Python Master*), for all you taught me about programming and for your invaluable contribution to all the work presented in this thesis. You are hands-down the best programmer that I have ever met and I am a much better programmer because of you. More than anything, thank you for being a very good friend. I will always cherish your friendship and getting to know your family; knowing that you guys were nearby always made me feel more at home in The Netherlands.

Thank you, **Julien**, for your friendship and your help in the lab. It was a pleasure getting to know you. I will never forget the train ride that we accidentally shared to London and how our conversation was so engaging that three and a half hours flew by. À bientôt !

Thank you, **Dorian**, for your friendship and support and for always bringing in a fresh chemist's perspective to the experimental problems that I had. Although I only started to work with surface-based techniques after you left the lab, your valuable contribution to surface passivation was very important in the later stages of my PhD.

Finally, I would like to thank my friend and former colleague, **Louis**, for many, so very many, different things. I do not even know where to start, you were there on the first day of my PhD, starting your official contract as a PhD student on the very same day that I started. Although we did not see each other much during

the first two years of my PhD, as you were often in a different lab, over time you became one of my closest friends in The Netherlands. Thank you for all the post-lunch coffees we had by the bicycle rack, for the two beers we had together, for all the dancing, and the Dutch karaoke singing. Above all, thank you for always being there and for making me feel like part of your family. I'm sure that this friendship will last for many years.

To be completely honest, the beginning of my stay in The Netherlands was not easy. Shortly after I moved, right when I was beginning to meet new people outside my lab, a global pandemic hit the world and forced us to isolate ourselves from others. The isolation prevented me from meeting a lot of new people outside my lab during the first years of my PhD. Nevertheless, after the restrictions were lifted, I had the honor of meeting a lot of kind and caring people in the Department of Bionanoscience who helped and supported me throughout this journey: **Ana**, **Fede**, **Christos**, **Ilya**, **Esengul, Brian**, **Theo, Ashmiani**, **Alejandro**, **James**, **Luci**, **SaFyre**, **Cecilia**, **Clémence**, **Natasha**, **Mike**, **Carlos**, and **Koushik**. You guys have made my experience at the TU Delft very memorable and I will miss you all dearly. I will always cherish your friendship and having spent these important years of growth and learning with you all.

I would like to specially thank **Joyce** for all her care and support during the past four years. Thank you, Joyce, for always looking after me and for always bringing a smile to my face. Thank you for all the home-cooked meals, for always ensuring I felt at home in the department and for always helping me take care of my plants. Having you in the department helped me feel more at home and I cannot thank you enough for all your support.

Finally, I would like to thank **Beatriz**. Although we only overlapped in the department towards the end of my PhD, you have become a very good friend and I will miss you a lot. Thank you for all the *cafecitos* and nice chats we had and for the nice conversations about life that we had all the time. Thank you very much as well for being my paranymph and for all your help during my defense. ¡Muchísimas gracias!

Quisiera también agradecer a mi familia mexicana en Delft, **David** y **Cass**. No puedo agradecerles con palabras por todo lo que han hecho por mí en todos estos años, por su apoyo incondicional, por todos los paseos, las comidas, las cenas, los viajes, y por todo su cariño. Conocerlos fue una de las cosas más lindas que me pasaron en los Países Bajos. Muchas gracias por haber sido una familia y un pedacito de México en Delft. Los voy a extrañar muchísimo, pero sé que nuestra amistad va para largo y que nos veremos seguido. A David le agradezco en particular por haber aceptado ser mi paranymph y por todo su apoyo durante mi defensa.

A **Leila** y **Jose** (¡y a Pablito y a Mina también!) les agradezco muchísimo su apoyo y compañía todos estos años. Tenerlos en Delft siempre me dio mucha

alegría y mucha paz. Muchas gracias por todo su apoyo y cariño. Ha sido muy lindo poder ver el inicio de su linda familia. ¡Nos vemos pronto!

Quisiera concluir dándole las gracias a mi familia, que me ha apoyado desde que tengo memoria, sin el apoyo de la cual no habría llegado a este punto.

Primero que nada, quiero agradecer a mis papás, **Alejandro** y **Victoria**, por siempre priorizar mi educación por encima de otros lujos, por inculcar en mí el valor de la educación desde una edad temprana y por siempre apoyarme incondicionalmente desde la distancia. Muchas gracias por haberme apoyado en mis decisiones profesionales y académicas, aunque algunas hayan sido difíciles, como el irme de casa tan joven. Muchas gracias por siempre alentarme cuando hubo momentos difíciles del doctorado y por siempre mandarme mensajes cariñosos. El sentir su cariño desde la distancia siempre me motivó a seguir y es en buena parte gracias a ustedes que llegué a este punto. Este título es en parte suyo, ¡los quiero mucho!

Me gustaría también darle las gracias a mis hermanos, **Álex**, **Shary** y **Ceci**, por apoyarme siempre y en particular en los años del doctorado. Ustedes saben que ha habido momentos difíciles, pero el sentir su cariño y apoyo apoyo a la distancia (¡y en persona!) me ha ayudado mucho. Una de las cosas que más me han gustado del doctorado es el poder estar más cerca de ustedes ahora que todos somos adultos. Siento que los he vuelto a conocer de una manera más profunda y me da mucho gusto poder pasar este día con ustedes. ¡Los quiero mucho!

Para terminar, me gustaría agradecer a la persona que más me apoyó durante los cuatro años y medio que duró este doctorado: mi compañero, **Diego**. Tú más que nadie sabes que hubo momentos difíciles en el trayecto. Sin embargo, aun en los días más difíciles, el darte un abrazo al llegar a casa siempre me despejaba la mente y me recordaba lo que es realmente importante en la vida. Se me acaban las palabras para darte las gracias por todo tu cariño y apoyo, por siempre escucharme, motivarme y alentarme, y por todos los viajes que hicimos juntos en estos años; no habría podido llegar hasta este punto si no te hubiera tenido a mi lado. En estos años también tuvimos que aprender lo que es estar separados. Sin embargo, la distancia nos ha acercado aún más y me ha hecho quererte más que nunca. ¡Me emociona mucho pensar en todo lo que haremos juntos en los años que vienen ahora que finalmente se acabó la distancia! Te quiero mucho.

Daniel Ramírez Montero, June 2024