Watching the Replisome: Single-molecule Studies of Eukaryotic DNA Replication

Daniel Duzdevich

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

ABSTRACT

Watching the Replisome: Single-molecule Studies of Eukaryotic DNA Replication

Daniel Duzdevich

The molecules of life are small to us—billionths of our size. They move fast too, and in the cell they crowd together impossibly. Bringing that strange world into ours is the trick of molecular biology. One approach is to harness many copies of a molecule and iterate a reaction many times to glimpse what happens at that small, foreign scale. This is a powerful way to do things and has provided major insights. But ultimately, the fundamental unit of molecular biology is the *individual* molecule, the *individual* interaction, the *individual* reaction. Single-molecule bioscience is the study of these phenomena.

Eukaryotic DNA replication is particularly interesting from the single-molecule perspective because the biological molecules responsible for executing the replication pathway interact so very intricately. This work is based on replication in budding yeast—a model eukaryote. The budding yeast genome harbors several hundred sequence-defined sites of replication initiation called origins. Origins are bound by the Origin Recognition Complex (ORC), which recruits the ring-shaped Mcm2-7 complex during the G1 phase of the cell cycle. A second Mcm2-7 is loaded adjacent to the first in a head-to-head orientation; this Mcm2-7 double hexamer encircles DNA and is generally termed the Pre-

Replicative Complex, or Pre-RC. Mcm2-7 loading is strictly dependent on a cofactor, Cdc6, which is expressed in late G1. Much less is known about the details of downstream steps, but a large number of factors assemble to form active replisomes.

Origin-specific budding yeast replication has recently been reconstituted *in vitro*, with cell cycle dependence mimicked by the serial addition of purified Pre-RC components and activating kinases. This work introduces the translation of the bulk biochemical replication assay into a single-molecule assay and describes the consequent insights into the dynamics of eukaryotic replication initiation. I have developed an optical microscopy-based assay to directly visualize DNA replication initiation in real time at the single-molecule level: from origin definition, through origin licensing, to replisome formation and progression. I show that ORC has an intrinsic capacity to locate and stably bind origin sequences within large tracts of non-origin DNA, and that ordered Pre-RC assembly is driven by Cdc6. I further show that the dynamics of the ORC-Cdc6 interaction dictate the specificity of Mcm2-7 loading, and that Mcm2-7 double hexamers form preferentially at a native origin sequence. This work uncovers key variables that control Pre-RC assembly, and how directed assembly ensures that the Pre-RC forms properly and selectively at origins.

I then characterize replisome initiation and progression dynamics. I show that replication initiation is highly precise and limited to Mcm2-7 double hexamers.

Sister replisomes fire bidirectionally and simultaneously, suggesting that previously unidentified quality control mechanisms ensure that a complete pair of replisomes is properly assembled prior to firing. I also find that single Mcm2-7 hexamers are sufficient to support processive replisome progression. Moreover, this work reveals that replisome progression is insensitive to DNA sequence composition at spatial and temporal scales relevant to the replication of an entire genome, indicating that separation of the DNA strands by the replicative helicase is not rate-limiting to replisome function.

I subsequently applied this replication assay to the study replisome-replisome collisions, a fundamental step in the resolution of convergent replication forks. I find that, surprisingly, active replisomes absolutely lack an intrinsic capacity to displace inactive replisomes. This result eliminates the simplest hypothesized mechanism for how the cell resolves the presence of un-fired replisomes and has prompted and guided the development of alternate testable hypotheses. Taken together, these observations probe the molecular basis of eukaryotic inheritance in unprecedented detail and offer a platform for future work on the many dynamic aspects of replisome behavior.

CONTENTS

LIST OF ILLUSTRATIONS

ACKNOWLEDGEMENTS

Prof. Eric Greene and the extraordinary members of his laboratory have been my well-wishers, guidance-givers, experiment-analyzers, question-askers, and answer-offerers. Thank you for indulging me. Prof. Ruben Gonzalez has always enlightened me with the most honest—and so the best—advice and members of his laboratory have heard my talks too many times but listened closely anyway. I learned an unreasonable amount of biology for one rotation with Prof. Lars Dietrich, and he has supported me in all my endeavors since. Prof. Songtao Jia gave me free reign in helping teach his molecular biology class, and in the process I was reminded never to take good teachers for granted. Profs. Max Gottesman and Ken Marians have seen me present my work formally and offered the compliments that make such moments memorable for a young scientist. I thank them for their time in rounding out my committee. Prof. Caterina Pizzigoni and Dr. Elliot Hertzberg have been mentors standing just outside my small world of science and I thank them for their much-needed support from a different perspective. This work began as a collaboration with the laboratory of Prof. Stephen P. Bell at MIT, which supplied many of the constructs and protein preparations used in this work.

At various stages this project has been directly or indirectly supported by funds from the Howard Hughes Medical Institute and the National Institutes of Health. Key work on Pre-RC dynamics was funded by an award from the Paul and

Daisy Soros Fellowships for New Americans, an organization and community that has meant more to me than I can relate here. The final stage of the project was funded by an award from the Josephine de Kármán Fellowship Trust. In both cases I was also supported by the Columbia University Graduate School of Arts and Sciences. I am deeply grateful to these funding agencies for recognizing the value of risk in experimental design.

Dedicated to the memory of Chinweike Okegbe, Ph.D.

PREFACE

The problems that lie ahead are not static ones, resolvable by static techniques, but concern the "machinery" of replication. This machinery, which is responsible for the rapid and accurate synthesis of a single and enormous polymer, promises to have a physical complexity that matches the scale of the operation. But I see that I have wandered into mixing prophecy with history. It is therefore time to stop.

- John Cairns, Former Director of Cold Spring Harbor Laboratories, 1966[1]

DNA replication connects all cells throughout time. It grants the genetic information of one generation to the next, and so it is the molecular heart of evolution. It enables reproduction—for single-celled organisms it effectively *is* reproduction—and multicellular development and growth. Replication, well-timed, is so important that every major cellular process answers to its correct execution. Viruses hijack it, and biologists too.

The imagery of the double helix embodies the relationship between form and function that only evolution and human creativity are known to spawn. DNA's structure practically predicts the dance of its duplication. Take a replication event in any cell and imagine that the proteins aren't there: a small region of DNA separates into its two constituent strands, exposing the template sequences. Two replication forks take off in opposite directions, with leading and lagging strands

---

[1] *Phage and the Origins of Molecular Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.

proceeding from each. (There is a satisfying symmetry to the reaction, a rare treat for the aesthetically-inclined experimentalist.) These essential snapshots of what happens to the DNA, old and new, were identified mostly through work with *E. coli*. But the machinery used by eukaryotes is not identical to what had been discovered in bacteria. Our understanding of the eukaryotic system is relatively fresh, and major advances have coincided with the development of another discipline important to this work: single-molecule bioscience.

Interactions among biological molecules are the fundamental unit of molecular biology. But these interactions are generally too fast for us to deal with. Besides which, everything involved is an order of magnitude smaller than the wavelength of visible light. Most experiments scale up by using huge numbers of reactions run for many minutes. This can yield enough of a readout (a band on a gel, for example) to infer what happened molecularly. These readouts are always the averaged product of all those reactions. But the ideal unit is one reaction, and that is the realm of single-molecule approaches.

Here I present the single-molecule analysis of the replication initiation and progression pathway using reconstituted components from budding yeast, a model eukaryote. I begin by introducing the single-molecule perspective on biology and the “DNA curtain” method developed by the Greene laboratory for the study of protein-DNA interactions. Subsequent sections explore how I have applied this experimental platform to study the replication pathway, including the interplay among initiation factors, the specification and activation of initiation sites, and the

assembly and procession of replisomes. I conclude with a description of experiments designed to measure collisions between replisomes and physiologically relevant roadblocks. All these experiments exploit the unique strengths of single-molecule analysis: the capacity to identify heterogeneity in reaction pathways that otherwise all access identical components, and to identify one component and know its fate. This approach yields surprising insights when combined with the capacity of DNA curtains to specify location on a DNA substrate.

To coax precise information from such complex and miniscule systems is the challenge of molecular biology. I've used optics—lasers and fluorescent probes, mirrors and lenses—to link with the molecular world by chains of photons. I've gotten to *see*, as directly as can be seen, my transient subjects. I hope that the visceral thrill of the experience courses through the figures, numbers, and words that are so many steps removed from the true hours of the lab.

CHAPTER 1

Single-molecule Biology and DNA Curtains

The Single-molecule Approach

The very first single-molecule observations of biological molecules and their enzymatic activity were made many decades ago (*1, 2*). The benefits of measuring behavior at this fundamental scale have always been apparent, but so has the challenge of marshaling the tools and techniques to actually make such measurements. Single-molecule bioscience has matured only recently as significant advances in technology became available (*3, 4*). Computing power and software sophistication are important for manipulating data, and microfluidics enable experiments with controlled buffer flow and composition. An entire field of chemistry has sprung up to supply and characterize biologically useable fluorescent dyes (*5*). The most essential advances have been in optics. Sensitive digital cameras, appropriate lasers, and quality optical filters are all integral to the type of research presented here. Understanding how to interpret single-molecule results is also important. The field of single-molecule bioscience really begins with the realization that there are new concepts at play when we study the discrete units of molecular biology (*6-9*).

The majority of techniques in molecular biology require scaling up from the individual units that are the ultimate subjects. The readouts of these techniques—

the observables available to the scientist at the end of an experiment—represent the averaged outcomes of the behavior of many molecules. There is usually an averaging across time too. Single-molecule techniques extend the reach of molecular biology with readouts that reflect the behavior of individual molecules and individual reactions, sometimes with high temporal resolution. Averaged readouts and single-molecule readouts are related because they offer different perspectives of the same phenomena, but the single-molecule perspective is entirely inaccessible to other methods and supplies unique insights into the logic of molecular pathways.

The salient feature of single-molecule data is that it preserves the identity of molecules and reaction trajectories: it's possible to tell which molecule does what as a function of location, time, or some other variable. The precision with which each variable can be measured depends on the specific technique, but the capacity to point out one molecule or reaction and assign some value or characteristic to it is at the core of the approach. This is vital information when studying biological systems, with their evolved intricacies. There are many interactions and many steps, often accompanied by precise timing: reactions happen at defined timepoints with respect to some external variable, steps of a reaction happen at different rates to control the fidelity of a pathway, or the components interact in a very particular order so that one step must happen before another. Complexity plus controlled timing equals *dynamics*, and it is

impossible to understand the dynamics so pervasive in molecular biology without identifying and tracking the individual components of a pathway.

Identity may not be an interesting observable for extremely homogenous systems, with just one or two components and very fast reaction times. This is the assumption at the heart of traditional bulk biochemistry: key experimental inputs, such as starting concentrations and temperature, and a few scaled readouts, such as reaction times and the chemical properties of substrate and product, offer a great picture of the pathway and how it works if the reaction is uniform, if every molecule and every reaction event is identical. The bulk biochemist controls the input of a reaction and measures the products. Inferring what happens on pathway to the products generally requires the assumption that the reaction and the molecules participating in the reaction are homogenous. But even the simplest cases are not truly homogenous, and if even just two biological macromolecules are involved—say, one type of DNA and one type of protein—then dynamics, and identity, become relevant to fully understanding the pathway.[1]

Many techniques have been developed and refined to access specific aspects of single-molecule biology. Some do away with the time variable and collect data on static molecules. These include versions of atomic force

[1] Appreciating that dynamics matter for understanding how biological molecules function does not taint the value of knowledge gained from techniques that require scaling up (bulk biochemistry, genetics, cell biology, *etc.*). But it is helpful to understand the strengths and limitations of each approach. The price of probing dynamics at a minute scale by doing single-molecule experiments can be high: extreme specialization, many technical hurdles, and limits to the quantity of raw data, so it is good practice to make sure that a single-molecule experiment is worth performing and to think carefully about what the resultant data mean in terms of single-molecule dynamics and in the context of results from other techniques that average over dynamics.

microscopy (AFM) that probe the topology of biological molecules on a surface, and some versions of electron microscopy (EM). Although time-dependent changes are not always accessible to these methods, the structural status of a molecule can be correlated to its origin or to some other structural feature. For example, DNA with specific sequences may contain non-B-form peculiarities (*10*), or DNA from a eukaryote may look like “beads on a string” from all those nucleosomes (*11*). Some measure molecular forces by pulling and twisting on DNA (or other linear molecules), or holding on to it as some translocase or helicase tugs back (*12*). Single-molecule Förester resonance energy transfer (smFRET) measures changes in correlated dye-pair intensities to infer structural rearrangements as a function of reaction coordinate: it is an exceptionally useful technique when studying fast biochemical events on large biomolecules such as the ribosome (*9*). Finally, a large and diverse class of techniques, including the one used in this work, has been built around the visualization of fluorescent probes attached to biological molecules. The common characteristic of all is that some interesting variable can be linked to individual molecules or interactions.

The identity data from a single-molecule experiment offers at least two major general insights into how a molecular pathway works. (i) It shows how individual molecules process the information of their surroundings, and (ii) highlights some of the heterogeneity intrinsic to the pathway and how that heterogeneity contributes to the outcome of a reaction.

All levels of biology integrate information. Organisms, cells, and sub-cellular processes detect and respond to a vast range of conditions and stimuli. Biological molecules change structure, mediate sensitivity to other molecules, and alter their capacity to follow one trajectory over another in response to many types of information. The information can be cellular: dependent on the cell cycle, a particular cellular compartment, or some unique aspect of the cellular environment. Or it can be experimental, including, for example, the typical variables of an *in vitro* biochemical assay. Sometimes these variables are meant to mimic cellular conditions, and sometimes they are manipulated to illuminate the basic biophysical properties of a reaction. All the common variables of molecular biology are relevant to single-molecule experiments too, but the way that an individual molecule responds to a source of information as simple as concentration may not be apparent from an averaged result. More importantly though, there are some more subtle types of information that can only be understood at the single-molecule level. These generally have to do with how the momentary status of one molecule influences how it interacts with other molecules, or how it continues to participate (or not) in the rest of a pathway. DNA-binding proteins play liberally with the control and nuance offered by interactions with other molecules because the DNA is information-rich and in a constant state of manipulation, a theme throughout much of this work.

Molecular heterogeneity is, perhaps, less intuitive than molecules' handling of information. It is also commonly invoked as the defining strength of single-molecule data, and yet few studies explore how one or another form of heterogeneity contributes to the essence of a pathway. To begin, consider a previously expounded hypothetical example of color-morphing small molecules (*8*). Imagine a mole of these molecules in solution. They occupy an energy landscape—an array of energetic states—and each state endows any molecules that are in it with a defined, visible color. The states are separated by relatively small energetic barriers and so the molecules interconvert freely among them at room temperature. A measurement of the exact color of the solution would yield the average color of all the molecules at that moment. Now imagine following the color changes of just one molecule with time: on some timescale—the relaxation time of the system—that one molecule will sample all the thermally accessible states. This means that the average color of one molecule with time will eventually equal the average color of the entire solution. This equivalence is known as the ergodic hypothesis.

In almost any real system the barriers among states and their respective stabilities (the depths of the energetic wells) are not the same. Different states are occupied to different degrees. The system is still ergodic because a single exploratory molecule spends more time in more stable states, effectively weighting the average. But the differences in state occupancy are an inherent form of

heterogeneity known as static disorder. *Static disorder is the minimum amount of heterogeneity in a population of molecules in solution.*

Now consider biological macromolecules. They occupy almost incomparably broader state-spaces than simple organic molecules. And in the context of the cell and many typical experiments, a molecule's energy landscape *itself* fluctuates. The barriers and wells defining states move around, constantly shifting local and global minima. This type of *dynamic disorder* is probably safe to assume in most biological systems. Interactions among different molecules require that multiple energy landscapes intersect in space and time, and when the molecules come into physical contact, effectively creating a novel molecular entity, a new landscape will pop into existence. For example, if the blue shade of the color-morphing molecules were an enzyme substrate, then catalysis would require the molecule to be in the correct "blue" state on its energy landscape when it collides with just the right part of the enzyme. The enzyme too will have to be in an amenable state, in a favorable conformation to recognize and act on the blue molecule. Together, blue plus enzyme, they fall onto a new landscape and travel along it until the reaction is over and the enzyme can return to its home landscape. The substrate will have been chemically altered into something different, so it will find itself on an entirely new landscape, one that does not intersect with the enzyme's: the basis of catalytic turnover and enzymatic specificity.

Systems that break or effectively break ergodicity are interesting here because only single-molecule analysis can directly reveal how the individual components of multiple subpopulations create the averaged behavior. This is extremely valuable to understanding how a reaction works in the cell where one or a few molecules can have drastic effects. Anything that changes the energy landscape of a molecule so that the states stop interconverting on physiological timescales breaks ergodicity and is ideal for single-molecule study. Sets of molecules will be separated from each other, defining different subpopulations. Different isomers or thermal states of a protein may affect overall sensitivity to some signal, different conformations of DNA may prove differently accessible to a signaling molecule, or different biophysical characteristics of disordered motifs may depend on local ionic strength. These and many other scenarios are readily apparent for almost any pathway and grow exponentially as the number of molecules is increased and the variety of components is elaborated. They are common and essential to systems with DNA-binding proteins, especially if there are multiple proteins and serial interactions involved. Applying a single-molecule approach to the eukaryotic DNA replication pathway, for example, should therefore offer deep biological revelations. But such experiments require a robust and sensitive technique that can measure where and when proteins bind DNA. They also have to yield data from across many molecules and many interactions. The DNA curtain technology was developed for exactly this type of experiment.

DNA Curtains

Dissecting the distinctions and relationships between a molecular ensemble and its constituents requires data on many individual molecular trajectories. Furthermore, applying a single-molecule technique to any system with more than two or three components is a distinct challenge because the experimental protocols tend to be delicate and difficult. The DNA curtain approach addresses these issues: it is high-throughput, and it has been refined to a mature enough level of sophistication to tackle a multi-component and multi-step pathway as intricate as replication. A DNA curtain experiment combines nanopatterning, microfluidic manipulation, and optics to measure how proteins interact with parallel arrays of DNA molecules. Each experiment occurs inside a reaction chamber, or "flowcell," built around a microscope slide with a supported lipid bilayer. Functionalized lipids within the bilayer are attached to individual DNAs, and nanometer-scale metal features on the slide surface disrupt the otherwise two-dimensionally fluid bilayer. These disruptions are responsible for fashioning the curtains, and generating them is the first step of experimental design.[2]

Nanofabrication (*13-15*) (**Fig. 1.1**) begins with the stringent cleaning of pure fused silicon dioxide glass[3] microscope slides to strip away all organic contaminants. The slide is then sequentially layered with low molecular weight

[2] Finer technical details are in Appendix II.

[3] Trace impurities in natural quartz or lime glass tend to fluoresce, contributing to noise in light microscopy.

Figure 1.1

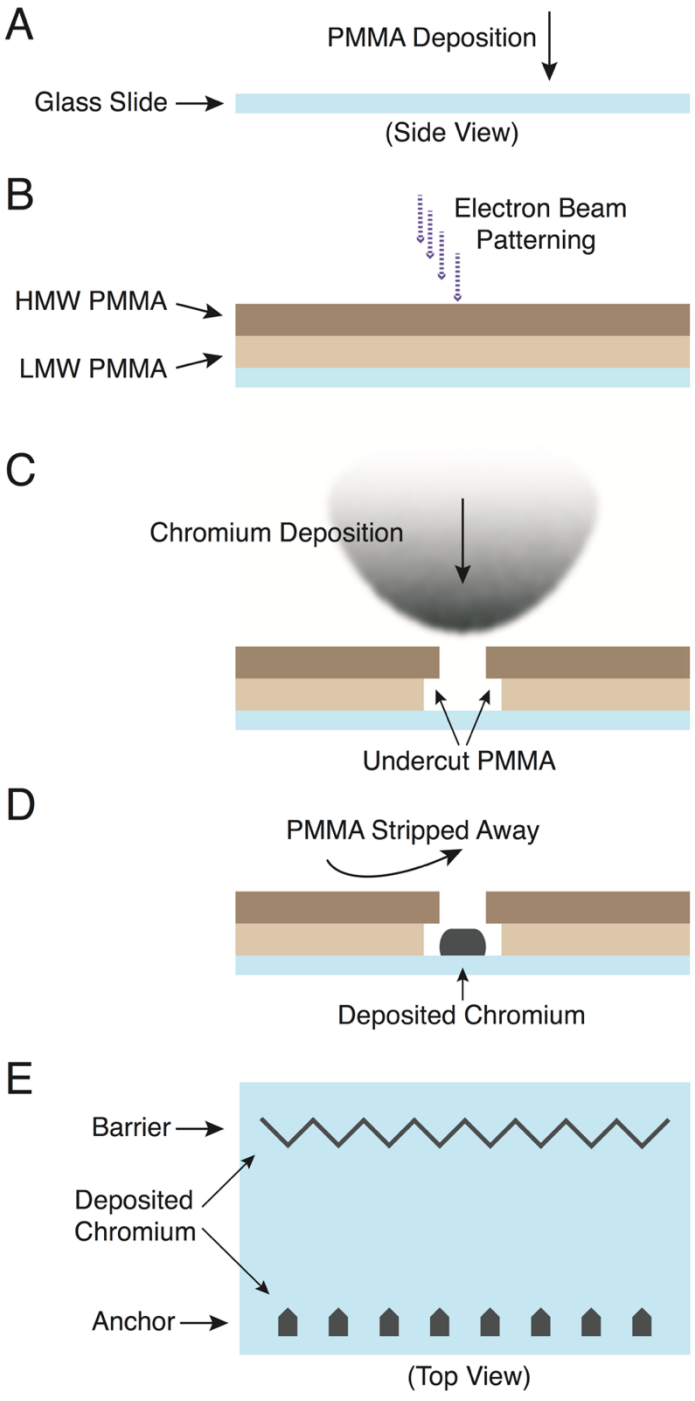
A
PMMA Deposition
Glass Slide
(Side View)
B
Electron Beam
Patterning
HMW PMMA
LMW PMMA
C
Chromium Deposition
Undercut PMMA
D
PMMA Stripped Away
Deposited Chromium
E
Barrier
Deposited
Chromium
Anchor
(Top View)

Figure 1.1 (Preceding Page) Nanofabrication
(A) A glass microscope slide is layered with PMMA plastic. (B) A directed beam of electrons de-crosslinks the plastic in a defined pattern. (C) A developing agent exposes the pattern negative. The LMW PMMA is developed more and leaves behind a crucial undercut. (D) Chromium metal is evaporated onto the slide and deposits directly onto the glass where the PMMA has been developed. (E) The PMMA is stripped off, revealing the pattern written in chromium.

(LMW) poly(methyl-methacrylate) (PMMA) and then high molecular weight (HMW) PMMA (**Fig. 1.1A**). E-beam lithography, which essentially amounts to patterning with the electron gun of an electron microscope, is then used to trace a defined pattern on the PMMA (**Fig. 1.1B**). The electron energy transfer de-crosslinks the plastic along the pattern. Importantly, the LMW PMMA is more susceptible to de-crosslinking for a given input of energy than the HMW PMMA. This yields an undercut in the lower layer of PMMA when the plastic is developed to reveal a negative of the pattern (**Fig. 1.1C**). A high energy electron beam is then used to generate gaseous chromium inside a vacuum; the metal deposits directly onto the glass surface of the microscope slide in the shape of the pattern, atomic layer by atomic layer, and the undercut ensures that the deposited chromium does not contact the PMMA. An organic solvent strips away the PMMA (**Fig. 1.1D**) to reveal the deposited chromium in the defined pattern (**Fig. 1.1E**). The chromium structures are approximately 250nm above the slide surface with feature dimensions on the order of tens of nanometers, and they efficiently disrupt lipid bilayers, which are only tens of nanometers above the surface. The chromium also nonspecifically adsorbs some proteins, which is crucial for certain applications (see below).

The nanopatterned slides are then assembled into flowcells (**Fig. 1.2**). A layer of double-sided tape with an excised channel is sandwiched between the slide and a coverslip. The channel forms a chamber of ~13µl. Two precisely drilled holes in the slide are attached to microfluidic ports that introduce buffer flow into the flowcell either through manual syringes for setting up experiments or regulated syringe pumps for running experiments. Flowcells are not reusable, but the patterned microscope slides can be salvaged after an experiment, cleaned, and reused approximately ten times before the glass or the chromium patterns becomes too degraded for quality data acquisition.

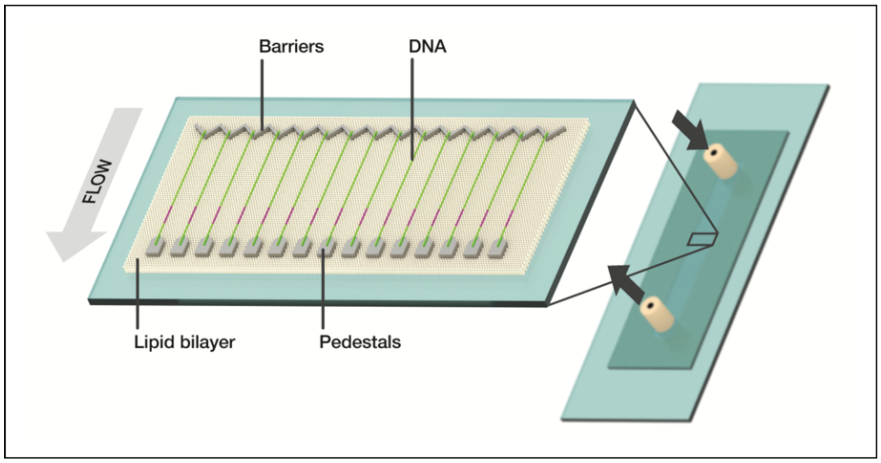

Figure 1.2 Flowcell and DNA Curtain Schematic
A flowcell (right) is a narrow microfluidic reaction chamber with a buffer inlet and outlet. DNA curtains (left) are formed atop a lipid bilayer inside the reaction chamber at the nanopatterned chromium. Cartoon courtesy of Myles Marshall.

A major technical feature of DNA curtains is the lipid bilayer used to organize DNA. The bilayer solves two issues common to fluorescent single-molecule techniques. First, it passivates the glass surface. Biological molecules tend to be sticky—they carry charges and readily adsorb to charged surfaces either directly or through ionic bridges. A lipid bilayer blocks easy access to the

glass. It also supplies a more physiologically reasonable environment for experiments than, for example, layers of aggregated protein or polyethylene glycol (PEG). Second, the bilayer is two-dimensionally fluid, and this property is harnessed to organize many DNA molecules. Each experimental run can consequently yield numerous individual trajectories, a major advantage for interpreting data and understanding the role of any heterogeneity.

The lipid bilayer is generated by introducing lipid vesicles into the flowcell at a high enough concentration for some vesicles to spontaneously rupture on the glass surface; the amphiphilic phosphatidylcholine head groups on our lipids of choice are essential for this process because they mediate the initial interaction force between vesicles and the charged surface. A small subpopulation of the lipids has head groups modified with biotin. Each biotinylated lipid is free to diffuse around the bilayer two-dimensionally, but it cannot pass over any chromium features. Streptavidin, which binds biotin exceedingly tightly, is then linked to the biotinylated lipids. Streptavidin is multivalent so DNA biotinylated at one end can be anchored to the bilayer by a lipid-biotin-streptavidin-biotin-DNA linkage. The DNA now has the same property as the lipid to which it is anchored: it can move two-dimensionally in the bilayer. In the presence of buffer flow the DNA functions as a sail and is pushed along until the lipid to which it is linked encounters a chromium barrier (**Fig. 1.2**). The lipid gets caught and the DNA stretches out in parallel to the slide surface. The chromium barrier is arranged so that many DNAs

can be positioned and aligned next to each other. Because each DNA is functionalized on the same end (see Chapter 2), each DNA in the curtain is aligned in the same orientation. Therefore, a given sequence within the DNA will occupy the same position in each DNA across the entire curtain. This is illustrated in **Figure 1.2** with a magenta-colored region which could represent, for example, a cloned origin of replication sequence. In this simplest "single-tethered" scenario, continuous buffer flow is required to keep the DNA in an extended curtain conformation. If buffer flow is stopped, then the DNA retracts and forms a more compact structure near the barriers as a function of its persistence length. (The flow conditions typical of DNA curtain experiments stretch DNA to ~85% of crystalline B-form length, and spatial resolution for most experiments is 800-1,000bp.) A variation is "double-tethered" curtains, used extensively in this work. Pedestal-shaped anchors can be patterned at defined positions downstream of the curtain barriers (**Figs. 1.1 and 1.2**). These are nonspecifically decorated with anti-digoxigenin (anti-DIG) antibodies, and the distal ends of the DNAs are functionalized with DIG. An initial burst of buffer flow is used to generate a curtain and bring the DIG and anti-DIG into contact, thereby anchoring the DNAs. With both DNA ends held in place, continuous buffer flow is not required to maintain the curtain configuration. The functionalization on either DNA end is linked to the DNA backbone through several single bonds. The DNA is therefore expected to be free to rotate and exhibit a relaxed topology. However, in the presence of excess DNA-

binding proteins or proteins that extensively coat the chromium barriers, the DNA may become locally attached to the chromium, precluding free rotation.

The particular configuration of DNA curtains also offers an advantage for fluorescent microscopy. Minimizing unwanted fluorescence is important for any optical microscopy technique, but especially so for single-molecule data acquisition where the signal-to-noise ratio can determine the success of an experiment. For curtain experiments we are only interested in signal from the DNA (if it is labeled) and/or from proteins bound to the DNA. Material in bulk solution (*i.e.*: most of the volume of the flowcell above the curtains) would only contribute to the noise. We use total internal reflection fluorescence microscopy (TIRFM) to selectively excite fluorophores near the slide surface (**Fig. 1.3**). A laser beam is reflected off of the interface between the glass and the buffer underneath, which have different optical densities. An evanescent wave extending several hundred nanometers into the buffer beneath the interface is the excitation field within which fluorophores absorb and emit photons. Any sources of fluorescent signal in bulk solution remain outside the shallow penetration depth of the field and are not excited.

DNA curtains remain stable for many hours and the lipid bilayer can accommodate a wide range of buffer conditions, buffer exchanges, and temperatures. The curtains and flowcells are versatile and robust with the high-throughput potential of many dozens or hundreds of DNA molecules all in one

optical field of view. It is an ideal platform to tackle the challenge of accommodating the eukaryotic replication initiation pathway.

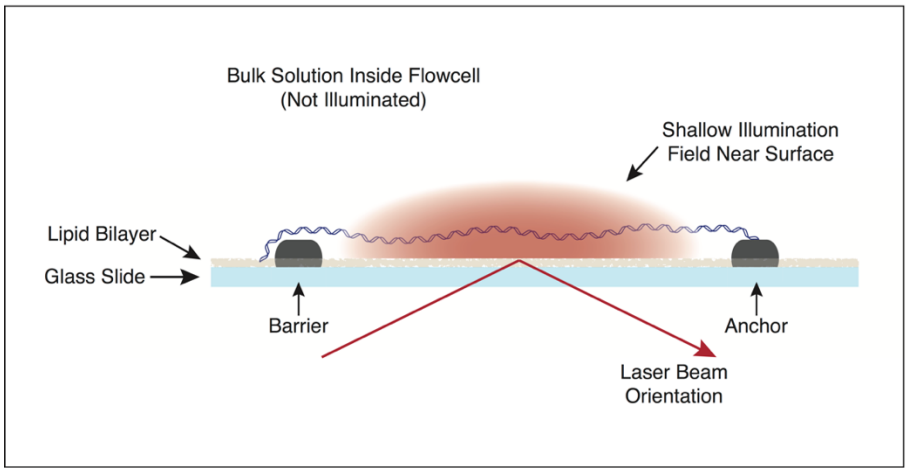

Figure 1.3 TIRFM and DNA Curtains
A laser beam is reflected at the interface between the glass surface and the buffer underneath to generate a shallow field of illumination within which fluorescent molecules are excited.

CHAPTER 2

The Origins of DNA Replication

Within a decade of Watson and Crick's casual remark that DNA's structure immediately suggests a mode for its replication (*16*), Jacob and Brenner proposed the replicon model as an explanation of how a bacterium could go about copying its circular chromosome (*17*). The concept is simple, with just three components. First, the chromosome must harbor a *replicator*, a defined sequence where replication begins. This is the key piece of DNA where the template is initially exposed. Second, an *initiator*, taken to be some protein factor, recognizes and binds the replicator. Third, the action of a replicator-initiator pair results in the replication of a defined region of DNA, the *replicon*. Extensive work on *E. coli* demonstrated that the model is correct, though naturally the details proved more complex. In eukaryotes, with their multiple chromosomes and cell cycles, the situation is more complex still. This chapter introduces the eukaryotic initiation pathway and reports the single-molecule behavior of the eukaryotic initiator protein complex.

Pre-replicative Complex (Pre-RC) Assembly in a Model Eukaryote

The work presented here is on the *Saccharomyces cerevisiae* budding yeast replication initiation and progression pathway. Yeast has served as a seminal

model of eukaryotic replication for many decades: the essential genes needed for replication were identified in classic genetic screens (*18-22*), and the otherwise extensive genetic work on yeast makes it an ideal subject. Although bacterial and eukaryotic replication factors differ drastically, among eukaryotes they are very highly conserved (*23*). (Archaeal replication proteins are eukaryotic-like and have been especially useful in structural and mechanistic studies (*24-27*).) Therefore, yeast is both a practical and generally instructive model system to study DNA replication.

Unlike most bacteria, eukaryotic cells carry multiple linear chromosomes and usually contain more DNA overall. They also have well-regulated cell cycles. These contingencies have shaped eukaryotic replication in at least two major ways: the use of many replicators (origins of replication), and control of the initiation pathway to ensure that the genome is duplicated only once per cell cycle. Multiple origins enable timely replication of the complete genome but also present some issues. The location of origin sites in the genome and how initiators use them, for example, is not trivial. Multiple origins also implies multiple replication forks which create an entire set of additional issues (see Chapter 6). The relationship between replication control and the cell cycle is more subtle, but in essence, the cell applies a range of mechanisms to ensure that all of the DNA is accurately replicated only once per cycle (see also Appendix I). This is important for organizing the chromosomes during mitosis and meiosis, when excess copies

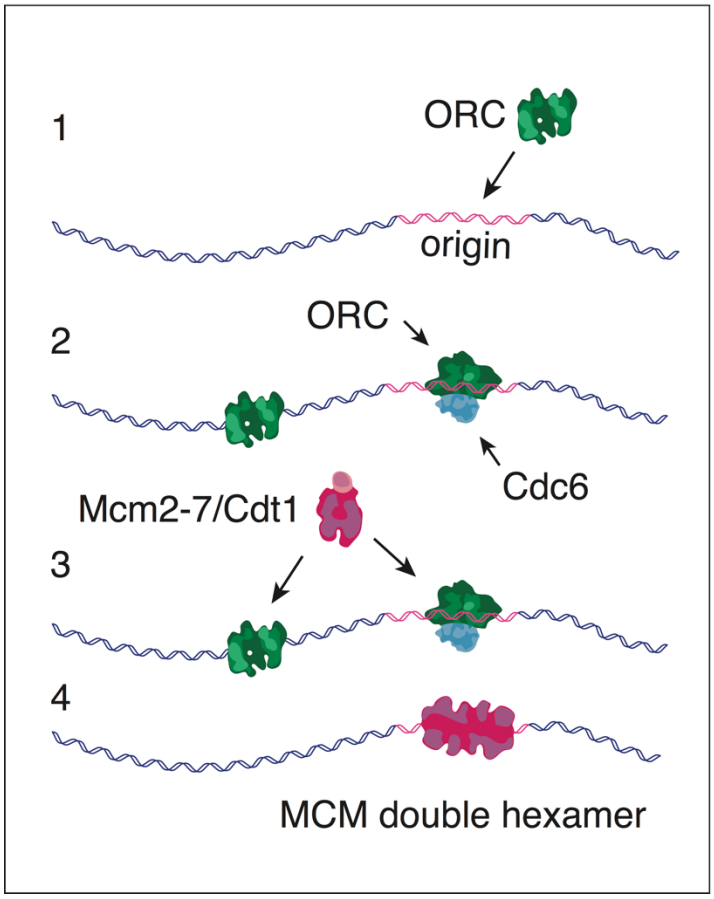

Figure 2.1 Pre-RC Assembly
(1) ORC binds the *ARS1* origin sequence.
(2) Cdc6 interacts with ORC.
(3) Together, ORC and Cdc6 recruit Mcm2-7/Cdt1.
(4) A pair of Mcm2-7s forms oppositely oriented hexamers.

of some chromosomes or some genomic regions could lead to genomic instability. From the perspective of many generations of a eukaryotic organism, maintaining only one full copy of the genome (if there is a selection pressure to do so, and we infer that there generally is) requires exquisite control of chromosomal copy number; keeping a tight reign over replication initiation is an ideal way to achieve this control. The most conserved of these mechanisms involve cell-cycle-specific kinase activities (*28*), and these will figure in later chapters.

The intersection of origin usage and cell cycle control begins at the earliest stages of the replication pathway: the assembly of the Pre-Replicative Complex (Pre-RC) (**Fig. 2.1**). The yeast genome contains many hundreds of origins of replication (*29-31*), and these can be bound by the Origin Recognition Complex (ORC) throughout the cell cycle (*32*). ORC is heterohexameric and interacts with

the important cofactor Cdc6 in G1 (*33-37*). Together, ORC and Cdc6 recruit Mcm2-7/Cdt1, a heptameric complex in solution (*22, 38-44*). Cdt1 is essential for Mcm2-7 recruitment but is subsequently released (as are ORC and Cdc6) (*28, 44*). In a poorly-understood step, a pair of Mcm2-7s is ultimately loaded onto double-stranded DNA (dsDNA) to form an Mcm2-7 double hexamer. Each Mcm2-7 is a closed ring in this structure and the pair face away from each other; in downstream steps during S-phase, each will form the core of a replicative helicase (*40, 42, 44-48*). Many regulatory mechanisms ensure that Pre-RC formation only happens in late M-phase and/or G1-phase, so that all the potential sites of replication initiation are defined by the end of G1 with loaded double hexamers (*28*). Each Pre-RC assembly component contains ATPase motifs, and many have been demonstrated to bind and actively hydrolyze ATP under a variety of conditions. However, the regulatory role of each ATP-binding and hydrolyzing activity has yet to be clearly illuminated (*36, 38, 39, 49-61*).

These components and steps were identified through a variety of genetic and biochemical assays. For example, all the genes were picked up in genetic screens, and the restructuring of ORC as a function of Cdc6 and Mcm2-7/Cdt1 recruitment were defined as changes in DNA footprints (*19-22, 35, 62*). However, almost none of the dynamics, as discussed in Chapter 1, were known until this work. Specifically I asked how does ORC interact with DNA? How does Cdc6 interact with ORC and in what order do they function? How does this interaction

affect Mcm2-7/Cdt1 recruitment? Each of these questions and their corollaries will be addressed in successive chapters, beginning with ORC behavior in the sections below. The importance of heterogeneities will be highlighted throughout. The properties of individual molecules and distinct subpopulations prove indispensable to understanding how replication initiation is controlled.

The Yeast Replicator

Origin sites in budding yeast are largely sequence-defined. This has been a convenient boon to dissecting the Pre-RC pathway because it enables straightforward experimental design. For example, the presence or absence of an origin sequence can lead to binary biochemical results or genetic readouts (several such experiments figure below). But other circumstances are also relevant for defining origins, most notably local chromatin context (*63-65*). Ultimately, origin function is determined by Pre-RC assembly, and so ORC binding is a prerequisite. Not every potential origin, as predicted by sequence, is occupied by ORC; some sites seem to remain always inaccessible or non-functional for Pre-RC assembly despite the potential for ORC binding. Other sites are only used occasionally. These distinctions have been mapped extensively in yeast (*29*), and yet a satisfactory set of rules for the overall scheme has remained elusive, other than the apparent requirement for regularly-spaced origins to ensure that no one replisome has to travel particularly far (*66*). Each replisome has an intrinsic failure

rate, and large inter-origin stretches increase the probability that both replisomes responsible for replicating a given stretch will fail. The situation is even more vexing in metazoa where determinants of origin sites have remained mostly elusive (*67*). In *Xenopus laevis* egg extracts, for example, Pre-RCs seem to form across exogenous DNA with few preferences (*41, 68, 69*). This may be because the genomes of the early embryo are meant to propagate very quickly during early development, before the complications of differentiation could affect origin usage (*70*). In *Drosophila melanogaster*, DNA topology has been reported as a significant determinant of ORC binding (*71*). There are other examples, but in any case it is reasonable to assume that origin definition and usage are important to the cell, even if that definition does not make reference to defined sequences.

Budding yeast, however, do make this reference, and this was the seed to most subsequent studies of eukaryotic origins. Following the logic of the very strictly-defined *E. coli oriC* origin (*72, 73*), early work on propagating yeast plasmids sought to identify an equivalent replicator. Yeast genome libraries were transformed into yeast cells and tested for their capacity to propagate in plasmid form. Plasmids can only be replicated if they carry the sequence necessary to initiate replication. Otherwise, the plasmid is not replicated and lost within the population of proliferating cells. Therefore, successfully propagated plasmids were inferred to carry sequences that confer replicative potential on the plasmid. These sequences were termed autonomously replicating sequences, or ARS sites (*74,*

75). Later work on yeast artificial chromosomes (YACs) agreed with these findings: an ARS sequence is one of three essential components needed for stable YAC propagation, the others being centromere and telomere sequences (*76*). The first origin sequence identified by the plasmid assay was designated *ARS1*, and although the primary designation is purely historical, it has turned out to be a relatively strong origin sequence in that it is used frequently by the cell (*29*). The precedent also means that it is the most-studied origin, and much of what we know about origin structure and Pre-RC formation is based on this site.

The *ARS1* site is nearly 300bp and contains a variety of defined sequence elements. However, the roles of these elements are not entirely understood, in that their molecular roles remain unknown (*63, 77-80*). Furthermore, the inclusion of sequences beyond the few defined elements, up to the complete 300bp, increases origin function in the plasmid retention assay, but it remains unknown why this is the case (although nucleosome positioning is the likeliest explanation (*64*)). Nonetheless, certain biochemical measurements have been correlated with some sequence elements of some ARS sites. Early analysis of *ARS1* and the similar 2µm *ARS* identified a core 11bp consensus sequence, termed domain A and later the ARS consensus sequence (ACS), as required for plasmid retention function (*81-83*). Mutations within domain A were found to reduce or abolish origin function in the plasmid retention assay, but as additional origin sequences were discovered it became apparent that not all ARS sites had identical or even similar A domains

(*82, 84-87*); many origins lack sequence similarity to ARS sites altogether (*29, 88*). Additional sequence elements were subsequently identified and named the B1, B2, and B3 sites, all adjacent to A (*89*). Some mutations in some B elements are tolerated for replication function, and some combinations of mutations in multiple B elements abolish origin function in certain ARS sequences (*77*). B3 was found to be a binding site for the transcription factor Abf1 and it enhances origin efficiency by unknown mechanisms for some ARS sequences (*89-91*). A and B1 were eventually identified with the ORC binding footprint (*49, 62, 92*) and B2 with the Pre-RC footprint (*79, 80, 93*). Despite these analyses, there is currently no clear relationship between origin sequence architecture and Pre-RC assembly other than the affinity of ORC for the ACS. It is therefore more useful to consider the general properties of origin sequences: ARS sites are AT-rich, which may facilitate initial DNA melting—AT pairing being energetically weaker than GC pairing by one hydrogen bond—or it may be related to the higher stiffness of AT sequences, which has been implicated in excluding nucleosomes (*93, 94*).

The first component needed to assay Pre-RC assembly at the single-molecule level is a DNA substrate both suitable for curtain experiments and harboring a cloned *ARS1*. The expectation is that a native origin with a defined sequence will yield functional Pre-RCs; these will be specifically identifiable as occupying the cloned *ARS1* site, which can be identified within a DNA curtain based on position (**Fig. 1.2**). We use the genome of the bacteriophage λ for many

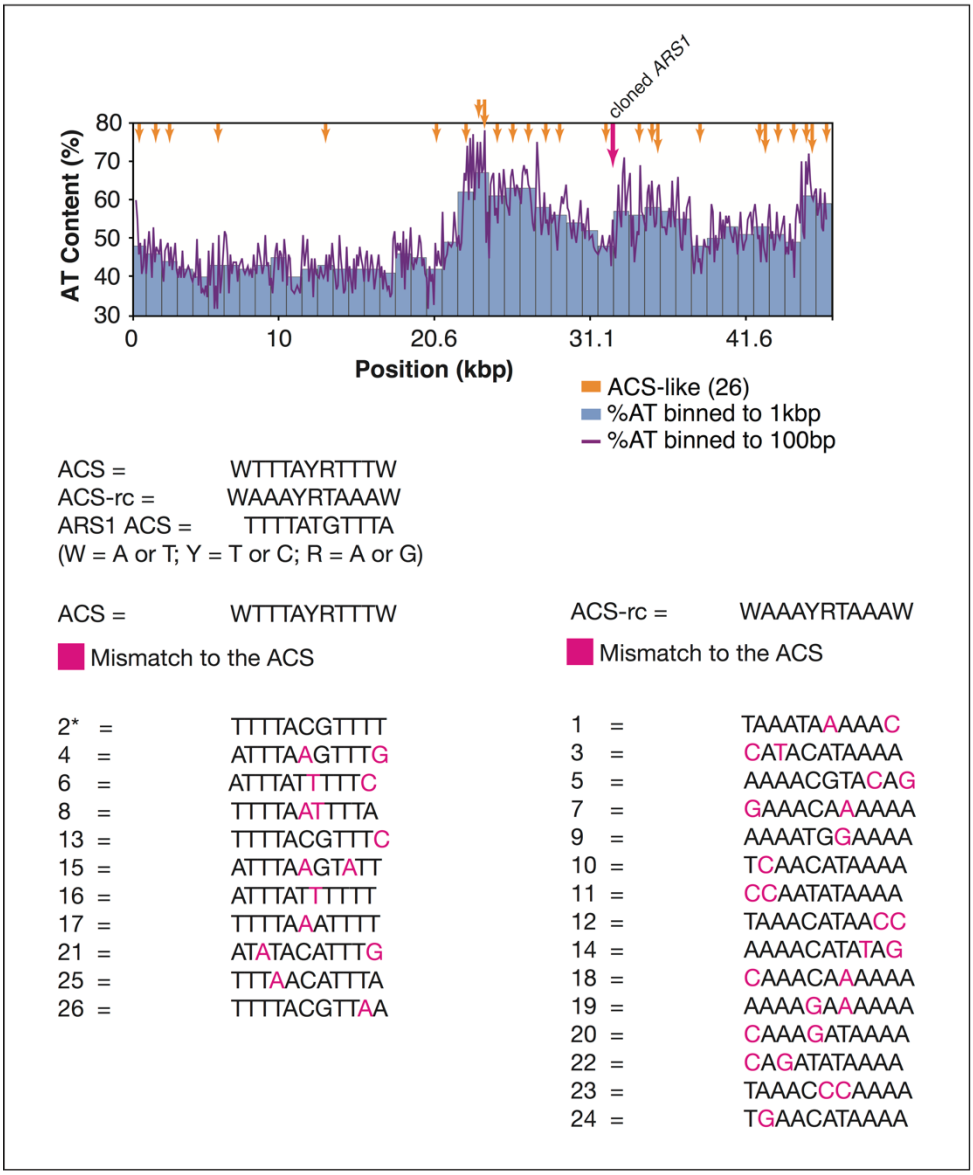

Figure 2.2 The λ_{ARS1} Substrate
The genome of the bacteriophage λ has an inherently skewed AT content. The histogram shows %AT content binned to 1kbp and the overlaid purple graph shows the same data binned to 100bp. The magenta arrow indicates the location of the cloned native *ARS1* sequence and the orange arrows indicate the locations of 26 near-ACS sites native to λ. Histogram reprinted from reference (*133*), with permission from Elsevier.

of our studies because it is relatively long and therefore ideal for microscopy (see below), can be propagated easily and purified to high concentrations, and carries natural 12bp overhangs that can be readily functionalized with biotin and DIG

handles (see Chapter 1 and Appendix II). I cloned a single native yeast *ARS1* origin into a naturally-occurring cloning site within λ to yield λ_{ARS1} (**Fig. 2.2**). The product is ~47.5kbp, which is coincidentally roughly the average size of inter-origin regions in budding yeast. λ has a skewed AT content, very apparent in the **Figure 2.2** histogram, and also contains 26 ACS-like sites, each with at least 9 matches to the 11bp canonical ACS (**Fig. 2.2**). This substrate enabled direct observation of the individual yeast replicators and initiators in action.

The Yeast Initiator

Yeast ORC was identified through its very capacity to bind *ARS1* (*49*). Orc1-Orc6, numbered according to size, form a stable heterohexameric complex in many organisms, though in metazoans Orc1 is independent and its cycling is used as an additional mechanism to regulate initiation (*95*). ORC binds DNA throughout the cell cycle, but its status for Pre-RC assembly is regulated (*96-98*). Orc1-5 and Cdc6 belong to one family: they contain AAA+ motifs (including Walker A and B and Sensor 1 and 2 motifs) and winged helix domains that bind DNA and in some cases mediate protein-protein interactions (*99*). In addition, Orc1 harbors an N-terminal bromo-adjacent homology (BAH) domain implicated in nucleosome interactions, and Orc2 harbors an AT-hook for binding AT-rich sequences (*99-101*). Orc6 does not share these homologies, but it is highly conserved as a

component of ORC (*23*). ORC is purified as a single, stable complex out of yeast (see Protein Constructs and Preparations).

ORC and *ARS1*

To see how ORC behaves when confronted with a region of target DNA in the context of a vast excess of non-target DNA, we first generated a version of the complex that could be fluorescently labeled. We designed a sortase tagged Orc1 that could be conjugated to biotin and in turn labeled with a streptavidinated quantum dot (QD) (*102*). These ORC$^{sort\text{-}bio}$ and ORC$^{sort\text{-}bio\text{-}QD}$ constructs were tested for biochemical activity using an established Pre-RC assay that measures Mcm2-7 loading (**Fig. 2.3**).[4] I preincubated ORC$^{sort\text{-}bio}$ with a four-times molar

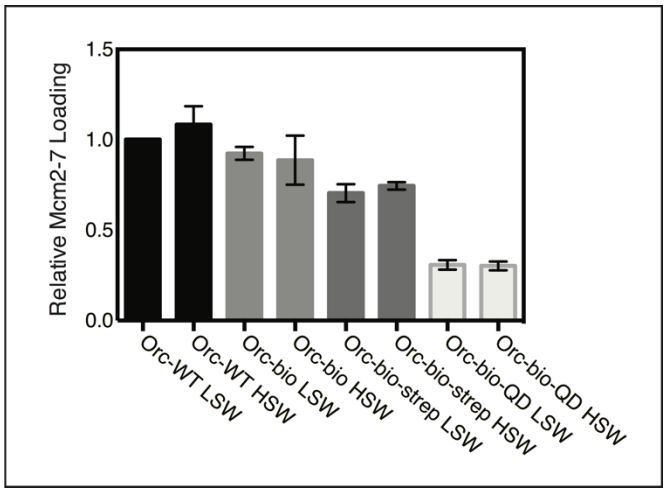

Figure 2.3 Tagged and Labeled ORC is Biochemically Functional
Sortase-mediated biotinylation of ORC does not affect its ability to load salt-stable Mcm2-7 as detected by western blot. Labeling the construct with streptavidin (strep) or quantum dot streptavidin conjugate (QD) reduces the efficiency of Mcm2-7 loading, but all loaded Mcm2-7 is salt stable under Low Salt Wash (LSW) and High Salt Wash (HSW), indicating that the labeling reduces efficiency but does not prevent correct loading. With Megan Warner. Reprinted from reference (*133*), with permission from Elsevier.

[4] QDs are exceptionally bright and photostable, but they are also relatively bulky at tens of nanometers in diameter. In some cases the extra bulk does not affect fundamental biological behavior, though it will affect biophysical parameters such as the diffusion constant and the rate of conformational changes in regions near the adduct. The choice of QD or organic dye was made on a case-by-case basis for all the experiments reported throughout this work with special attention to the advantages and disadvantages of each. In some cases, using QD labeling was eliminated outright because it fundamentally inhibited activity.

excess of QD streptavidin conjugate. ORC$^{sort\text{-}bio\text{-}QD}$ was then incubated with single-tethered λ_{ARS1} DNA curtains for two minutes and excess protein flushed out (**Fig. 2.4**). For these initial experiments, the DNA was labeled with the intercalating dye YOYO-1. Subsequent measurements of all protein activities did not include YOYO-1. The cloned *ARS1* occupies a position approximately two-thirds downstream of the leading barriers (**Figs. 2.3** and **2.4**), and ORC$^{sort\text{-}bio\text{-}QD}$ binding to *ARS1* is readily apparent. Note that many DNA molecules are unoccupied. At a slightly higher protein concentration (**Fig. 2.4B**) ORC binds non-*ARS1* sites, though the *ARS1* binding remains apparent (see below). To confirm the identity of the fluorescent signals, buffer flow could be transiently paused. Single-tethered curtains require continuous buffer flow for the curtain to remain within the TIRF field (see Chapter 1); when buffer flow stops, the DNA molecules take on an energetically preferred and significantly more compact conformation and diffuse out of the TIRF field toward the leading barriers. This manifests as the buffer-flow-

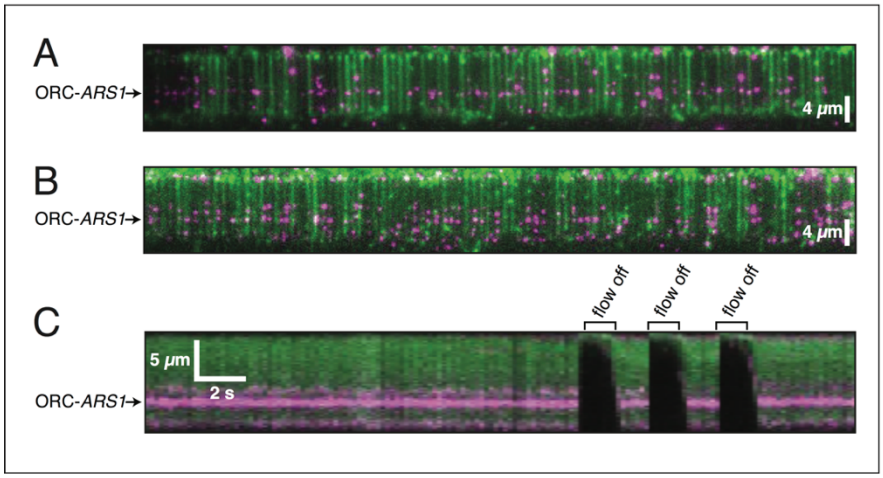

Figure 2.4 DNA Curtains and ORC Binding
(A) Wide-field image of a DNA curtain following incubation with 0.5nM ORC$^{sort\text{-}bio\text{-}QD}$ and (B) 1.0nM ORC. (C) Kymogram of a single-tethered DNA molecule with bound ORC, showing flow on/off events. Reprinted from reference (*133*), with permission from Elsevier.

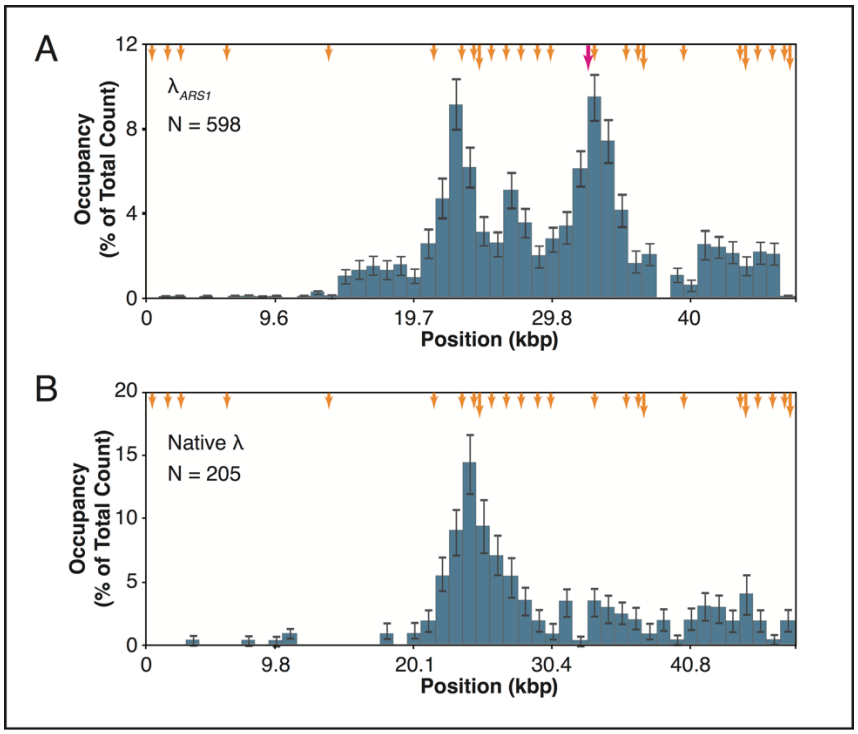

Figure 2.5 ORC Binding Distributions
(A) 1nM ORC$^{sort\text{-}bio\text{-}QD}$ binding distribution histogram on λ_{ARS1}. Binding positions were scored at a single time point following a 2 minute incubation and the removal of excess protein. Error bars indicate an 85% confidence interval based on 300 bootstrapping samples. N = number of ORC molecules. (B) 1nM ORC$^{sort\text{-}bio\text{-}QD}$ binding distribution histogram on native λ. The magenta arrow indicates the location of the cloned native *ARS1* sequence and the orange arrows indicate the locations of near-ACS sites native to λ. Reprinted from reference (*133*), with permission from Elsevier.

dependent loss of fluorescent signal from any fluorescently labeled DNA and any fluorescently labeled proteins bound to that DNA. **Figure 2.4C** shows a kymogram of one DNA molecule with several ORC$^{sort\text{-}bio\text{-}QD}$ molecules bound to it against time. When buffer flow is stopped, both the DNA and ORC$^{sort\text{-}bio\text{-}QD}$ signals are lost, confirming that ORC$^{sort\text{-}bio\text{-}QD}$ is bound to curtain DNA.

To quantify ORC binding preferences, I visualized ORC$^{sort\text{-}bio\text{-}QD}$ on a double-tethered λ_{ARS1} curtain as for the higher concentration case above. I then measured ORC$^{sort\text{-}bio\text{-}QD}$ binding positions as a function of pixel (px) distance form the leading barrier, converted to base pairs (**Fig. 2.5**). The spatial resolution here and throughout is 1px, or ~1kbp; additional error arises from variability in the

absolute position of each DNA molecule with respect to the barrier (see Appendix II for additional details). The position distribution histogram reveals a defined peak at *ARS1*, general binding across the AT-rich region, and a defined peak coincident with the particularly AT- and ACS-rich central region of λ_{ARS1}. The position distribution histogram on native λ clearly shows that the *ARS1*-coincident peak on λ_{ARS1} can confidently be assigned to the influence of the origin sequence itself.

ORC Binding Dynamics

All proteins involved in site-specific binding reactions must locate their binding targets through some mode of diffusion.[5] I assayed how ORC finds its target by observing binding events in real-time (**Fig. 2.6**). 67% of *ARS1*-binding events occur directly out of solution by 3-dimensional diffusion while 33% occur after some amount sliding by 1-dimensional diffusion after initial contact with DNA. Naturally, a

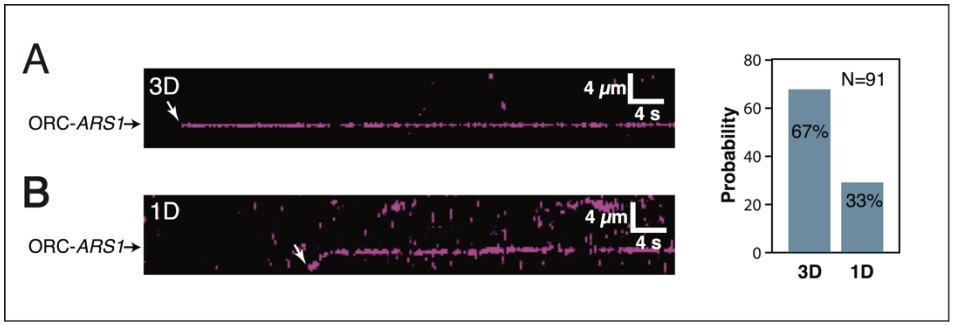

Figure 2.6 ORC Can Slide to its Target Site
Kymograms showing examples of (A) ORC binding to *ARS1* directly from solution and (B) 1D sliding of $ORC^{sort\text{-}bio\text{-}QD}$ along λ_{ARS1} to its target. At the temporal and spatial resolution reported here, most ORC molecules bind their target directly out of solution, but ORC is also capable of sliding along DNA. Reprinted from reference (*133*), with permission from Elsevier.

[5] The target search problem applies not only to proteins that search genomes for specific DNA sequences, but also to proteins that search for already-extant protein-nucleic acid complexes, although these higher order search processes have yet to be explored in-depth (see Chapters 3 and 4).

significant number of DNA-encounter events do not result in *ARS1*-binding. The resolution limits, temporally 200ms in this case, dictate that *at least* 33% of events occur by sliding under these conditions. It is unsurprising that ORC has the capacity to slide on DNA, as explored below, but any potential role for this sliding is unlikely in the *in vivo* context where protein obstacles on the DNA are common and protein density is much higher. Importantly, however, this result shows that ORC, on its own, can convert into a state stable enough to observe microscopically if it encounters *ARS1* or the ACS-like sequences in the AT-rich portion of λ_{ARS1}, and this is synonymous with the nature of ORC specificity.

An important aspect of dynamics is the stability of particular subpopulations of interactions. ORC is the first protein component of the Pre-RC, and subsequent steps were expected to play out as a function of its basal behavior, so I measured the stability of ORC binding. ORC molecules bound across all sites have a variety of lifetimes, as expected, with some very short lifetimes observed during incubation and some short lifetimes across non-ACS sites. Surprisingly and most interestingly, all stably-bound ORC molecules—those still present on the DNA ten minutes after excess protein is flushed out of the flowcell—show a very long lifetime of almost 37 minutes (**Fig. 2.7**). The lifetime data are best fit by a single exponential decay function, suggesting that with respect to stable DNA binding, all ORC molecules are equivalent. This will prove an interesting point in the next

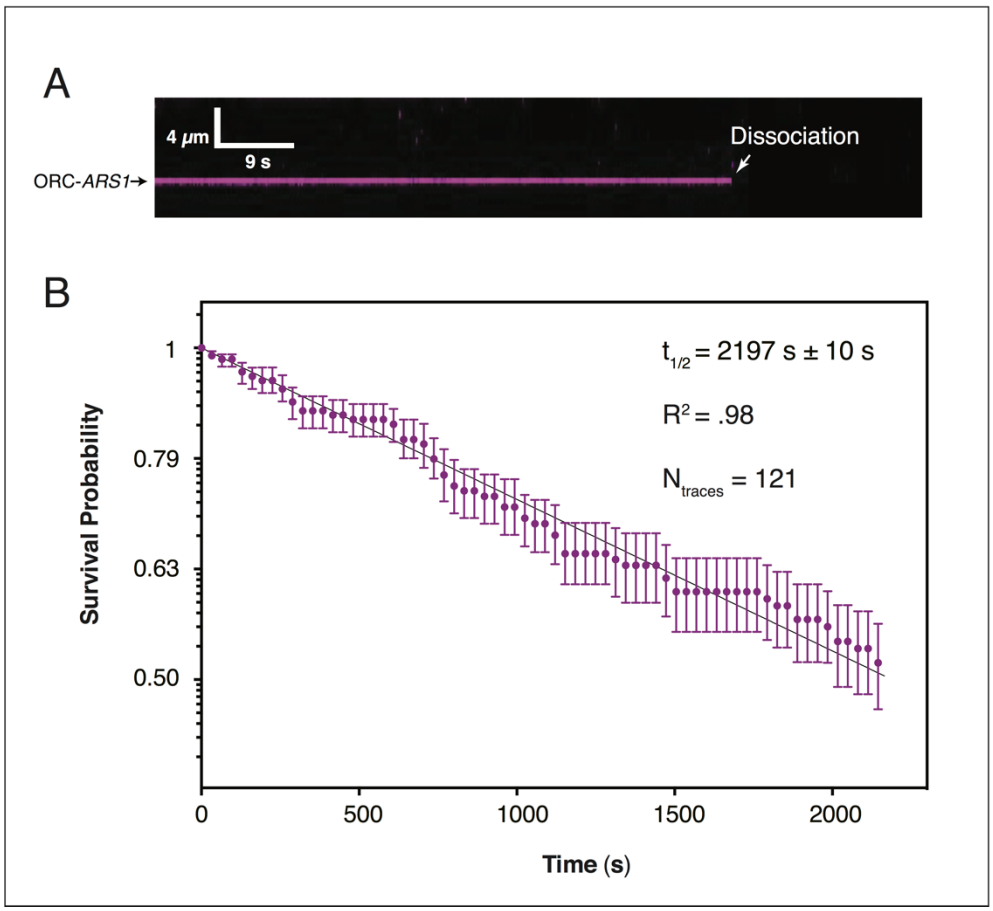

Figure 2.7 ORC Can Bind DNA Very Stably
(A) ORC$^{\text{sort-bio-QD}}$ bound to preferred sites—predominantly *ARS1* but also other AT-rich sites—can remain bound for many minutes. To measure this lifetime we tracked the length of ORC occupancy beginning 10 minutes after excess protein was flushed out. The arrow indicates spontaneous dissociation of ORC$^{\text{sort-bio-QD}}$ into solution. (B) Survival probability plot of stably bound ORC at all sites ($n_{ARS1\text{-ORC}}$ = 90; $n_{\text{non-}ARS1\text{-ORC}}$ = 31). 300 bootstrapping samples set to an 85% confidence interval were used to calculate the best-fit single exponential decay. The graph depicts every 15$^{\text{th}}$ bootstrapped data point for clarity. Reprinted from reference (*133*), with permission from Elsevier.

chapter where I demonstrate that with respect to downstream steps, these ORCs are not equivalent even though they bind DNA in the same stable mode.

Biochemical studies have shown that ORC is released from the DNA during Pre-RC formation. It seems that after Mcm2-7 double hexamer loading is complete, the ORC releases (or is displaced from) the DNA. Therefore, in co-

incubation experiments with all Pre-RC components present, ORC will not exhibit a long lifetime at *ARS1* if the other factors are saturating (*103*). In this context, the long lifetime measured here has two implications. First, ORC will still occupy non-*ARS1* sites with a long lifetime if double hexamers are not loaded there. Second, ORC does not cycle on origin sites during Pre-RC formation. That is, its molecular mode of operation is to bind an origin and wait until the other components arrive. This is in contrast to another possible mode where ORC binds the origin frequently but with a short lifetime, which would give the same biochemical signature of high occupancy (*104*). Single-molecule analysis can distinguish among these, and ORC's stable binding suggests that it has evolved to stay out of solution and maximize the probability that origin sites will be found ORC-occupied before S-phase begins.

Protein Concentrations, Salt, and ATP

The nature of ORC specificity is a point of frequent debate, and understanding ORC behavior requires an understanding of some parameters that influence the behavior of DNA-binding proteins.

The only distinction between **Figure 2.4A** and **Figure 2.4B** is the higher concentration of ORC protein used to generate the data in **B**. Both curtains were incubated with ORC for the same amount of time. There is greater *ARS1*-specificity in **A** (this will be quantified in Chapter 3), but fewer of the DNAs are

occupied. There is lower specificity in **B**, but more of the DNAs are occupied. Apparently, a higher concentration of ORC results in a higher probability that off-target sites will be bound in a given unit of time. It is also apparent that in the concentration and incubation time regime needed for ORC to occupy most of the available origin sites, many non-origin sites are also occupied. This type of behavior is sometimes taken as obvious, and sometimes ignored. In any case it is absolutely essential for understanding these and any results with ORC or similar factors.

When a DNA-binding protein is incubated with a DNA substrate, many proteins collide continuously with each DNA. The higher the protein concentration, the higher the collision rate. The vast majority of the collisions are very fast and not productive because the protein-DNA configuration is wrong at the moment of collision. Some collisions occur with the correct configuration and grant the protein momentary access to the characteristics of the underlying DNA—its topology or sequence—after which it can enter a pathway to either bind or disengage. These collisions can be fast too, but sometimes just stable enough to characterize at the single-molecule level (see Chapter 3). The capacity of DNA-binding proteins to probe DNA and sometimes bind it stably means that all DNA-binding proteins, by definition, can bind *any* DNA sequence with *some* affinity. What makes the target sequence stand out is that after the initial encounter its characteristics can funnel

the protein-DNA structure to an energetically preferred state wherein the protein is "bound."

The ORC collision frequency is assumed to be essentially uniform along the entire length of the 47.5kbp DNA substrate (at an ORC concentration of 1nM it should be on the order of ~300 sec^{-1} across one DNA in the curtain (*105*)). Some subset of these transient intermediates is converted to the stably bound state as a function of DNA sequence at the collision site. This same logic applies in other contexts, including a test tube with many more molecules than a single-molecule experiment, and the cell, with its layers of complications. The histogram distributions reported above reflect the intrinsic binding landscape of the ORC-DNA interaction across λ_{ARS1}. Each bin represents an approximately 1kbp bit of sequence space; each one presents a potential binding site. The probability of binding the sequence in any bin is necessarily non-zero. The histograms report on stably-bound ORC, so they show binding to the relatively favored regions. If ORC is incubated with the DNA for a longer time or at a higher concentration, then the probability of binding any bin increases. If ORC is incubated with the DNA for a shorter time or at a lower concentration, then the probability of binding any bin decreases. (Both these predictions are validated by experiments presented in Chapter 3.) *ARS1* is simply a region of the DNA substrate that represents a higher probability of being bound by ORC in a given unit of time. Note that if the substrate DNA is very short—on the order of the size of the sequence being tested—then

the presence or absence of the target sequence can cause huge changes in the apparent preference of the protein for the substrate. This is because the available sequence space becomes almost negligible, so using short DNA probes to understand protein specificity is very challenging at best (*104*).

The corollary to the above scheme is that any variable that changes the protein-DNA collision frequency also changes the overall DNA occupancy and specificity. This applies *in vitro* and *in vivo* because it is a basic feature of biomolecular interactions, but it is easier to measure *in vitro* because the interactions among other factors and the DNA can be ignored. Several research programs on Pre-RC assembly have reported that ORC specificity is inherently altered by competitor DNA and salt concentrations (*78, 106-108*). Adding competitor DNA is equivalent to lowering the ORC concentration because the competitor DNA simply soaks up ORC molecules. At lower ORC concentrations, the average binding specificity is much higher, as explained above. If the relatively few available ORCs are likely to bind anywhere on the experimental DNA, it's going to be on the patch with the highest probability of being bound, the patch that ORC prefers binding: the origin. This effect becomes even more apparent if the experimental DNA is relatively short and the competitor DNA long.

Ionic strength functions similarly. Many protein-DNA interactions are dominated by electrostatics, including initial encounters and stable binding (*104*). This explains why many DNA-binding proteins can slide on DNA at some ionic

strength: electrostatic interactions are made and broken along the highly charged polymer that is DNA. Ions stabilize protein and DNA structures in physiological states when the ions are present at physiological concentration ranges because biomolecules have evolved to function under physiological ionic conditions. This is important for the experimentalist to consider. Furthermore, at low ionic strengths additional electrostatic interactions are expected to form among all molecules. Some of these interactions may only “appear” when ions are rare, others may be physiologically relevant, though weak, but become stronger in the absence of ionic shielding. In the simplest terms, proteins and DNA become sticky at low ionic strengths and their interactions are generally disrupted at high ionic strengths. It is therefore entirely unsurprising that ORC specificity is perturbed at low salt concentrations: it sticks to the DNA! More precisely, collisions, encounters, and binding all become more frequent across the entire DNA. Binding stability is also expected to increase. All the experimental protocols developed and applied throughout this work use physiologically relevant ions and ionic concentrations (see Appendix II) (*109*) that are significantly higher than used for all other currently published work on budding yeast Pre-RC assembly, to the best of my knowledge. Furthermore, I characterize the nature of Pre-RC specificity with reference to the internal conditions of the single-molecule experiment. For example, in this chapter I identified conditions that yield curtains with most origin sites occupied. This is analogous, but not equivalent, to the *in vivo* situation where available strong

origins such as *ARS1* are constitutively occupied (*29, 98*). These conditions lead to ORC binding elsewhere, which means that high origin occupancy implies conditions that could lead to ORC binding elsewhere. I use these same conditions in other experiments that measure the specificity of other steps in the pathway. I can easily override the determinants of specificity by altering any parameter that changes the collision frequency, and this approach is applied to address specific experimental issues in Chapter 3 and in Chapter 6. My ability to easily override those determinants does not mean that they are biologically irrelevant, merely that I can design an *in vitro* experiment that exploits the way biological macromolecules behave to manipulate how the entire system behaves. Simply because I can generate ORC binding distributions that are highly non-specific, as I do for other experiments, does not mean that ORC itself is not a specific protein.

The exact conditions that will yield a given level of ORC specificity in the cell are not directly measureable *in vitro*, but the same concepts do apply. For example, reducing the number of available origins should increase the number of free ORC molecules in the cell, leading to lower-specificity binding overall and therefore ORC occupancy at non-origin sites. Just as in the experiments reported here, tuning ORC concentrations, even inside the cell, tunes specificity, and this is exactly what has been found (*88, 110*).

A final general point of contention in the field is the role of ATP in Pre-RC assembly. Extensive citations listed above identify and measure a variety of

ATPase activities for almost all Pre-RC steps and components, though there are frequent contradictions within the literature as to their importance. The slowly-hydrolyzable ATP analogue, ATPγS, has been used to capture hypothesized Pre-RC intermediates in Cryo-EM studies and Western blots (*38, 39, 44*). I have found that no step of Pre-RC assembly functions in any detectable way at the single-molecule level without ATP. In fact, leaving out the ATP-regeneration system prevents robust double hexamer loading (see Appendix II), suggesting that not only is ATP binding important, but hydrolysis and turnover are also crucial for the pathway. (Efficient Mcm2-7 loading occurs over 10-20 minutes under the experimental conditions used in this work, unlike efficient ORC binding, which requires only 2 minutes. In both cases, buffer flow is off during the binding steps, and fresh buffer does not replenish otherwise spent ATP.) If I leave out ATP or saturate the system with ATPγS, ORC does not bind DNA even if I increase the ORC concentration by two orders of magnitude relative to standard experimental conditions. In the presence of ATP, increasing the concentration by even a factor of two results in protein coating the DNA so extensively that the DNA curtains are compacted and ripped from their tethers (see for example **Fig. 3.1A**). Similarly, downstream steps fail if I bind ORC in the presence of ATP, wash away free ATP, and replace it with no nucleotide or ATPγS. The ATPase motifs of all Pre-RC components are highly conserved and many have been measured as active *in vitro* (*23, 99*). It would be surprising if their capacity to hydrolyze ATP were

dispensable for Pre-RC formation, and indeed such mutants are mostly inviable or grossly defective for growth (*50, 111, 112*). Pre-RC proteins, particularly Mcm2-7/Cdt1, are generally purified and stored in ATP because it stabilizes the complexes (see Appendix III). In bulk experiments that measure intermediates, free ATP is not washed away as it is in a flowcell. Residual ATP combined with ATPγS could yield the measured subpopulations. Regardless of the causes behind the many contradictions in the literature concerning ATP usage during Pre-RC assembly, the experiments reported here are internally consistent, generate functional double-hexamers that are competent for replication, and absolutely require a free pool of hydrolyzable ATP.

This chapter demonstrates that ORC can bind a single *ARS1* embedded within 47.5kbp of nonspecific DNA: ORC has an intrinsic capacity to recognize origin sequences, something, as we will see, that no other Pre-RC component can do. Consequently, the locations of Mcm2-7 double hexamer loading are inherently delimited by ORC binding distributions. The relationship between these two bookends of the Pre-RC assembly pathway—ORC binding and double hexamer loading—is explored in the next two chapters.

CHAPTER 3

Cdc6 and the Dynamics of Pre-RC Assembly

Cdc6 is an essential cofactor in the Pre-RC assembly pathway, a bridge between ORC and Mcm2-7 double hexamer formation. The availability of Cdc6 is cell-cycle regulated and ensures one round of replication during S-phase. In this chapter I explain the mechanism behind previously-identified Cdc6-mediated changes in ORC specificity and relate it to conclusions from the previous chapter. I also propose a new function for Cdc6 and offer that the *in vitro* specificity result is a secondary phenomenon. I conclude with the identification of two distinct subpopulations of Cdc6 that are defined by sensing whether individual ORCs are bound to *ARS1* or not.

Cdc6 is Central to Pre-RC Assembly

ORC is hypothesized to function by grabbing the open Mcm2-7/Cdt1 ring and locking it down on dsDNA. The RFC clamp loader works this way during replication: it is responsible for loading the PCNA processivity factor, also a ring-like structure like closed Mcm2-7, onto DNA (*113*). Despite a recent crystal structure of ORC (*114*), the mechanism it uses to load Mcm2-7 remains entirely unknown. But ORC cannot execute the loading reaction on its own: it requires Cdc6, which changes ORC's overall structure and DNA footprint (*35, 37*). These

changes are presumably correlated with some increased capacity to interact with Mcm2-7/Cdt1. Cdc6 also integrates cell-cycle information, expressed and synthesized so as to be available only during G1, with S-phase dependent phosphorylation targeting it for degradation (*115-119*). This is one mechanism alluded to in Chapter 1 by which the cell ensures that Pre-RCs assemble prior to S-phase but not during S-phase, which could result in re-replication. ORC's phosphorylation status also varies as a function of the cell cycle, similarly enabling it for Pre-RC function in M and G1 but disabling it in S (*96, 97*). But unlike Cdc6, it is not degraded, and actually remains DNA-bound throughout the cell cycle (*98*). This is probably because ORC has several other functions unrelated to Pre-RC assembly (*120-124*). This distinction between ORC and Cdc6—the one cycling in phosphorylation status, the other in phosphorylation status and existence—is key for the interpretation of Cdc6's single-molecule behavior. The dynamics of the ORC-Cdc6 interaction had not been studied before, and understanding its order and timing, and any heterogeneities, were the primary goals of the experiments related here.

Cdc6 Changes ORC Binding Distributions Without Actively Changing ORC's Inherent Specificity

In the previous chapter I introduced the concept that ratcheting up a DNA-binding protein's concentration increases the collision frequency between the protein and

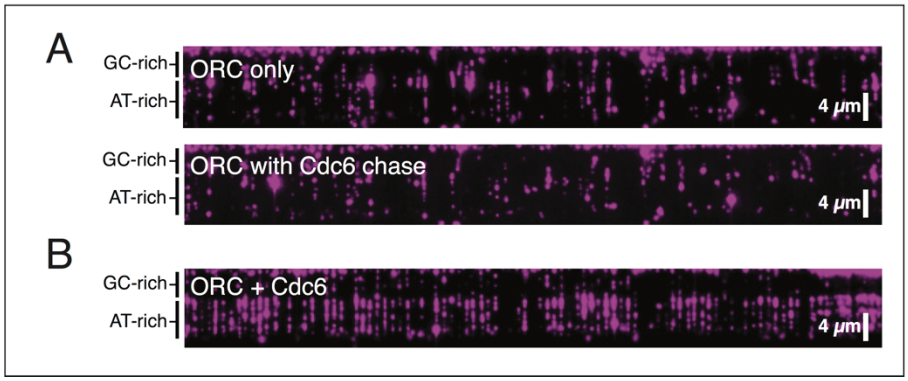

Figure 3.1 ORC Overloading and Cdc6
(A) Incubating 2nM ORC with a DNA curtain results in ORC aggregation across the DNA (top panel). Many of the DNA molecules have been ripped from their tethers by the compacting effect of excessive ORC binding. Subsequent addition of 4nM Cdc6 does not reverse ORC binding (bottom panel). (B) 2nM ORC in the presence of 8nM unlabeled Cdc6 binds across the AT-rich region of λ_{ARS1}. Reprinted from reference (*133*), with permission from Elsevier.

DNA, and this can lead to relatively high occupancy of all DNA sites. This applies to ORC as for any DNA-binding protein, and when even a slightly higher concentration of ORC is used in the curtain experiments the DNA becomes so heavily coated that the protein molecules form aggregates and compact the DNA (**Fig. 3.1A**, top panel). Although such overloaded curtains cannot be quantitatively analyzed, they are ideal for initial tests of how Cdc6 affects ORC binding to DNA. Several previous reports claim that Cdc6 actively directs ORC binding to correct sites (*e.g.*: *ARS1*) or actively removes ORC bound to incorrect sites (*35, 36, 61*). In those models, either the ORC-Cdc6 complex forms in solution and is more *ARS1*-specific than ORC on its own, or the ORC-Cdc6-non-*ARS1* complex has a low binding strength and readily falls off the DNA. Note that these models require either that ORC and Cdc6 interact when they bind DNA, or that ORC and Cdc6 interact when they are released from non-specific DNA. For convenience I will refer to these models as "active" in that Cdc6 actively

directs the specificity of ORC binding to DNA. As an initial follow-up to these claims, I chased an overloaded curtain with free excess unlabeled Cdc6 in the absence of any additional ORC (**Fig. 3.1A**, bottom panel). The aggregation remained entirely unaffected. Preliminarily, this suggests that Cdc6 cannot actively de-aggregate or remove ORC molecules that are incorrectly bound to the DNA. To test whether both factors have to be present for the claimed specificity effect, I coincubated the same relatively high concentration of ORC with excess Cdc6 (**Fig. 3.1B**). I expected either no change in ORC binding behavior and therefore another instance of protein aggregation and curtain ripping, or extremely high specificity with mostly *ARS1*-binding. Remarkably however, under these conditions λ_{ARS1} becomes heavily ORC-coated across the entire AT-rich half. (Recall that ARSs are AT-rich, and ORC shows a quantifiable AT-preference at lower concentrations [**Fig. 2.5**].) There is clearly less aggregation, but the ORC binding remains highly non-specific. *There appears to be an overall decrease in ORC binding, as though the ORC concentration were simply lower.* Recall that lower protein concentrations imply fewer protein-DNA collisions in a given unit of time.

Exploring this phenomenon quantitatively requires a lower ORC concentration regime (individual molecules cannot be distinguished in the **Figure 3.1** data, though the qualitative trend is obvious). I repeated the ORC binding experiment reported in **Figure 2.5A**, but in the presence of Cdc6: 1nM labeled ORC and 4nM Cdc6 were coincubated with a DNA curtain for two minutes and

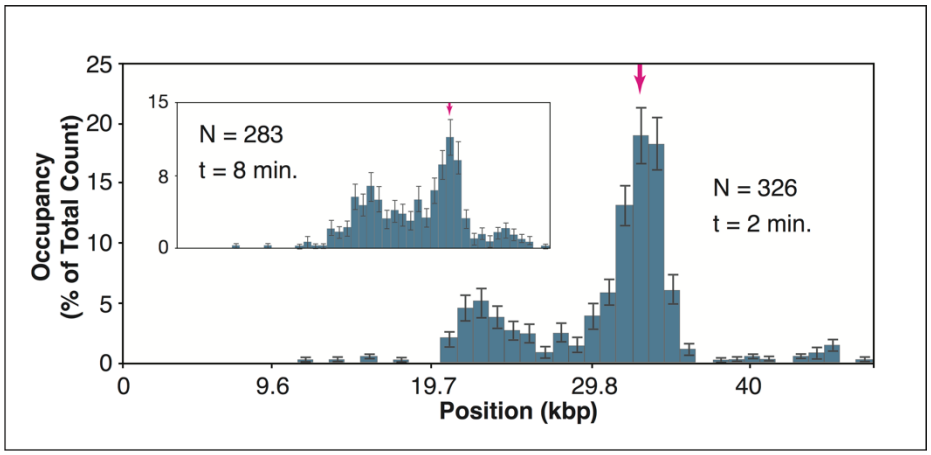

Figure 3.2 ORC Distributions in the Presence of Cdc6
1nM ORC binding distribution histogram on λ_{ARS1} in the presence of 4nM Cdc6. (Inset) in the presence of 40nM Cdc6 with a longer incubation time of 8 minutes. Reprinted from reference (*133*), with permission from Elsevier.

excess protein flushed out. The resulting ORC distribution is noticeably more *ARS1*-specific than the distribution of ORC on its own under otherwise equivalent conditions (**Fig. 3.2**; compare to **Fig. 2.5A**). There remains a clear sub-peak at the off-target central region, and it is interesting to note that if the DNA substrate were significantly shorter, as is common practice in bulk experiments, the difference between "specific" and "nonspecific" would be more pronounced simply because of the much smaller available sequence space. Very curiously, however, allowing ORC and Cdc6 to coincubate for longer almost completely recovers the ORC-only distribution (**Fig. 3.2** inset; compare to **Fig. 2.5A**). This too suggests that Cdc6 is not playing an active role: it is not systematically removing nonspecifically bound ORCs, nor is it preventing their binding to DNA. Again, its presence appears to be analogous to reducing the ORC concentration. A lower ORC concentration yields a more specific occupancy (see **Fig. 2.4A**). However, that specificity effect is lost if the incubation time is increased: a longer incubation time means a higher

cumulative collision frequency. This is exactly how ORC behaves in the presence of Cdc6.

The data presented in the inset to **Figure 3.2** does not rule out the possibility that Cdc6 does in fact actively direct ORC specificity but that the ORC collision frequency simply outstrips Cdc6's ability to carry out this function. To test this directly requires a single-molecule pulse-chase experiment wherein ORC is bound on its own first, followed by Cdc6 without additional ORC (a quantitative and non-aggregated-ORC equivalent of the experiment in **Figure 3.1A**). First I incubated 0.5nM ORC on a λ_{ARS1} double-tethered curtain and measured the resulting distribution after excess protein was flushed out (**Fig. 3.3**, blue histogram).[6] Then I allowed the distribution to decay for five

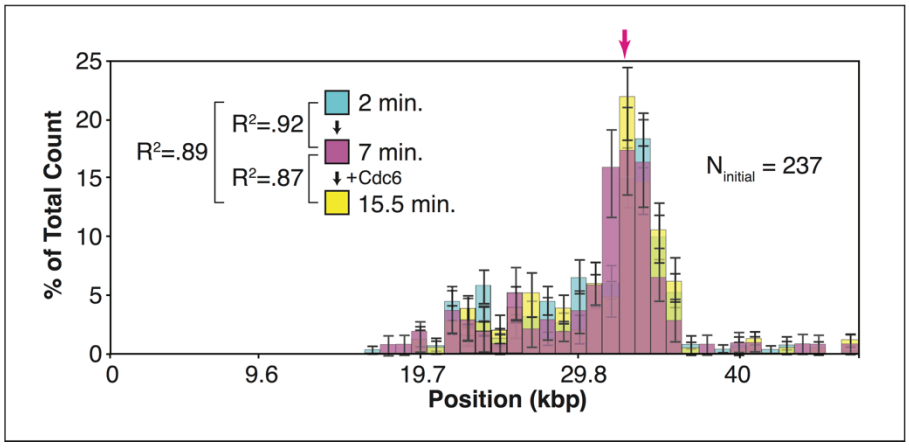

Figure 3.3 ORC Distribution Decay in the Presence of Cdc6
The initial 0.5nM ORC distribution (cyan) does not change 7 minutes after excess protein is flushed out (magenta), nor after an additional 7 minutes after Cdc6 is introduced (yellow). (The dead volume of the microfluidics requires 1.5 minutes for the controlled introduction of Cdc6.) Reprinted from reference (*133*), with permission from Elsevier.

[6] Note that this histogram quantitatively demonstrates the relationship between low ORC concentration and a relatively high specificity across a curtain, as is qualitatively apparent from **Figure 2.4A**. This experiment requires a distribution from the low concentration and therefore relatively specific ORC condition to ensure that only stable ORCs participate in the downstream measurement. Less stably bound ORCs at some nonspecific sites show a convolution of short lifetimes, as reported in Chapter 1, and would yield a false-positive result in measuring Cdc6's effects.

minutes but observed no change in the distribution (**Fig. 3.3**, purple histogram). This means that all stably-bound ORCs have the same, or very nearly the same, off-rate, and are therefore equivalent with respect to DNA binding stability (an independent validation of the measurement in **Figure 2.7**). Finally, I introduced free Cdc6 and allowed it to incubate on the ORC curtain for 7 minutes and observed no change in the distribution (**Fig. 3.3**, yellow histogram). Note that in all three distributions reported in **Figure 3.3**, the total proportion of nonspecifically bound ORCs is almost 50%. I conclude that Cdc6 is wholly incapable of actively removing ORC molecules bound nonspecifically. This predicts that ORC and Cdc6 do not interact when ORC falls off of the DNA, and I will test this prediction explicitly below.

These data all point away from any active mechanism for the ORC specificity effect observed here and in bulk. Instead, I propose that Cdc6 can interact with ORC in solution, and that this ORC-Cdc6 complex has a very highly reduced or completely inhibited capacity to bind any DNA (see below). This effectively reduces the ORC concentration in any experiment where Cdc6 and ORC are both present in solution and accounts for all the observed data. By inhibiting simultaneous ORC-Cdc6 binding to DNA, Cdc6 dictates that successful Pre-RC assembly happen in a defined order, with ORC binding first, followed by Cdc6, and then Mcm2-7/Cdt1 recruitment. This imposed order may be useful enough to have evolved independently, but additional dynamics among ORC,

Cdc6, and DNA uncovered in the following section make for an even stronger case.

Cdc6 Senses the Status of Underlying ORC

Given these intriguing results, I sought to better understand the relationship between ORC and Cdc6, and in particular their dynamics with respect to DNA. This required a labeled Cdc6 and we ultimately opted for the sortase tagging strategy, which had been so useful with ORC, but switched to fluorescent streptavidin (flSTA) rather than a bulkier QD. An established bulk biochemical Pre-RC assembly assay confirms that Cdc6$^{sort\text{-}bio}$ and Cdc6$^{sort\text{-}bio\text{-}flSTA}$ can load stable Mcm2-7s (**Fig. 3.4**). Cdc6$^{sort\text{-}bio\text{-}flSTA}$ does not interact with naked DNA stably enough to observe any fluorescent signal on a λ_{ARS1} double-tethered curtain. However, under conditions equivalent to those used to generate the inset to **Figure 3.2**, in the presence of

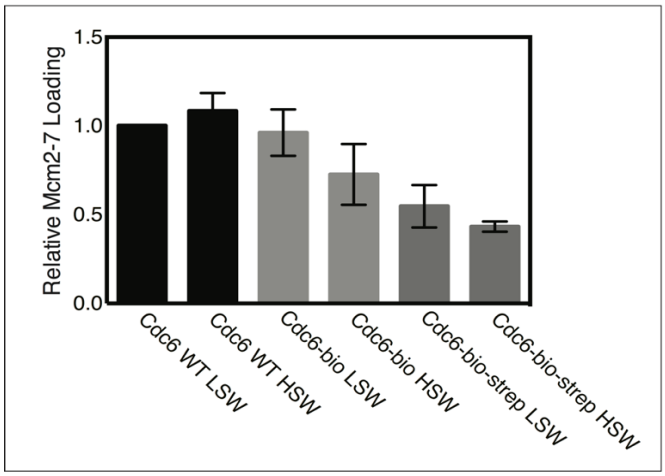

Figure 3.4 Tagged and Labeled Cdc6 is Biochemically Functional
Sortase-mediated biotinylation of Cdc6 does not affect its ability to load salt-stable Mcm2-7 as detected by western blot. Labeling the construct with streptavidin (strep) somewhat reduces the efficiency of Mcm2-7 loading, but loaded Mcm2-7 is salt stable under Low Salt Wash (LSW) and High Salt Wash (HSW), indicating that the labeling reduces efficiency but does not prevent correct loading. Experiments with error bars are in duplicate and error bars indicate 1σ. With Megan Warner. Reprinted from reference (*133*), with permission from Elsevier.

unlabeled ORC, Cdc6 interacts frequently and transiently with DNA (**Fig. 3.5A**). Data presented below confirms that the interaction is in fact between DNA-bound ORC and Cdc6, as expected. Cdc6 also exhibits no sliding behavior with all binding events happening directly out of solution. This suggests that Cdc6 recognizes the unique DNA-ORC complex, distinct from both ORC alone and DNA alone. I cannot rule out that elements of the DNA contribute directly to Cdc6 binding, but unlike ORC, Cdc6 is not a DNA-binding protein. Initial encounters are very probably between Cdc6 and DNA-ORC rather than Cdc6 and DNA.

The transience of the Cdc6 binding events is striking, and in contrast to stably-bound ORC molecules. Cdc6 samples ORC molecules bound to DNA; it is in rapid equilibrium between free and DNA-ORC-bound states. I measured the location distribution of these transient events—the locations on DNA where Cdc6 binding occurs—and found a pronounced *ARS1*-specificity (**Fig. 3.5B**). The ORC distribution under these conditions is significantly less *ARS1*-specific (**Fig. 3.2**, inset) and cannot account for the shape of the Cdc6 sampling location histogram. These data reveal that Cdc6 is likelier to convert into a state stable enough for observation within our resolution limits if the underlying ORC is at *ARS1*. Note that this histogram is not equivalent to a position-distribution histogram, such as have been shown for ORC, because it does not report on a stable binding distribution. Instead, it shows the cumulative set of sampled sites without reference to stability. To characterize these stability dynamics I measured the lifetime of the Cdc6

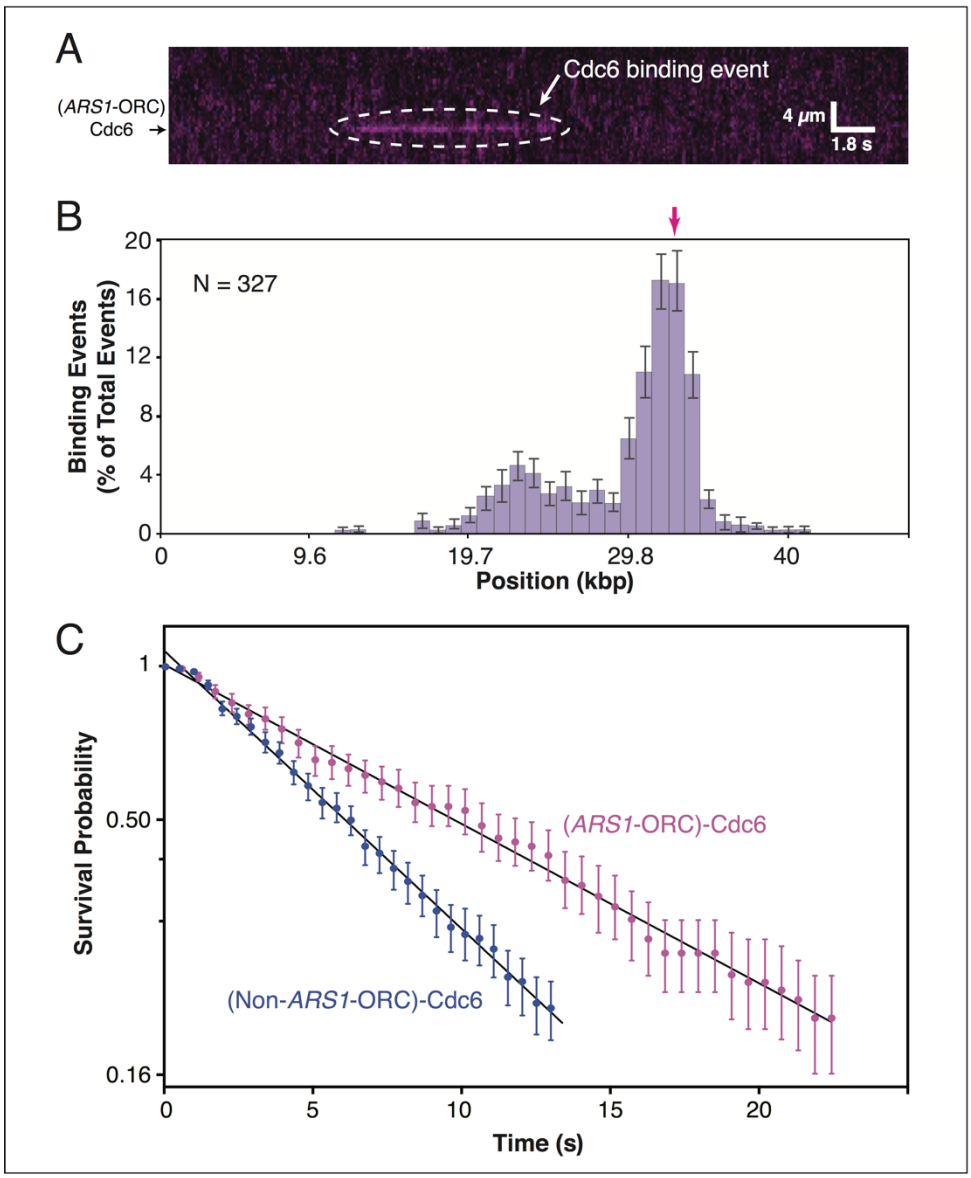

Figure 3.5 Cdc6 Binding Dynamics
(A) Cdc6 labeled with fluorescent streptavidin binds transiently to DNA-ORC (unlabeled). (B) Cdc6 interacts preferentially with *ARS1*-ORC. (C) The magenta survival probability plot of Cdc6 interacting with *ARS1*-ORC (every 5th 85% CI bootstrapped data point shown, for clarity), and the blue survival probability plot of Cdc6 interacting with non-*ARS1*-ORC (every 6th 85% CI bootstrapped data point shown) are fit with single exponential decay functions. The existence of two lifetimes was ascertained by a statistical F-test (F = 5024, P < .0001; see Appendix II). Reprinted from reference (*133*), with permission from Elsevier.

sampling events and discovered two distinct subpopulations (**Fig. 3.5C**; see Appendix II). (*ARS1*-ORC)-Cdc6 exhibits a $t_{1/2}$ of 9.6 ± 0.1s whereas (non-*ARS1*-

ORC)-Cdc6 exhibits a $t_{1/2}$ of 5.5 ± 0.1s. Not only is Cdc6 likelier to bind *ARS1*-ORC, as demonstrated by the sampling location distribution, but it spends more time interacting with an ORC molecule bound to *ARS1* than with an ORC molecule bound to a non-*ARS1* site. Using the data from **Figure 3.5B** and **Figure 3.5C**, I estimate that the ratio of the Cdc6 on-rate at *ARS1*-ORC to the on-rate at non-*ARS1*-ORC is ~130 (see Appendix II for details). I conclude that Cdc6 can very effectively sense the underlying status of the DNA-ORC complex as *ARS1*-bound or non-*ARS1*-bound.

Pre-RC Assembly is Strictly Ordered

The labeled Cdc6 construct enabled me to directly test two strong predictions introduced earlier. I have presented several lines of evidence that Cdc6 does not actively mediate ORC specificity, and that the specificity effect results from an effective reduction in ORC concentration when Cdc6 is present. If Cdc6 does not actively direct ORC specificity, then it should not be found to interact with ORC that is in the act of binding DNA or in the act of releasing DNA. Alternatively, we can consider the inverse prediction: several reports have claimed that Cdc6 actively increases ORC specificity, possibly by removing nonspecifically bound ORC from DNA. If this is correct, then Cdc6 must interact with ORC when it binds DNA or when it releases DNA. To test this directly I performed a two-color experiment and coincubated fluorescently labeled ORC and fluorescently labeled

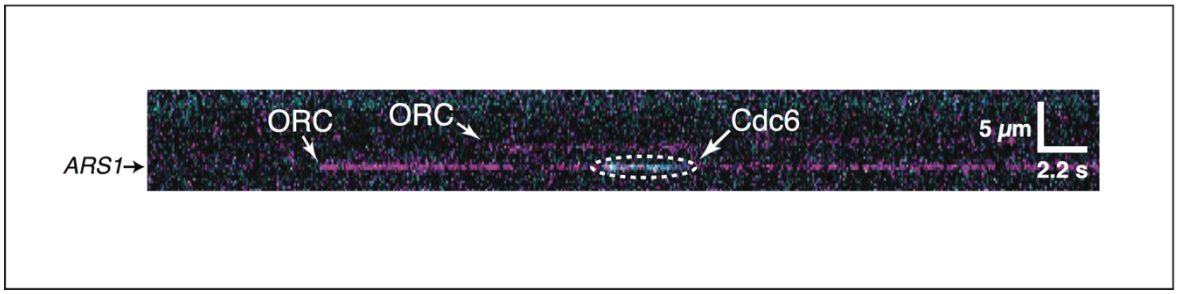

Figure 3.6 Cdc6 Binding Dynamics With Labeled ORC
Kymogram showing sequential binding of ORC (magenta) and Cdc6 (cyan) to DNA. Reprinted from reference (*133*), with permission from Elsevier.

Cdc6. ORC always binds DNA on its own first followed by Cdc6 sampling events, as observed for the single-color Cdc6 experiments above. Traces that capture ORC falling off show no interaction with Cdc6 at the moment of release. **Figure 3.6** shows a sample kymogram. ORC is depicted in magenta and can be seen bound to DNA for an extensive proportion of the trace before Cdc6 binds it transiently, falling off several seconds later. ORC never falls off concomitantly with Cdc6.

Cdc6 cannot dictate ORC structure if it is not bound to ORC. Note also that the implausible possibility that Cdc6 is somehow catalytic, changing the DNA-ORC interaction in such a way that the change persists after Cdc6 falls off, is eliminated by the lifetime data in **Figure 3.3**: Cdc6 cycling on DNA-ORC does not change the ORC dissociation rate at any site. Similarly, the possibility that Cdc6 changes ORC before ORC binds the DNA is eliminated by the inset to **Figure 3.2**: exposing ORCs to an excess of Cdc6 both prior to and during coincubation with DNA does not change the capacity of those ORCs to bind extensively across nonspecific

DNA. It is interesting to note in connection with this point the measurement of Cdc6 binding frequency to off-target ORCs (**Fig. 3.5B**), which is very decidedly low. Any model that invokes Cdc6 as a selective remover of nonspecifically bound ORC would predict high frequency binding to those nonspecific ORCs. These data thoroughly rule out any active role for Cdc6 in directing ORC specificity, and the two-color experiment in particular is definitive.

This conclusion raises two major questions: Why does coincubation increase the specificity of the ORC binding distribution, and what does this phenomenon reveal about the Pre-RC pathway? I propose a new model in which Cdc6 dictates the order of Pre-RC assembly (**Fig. 3.7**) and the consequent changes to the ORC binding distribution are secondary. The most straightforward hypothesis is that ORC and Cdc6 interact in solution, and that this ORC-Cdc6 complex has a reduced or inhibited DNA-binding affinity (**Fig. 3.7**, top schematic). It is presumably short-lived so as not to permanently sequester ORC, and ATP hydrolysis mediates release. Indeed, ORC and Cdc6 stimulate each other's ATPase activities in the absence of DNA while ATPγS allows their coimmunoprecipitation, which according to the same reports cannot be done if ATP is available (*36, 61*). (My data agrees with observations that ATP hydrolysis is required for all steps, including ORC and Cdc6 binding, and that inhibiting ATP hydrolysis at early steps prevents ORC binding to DNA and also sequesters it in an ORC-Cdc6 complex.) This explains why coincubating ORC with Cdc6 yields

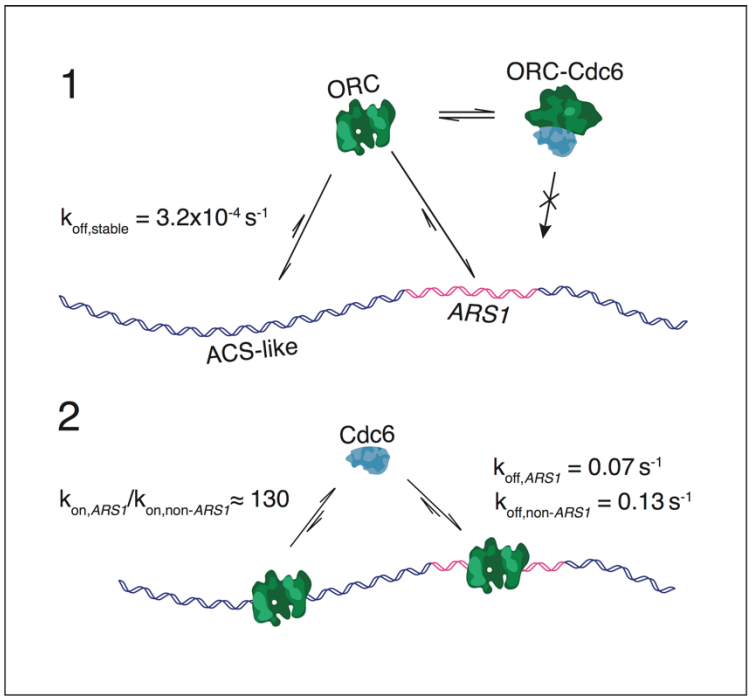

Figure 3.7 Cdc6 Dynamics and Pre-RC Assembly
(1) When ORC and Cdc6 interact in solution, the ORC-Cdc6 complex cannot bind DNA. This reduces the effective ORC concentration and ensures that ORC must bind DNA on its own, first. (2) Cdc6 differentiates among *ARS1*-ORC and non-*ARS1*-ORC. Reprinted from reference (*133*), with permission from Elsevier.

more specific ORC distributions: *the effective ORC concentration is lower when there is Cdc6 in solution.*

This would be a bizarre way to control ORC specificity, especially given that Cdc6 has significantly more effective mechanisms to identify ORC molecules bound to origins (**Fig. 3.5**). The DNA-ORC-Cdc6 interaction has evolved nuanced dynamics that regulate which ORC molecules are likely to be Cdc6-occupied and thereby likely to recruit Mcm2-7/Cdt1 (**Fig. 3.7**). *Cdc6 can tell which ORCs are in the correct place, so the ORC distribution does not need to be specific to yield specific Pre-RC assembly.* This prediction will be measured directly in the following chapter on double hexamer assembly. I will also find additional steps in Pre-RC site selection, which is unsurprising because ORC is primarily a DNA-binding protein, unlike Cdc6 and Mcm2-7/Cdt1, which, from the perspective of recruitment during Pre-RC assembly, are DNA-protein complex binders. Therefore, ORC has a capacity to bind other

DNA sequences and if Pre-RC assembly is to be regulated and specific then there must be some mechanism(s) to discriminate among those ORCs. On a physiological level, ORC has several other functions unrelated to replication and therefore unrelated to its interaction with Cdc6 (these include transcription regulation, heterochromatin maintenance, and possibly nucleosome remodeling) (*120-124*). By separating the DNA-binding step and the subsequent Cdc6-interacting step needed for Mcm2-7/Cdt1 recruitment, the Pre-RC pathway is made distinct from other pathways with respect to ORC and does not interfere with them. It would presumably be a problem for ORC's heterochromatin maintenance role, for example, if it were kicked off from regulatory non-origin sites every time Cdc6 levels spiked. It would also be a problem if all ORC molecules that had to bind non-origin DNA were already interacting with the very factor, Cdc6, that defines them as origin-bound for downstream steps.

An interesting property of this model is that Pre-RC assembly can only happen in one very well-defined order: ORC binds DNA, Cdc6 binds DNA-ORC, Mcm2-7/Cdt1 binds DNA-ORC-Cdc6. This is a consequence of the ORC-Cdc6 dynamics described above, and it may be a defined feature of the system. Complete Pre-RC assembly involves 21 polypeptides, two multiprotein complexes, and at least six, though probably more, on-pathway multi-component subcomplexes. And yet the end result is something very well-defined: a precisely loaded and oriented double hexamer. If this substrate is not loaded correctly then

the replication program is compromised. By dictating an order, each step can be highly controlled. The pathway has no ambiguities and therefore the minimal number of possible incorrect end-products or dead ends. Furthermore, all Pre-RC factors interact with one another on DNA and may have some capacity to interact with one another in solution. This certainly seems to be an issue for ORC and Cdc6, which is unsurprising given that Orc1-5 and Cdc6 belong to the same family of co-interacting proteins. Any mechanism that leverages cycling of one component needs to ensure that the one and the whole interact in the right place at the right time. (In humans, Orc1 cycles too, and it would be interesting to look for analogous phenomena in that system.) If ORC and Cdc6 remained bound in solution without DNA for any appreciable time, then the resources needed for replication would be wasted and the capacity of ORC to carry out its other roles would be compromised. This could also open the possibility for additional off-pathway dead ends. ATP turnover is a useful mechanism to kick apart complexes that have no function, freeing them for their proper roles. The cell needs to get double hexamer formation right, and ensuring that it happens the same way every time, despite the complexity involved, is an excellent form of quality control.

CHAPTER 4

Directed Assembly of Mcm2-7 Double Hexamers

The interactions among DNA, ORC, and Cdc6 ultimately function to correctly recruit and load Mcm2-7 double hexamers. ORC requires Cdc6 to achieve loading, but Cdc6 is not merely a passive co-factor: it actively dictates an order to Pre-RC assembly and samples DNA-bound ORC, differentiating between those at an origin site and those at other sequences. The interpretation of these results as depicted in **Figure 3.7** offers several predictions, foremost of which is that Mcm2-7 recruitment should be origin-specific. Here I test the predictions of the previous chapter and also demonstrate that double hexamer formation, a step distinct from recruitment, is also specific. I conclude with a summary of the dynamics that characterize the complete Pre-RC assembly pathway, and with double hexamers in place, set the stage for the action of replication.

Mcm2-7 is a Helicase Awaiting Activation

MCM genes were first identified in a genetic screen for minichromosome maintenance (*22*). Like the components of ORC, the proteins that make up the Mcm2-7 complex all belong to the wide-ranging AAA+ family of ATPases (*56, 125, 126*). They were promptly pegged as the likely constituents of the long sought after eukaryotic replicative helicase, since then confirmed as such by genetic,

biochemical, and now single-molecule evidence. That it should have turned out to be heterohexameric is curious: Mcm2-7 is the only known replicative helicase that is not a homomultimer. Having distinct subunits is not an essential feature of this type of helicase because the archaeal MCM is homohexameric (*25*). The divergence of the subunits in eukaryotes could be a consequence of the heavy regulation needed to integrate replication with the cell cycle and other processes. Many signals feed into Mcm2-7 behavior, from replication initiation through checkpoint activation to replication termination (*28, 127, 128*). Multiple non-equivalent subunits enable a greater variety of inputs into a macromolecule that is at the center of the replication pathway than would otherwise be possible, if each subunit were identical. Mcm2-7 is well-suited for this role given that almost every replisome component interacts with it at various stages of replication.

Despite the distinct subunits, some of which are more similar across species than to the other Mcm2-7 proteins (*23, 126*), each Mcm2-7 within a double hexamer loaded on dsDNA displays a 6-fold symmetry about an overall ring-like structure (*40, 42, 45, 125*). A gate between Mcm2 and Mcm5 is thought to open during loading and then close to topologically encircle the DNA (*43, 51, 56*). The double hexamer looks like four doughnuts stacked one on top of the other, each hexamer consisting of two: one outward-facing with six C-terminal domains and one inward-facing with six N-terminal domains. The hexamers are oriented in opposite directions along the DNA axis, with the C-terminal ends facing away from

each other and the N-terminal ends holding the double hexamer together. The numbered subunits of the sandwiched-together N-terminal ends are not initially oriented symmetrically with respect to the opposite hexamer (*45*). However in downstream steps the two hexamers rotate to almost 180° with respect to each other, possibly facilitating initial DNA melting (*47, 129*). The central pore of the double hexamer is thought to accommodate dsDNA, which should just about fit according to Cryo-EM structures. A series of hairpin loops and an overall positive charge are thought to mediate the protein-DNA interaction. The C-terminal domains carry ATPase motifs, which have poorly-defined regulatory roles and are essential for helicase activity, and zinc fingers throughout the N-terminal domains hold the double hexamer together until activation (*45, 47*). The conformational changes involved in the transition between DNA-ORC and DNA-ORC-Cdc6 potentiate the complex to recruit Mcm2-7/Cdt1 out of solution through a winged helix domain on the C-terminus of MCM3. These structural observations do not address how the identity of the underlying DNA affects the outcome of this pathway. The single-molecule DNA curtain assay is ideal for mapping the role of the sequence, as already demonstrated for ORC and Cdc6.

Interestingly, purified Mcm2-7 has no appreciable helicase activity *in vitro*. When the Mcm2-7 genes were cloned and the proteins individually purified, dimers of Mcm4,6,7 were found to exhibit helicase activity, and this led to a long line of research into whether there are two forms of the helicase (*126*). However,

significant evidence from multiple sources established that only Mcm2-7 is physiological and, in budding yeast at least, the association of Cdt1 with the complex seems dependent upon the presence of the correct heterohexameric stoichiometry. Cdt1 also harbors the necessary nuclear localization signal (*130*). Mcm2-7 activation requires both DDK phosphorylation, with its presumed consequent conformational changes, and association of a host of additional replisome factors (*28*). The crucial additional components from the perspective of Mcm2-7 are Cdc45 and GINS. Cdc45-Mcm2-7-GINS (CMG) is the active replicative helicase, and Mcm2-7 is at the core of the complex with Cdc45 and GINS bridging the Mcm2-Mcm5 gate (*46, 131*). The details of activation and the mechanism by which CMG pries apart DNA remain unknown, although structural work suggests that (i) conformational changes in the double hexamer melt the DNA inside the pore, possibly aided by some translocation activity by each helicase without decoupling, (ii) the Mcm2-Mcm5 gates open and ssDNA is extruded either out and around CMG or through a secondary pore formed by Cdc45 and GINS, (iii) and then full and processive helicase activity begins (*131, 132*).

Mcm2-7 is the only component of the replication pathway present from initiation to termination, the figurative core as well as the literal. It must be loaded correctly, in the right place and in the right way. Bidirectionality is especially important: if replication initiation were to occur unidirectionally at even a low rate it

would introduce the probability that stretches of DNA would remain unreplicated and S-phase would fail. Here I present the determinants of correct double hexamer formation and show that the dynamics of Pre-RC assembly, including the origin sequence at the heart of it, is key.

Mcm2-7 is Stably and Specifically Recruited to *ARS1*

To visualize Mcm2-7 recruitment at the single-molecule level we generated a SNAP-tagged MCM4 that was biotinylated to yield Mcm2-7$^{MCM4\text{-}SNAP\text{-}bio}$/Cdt1. This construct could subsequently be labeled with streptavidinated QD prior to microscopy. Replacing the endogenous *MCM4* gene with MCM4^{SNAP} yields viable cells, indicating that the SNAP domain does not inhibit normal helicase function *in vivo* (*103, 133*). We tested the bulk biochemical activity of this construct using a standard loading reaction in which ORC, Cdc6, and Mcm2-7$^{MCM4\text{-}SNAP\text{-}bio}$/Cdt1 were coincubated on a bead-coupled 7.4kb PCR product harboring a single *ARS1* origin (**Fig. 4.1A**; see Appendix II for details). This yields high-salt-wash resistant Mcm2-7s, historically interpreted as a measure of Mcm2-7 double hexamer loading (see below). Using the single-molecule approach, 10nM Mcm2-7$^{MCM4\text{-}SNAP\text{-}bio}$/Cdt1 pre-labeled with a four-times molar excess of streptavidinated QD and coincubated with 1nM ORC and 4nM Cdc6 on a λ_{ARS1} curtain for 15 minutes yields salt-stable loading of Mcm2-7 at *ARS1* (**Fig. 4.1B**). Here and throughout this work I have found that loaded Mcm2-7 is extremely stable and salt resistant. The retention of

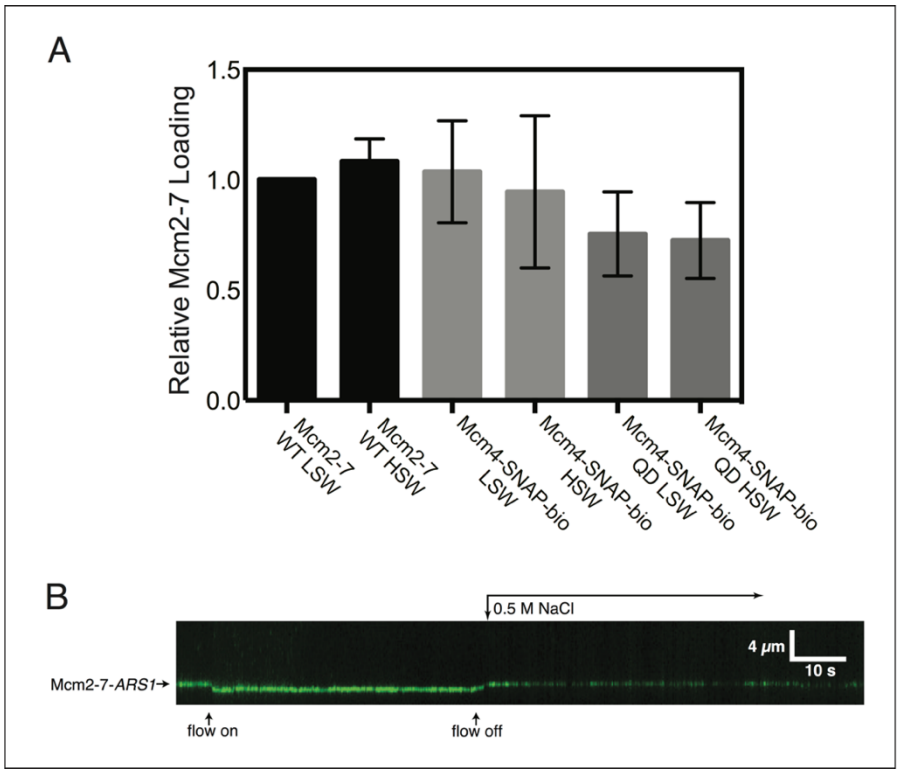

Figure 4.1 Tagged and QD-Labeled Mcm2-7 is Salt Stable
(A) SNAP-mediated biotinylation of Mcm2-7 does not affect its ability to load as a salt-stable complex as detected by western blot. Labeling the construct with quantum dot streptavidin conjugate (QD) also does not affect stability under Low Salt Wash (LSW) and High Salt Wash (HSW). Experiments with error bars are in duplicate and error bars indicate standard deviation of the mean. (B) Loaded Mcm2-7 in single molecule experiments is resistant to high salt wash. The kymogram shows continuous binding of Mcm2-7 as 0.5M NaCl is introduced into the flow chamber. The change in Mcm2-7 position is due to extension of the slack in the DNA molecule with buffer flow and the change is QD fluorescence is due to the change in the buffer composition. (A) with Megan Warner. Reprinted from reference (*133*), with permission from Elsevier.

Mcm2-7 on DNA when challenged with a high ionic strength buffer (typically 0.5M NaCl) has been used as an indirect measure of Mcm2-7 loading. Salt challenge has also been claimed to result in free double hexamer diffusion on DNA, which I do not observe. This minor technical point is useful for expounding the relationship between single-molecule and bulk biochemical results and is discussed separately at the end of this chapter. It is sufficient here to highlight that Mcm2-7 loaded in the single-molecule assay is salt-resistant and this agrees with the interpretation that

once Mcm2-7 is loaded by ORC and Cdc6 it is very stable on DNA (see Chapter 6).

Cdc6 dynamics predict that Mcm2-7 loading should be very *ARS1*-specific because Cdc6 is likelier to bind *ARS1*-bound ORC and once there exhibits a longer lifetime than on ORC bound at other sites. Therefore, free Mcm2-7 is likelier to encounter a DNA-ORC-Cdc6 complex at *ARS1* than any other site on λ_{ARS1} and should load preferentially at *ARS1* under the conditions tested for ORC and Cdc6 behavior in previous chapters. The Mcm2-7$^{\text{MCM4-SNAP-bio-sta-QD}}$ position distribution histogram in **Figure 4.2A** bears out this prediction. Despite the extensive binding of ORC across much of the DNA substrate under these incubation conditions, Mcm2-7 forms a prominent peak at *ARS1*. It is the most origin-specific localization of the Pre-RC pathway. Note that these experiments make no reference to the single or double hexamer status of loaded Mcm2-7, which is measured below.

Mcm2-7 loading in the single-molecule assay depends almost entirely on the presence of *ARS1* and entirely on the presence of ORC and Cdc6, as expected (**Fig. 4.2B**). This enabled a test of another prediction made by the interpretation of results presented in the previous two chapters: ORC has a relatively long lifetime of many minutes on favored sequences whereas Cdc6 has a relatively short lifetime of seconds on DNA-ORC. Therefore the DNA-ORC-Cdc6 complex should be absent very soon after Cdc6 is flushed out of a reaction and Mcm2-7 should consequently fail to load if it is introduced without additional Cdc6.

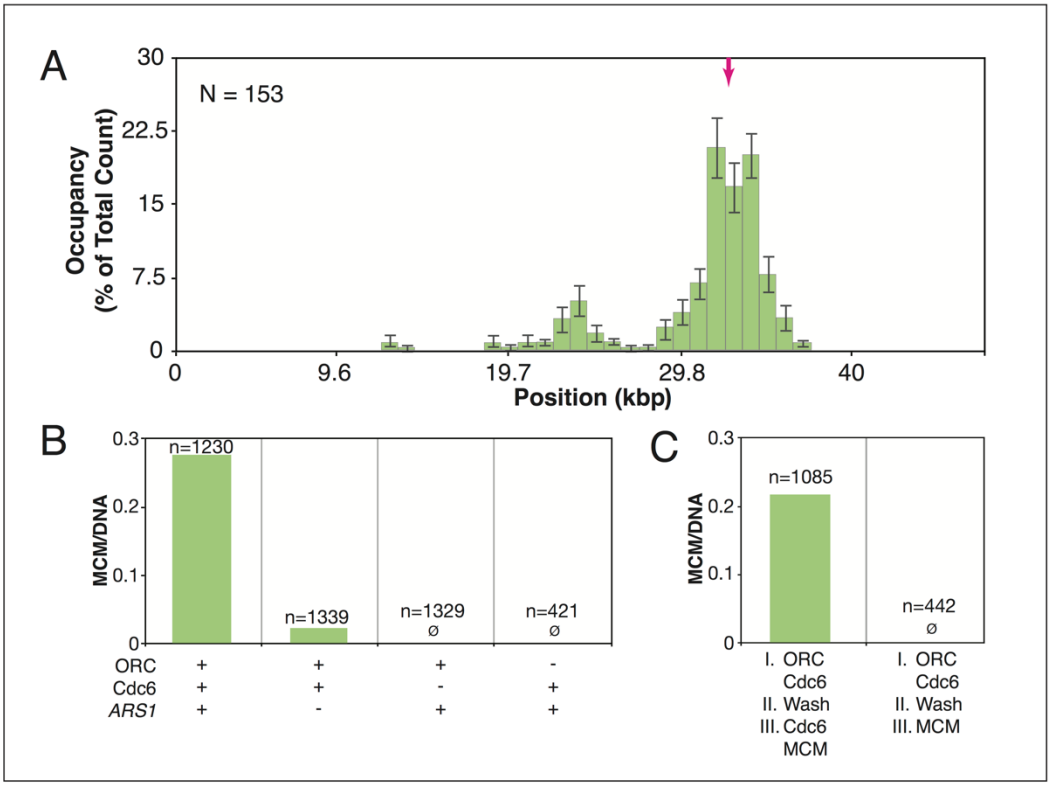

Figure 4.2 Cdc6 Dynamics Control Mcm2-7 Specificity
(A) Mcm2-7 binding distribution histogram on λ_{ARS1}. (B) Mcm2-7 loading efficiency under various conditions. Mcm2-7 molecules per DNA was quantified by labeling the DNA with the intercalating dye YOYO-1 after the completion of Mcm2-7 loading. n = number of DNA molecules. (C) Order-of-addition experiment showing that Cdc6 must be present in solution for Mcm2-7 loading. Reprinted from reference (*133*), with permission from Elsevier.

An order-of-addition experiment confirms this effect (**Fig. 4.2C**). First ORC and Cdc6 were incubated on a λ_{ARS1} curtain exactly as for experiments in Chapter 3. Then excess proteins were flushed out and chased with Cdc6 and Mcm2-7$^{MCM4\text{-}SNAP\text{-}bio\text{-}sta\text{-}QD}$, yielding robust Mcm2-7 loading (**Fig. 4.2C**, left panel). For the experimental case, ORC and Cdc6 were incubated on the curtain as above. Then excess proteins were flushed out and chased with only Mcm2-7$^{MCM4\text{-}SNAP\text{-}bio\text{-}sta\text{-}QD}$, and no free Cdc6. Under these conditions Cdc6 from the first step has been washed out of the reaction chamber, with even those initially bound to DNA-ORC having fallen off by the time Mcm2-7$^{MCM4\text{-}SNAP\text{-}bio\text{-}sta\text{-}QD}$ appears in the flowcell

several minutes later. Therefore, as predicted, Mcm2-7 does not load on DNA (**Fig. 4.2C**, right panel). I conclude that the transience of the DNA-ORC-Cdc6 complex requires the presence of Cdc6 in solution to recruit and load Mcm2-7. This also agrees with the *in vivo* state where ORC and Mcm2-7 protein levels remain fairly constant throughout G1 and S-phase whereas Cdc6 protein levels are regulated and fluctuate (*119, 134*). ORC and Mcm2-7 can be synthesized and prepared in advance of S-phase, even to excess, thereby serving as a primed switch for Pre-RC assembly, ready to be thrown by the introduction of Cdc6. Cdc6, being only one polypeptide, may be easier to synthesize than heteromultimeric complexes. And its capacity for turnover may enable one molecule to interact with many ORCs and Mcm2-7s in succession.

Measuring Mcm2-7 loading at this scale is also interesting from a technical perspective because of the unique complexity of the pathway for a single-molecule experiment. Observing multi-component systems is a new frontier in the field, especially for protein-DNA interactions (*8*). Mcm2-7 is comparable to Cdc6 in this respect because although it contacts DNA, it recognizes only a protein-DNA complex (namely, DNA-ORC-Cdc6). The Pre-RC assembly pathway manifestly incorporates a variety of mechanisms to ensure that each step is regulated, requiring that each step recognize the prior steps. This explains why Cdc6 does not bind DNA on its own and recognizes only DNA-ORC (**Fig. 3.5A**). Mcm2-7 is expected to exhibit a similar capacity to identify the DNA-ORC-Cdc6 substrate

during recruitment and fail to linger on naked DNA. This is demonstrated by the direct visualization of individual loading events: Mcm2-7$^{MCM4\text{-}SNAP\text{-}bio\text{-}sta\text{-}QD}$/Cdt1 does not interact with DNA on its own and all binding events occur directly out of solution onto DNA-ORC-Cdc6 **(Fig. 4.3**). Unlike ORC, both Cdc6 and Mcm2-7 recognize molecular species entirely distinct from just DNA and thereby funnel the pathway to its goal of loading double hexamers. Each successive step is well-defined and unambiguous. Interactions of any appreciable stability among DNA, Cdc6, and Mcm2-7 would merely waste resources and possibly introduce off-target subpopulations. Importantly, the major role of protein-protein contacts in recruitment does not preclude the possibility that the DNA plays a direct role in subsequent steps, especially given that Mcm2-7 comes to fully encircle dsDNA.

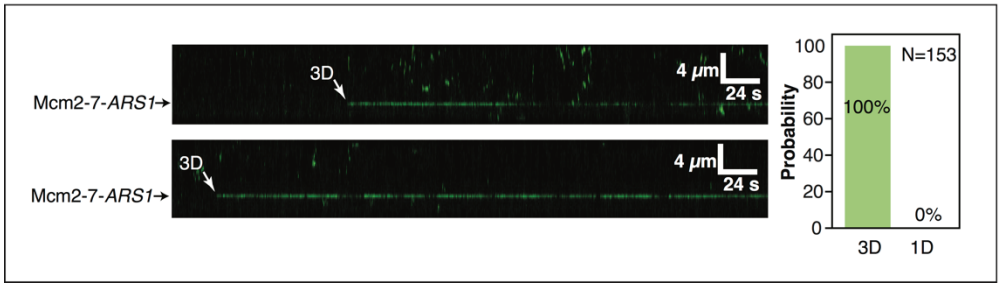

Figure 4.3 Mcm2-7/Cdt1 Binds DNA-ORC-Cdc6 Directly out of Solution
These kymograms show direct binding of Mcm2-7/Cdt1 to DNA-ORC-Cdc6. All Mcm2-7/Cdt1 molecules bind DNA-ORC-Cdc6 directly out of solution (right panel). Reprinted from reference (*133*), with permission from Elsevier.

Mcm2-7 Double Hexamers Form Preferentially at *ARS1*

The single-molecule Mcm2-7 reaction can also be used to illuminate the final step of Pre-RC assembly: double hexamer formation. Experiments using QD-labeled

Mcm2-7 do not show colocalization of multiple Mcm2-7s because the bulky QD sterically prevents two hexamers from loading. (It is interesting to note that what was in retrospect recognized to be single Mcm2-7 hexamers are as stable as the double hexamers that form using other constructs.) Furthermore, QDs are not necessarily ideal for quantifying the formation of higher-order structures due to labeling efficiency ambiguities and blinking. Instead we developed a sortase-tagged MCM4 and conjugated it to the fluorescent dye DY549, ultimately yielding Mcm2-7$^{MCM4\text{-}sort\text{-}DY549}$/Cdt1. The dye labeling efficiency here is ~90%. This and other dye-labeled Mcm2-7s show the same *ARS1*-specific behavior as the QD-labeled construct so that all results concerning specificity are internally consistent (see also Chapter 5 and Chapter 6). To quantify the number of Mcm2-7 hexamers loaded in each point of fluorescent resolution as a function of location on DNA, I first loaded Mcm2-7$^{MCM4\text{-}sort\text{-}DY549}$ on a λ_{ARS1} curtain as described above and flushed out excess proteins. I then intentionally photobleached the dye-labeled Mcm2-7 and characterized the number of photobleaching steps for each point (**Fig. 4.4**). I interpret one-step photobleaching as indicating the presence of a single hexamer and two-step photobleaching as indicating the presence of a double hexamer. Remarkably the relatively few Mcm2-7s loaded at non-*ARS1* sites are predominantly single hexamers whereas the Mcm2-7s loaded at *ARS1* are predominantly double hexamers (**Fig. 4.4A**). This strongly suggests that the sequence underlying Pre-RC assembly plays a role in defining specificity even at

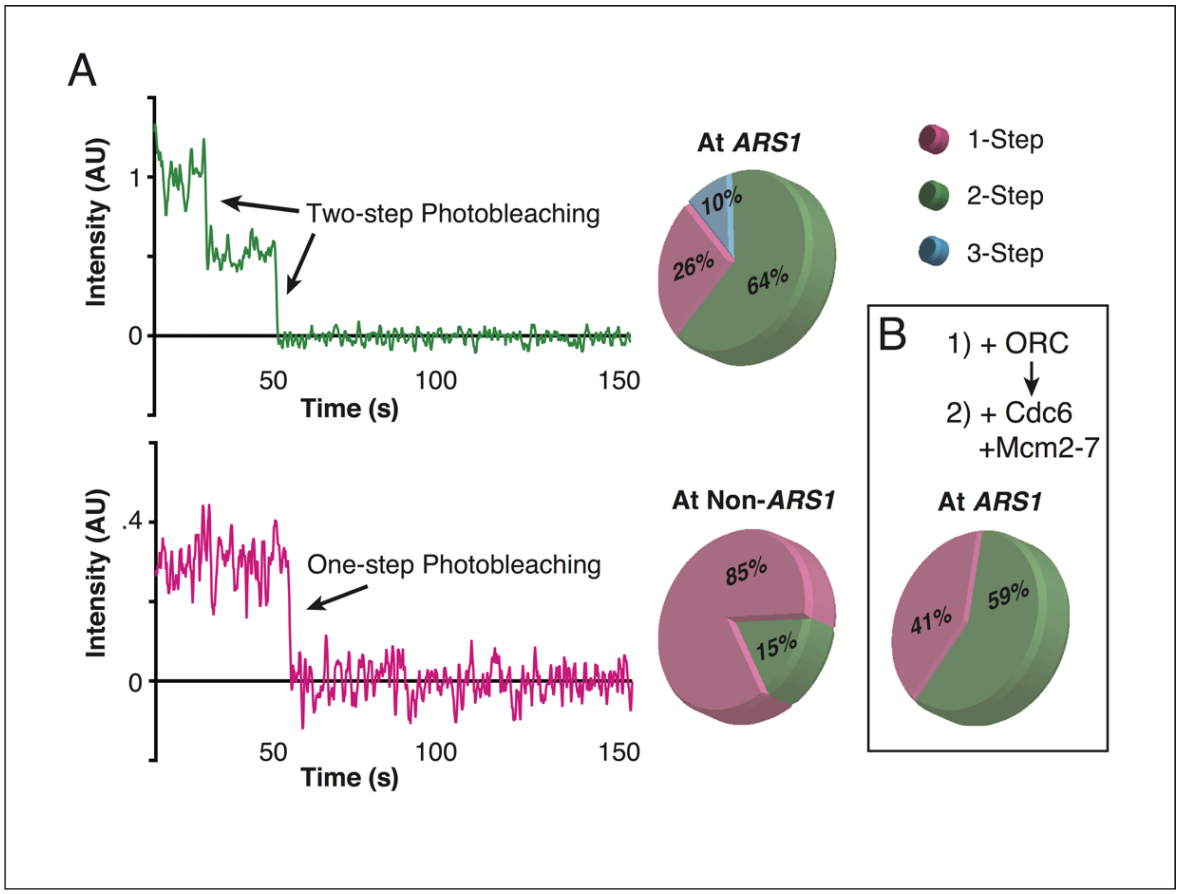

Figure 4.4 Double Hexamer Formation is Specific
(A) Examples of one-step and two-step photobleaching curves of *ARS1*-localized Mcm2-7$^{Mcm4\text{-}DY549}$. Pie charts show the proportions of 1-step, 2-step, and 3-step photobleaching at *ARS1* (top) and non-*ARS1* sites (bottom). (B) An order-of-addition experiment shows that ORC is not required in solution to load Mcm2-7 double hexamers. Reprinted from reference (*133*), with permission from Elsevier.

the final step of the reaction. This could be through the direct influence of particular *ARS1* sequences after Mcm2-7/Cdt1 recruitment or it could be the influence of *ARS1* on the DNA-ORC-Cdc6 structure. Interestingly, double hexamers can form at the ectopic sites, though very inefficiently. I interpret the small percentage of 3-step photobleaching traces as indicative of multiple, unrelated Mcm2-7 loadings in a single point of resolution (which may also explain some percentage of the two-step events at ectopic sites). There is no reiterative

Mcm2-7 loading here or under conditions with significantly higher protein concentrations or longer incubation times (see Chapter 6 and Appendix II) (*135*).

The transition from a single loaded Mcm2-7 hexamer to the recruitment and loading of a double hexamer is not well characterized. In particular, the role of ORC in loading the second Mcm2-7 remains unknown. I applied the photobleaching assay to test whether free ORC in solution is required to load the second Mcm2-7. This scenario would involve the recruitment of a second ORC following the loading of the first Mcm2-7, which is an attractive possibility given that the first ORC is expected to be physically separated from the site where the second Mcm2-7 hexamer is ultimately loaded; the first Mcm2-7 blocks the way (*38, 39*). I incubated the DNA with ORC alone, flushed out excess protein, and then introduced Cdc6 and Mcm2-7 (**Fig. 4.4B**). Under these conditions Mcm2-7 does not have access to free ORC, leaving only DNA-bound ORC to mediate double hexamer assembly.[7] Nonetheless, Mcm2-7s readily assemble into double hexamers despite the absence of free ORC. This demonstrates that DNA-bound ORC is sufficient to mediate recruitment and loading of the second Mcm2-7. The topological constraints of the reaction suggest that either a second ORC molecule is involved, though not recruited specifically after the first hexamer is loaded, or extensive DNA bending is required. Several studies suggest that one ORC is sufficient for double hexamer formation (*38, 39, 103*), though it remains entirely

[7] The concentration of free ORC in the reaction chamber if all DNA-bound molecules were to dissociate from the curtain would be on the order of 10^{-14}M and is therefore entirely negligible.

Figure 4.5

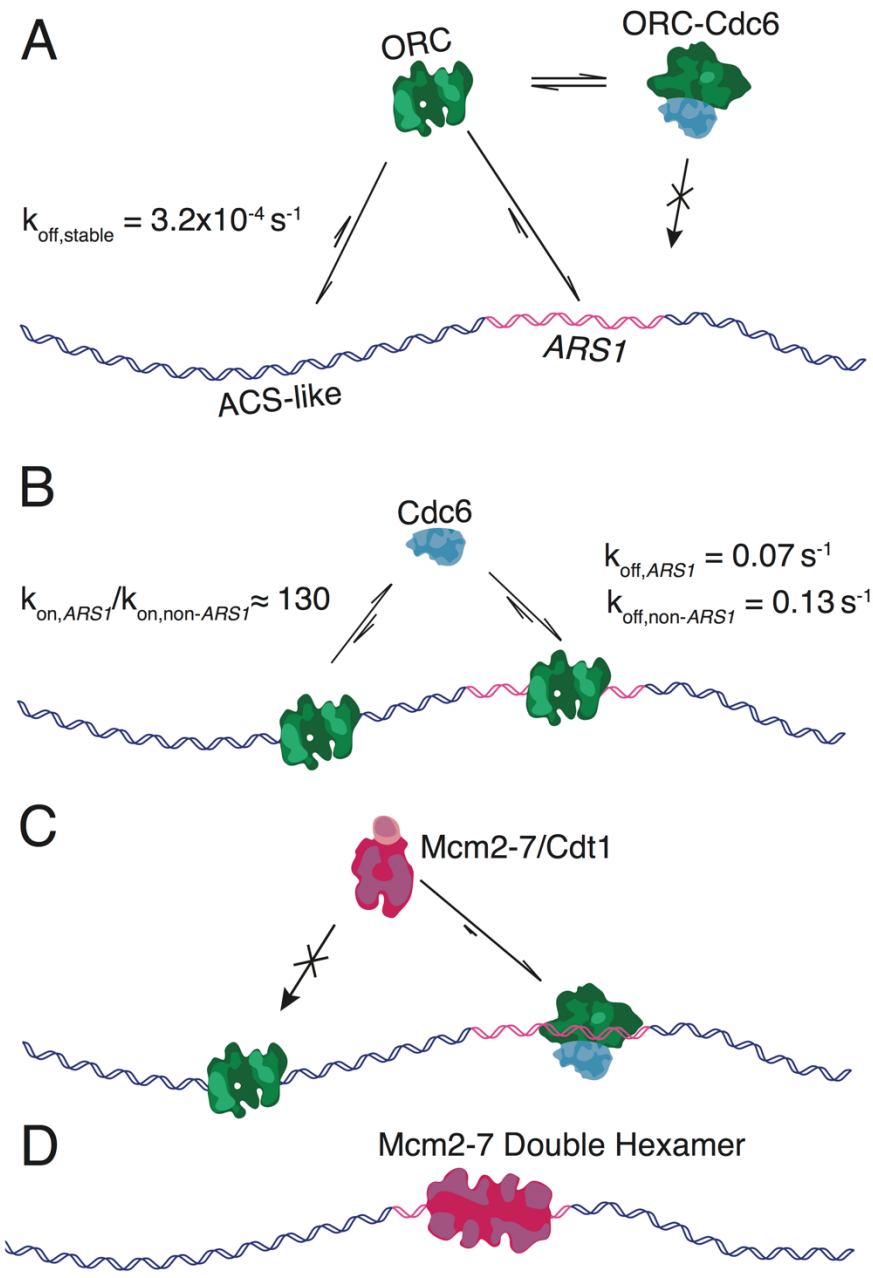
A
ORC
ORC-Cdc6
$k_{off,stable}$ = 3.2x10^{-4} s^{-1}
ARS1
ACS-like
B
Cdc6
$k_{on,ARS1}/k_{on,non\text{-}ARS1} \approx 130$
$k_{off,ARS1}$ = 0.07 s^{-1}
$k_{off,non\text{-}ARS1}$ = 0.13 s^{-1}
C
Mcm2-7/Cdt1
D
Mcm2-7 Double Hexamer

Figure 4.5 (Preceding Page) The Dynamics of Pre-RC Assembly
(A) ORC has a moderate to high preference for ACS-like sequences and a high preference for *ARS1*. (B) Cdc6 is likelier to bind *ARS1*-ORC than non-*ARS1*-ORC, and once there, bind for longer. (C) Mcm2-7 interacts only with the DNA-ORC-Cdc6 species. (D) Mcm2-7 double hexamers form preferentially at *ARS1*. For clarity, the figure depicts only the Mcm2-7 double hexamer. However, the data do not address the presence or absence of other Pre-RC components at this stage. Reprinted from reference (*133*), with permission from Elsevier.

unclear how such a mechanism functions. Further structural and single-molecule work will be needed to capture the elusive mechanism of double hexamer formation.

The Pre-RC Assembly Pathway

These analyses of Mcm2-7 double hexamer formation complete the single-molecule dissection of Pre-RC assembly (**Fig. 4.5**). Chapter 2 introduced the concentration-dependent nuances of ORC binding to DNA, the key to understanding why Cdc6 apparently increases ORC specificity in Chapter 3 (**Fig. 4.5A**). That same Cdc6 behavior dictates an order to Pre-RC assembly: ORC first, then Cdc6, and finally Mcm2-7. Chapter 3 also explored a strong example of how dynamics and heterogeneity can control a multistep pathway that requires precision: Cdc6 is found to discriminate among different subpopulations of DNA-bound ORC (**Fig. 4.5B**). This explains the central measurement of this chapter: Mcm2-7 loading is highly *ARS1*-specific (**Fig. 4.5C**). I also discovered another instance of molecular heterogeneity in the marked *ARS1*-preference of double hexamer formation (**Fig. 4.5D**). The Mcm2-7 double hexamer is the substrate for

replication: it remains inactive until S-phase, awaiting kinase signals and the multitude of replisome components and assembly factors required for the delicate molecular play of replication initiation. The following chapters explore these steps and go on to probe how active replisomes handle physiologically relevant obstacles on DNA.

Postscript

On the Free Diffusion of Mcm2-7 Along dsDNA

An array of studies cited throughout this work has demonstrated that Mcm2-7 double hexamers are ring-like and very probably encircle dsDNA. Additional experiments, including those presented here, are wholly consistent with that interpretation. Two simultaneous reports demarcate the moment when the research community accepted this hypothesis, and both correlate Cryo-EM-visualized double hexamers with topological stability on DNA in the face of high salt challenge (*40, 42*). Prior to that work, high salt challenge was used to differentiate between "bound" and "loaded" Mcm2-7. A typical experiment along these lines involves co-incubation of Pre-RC components on bead-coupled DNA with an origin sequence, followed by several washes in reaction buffer, and then high salt challenge. What remains on the beads is usually assayed by Western blot against various Pre-RC components. A significant amount of Mcm2-7 is usually washed away by the high salt, and variables such as the presence or

absence of ATP will yield differential results in the face of salt challenge (*60, 135, 136*). To date no published work has demonstrated what the "bound" subpopulation could be, let alone its potential biological significance. The salt stable subpopulation was ultimately re-interpreted as loaded Mcm2-7 that cannot be washed away because the Mcm2-7s encircle the DNA, the hexamers being closed rings around the rope of DNA. However, those same studies introduced a new conclusion, which I would rather categorize as a hypothesis, that Mcm2-7 can freely diffuse along dsDNA, especially in high salt. All the single molecule data presented in this work disagrees with that conclusion.

The bulk experiments took two forms. In the first case Mcm2-7 was loaded on closed circular DNA or linear DNA and found to fall off the linear DNA when challenged with salt (*42*). In the second case Mcm2-7 was loaded on linear DNA with free ends or blocked ends and found to fall off the DNA with free ends when challenged with salt (*40*). The interpretation in both cases was that Mcm2-7 has the capacity to freely diffuse on dsDNA. This has subsequently been used as the basis of a variety of arguments concerning the redistribution of Mcm2-7 after loading (see Chapter 6). Several major and minor variables were not addressed in these experiments.

Almost every protein and protein complex we have studied in the Greene laboratory forms aggregates under low salt and high protein concentration conditions. This is generally true, but we happen to have direct and immediate

access to recognizing aggregation events because they are very apparent under the microscope as large bright fluorescent globs on DNA that photobleach as a clean single exponential curve against laser exposure (time). The Pre-RC protein concentrations used in the cited studies immediately form massive aggregates on DNA curtains, even when competitor DNA is present to a concentration equivalent to the substrate DNA used in the bulk studies. Similarly, Pre-RC components aggregate on the DNA at the low salt concentrations used in the bulk studies at even low protein concentrations. All experiments reported in this present work use physiological salt concentrations and all aspects of Pre-RC behavior are effectively short-circuited at lower salt concentrations because of aggregates. Aggregates of many varieties readily diffuse on DNA. (We do not report these phenomena as they are physiologically irrelevant, and we subsequently work to develop assay conditions that yield discrete binding events.) This is entirely unsurprising given that individual proteins make many electrostatic contacts with DNA; a large irregular sphere of those proteins would interact with the DNA more extensively than just one protein and the electrostatic contacts would be more stable overall in high salt. Furthermore, it is unsurprising that even an aggregate would retain some sequence specificity. The DNA-aggregated protein contacts would have the same individual identities as in the DNA-single protein case, and should exhibit at least some specificity. Therefore, the inferred behavior may remain, say, *ARS1*-specific, and yet no further meaning should be deduced from the result.

The use of beads is also problematic. Both magnetic and glass beads coupled to DNA are extremely difficult to passivate, and they exhibit their own electrostatic properties with respect to DNA-binding proteins. Salt concentrations, protein concentrations, bead concentrations, reaction volumes, and incubation times all influence what binds to the beads and how stably. In at least some of the published experiments, the beads are clearly acting as a sink for Cdc6 because despite the absence of Cdc6 during the Mcm2-7 loading step, Mcm2-7 is loaded onto DNA (*40*). (This is the low-salt, high-protein-concentration, and bead-coupled-DNA bulk equivalent of the order-of-addition experiment presented in **Figure 4.1C** which shows an absolute requirement for Cdc6 in solution because of its relatively short lifetime on DNA-ORC, as characterized by several experiments in Chapter 3.)

There are several additional, though more subtle, issues with using circular and linear but blocked DNA substrates to infer Mcm2-7 diffusion. Circular DNA is not equivalent to linear DNA in its basic biophysical properties. All Pre-RC components could very well behave differently when DNA is, for example, slightly supercoiled (*71*) or more locally concentrated. Comparing salt-resistance on circular DNA to salt resistance on linear DNA and then inferring that in the one case a ring-like structure can slide off the end is a significant leap. For one, it implies that Mcm2-7 can simply fall off of a free end, which is not at all a trivial assumption. In addition, ORC readily binds free DNA ends, even when they are

blocked with streptavidin (present study, unpublished results). This is almost certainly not physiological behavior, but it has never been controlled for in bulk (or recognized, to my knowledge). End-binding ORC could, for example, preferentially misload Mcm2-7s that readily fall off the DNA. This again does not imply free diffusion. Most crucially, no bulk study has ever definitively identified the distinction among the several subpopulations of Mcm2-7s involved in these studies (there are almost certainly more than the two typically assumed). In bulk, we do not know the identity of Mcm2-7s that are salt resistant, where or how they ended up on the DNA, and how they differ from the non-salt resistant ones.

This present work directly measures the stability of loaded Mcm2-7s under high salt challenge and also long incubation times in physiological salt. In all cases the Mcm2-7s are highly stable and do not freely diffuse. Single Mcm2-7 hexamers loaded at *ARS1* by ORC and Cdc6 do not diffuse when challenged with salt (**Fig. 4.1B**). Three distinct measurements of replication initiation with orthogonal fluorescent readouts show that Mcm2-7s do not freely diffuse over the course of several hours across several buffer conditions, all with significantly higher ionic strengths than routinely applied in bulk (Chapter 5, Chapter 6, and Appendix II). The DNA curtain configuration is not responsible for preventing Mcm2-7 diffusion because double hexamers can be loaded on λ_{ARS1} prior to curtain assembly and still yield an *ARS1*-specific distribution. These single-molecule measurements are the first to directly test the diffusive behavior of loaded Mcm2-7 and they

conclusively show that Mcm2-7 does not readily transition into a freely diffusive state even when challenged with high salt. I cannot rule out that Mcm2-7 is somewhat diffusive under certain conditions, though the diffusive subpopulation would have to be very small or exhibit a very low diffusion constant; furthermore, if such a subpopulation exists, it is not essential to the replication reaction, as demonstrated by the results reported here. Note also that just because a protein can be pushed along DNA does not mean it can diffuse along DNA (see Chapter 6). Inferring Mcm2-7 behavior or structure using salt challenge in bulk is fraught and the field has advanced enough to retire it as a measure of anything meaningful.

CHAPTER 5

Replication Initiation

Mcm2-7 double hexamers are scaffolds for replication initiation; they are the link between the Pre-RC assembly of G1-phase and the replisome assembly of S-phase. The previous chapters explored the single-molecule dynamics that lead to double hexamer formation and in the process established an experimental protocol for generating DNA curtains primed for replication. Here I introduce an assay for the direct visualization of individual replisomes and measure the determinants of firing.

The Replication Reaction

Two cell-cycle-regulated kinases are the gatekeepers of S-phase and the duplication of the genome. Dbf4-dependent kinase (DDK) phosphorylates multiple subunits on Mcm2-7 and the S-phase cyclin-dependent kinase (S-CDK) has an array of targets; these include ORC and Cdc6, which it inactivates for their Pre-RC assembly capacity, and Sld2 and Sld3, which recruit a variety of replisome factors (*28, 137*) (see also Appendix I). Each replisome ultimately consists of many protein complexes (*46, 47, 138*), including, at its core, the CMG (*139*); the Polα-primase complex needed to initiate leading strand replication and initiate each okazaki fragment; the leading strand polymerase Polε; the lagging strand

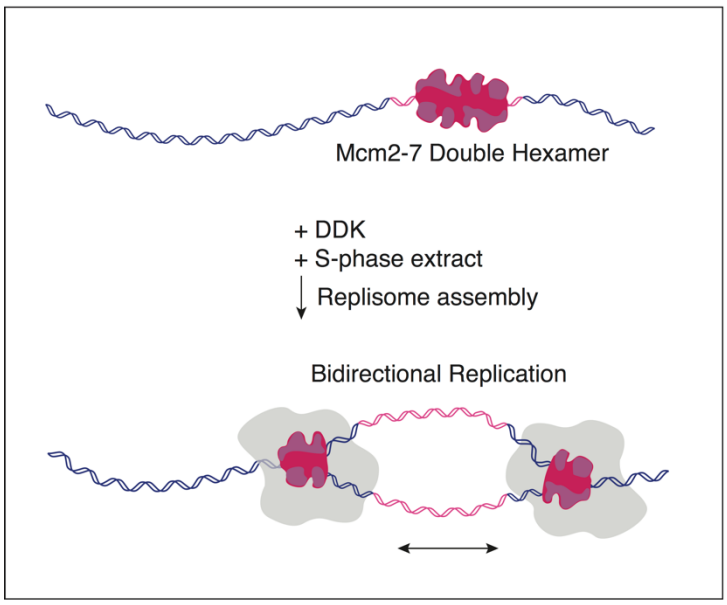

Figure 5.1 An Assay for Initiating Replication
Mcm2-7 double hexamers can be activated for replisome assembly and replication initiation by the addition of DDK kinase and an S-phase whole cell extract. Reprinted from reference (*133*), with permission from Elsevier.

polymerase Polδ (*47, 140*); the RFC clamp loader which loads the polymerase processivity factor PCNA (*113*); the Ctf4 trimer complex which holds together several replisome components (not all of them identified, though known to include Polα) (*141, 142*); and the eukaryotic single-stranded binding protein RPA (*143*). Additional factors are also present, and more are likely to be found. It is challenging to determine the complete ensemble partly because the machinery is inherently complex and partly because many factors probably only interact with the replisome at specific times or under specific conditions, transiently.

The purification of all the core replisome factors has only been achieved very recently (*47, 108*). For this present work I used an S-phase extract from cells that overexpress S-CDK, Cdc45, Sld2, and Sld3 (*28, 106, 107*). The replication assay (**Fig. 5.1**) begins where Pre-RC assembly ends: with Mcm2-7 double hexamers. These are treated with purified DDK and then with S-phase extract to

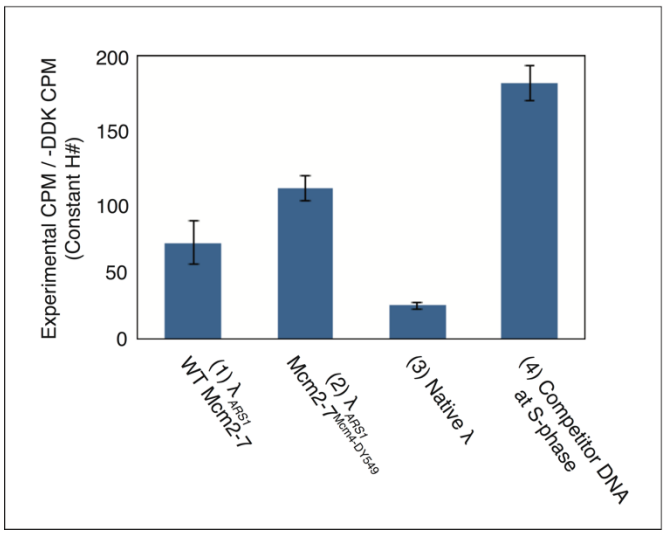

Figure 5.2 Mcm2-7$^{MCM4\text{-}sort\text{-}DY549}$ and λ_{ARS1} Are Functional For Bulk Replication
The single-molecule experimental conditions (1) and Mcm2-7$^{MCM4\text{-}sort\text{-}DY549}$ and λ_{ARS1} (2) support origin-dependent replication (3). Furthermore, competitor DNA (4) enhances the reaction by sequestering DNA-binding proteins. All experiments are in triplicate, and each experimental run was normalized to a -DDK control. Errors bars indicate 1σ. Reprinted from reference (*133*), with permission from Elsevier.

initiate semiconservative replication (*28*) (see appendices I-III). I tested my experimental conditions and substrates for replication competence with a radioactive nucleotide incorporation assay. I found that λ_{ARS1} and Mcm2-7$^{MCM4\text{-}sort\text{-}DY549}$ support DDK-dependent and largely origin-dependent DNA replication in bulk (**Fig. 5.2**). Competitor DNA increases replication efficiency by sequestering DNA-binding proteins from the substrate DNA. The amount of competitor DNA used here was determined by gauging the minimum concentration needed in single-molecule experiments to prevent DNA curtain compaction in the presence of S-phase extract. To translate this assay to the single molecule level, Mcm2-7$^{MCM4\text{-}sort\text{-}DY549}$ double hexamers were assembled on a double tethered λ_{ARS1} DNA curtain as in the previous chapter, treated with DDK, and then with S-phase extract supplemented with additional DDK (see below). Bidirectional replication should appear as the divergence of the dye-labeled Mcm2-7s (**Fig. 5.3A**). At the S-phase extract step I can observe individual

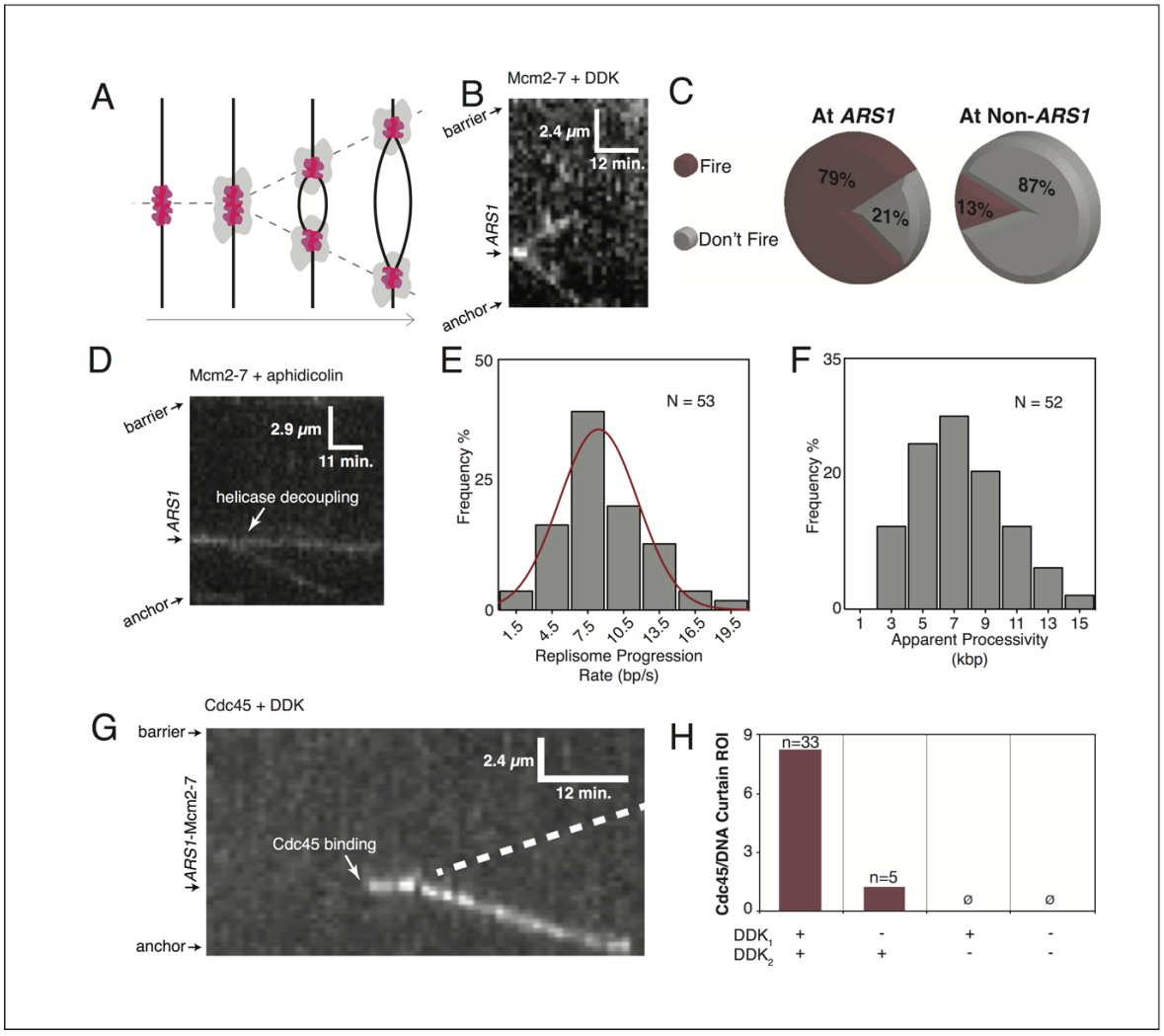

Figure 5.3 Replication Initiation
(A) A cartoon of replisome assembly around loaded Mcm2-7 double hexamers and subsequent bidirectional replication. (B) A kymogram of bidirectional Mcm2-7 progression during replication. (C) Mcm2-7 double hexamer firing is highly efficient within the observation window of the replication experiment (1.5-3 hours). (D) In the presence of aphidicolin, 18% of CMG complexes become decoupled from the replisome. (E) The replisome progression rate is 8 ± 0.4bp/s (± S.E.M.; σ = 3.1bp/s). (F) The median apparent processivity is 7.4kbp. (G) Cdc45 binds Mcm2-7 and progresses with the replisome. Poor labeling efficiency prevents visualization of the second Cdc45, presumed to progress in the opposite direction (dotted line). (H) Efficient Cdc45 binding requires prior DDK phosphorylation of Mcm2-7 (DDK_1) and supplemented DDK in the S-phase extract (DDK_2). The bar graph shows Cdc45 binding efficiency (as number of Cdc45 binding events per Region of Interest, which amounts to four DNA curtains) under different DDK conditions. Reprinted from reference (*133*), with permission from Elsevier.

bidirectional replication forks as each fluorescent dye-labeled Mcm2-7 becomes part of a CMG and then an active replisome (**Fig. 5.3B**).[8]

Remarkably, origin firing is highly efficient (**Fig. 5.3C**). 79% of *ARS1*-localized Mcm2-7$^{MCM4\text{-}sort\text{-}DY549}$ complexes go on to yield observable replisome progression. However, only 13% on non-*ARS1*-localized complexes fire. Interestingly, these proportions are correlated with the frequency of at-least-two-step and one-step photobleaching events used to measure double hexamer formation in Chapter 4 (**Fig. 4.4**). I conclude that double hexamers are highly competent for replication initiation regardless of location whereas single hexamers, significantly more prevalent at non-*ARS1* sites, fail to initiate. Furthermore, bidirectional firing happens simultaneously for both sister replisomes within the temporal resolution of these experiments (60-80s; see also Chapter 6). This suggests that quality control mechanisms ensure that both replisomes are fully assembled before either one fires.

If DDK is eliminated from the reaction, or if 1μM of the S-CDK inhibitor Sic1 is supplemented to the S-phase extract, then Mcm2-7$^{MCM4\text{-}sort\text{-}DY549}$ remains stationary during the S-phase extract incubation. This confirms the strict kinase-dependence of replication initiation without which the reaction is arrested at some stage prior to functional replisome assembly and activation. Furthermore, the addition of the DNA polymerase inhibiting drug aphidicolin (*73*) poisons

[8] The single-molecule assay was subsequently refined to yield higher-quality fluorescent data (see Chapter 6).

bidirectional firing. The vast majority of Mcm2-7$^{MCM4\text{-}sort\text{-}DY549}$ complexes remain stationary in the presence of aphidicolin, and 18% decouple from the double hexamer complex (N = 103) (**Fig. 5.3D**). This suggests that the reaction depends on DNA synthesis and that polymerase inhibition can decouple the CMG from the replisome, as previously reported (*143*). It is interesting that proper polymerase function is necessary to ensure the quality control mechanism involved in simultaneous firing.

There is a significant lag time of up to 50 minutes after the introduction of S-phase extract into the reaction chamber and prior to replication initiation. The extent of the lag depends on the concentration of supplemented DDK (see below). Interestingly however, the Mcm2-7 double hexamers remain stationary during this lag time with no detectable diffusion or movement away from *ARS1*. This agrees with *in vivo* studies that have mapped initiation to specific locations within origin sites (*31, 144, 145*) and with the direct measurements of loaded Mcm2-7 stability made in the previous chapter. Once replication begins however, each individual replisome is found to contain a single Mcm2-7: all loss-of-signal events, interpreted as photobleaching, happen in a single step (**Fig. 5.3**; also see Chapter 6). This suggests that a single CMG is sufficient for replication initiation and replisome progression and that multiple replicative helicases are not necessary for replisome function. In agreement with this single-molecule data, a single CMG has

been found to account for the footprint of a stalled replisome (*48*) and structural work shows that the CMG contains one Mcm2-7 hexamer (*131*).

The average replisome progression rate of 8 ± 0.4bp/s is in very good agreement with *in vitro* and *in vivo* studies (**Fig. 5.3E**) (*146, 147*). Processivity has an apparent median value of 7.4kbp (**Fig. 5.3F**). This is a lower bound as it is not possible to assess the true processivity of the replisomes in this assay because of photobleaching and collision with the barrier; additional work in the next chapter suggests that only direct collisions with physical blocks stop the replisome in this assay.

The large number of replisome components will each display their own complex dynamics. To begin analyzing these downstream steps, I developed an assay to visualize Cdc45, which is overexpressed in the S-phase extract with a 3x HA-tag (*28*). I first generated Mcm2-7 double hexamers using unlabeled, wild-type Mcm2-7. Then I pre-incubated the S-phase extract with biotinylated anti-HA F_{ab} fragment antibodies coupled to fluorescent streptavidin, yielding a subpopulation of dye-labeled Cdc45 molecules. Using this strategy I can visualize Cdc45 association with Mcm2-7, followed by replisome firing (**Fig. 5.3G**). This result is yet another confirmation, using an orthogonal approach, that Mcm2-7 double hexamers remain stationary prior to replication initiation.

DDK is essential for replication initiation. It is coincubated with Mcm2-7 double hexamers prior to the introduction of S-phase extract (DDK_1). Mcm2-7 is

the only known target of DDK activity, with a preference for phosphorylating double hexamers (*28*). The development of the single-molecule replication protocol led to the discovery of an absolute requirement for additional DDK supplemented to the S-phase extract (DDK_2). I tested whether this secondary activity is required before or after Cdc45 associates with the replisome. Leaving out either DDK_1 or DDK_2 abolishes replication initiation regardless of the single-molecule readout used. With only the DDK_2 incubation there is some Cdc45 association, but at a significantly lower level than with both DDK steps. With only DDK_1 and no DDK_2, all Cdc45 association is abolished (**Fig. 5.3H**). I conclude that both prior phosphorylation of Mcm2-7 and the continued presence of DDK during replisome assembly are required for robust Cdc45 association with the replisome. Additional work will be required to determine the mechanism behind the DDK_2 requirement: it may have additional targets, the S-phase extract may dephosphorylate Mcm2-7 at some rate, or DDK may be a physical component of replisomes.

The high processivity of the replisomes enables direct measurement of whether DNA sequence is rate-limiting to replication fork progression. Recall that λ_{ARS1} has AT-rich and GC-rich halves (**Fig. 5.4A**). The energy required to separate GC base pairs is greater than the energy required to separate AT base pairs, and at some scale of the replication reaction this will require greater or lesser work by the helicase. However, it is unknown whether this limits the replication rate of the

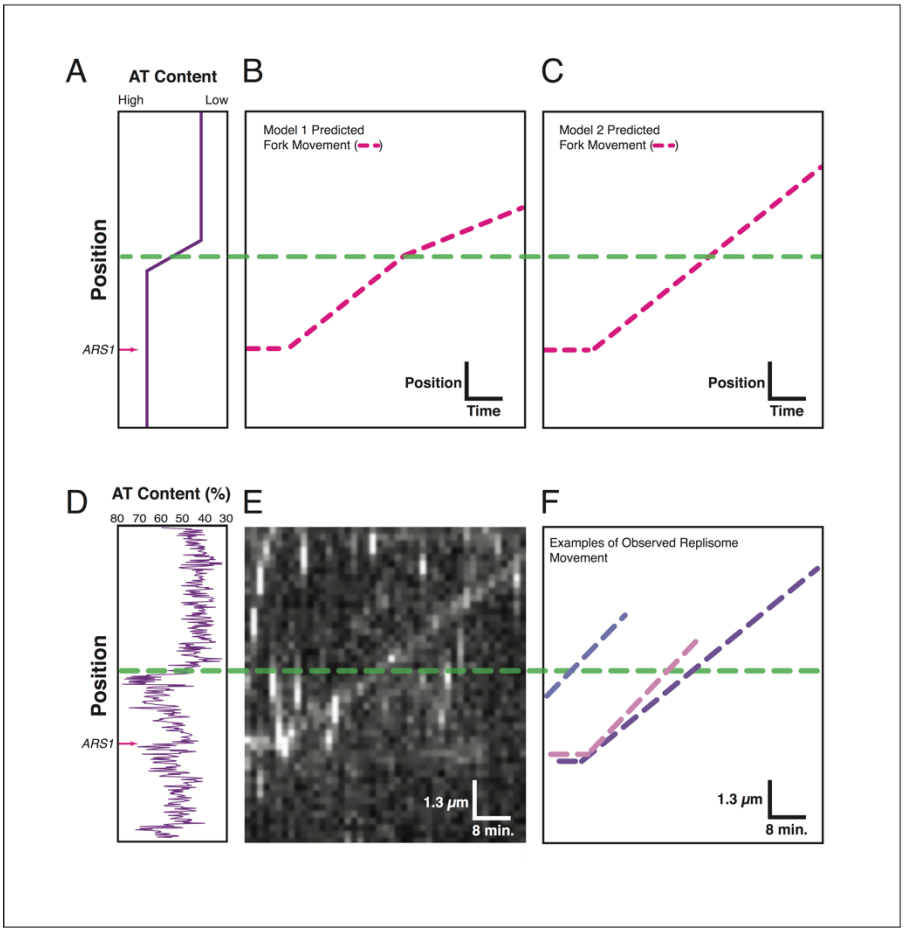

Figure 5.4 DNA Strand Separation is Not Rate Limiting to Replisome Progression
(A) A cartoon of AT content along the length of λ_{ARS1}. AT content jumps from ~45% to almost 70% in the space of less than 1kbp, the transition indicated by the green dotted line. (B) If AT content affects replisome progression rate (Model 1), then a single labeled Cdc45 moving with the replisome should slow down at the transition. However, if AT content does not affect replisome progression rate (Model 2), then the rate should remain constant through the transition region (C). (D) The AT content of λ_{ARS1}. (E) Replisome progression rate, as tracked by fluorescently labeled Cdc45, does not change at the high-to-low AT content transition region, indicated by the green dotted line. (F) The three traces are of tracked Cdc45 molecules overplayed and normalized to the kymogram shown in (E) (purple track), all showing the same monotonic behavior regardless of AT content. Reprinted from reference (*133*), with permission from Elsevier.

entire replisome at the resolution of hundreds of base pairs and tens of seconds. If the melting temperature of the DNA affects the replication rate then there should be a marked change in the rate at the AT/GC transition point roughly in the middle of λ_{ARS1} (**Fig. 5.4B**). If the melting temperature is not rate-limiting, however, then there should be no observable change in the rate at the transition point (**Fig. 5.4C**). All kymograms of replisomes passing through

the transition point show no overt change in progression rate (**Fig. 5.4D-F**; see also Chapter 6). This suggests that melting temperature is not a significant determinant of replication rate, in agreement with time-resolved ChIP experiments (*147*). This result also implies that within the context of the CMG, Mcm2-7 actively melts DNA rather than functioning through a passive, steric exclusion mechanism (*132, 148*).

Discussion and Prospects

Previous bulk replication experiments have reported that replication initiation is inefficient (*28, 106, 107*). In contrast, the single-molecule approach reveals that once an Mcm2-7 double hexamer is formed, regardless of location, it is highly likely to fire (**Fig. 5.5**). This shows that Mcm2-7 double hexamers are highly competent for their only function and attests to the importance of correct assembly, which was explored in previous chapters. The only subpopulation that fails to initiate very probably consists of single hexamers. It is possible that in bulk assays certain essential components are limiting. It is also important to note that no direct measure of firing efficiency has been reported prior to the data presented in this chapter, and the error involved in back-calculating firing efficiency based on radioactive incorporation may well have yielded an underestimation.

The failure of (the small population of) single hexamers to fire is particularly interesting given the additional finding that sister replisomes fire simultaneously

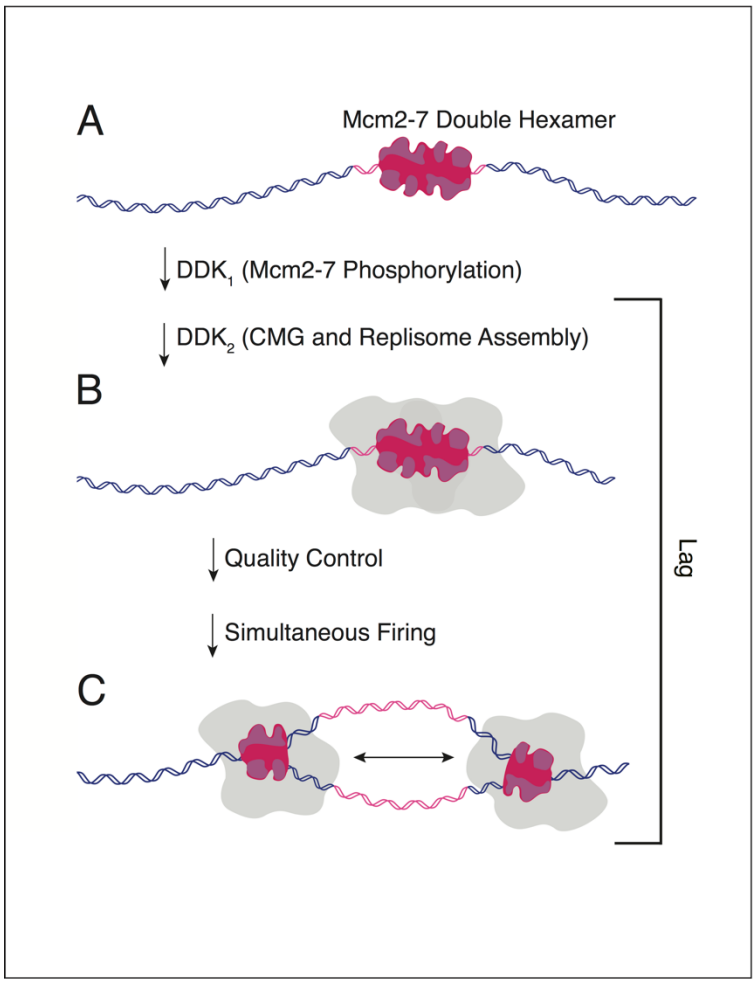

Figure 5.5 The Elements of Replication Initiation
(A) Mcm2-7 double hexamers are substrates for DDK phosphorylation, which leads to CMG and replisome assembly. (B) Quality control mechanisms ensure that both replisomes are fully assembled before either one fires. (C) Once fully assembled, sister replisomes fire simultaneously and initiate processive, bidirectional replication. Reprinted from reference (*133*), with permission from Elsevier.

within the 1 to 1.5 minute temporal resolution of the replication experiment (**Fig. 5.5B-C**). Some mechanism ensures that both replisomes are fully assembled before they are released and fire bidirectionally. It will be very interesting to uncover how the replisomes are held in check until each is whole.

Bidirectionality was expected, but not necessarily a foregone conclusion (*149*). The results presented here and in the next chapter strongly suggest that sister replisomes release each other during processive replication. However, the DNA is maintained in an extended conformation in these assays, and formally it is not possible to exclude the possibility that the replisomes separate as a consequence of tension. But several

circumstantial arguments suggest that this unlikely. Every aspect of the reaction behaves as would be expected under a variety of control conditions: replication initiation is Pre-RC dependent, DDK dependent, Cdc45 dependent, and inhibited by the polymerase poison aphidicolin. The replication rate also agrees very well with previously made measurements, which would be surprising if the individual replisomes were functioning in an entirely non-physiological conformation. Furthermore, the probability that a DNA will be ripped from its tethers if replication initiates is not higher than if replication does not initiate (Corentin Moevus, personal communication). A similar argument applies to the action of topoisomerase, which is present in the extract as measured by mass spectrometry, though its activity has not been verified (*107*) (Corentin Moevus, personal communication). Individual DNAs are not assumed to be free to rotate in the presence of S-phase extract. The functionalization used for tethering DNA at each end includes single bonds about which the entire molecule could in principle rotate, but the prevalence of DNA-binding proteins and the proteinaceous material that accumulates on the chromium very probably lock down both DNA backbones at the barriers. Formally, however, this has not been tested and it will be interesting to look at topoisomerase dynamics in the future.

This work has only begun to explore the dynamics at the most exciting stage of replication: replisome assembly and firing. The large number of essential factors needed to orchestrate these events will exhibit many interesting

mechanisms for ensuring accuracy in the pathway, or enabling plasticity to accommodate the system to the complexity of the cell. The quality control mechanism that ensures simultaneous firing and the preliminary measurements of Cdc45 dynamics, for example, are tantalizing previews of the power of this assay.

CHAPTER 6

Replisome-replisome Collisions

The DNA substrate navigated by an active replisome is coated with many DNA-binding proteins, including transcription factors, nucleosomes, and other replisomes. The single-molecule assay presented in the previous chapter is a powerful tool for probing collisions between active replisomes and physiologically relevant obstacles. Here I introduce an experimental platform to observe collisions between replisomes and find that inactive replisomes are a powerful block to oncoming active replisomes. I explore the implications of this unexpected result and offer several explanations for its experimental and physiological relevance.

The Stability of Mcm2-7 Double Hexamers on DNA May be a Problem for the Cell

A single active origin of replication could in principle serve an entire eukaryotic chromosome (*110*), but that would be a risky strategy. It could take too long to replicate the entire template, and if one replisome were to fail then the entire replication program would fail because at least one stretch of DNA would remain uncopied. Multiple active origins alleviate these challenges (*150*) and also enable the cell to orchestrate early and late domains of replication, which, among other roles, may be important for preventing the exhaustion of limiting factors if too many

replication forks were active simultaneously (*151-155*). In addition, several lines of research suggest that many more Mcm2-7 double hexamers are loaded than are ultimately used, possibly as a pool of "potential replisomes" under stress conditions (*156, 157*). (Notably however, the results of the previous chapter suggest that this may not be necessary because replisome firing is inherently highly efficient provided that replication factors are not limiting.) Taken together, these observations raise a crucial question: How does the cell remove unused and/or inactive replisome components from DNA?

In vivo measurements show that loaded Mcm2-7 is extremely stable. In a seminal report, Kuipers *et al.* find that passage through S-phase and DNA replication are necessary to remove loaded Mcm2-7 from chromatin (*158*). Furthermore, structural work shows that loaded double hexamers are topologically locked down on dsDNA, encircling it with the overall stacked doughnut arrangement described in Chapter 4, and high positive charge in the predicted DNA channel (*42, 45, 107, 159*). Data presented in Chapter 4 and Chapter 5 agrees with these results and shows the remarkable persistence of loaded double hexamers on DNA (*133*). However, the mechanism by which all stably loaded Mcm2-7 is removed from chromatin remains unknown. There are currently only two known and characterized mechanisms by which an Mcm2-7 that has been loaded onto DNA through the canonical ORC-Cdc6-Cdt1 pathway can be removed, but both absolutely require the incorporation of Mcm2-7 into an active

replisome. When two active replisomes collide, the CMG replicative helicases first run into each other and then past each other, each transitioning from single-stranded DNA (ssDNA) onto newly synthesized dsDNA (*160*). A post-replicative CMG on dsDNA is a unique complex and a substrate for a specific ubiquitin-mediated disassembly pathway (*161-163*). Similarly, when two replisomes collide at an interstrand crosslink, a specific and regulated BRCA1-mediated pathway is needed to disassemble the CMGs (*164*). The collisions inferred to happen between active replisomes and inactive replisomes are an entirely different molecular encounter. The currently prevalent hypothesis is that “passive replication” removes the inactive obstacles (*165*), but no experiment has directly tested this model. Here I describe a single-molecule assay to directly visualize the outcome of collisions between active replisomes and inactive (not-yet-fired) replisomes.

Refining the Single-molecule Replication Assay

Organic dyes are challenging to use in the complex redox environment of a cellular extract. Preventing photobleaching or other unusual photophysical behavior is significantly more difficult than in typical experimental buffers. Nonetheless, I evaluated several alternatives to the DY549 dye used in Chapter 5. Both dye-coupled streptavidin and QDs proved too bulky, and so only an organic dye could ultimately be used. I found that the JF646 dye yields a superior signal

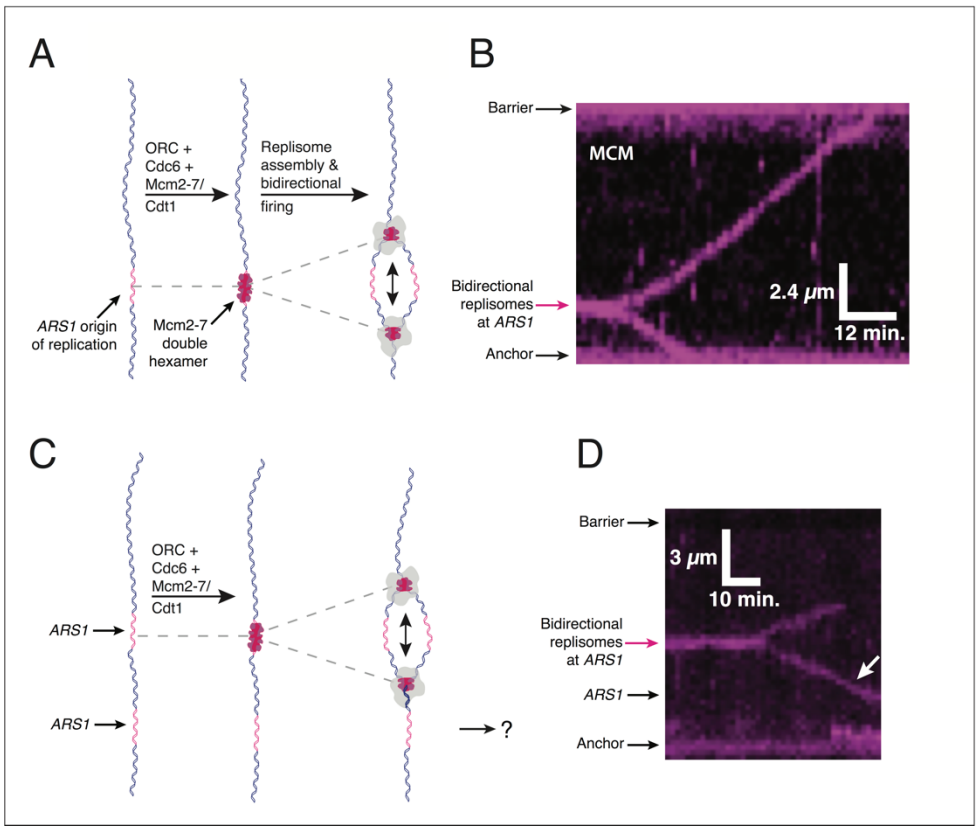

Figure 6.1 A Refined Single-molecule Replication Assay
(A) Schematic of the single-molecule DNA replication assay. Lambda phage DNA with a cloned *ARS1* origin of replication sequence is treated with initiation proteins to load Mcm2-7 double hexamers. The double hexamer is a substrate for replisome assembly. Bidirectional replication is visualized by tracking individual dye-labeled Mcm2-7 hexamers, each of which is at the core of a replicative helicase. (B) Kymogram showing bidirectional replication of a single DNA substrate (not visible) using a pair of dye-labeled Mcm2-7 hexamers. Note that fluorescent material can nonspecifically adhere to the upstream barrier and downstream anchor, which hold the DNA in place. (C) The replication assay using DNA with two cloned origins. (D) An active replisome readily replicates through an unoccupied origin site (white arrow).

for the replication experiment when compared to several other tested dyes and the Mcm2-7$^{MCM4\text{-}SNAP}$ construct has been validated *in vitro* and *in vivo* (*103, 133, 166*). I can load dye-labeled double hexamers on λ_{ARS1} as in the previous chapter by coincubating ORC, Cdc6, and Mcm2-7$^{MCM4\text{-}SNAP\text{-}JF646}$/Cdt1 in the presence of ATP on a DNA curtain (**Fig. 6.1A**). I used Mcm2-7$^{MCM4\text{-}SNAP\text{-}JF646}$ for all remaining replication experiments and refer to it hereafter as simply Mcm2-7. The readout to these experiments, as before, is the visualization of a pair of bidirectional

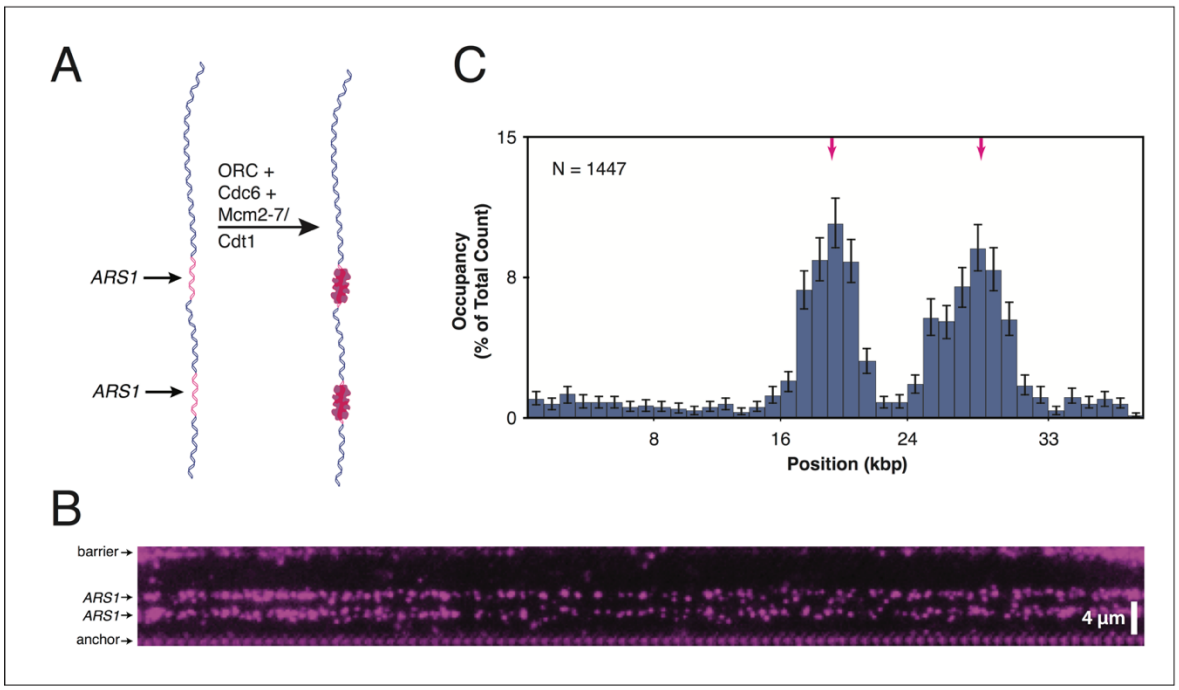

Figure 6.2 A DNA Substrate With Two Origins of Replication
(A and B) Lambda DNA with two *ARS1* origins of replication (A) behaves as expected for Mcm2-7 loading, with very high origin specificity across a large population of individual DNA molecules (B). (C) A position-distribution histogram shows the two high-occupancy peaks at the cloned origin sites (magenta arrows). N signifies number of scored individual fluorescent points, from triplicate experiments. (C) with Corentin Moevus.

replisomes using the dye on Mcm2-7 (**Fig. 6.1B**).

The probability of observing a replisome-replisome collision on λ_{ARS1} using the established conditions is exceedingly small. To begin to address this limitation I cloned a second *ARS1* site into λ_{ARS1} to yield λ_{2XARS1} (**Fig. 6.1C**). Under loading conditions that achieve high origin specificity, not all origins will be occupied. (This is analogous to the situation with ORC binding discussed in Chapter 2.) I used this to confirm that the origin sequence itself does not affect replisome progression: **Figure 6.1D** shows that a replisome can replicate through an unoccupied origin site without hindrance.

All subsequent experiments use λ_{2XARS1}, so it was important to establish that

the substrate behaves as expected (**Fig. 6.2**). I performed the same loading reaction as used for experiments with λ_{ARS1} and found that each origin site functions autonomously and mediates high Mcm2-7 loading specificity across a population of DNA molecules (**Fig. 6.2A-B**). **Figure 6.2B** shows a still image of a DNA curtain with loaded Mcm2-7. The locations of the two origin sequences are readily apparent from the two rows of loaded Mcm2-7s across the DNA curtain. A position-distribution histogram quantitatively confirms the high specificity of Mcm2-7 loading under these conditions (**Fig. 6.2C**). I conclude that λ_{2XARS1} behaves as expected and is a suitable substrate for single-molecule replication experiments.

Visualizing Replisome-replisome Collisions

The two origins in λ_{2XARS1} increase Mcm2-7 loading overall, but individual DNA molecules are still likely to have only one Mcm2-7 double hexamer, with one at either site. This, combined with the fluorescent lifetime of the dye, limits the probability of observing collision events. To further increase Mcm2-7 loading I increased the concentrations of the initiation factors ORC and Cdc6 during the loading reaction by up to an order of magnitude (**Fig. 6.3A** and Appendix II). **Figure 6.3B** and **Figure 6.3C** show that the mechanisms that mediate high origin specificity are overcome by the high protein concentrations, as expected. The influence of each origin sequence is nonetheless still apparent despite significant off-target loading (note the *ARS1*-specific peaks in **Figure 6.3C** and compare to

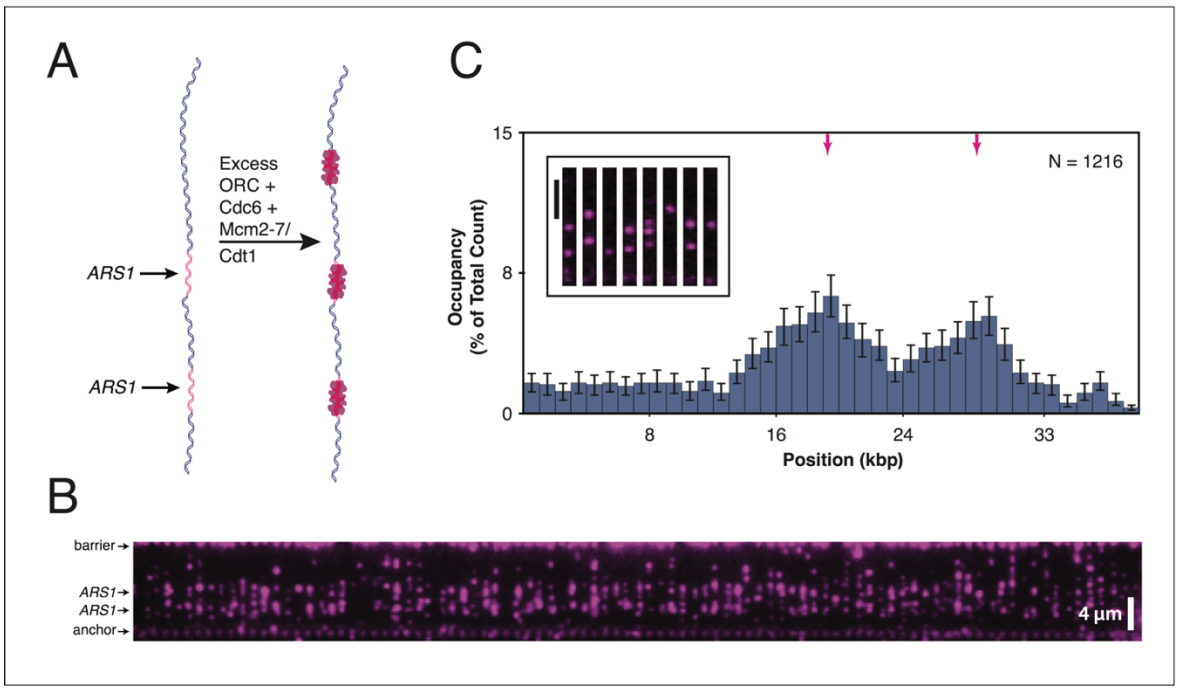

Figure 6.3 Using High Protein Concentrations to Increase Mcm2-7 Loading
(A and B) Increasing the concentration of initiation factors (A) increases the average number of Mcm2-7 double hexamers per DNA and decreases average specificity (B). (C) A position-distribution histogram shows the spread in loaded Mcm2-7 positions (compare with Figure 2C). Occupancy peaks are still apparent at the cloned origin sites (magenta arrows). N signifies number of scored individual fluorescent points, from triplicate experiments. The inset to (C) shows examples of well-isolated individual DNAs (not visible) prepared as outlined in (A). (C) with Corentin Moevus.

Figure 6.2C). Crucially, the combined use of λ_{2XARS1} and increased loading markedly increases the probability of multiple Mcm2-7s on one DNA molecule (**Fig. 6.3C** inset).

The higher density of loaded double hexamers enabled a test of the assay's validity: I looked for active replisomes colliding with each other. A collision between two active replisomes is known to result in the two CMG helicases bypassing each other by tens of base pairs onto dsDNA before coming to a stop (*160*). In our assay, with a spatial resolution on the order of hundreds of base pairs to 1kbp, this would appear as two replisomes colliding and stopping. I observe

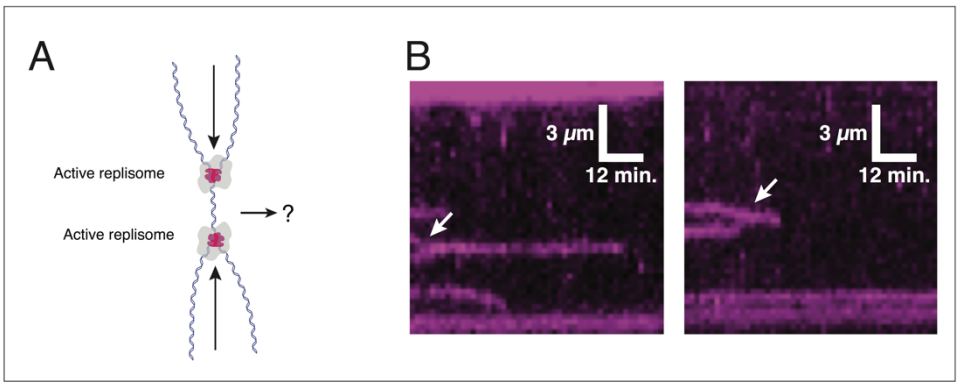

Figure 6.4 Collisions Between Active Replisomes
When active replisomes collide (A) they both stop, as expected, within the spatial resolution reported here (B). The arrows indicate collision events between active replisomes. Note that the dye labeling efficiency varies between 40-60% so that some Mcm2-7 hexamers are not visible.

exactly the predicted outcome in cases where two active replisomes are observed to collide, with no bypass, eviction, or pushing of one replisome by the other but rather two Mcm2-7-based signals merging into one (**Fig. 6.4**). The signal loss post collision is consistent with either photobleaching, CMG disassembly by the established ubiquitinylation pathway, or some combination of both. I conclude that this assay yields the expected readout for a known pathway involving replisome-replisome collisions and is a valid experimental platform to study such events. Note that topological stress could only build up between two oncoming replisomes if each DNA were topologically constrained along its entire length. However, if this were the case, then replication bubbles would be incapable of stretching across thousands to tens of thousands of base pairs, as already documented. Therefore, either the DNA is not topologically constrained in this assay (perhaps nicked) or the S-phase extract topoisomerases are at high enough a concentration and active.

Collisions Between Active and Inactive Replisomes

I next used the assay to observe collisions between active and inactive replisomes, relying on the stochastic nature of replisome firing to screen for these events. Each loaded double hexamer is very likely to assemble into a bidirectional pair of replisomes that will fire within the typical observation time and conditions of the experimental setup (*133*). For each DNA with at least two loaded double hexamers, one pair of replisomes could fire well before the other is activated (**Fig. 6.5A**). I define an active replisome as one that has visibly transitioned into a translocating state, having separated from its sister replisome (**Fig. 6.1**). I define inactive replisomes as any loaded Mcm2-7 that has not yet begun to translocate within the context of a fully-formed or forming replisome (see Discussion).

Each collision between an active replisome and inactive replisomes could result in one of several readouts (**Fig. 6.5A**). (i) The active replisome could directly evict the inactive ones. This would appear as continuous active replisome translocation through the fluorescent signal of inactive Mcm2-7s on the same DNA. (ii) The active replisome could stop translocating when it encounters the inactive replisomes. This would appear as an active replisome stopping when it encounters the inactive Mcm2-7s, similar to the data for collisions between two active replisomes (**Fig. 6.4B**). (iii) The active replisome could push the inactive replisomes along dsDNA. Several intermediate outcomes are also possible.

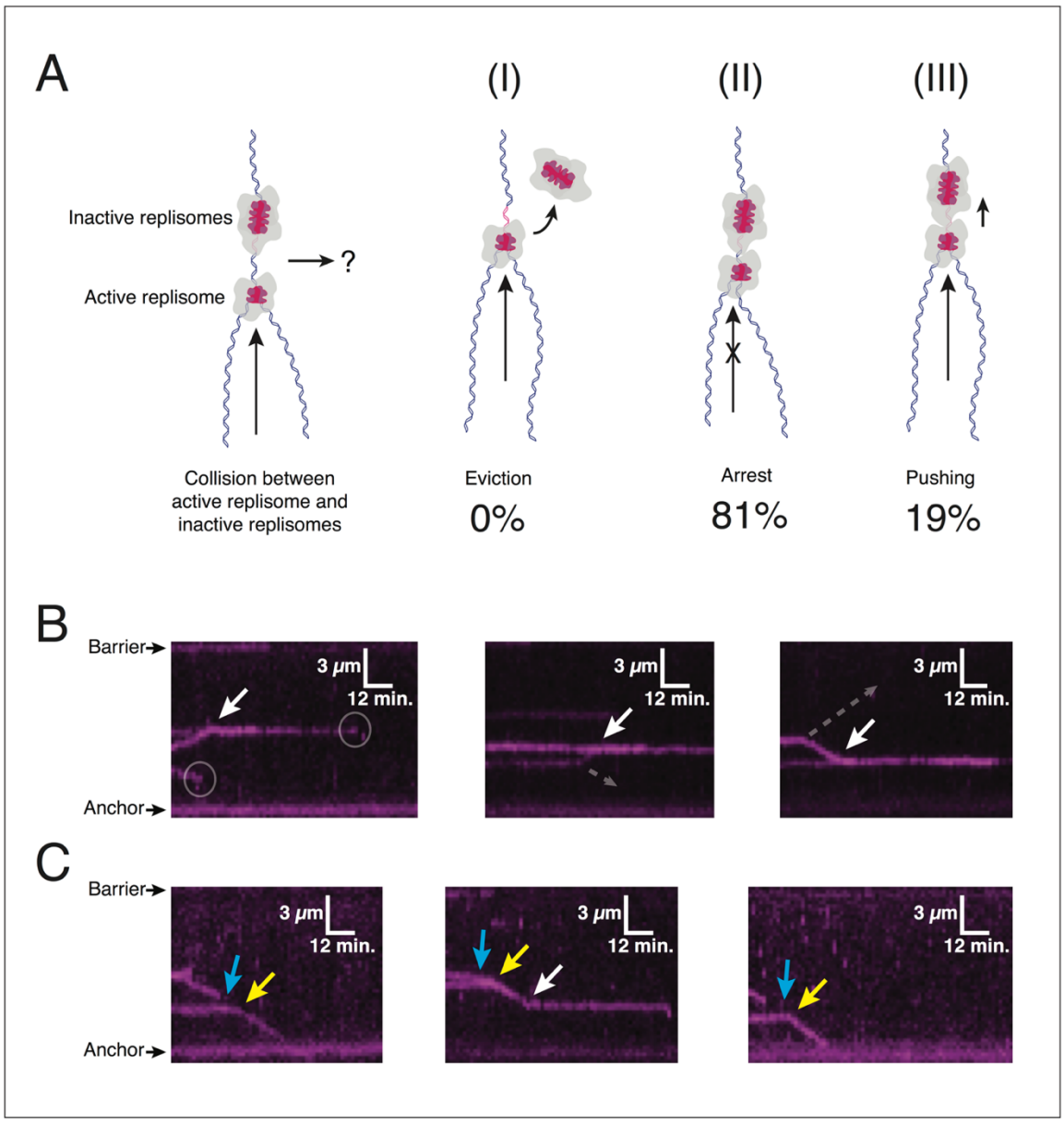

Figure 6.5 Active Replisomes Do Not Evict inactive Replisomes
(A) What are the possible outcomes of a collision between a translocating replisome and inactive replisomes? In model (I) the inactive replisomes are evicted by the passing active replisome. In (II) the active replisome is arrested by the encounter with inactive replisomes; the inactive replisomes could still activate later. In (III) the active replisome pushes the inactive Mcm2-7s. (B) Kymograms showing outcomes consistent with model (II). Opaque circles indicate photobleaching events (omitted from other kymograms for clarity), gray dotted lines indicate the inferred trajectory of unlabeled replisomes (omitted from other kymograms for clarity), and white arrows indicate collision events throughout. (C) Kymograms showing outcomes consistent with model (II) or (III). Cyan arrows indicate initial collision events and yellow arrows indicate either pushing (model (III)) or firing of the initially inactive replisomes after collision (model (II)).

I observed a total of 88 well-isolated active replisomes from across six independent experiments, including 31 collisions. Four collisions were between pairs of active replisomes and all resulted in arrest as described above. The remaining collisions were between active replisomes and inactive replisomes, and there were no evictions. The majority of collisions (81%) resulted in the arrest of the active replisome and some collisions (19%) resulted in possible pushing of the inactive replisomes (**Fig. 6.5B-C** and **Fig. 6.6**).

This result is inconsistent with models that invoke “passive replication” through inactive replisomes: the act of replication in and of itself is insufficient to cause the stationary obstacles, with Mcm2-7s at their core, to release the DNA. It is important to note that in this assay the identity of the inactive obstacle could be any structure on pathway to forming a pair of bidirectionally oriented replisomes (see Discussion). This result further suggests that once Mcm2-7 has been loaded by the canonical pathway and S-phase starts, removing that Mcm2-7 from DNA becomes a singular challenge.

Discussion

The genomic material of a eukaryotic cell is a complicated template to access along its entire length, and yet every successful round of DNA replication accomplishes this feat completely and efficiently. One salient challenge for a translocating replisome is the set of DNA-bound proteins occupying its path.

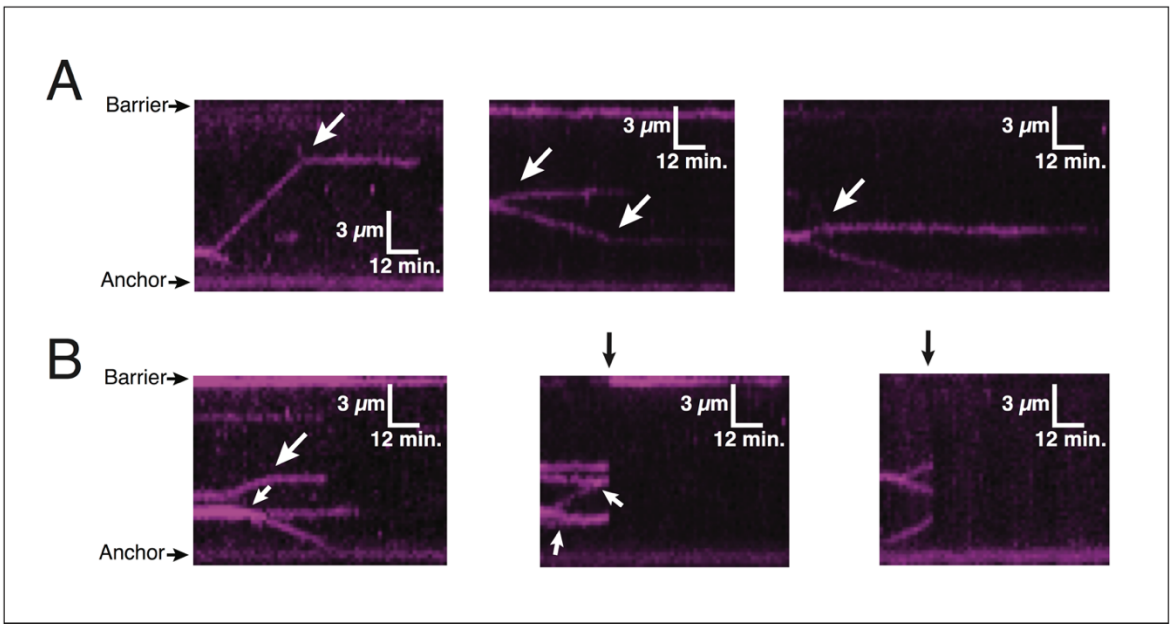

Figure 6.6 Additional Collision Data Categories
(A) Kymograms showing additional outcomes consistent with model (II) but with the inactive replisomes unlabeled. (B) The first kymogram includes a labeled-unlabeled replisome collision event, a labeled-labeled collision event, and a possible pushing event. The second and third kymograms show DNA breakage and concomitant loss of all signal. This demonstrates that we are observing events on individual DNAs.

Loaded Mcm2-7 in particular has a very long lifetime on DNA and is expected to be a common obstacle. Current models of the DNA replication program hypothesize that an active replisome can “passively replicate” through loaded Mcm2-7s that have not been activated. This hypothesis can only be addressed directly if the identity of each set of molecules can be tracked over the course of an experiment. Subpopulations of replisomes have to be categorized as active or inactive, a requirement very well suited to single-molecule assays, which can differentiate among distinct subpopulations.

Here I used the single-molecule replication assay to observe the outcome of collisions between active replisomes and inactive replisomes to find that a

translocating replisome is incapable of evicting inactive replisomes, with the majority of replisomes arresting at the encounter. A subpopulation of events could be interpreted as pushing of the inactive replisomes, but is also consistent with the photobleaching of one dye on a labeled Mcm2-7 followed by firing of the initially inactive replisomes (**Fig. 6.5C** and **Fig. 6.6**). Collision events do not appear to necessarily prevent the initially inactive replisomes from firing (**Fig. 6.6**). I conclude that the act of replication alone is not sufficient to displace inactive replisomes from the DNA.

The direct cause of replisome arrest in this system could simply be that the inactive obstacle is tightly bound to DNA. The replisome exerts some finite maximal force, and a large number of electrostatic interactions between the inactive replisomes and DNA could be sufficient to arrest translocation. This interpretation has well-established precedents. In a *Xenopus laevis* cell-free extract based replication assay, a sequence-targeted array of bacterial Lac repressor proteins can be used to reversibly arrest replisome progression (*167-169*). The Lac repressor is orthogonal to the system so the likeliest explanation is that the cumulative binding strength of multiple Lac proteins is sufficient to stop, or at least significantly slow down, replisome progression. Similarly, the bacterial terminator Tus-Ter, which functions by progressively locking down on the DNA in response to applied force, can also stop eukaryotic replisome progression (*170, 171*). Finally, a substantial body of work shows that collisions between a replisome

and the transcription machinery can stop fork progression (*172*). If arrest is due to obstacle binding strength then the putative pushing events could be the result of collisions with a distinct subpopulation of obstacles—perhaps more incompletely assembled replisomes—each making fewer contacts with the DNA. Interestingly, loaded Mcm2-7s can be readily pushed, but not evicted, by the bacterial molecular motor RecBCD (experiments in progress with Tsuyoshi Terakawa and Corentin Moevus), and so in principle an Mcm2-7 can be forced to move along dsDNA. This suggests an interesting alternative mechanism for the outcome of physiological collisions between replisomes: the arrest could be caused by some specific interaction between an active replisome and an inactive one. Structural signals could conceivably cue the oncoming replisome to stop when it encounters a specific Mcm2-7-derived complex. This may be advantageous in some contexts as explained below.

These results do not rule out the possibility that the extract-based assay is missing key factors needed to remove inactive replisomes during an encounter with a fork. This would not invalidate my claim that replication alone cannot achieve this: extensive evidence supports that the active replisomes observed in this assay are at least competent to carry out processive bidirectional replication of both daughter strands (*28, 106, 107, 133*). This requires the presence of at least a minimal replisome with all essential components present. In addition, there are multiple lines of evidence that the active replisomes readily displace a host of

DNA-bound proteins without arresting. Competitor DNA is included in our assay to prevent the DNA curtains from compacting and ripping off their tethers from the action of DNA-binding proteins (*133*). Mass spectroscopy has also shown the presence of many DNA-interacting factors in a budding yeast extract system very similar to the one used here (*107*); the DNA curtain substrates are also coated in DNA-binding proteins. This suggests that only inactive replisomes can arrest a translocating replisome in this assay.

The precise makeup of an inactive obstacle is unknown. Formally it could be any structure on pathway from an unphosphorylated Mcm2-7 double hexamer to a fully assembled pair of bidirectionally oriented replisomes. However, the Mcm2-7s are likely to be at least phosphorylated (see Chapter 5). The efficiency of firing and the overexpression of key replication factors in the S-phase extract also lead me to expect that the inactive obstacles are very likely partially assembled pairs of replisomes. This is crucial when evaluating the result with respect to the *in vivo* state where several replication factors are limiting and there could be a prevalence of Mcm2-7 double hexamers with few or no associated replisome components during S-phase (*153*). It is possible that unphosphorylated Mcm2-7 double hexamers are more easily pushed than downstream structures, and this has been invoked to explain the RNA polymerase-mediated redistribution of a small subpopulation of double hexamers *in vivo* (*173*). Pushing, however, is far from eviction and requires the inactive structure to move ahead of the fork and

somehow handle all the protein obstacles in its way—including nucleosomes—despite not being a functional replication fork.

Invoking pushing to explain double hexamer behavior fits with our current understanding of its structure: dsDNA is thought to thread through its central channel. However, invoking eviction—the wholesale removal of loaded Mcm2-7s because of a collision with a translocase—would also require invoking either an extreme instability of the double hexamer or highly convoluted DNA-protein acrobatics. A collision with a replisome could possibly cause some Mcm2-7 subunits to dislodge or change structure, but all current work on loaded double hexamers suggests that they are extremely stable on DNA. The MCM2-MCM5 DNA entry gate is closed for both hexamers within a double hexamer in published structures, and the gates do not align with each other (*45, 129*). Therefore, removal of an inactive double hexamer would nominally require opening of both gates, rotation of the oppositely oriented hexamers with respect to each other (despite their apparently stable interaction in the inactive state), and removal of the DNA from the positively charged channel. Given this information, a reasonable null hypothesis is that the removal of loaded Mcm2-7 requires some specific pathway and that direct eviction, as could be the case for a non-topologically closed protein-DNA interaction, is unlikely. This criteria certainly applies to this system where all Mcm2-7 double hexamers are loaded by the canonical ORC-Cdc6-Cdt1-mediated pathway. But some data suggest that other pathways could be involved

in loading excess Mcm2-7 onto chromatin, and the structural identity and general nature of this distinct subpopulation remain largely uncharacterized (*158, 174*).

The data in this chapter eliminate the simplest possible explanation for how active replisomes remove inactive ones. However, this is an exceptionally complicated problem, and the result also highlights a series of key questions. How many fully-formed Mcm2-7 double hexamers are loaded *in vivo* by the canonical pathway? Are these in excess to the number needed for replication, and if so, what is the mechanism by which the unnecessary ones are removed from DNA? Does excess loaded Mcm2-7 imply excess inactive replisomes? What are all the pathways to Mcm2-7 loading and are the differently loaded Mcm2-7s all equivalent?

The replisome arrest pathway identified here has some attractive features that may address some of these questions. When an active replisome collides with inactive replisomes, the subsequent activation of the initially inactive obstacle should shunt the collided pair of replisomes (the two that face each other in the collision) into the previously-characterized ubiquitin-mediated disassembly pathway. There is no need to invoke a novel Mcm2-7 eviction pathway. I suggest that removing a loaded hexamer or any downstream structure is a particular molecular challenge, and by arresting oncoming forks, the obstacle is given some time to enter the established ubiquitin-mediated pathway for one arm of the fork, and continue replicating with the other. This ensures that all the DNA is replicated

and that no Mcm2-7s are left on the DNA. Furthermore, the arrest mechanism is the most straightforward way to prevent early replication domains from impinging on late replication domains. The use and regulation of timing domains is an important feature of the eukaryotic replication program (*151, 152, 155, 175*), and if active replisomes were able to indiscriminately plow through all inactive Mcm2-7s then the expectation would be that late domains could be systematically disrupted. I conclude that not only does the arrest mechanism suggest that other pathways are not needed to remove inactive Mcm2-7s, but that arrest may itself be an important feature of replication control.

CONCLUSION

Eukaryotic DNA replication is among the most important and also among the most complex cellular pathways. Besides its many highly dynamic components, it integrates information from many local and cell-wide sources. The reaction itself is a challenge to navigate too: DNA must be melted, the leading and lagging strands polymerized in opposite directions, the ssDNA protected, the polymerases kept processive, the primase turned over but kept nearby for that next okazaki fragment... the orchestration is dizzying. Even the initiation pathway—so much less complex by comparison—is dynamic and intricate. Despite decades of work on just Pre-RC assembly, important questions remain unanswered.

Single-molecule approaches are ideal for studying such systems because they can reveal the steps, control mechanisms, and heterogeneities that determine how biological complexes operate. However, studying anything involving more than a few molecular species is a major technical challenge: single-molecule bioscience has just now entered a phase in which it can tackle physiologically relevant and therefore more intricate scenarios. I have presented a series of experiments that push the DNA curtain method to its very limit and will hopefully guide future work to plumb even denser phases of complexity.

I began by showing in Chapter 2 that ORC has an inherent capacity to recognize its target sequence even if it is buried in an excess of non-specific DNA.

I demonstrated that as with all DNA-binding proteins, ensemble measurements of specificity are concentration-dependent and can be fully explained through the single-molecule perspective. I then found, as explored in Chapter 3, that Cdc6 exhibits a remarkable capacity to differentiate among ORC molecules as a function of the underlying sequences. It introduces specificity to the system using dynamics: Cdc6 is likelier to bind ORC at *ARS1*, and once there, has a longer lifetime. Cdc6 also ensures that Pre-RC formation happens only in a defined order by limiting ORC's capacity to bind DNA if Cdc6 and ORC interact in solution. ORC and Cdc6 are thereby prevented from binding the DNA simultaneously and forced to follow a defined, ordered mode of DNA binding: ORC binds DNA first, then Cdc6 binds DNA-ORC. This phenomenon also explains the apparent increase in ORC specificity in the presence of Cdc6. Cdc6 does not direct ORC to specific sites, but rather effectively reduces the DNA-binding-competent ORC concentration in solution.

In Chapter 4 I show how Cdc6 behavior mediates very high Mcm2-7 loading specificity. ORC cannot recruit Mcm2-7 without Cdc6, and the likeliest place for Mcm2-7/Cdt1 to encounter a DNA-ORC-Cdc6 complex is at an origin. Interestingly, double-hexamers are also likelier to form at *ARS1*, yet another level of control for accuracy. Chapter 5 describes how the experimental system is elevated to the next level: Mcm2-7 double hexamers loaded in the single-molecule assay can serve as substrates for replication initiation. This assay reveals that

sister forks fire simultaneously, suggesting that control mechanisms ensure each replisome is fully assembled before either can fire. I also found that DDK is required prior to Cdc45 association, one Mcm2-7 is sufficient to support processive replication in the context of the CMG, and DNA sequence is not rate-limiting to replisome progression.

This assay renders a huge array of fascinating questions immediately apparent, and the formidable dynamics of replisome assembly and progression loom before many future projects. The replisome progression assay enables the direct observation of collision events, and in Chapter 6 I discuss the surprising result that replisomes cannot evict inactive replisomes. This may point to important regulatory mechanisms built into active replisomes to specifically sense inactive ones. It will be very interesting to explore the mechanisms behind the replisome's capacity to handle so complex a substrate; DNA is coated with proteins, and all are important for one or another function, and all must be crashed aside by replisomes to expose the template they absolutely need.

REFERENCES

1. B. Rotman, Measurement of activity of single molecules of beta-D-galactosidase. *Proceedings of the National Academy of Sciences of the United States of America* **47**, 1981-1991 (1961).

2. T. Hirschfeld, Optical microscopic observation of single small molecules. *Applied Optics* **15**, 2965-2966 (1976).

3. E. Brooks Shera, N. K. Seitzinger, L. M. Davis, R. A. Keller, S. A. Soper, Detection of single fluorescent molecules. *Chemical Physics Letters* **174**, 553-557 (1990).

4. E. Betzig, R. J. Chichester, Single molecules observed by near-field scanning optical microscopy. *Science* **262**, 1422-1425 (1993).

5. A. Schepartz, R. L. Gonzalez, Molecular imaging: Sine labore nihil. *Current Opinion in Chemical Biology* **15**, 749-751 (2011).

6. S. Wennmalm, S. M. Simon, in *Annual Review of Biochemistry*. (2007), vol. 76, pp. 419-446.

7. W. J. Greenleaf, M. T. Woodside, S. M. Block, in *Annual Review of Biophysics and Biomolecular Structure*. (2007), vol. 36, pp. 171-190.

8. D. Duzdevich, E. C. Greene, Towards physiological complexity with in vitro single-molecule biophysics. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **368**, 20120271 (2013).

9. I. Tinoco Jr, R. L. Gonzalez Jr, Biological mechanisms, one molecule at a time. *Genes and Development* **25**, 1205-1231 (2011).

10. K. J. Neaves, J. L. Huppert, R. M. Henderson, J. M. Edwardson, Direct visualization of G-quadruplexes in DNA using atomic force microscopy. *Nucleic Acids Research* **37**, 6269-6275 (2009).

11. V. E. Foe, L. E. Wilkinson, C. D. Laird, Comparative organization of active transcription units in Oncopeltus fasciatus. *Cell* **9**, 131-146 (1976).

12. A. N. Kapanidis, T. Strick, Biology, one molecule at a time. *Trends in Biochemical Sciences* **34**, 234-243 (2009).

13. T. Fazio, M. L. Visnapuu, S. Wind, E. C. Greene, DNA curtains and nanoscale curtain rods: High-throughput tools for single molecule imaging. *Langmuir* **24**, 10524-10531 (2008).

14. J. Gorman, T. Fazio, F. Wang, S. Wind, E. C. Greene, Nanofabricated racks of aligned and anchored DNA substrates for single-molecule imaging. *Langmuir* **26**, 1372-1379 (2010).

15. M. L. Visnapuu, T. Fazio, S. Wind, E. C. Greene, Parallel arrays of geometric nanowells for assembling curtains of DNA with controlled lateral dispersion. *Langmuir* **24**, 11293-11299 (2008).

16. J. D. Watson, F. H. C. Crick, Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).

17. F. Jacob, S. Brenner, On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon. *Comptes rendus hebdomadaires des séances de l'Académie des sciences* **256**, 298-300 (1963).

18. L. H. Hartwell, J. Culotti, J. R. Pringle, B. J. Reid, Genetic control of the cell division cycle in yeast. *Science* **183**, 46-51 (1974).

19. P. Nurse, Genetic control of cell size at cell division in yeast. *Nature* **256**, 547-551 (1975).

20. P. Nurse, P. Thuriaux, K. Nasmyth, Genetic control of the cell division cycle in the fission yeast Schizosaccharomyces pombe. *MGG Molecular & General Genetics* **146**, 167-178 (1976).

21. P. Thuriaux, P. Nurse, Regulation of cell size at division in the fission yeast Schizosaccharomyces pombe. *Experientia* **33**, 137 (1977).

22. G. T. Maine, P. Sinha, B. W. Tye, Mutants of S. cerevisiae defective in the maintenance of minichromosomes. *Genetics* **106**, 365-385 (1984).

23. S. J. Aves, Y. Liu, T. A. Richards, Evolutionary diversification of eukaryotic DNA replication machinery. *Sub-cellular biochemistry* **62**, 19-35 (2012).

24. N. Chia, I. Cann, G. J. Olsen, Evolution of DNA replication protein complexes in Eukaryotes and Archaea. *PLoS ONE* **5**, (2010).

25. J. M. Miller, E. J. Enemark, Archaeal MCM Proteins as an Analog for the Eukaryotic Mcm2-7 Helicase to Reveal Essential Features of Structure and Function. *Archaea* **2015**, (2015).

26. S. D. Bell, Archaeal orc1/cdc6 proteins. *Sub-cellular biochemistry* **62**, 59-69 (2012).

27. I. M. Slaymaker, X. S. Chen, MCM structure and mechanics: what we have learned from archaeal MCM. *Sub-cellular biochemistry* **62**, 89-111 (2012).

28. R. C. Heller *et al.*, Eukaryotic origin-dependent DNA replication in vitro reveals sequential action of DDK and S-CDK kinases. *Cell* **146**, 80-91 (2011).

29. C. C. Siow, S. R. Nieduszynska, C. A. Müller, C. A. Nieduszynski, OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Research* **40**, D682-D686 (2012).

30. J. J. Wyrick *et al.*, Genome-wide distribution of ORC and MCM proteins in S. cerevisae: High-resolution mapping of replication origins. *Science* **294**, 2357-2360 (2001).

31. S. R. McGuffee, D. J. Smith, I. Whitehouse, Quantitative, Genome-Wide Analysis of Eukaryotic Replication Initiation and Termination. *Molecular Cell* **50**, 123-135 (2013).

32. Z. Lygerou, P. Nurse, The fission yeast origin recognition complex is constitutively associated with chromatin and is differentially modified through the cell cycle. *Journal of Cell Science* **112**, 3703-3712 (1999).

33. J. H. Cocker, S. Piatti, C. Santocanale, K. Nasmyth, J. F. X. Diffley, An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. *Nature* **379**, 180-182 (1996).

34. C. Liang, M. Weinreich, B. Stillman, ORC and Cdc6p interact and determine the frequency of initiation of DNA replication in the genome. *Cell* **81**, 667-676 (1995).

35. T. Mizushima, N. Takahashi, B. Stillman, Cdc6p modulates the structure and DNA binding activity of the origin recognition complex in vitro. *Genes and Development* **14**, 1631-1641 (2000).

36. C. Speck, Z. Chen, H. Li, B. Stillman, ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nature Structural and Molecular Biology* **12**, 965-971 (2005).

37. J. Sun *et al.*, Cdc6-induced conformational changes in ORC bound to origin DNA revealed by cryo-electron microscopy. *Structure* **20**, 534-544 (2012).

38. C. Evrin *et al.*, The ORC/Cdc6/MCM2-7 complex facilitates MCM2-7 dimerization during prereplicative complex formation. *Nucleic Acids Research* **42**, 2257-2269 (2014).

39. A. Fernández-Cid *et al.*, An ORC/Cdc6/MCM2-7 Complex Is Formed in a Multistep Reaction to Serve as a Platform for MCM Double-Hexamer Assembly. *Molecular Cell* **50**, 577-588 (2013).

40. C. Evrin *et al.*, A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20240-20245 (2009).

41. A. Gambus, G. A. Khoudoli, R. C. Jones, J. J. Blow, MCM2-7 form double hexamers at licensed origins in Xenopus egg extract. *Journal of Biological Chemistry* **286**, 11855-11864 (2011).

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

43. S. A. Samel *et al.*, A unique DNA entry gate serves for regulated loading of the eukaryotic replicative helicase MCM2-7 onto DNA. *Genes and Development* **28**, 1653-1666 (2014).

44. J. Sun *et al.*, Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA. *Nature Structural and Molecular Biology* **20**, 944-951 (2013).

45. N. Li *et al.*, Structure of the eukaryotic MCM complex at 3.8 Å. *Nature* **524**, 186-191 (2015).

46. L. D. Langston *et al.*, CMG helicase and DNA polymerase ε form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 15390-15395 (2014).

47. J. Sun *et al.*, The architecture of a eukaryotic replisome. *Nature Structural and Molecular Biology* **22**, 976-982 (2015).

48. Y. V. Fu *et al.*, Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell* **146**, 931-941 (2011).

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

50. M. L. Bochman, S. P. Bell, A. Schwacha, Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability. *Molecular and Cellular Biology* **28**, 5865-5873 (2008).

51. M. L. Bochman, A. Schwacha, The Saccharomyces cerevisiae Mcm6/2 and Mcm5/3 ATPase active sites contribute to the function of the putative Mcm2-7 'gate'. *Nucleic Acids Research* **38**, 6078-6088 (2010).

52. F. Chang *et al.*, Cdc6 ATPase activity disengages Cdc6 from the pre-replicative complex to promote DNA replication. *eLife* **4**, (2015).

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

54. C. Evrin *et al.*, In the absence of ATPase activity, pre-RC formation is blocked prior to MCM2-7 hexamer dimerization. *Nucleic Acids Research* **41**, 3162-3172 (2013).

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

56. S. Kang, M. Warner, S. Bell, Multiple Functions for Mcm2-7 ATPase Motifs during Replication Initiation. *Molecular Cell*, (2014).

57. R. D. Klemm, R. J. Austin, S. P. Bell, Coordinate binding of ATP and origin DNA regulates the ATPase activity of the origin recognition complex. *Cell* **88**, 493-502 (1997).

58. R. D. Klemm, S. P. Bell, ATP bound to the origin recognition complex is important for preRC formation. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8361-8367 (2001).

59. D. G. Lee, A. M. Makhov, R. D. Klemm, J. D. Griffith, S. P. Bell, Regulation of origin recognition complex conformation and ATPase activity: Differential effects of single-stranded and double-stranded DNA binding. *EMBO Journal* **19**, 4774-4782 (2000).

60. J. C. W. Randell, J. L. Bowers, H. K. Rodríguez, S. P. Bell, Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Molecular Cell* **21**, 29-39 (2006).

61. C. Speck, B. Stillman, Cdc6 ATPase activity regulates ORC·Cdc6 stability and the selection of specific DNA sequences as origins of DNA replication. *Journal of Biological Chemistry* **282**, 11705-11714 (2007).

62. H. Rao, B. Stillman, The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 2224-2228 (1995).

63. M. A. Palacios DeBeer, U. Müller, C. A. Fox, Differential DNA affinity specifies roles for the origin recognition complex in budding yeast heterochromatin. *Genes and Development* **17**, 1817-1822 (2003).

64. T. Hoggard, E. Shor, C. A. Müller, C. A. Nieduszynski, C. A. Fox, A Link between ORC-Origin Binding Mechanisms and Origin Activation Time Revealed in Budding Yeast. *PLoS Genetics* **9**, (2013).

65. K. Hizume, M. Yagura, H. Araki, Concerted interaction between origin recognition complex (ORC), nucleosomes and replication origin DNA ensures stable ORC-origin binding. *Genes to Cells* **18**, 764-779 (2013).

66. T. J. Newman, M. A. Mamun, C. A. Nieduszynski, J. J. Blow, Replisome stall events have shaped the distribution of replication origins in the genomes of yeasts. *Nucleic Acids Research* **41**, 9705-9718 (2013).

67. S. Costa, J. J. Blow, The elusive determinants of replication origins. *EMBO Reports* **8**, 332-334 (2007).

68. S. Stanojcic, J. M. Lemaitre, K. Brodolin, E. Danis, M. Mechali, In Xenopus egg extracts, DNA replication initiates preferentially at or near asymmetric AT sequences. *Molecular and Cellular Biology* **28**, 5265-5274 (2008).

69. S. Waga, A. Zembutsu, Dynamics of DNA binding of replication initiation proteins during de Novo formation of pre-replicative complexes in Xenopus egg extracts. *Journal of Biological Chemistry* **281**, 10926-10934 (2006).

70. J. J. Blow, Control of chromosomal DNA replication in the early Xenopus embryo. *EMBO Journal* **20**, 3293-3297 (2001).

71. D. Remus, E. L. Beall, M. R. Botchan, DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. *EMBO Journal* **23**, 897-907 (2004).

72. S. Yasuda, Y. Hirota, Cloning and mapping of the replication origin of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5458-5462 (1977).

73. A. Kornberg, T. Baker, *DNA Replication, Second Edition*. (University Science Books, Sausalito, California, 1992).

74. C. S. M. Chan, B. K. Tye, Autonomously replicating sequences in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 6329-6333 (1980).

75. D. T. Stinchcomb, M. Thomas, J. Kelly, E. Selker, R. W. Davis, Eukaryotic DNA segments capable of autonomous replication in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 4559-4563 (1980).

76. A. W. Murray, J. W. Szostak, Construction of artificial chromosomes in yeast. *Nature* **305**, 189-193 (1983).

77. Y. Marahrens, B. Stillman, Replicator dominance in a eukaryotic chromosome. *EMBO Journal* **13**, 3395-3400 (1994).

78. A. Rowley, J. H. Cocker, J. Harwood, J. F. X. Diffley, Initiation complex assembly at budding yeast replication origins begins with the recognition of a bipartite sequence by limiting amounts of the initiator, ORC. *EMBO Journal* **14**, 2631-2641 (1995).

79. G. M. Wilmes, S. P. Bell, The B2 element of the Saccharomyces cerevisiae ARS1 origin of replication requires specific sequences to facilitate pre-RC formation. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 101-106 (2002).

80. F. J. Chang *et al.*, High-resolution analysis of four efficient yeast replication origins reveals new insights into the ORC and putative MCM binding elements. *Nucleic Acids Research* **39**, 6523-6535 (2011).

81. J. R. Broach *et al.*, Localization and sequence analysis of yeast origins of DNA replication. *Cold Spring Harbor Symposia on Quantitative Biology* **47**, 1165-1173 (1982).

82. S. E. Celniker, K. Sweder, F. Srienc, J. E. Bailey, J. L. Campbell, Deletion mutations affecting autonomously replicating sequence ARS1 of Saccharomyces cerevisiae. *Molecular and Cellular Biology* **4**, 2455-2466 (1984).

83. F. Srienc, J. E. Bailey, J. L. Campbell, Effect of ARS1 mutations on chromosome stability in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **5**, 1676-1684 (1985).

84. H. Feldmann, J. Olah, H. Friedenreich, Sequence of a yeast DNA fragment containing a chromosomal replicator and a tRNAGlu<inf>3</inf> gene. *Nucleic Acids Research* **9**, 2949-2959 (1981).

85. S. Kearsey, Analysis of sequences conferring autonomous replication in baker's yeast. *EMBO Journal* **2**, 1571-1575 (1983).

86. A. H. Bouton, M. M. Smith, Fine-structure analysis of the DNA sequence requirements for autonomous replication of Saccharomyces cerevisiae plasmids. *Molecular and Cellular Biology* **6**, 2354-2363 (1986).

87. A. H. Brand, G. Micklem, K. Nasmyth, A yeast silencer contains sequences that can promote autonomous plasmid replication and transcriptional activation. *Cell* **51**, 709-719 (1987).

88. A. Dershowitz *et al.*, Linear derivatives of Saccharomyces cerevisiae chromosome III can be maintained in the absence of autonomously replicating sequence elements. *Molecular and Cellular Biology* **27**, 4652-4663 (2007).

89. Y. Marahrens, B. Stillman, A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* **255**, 817-823 (1992).

90. S. S. Walker, S. C. Francesconi, S. Eisenberg, A DNA replication enhancer in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 4665-4669 (1990).

91. S. S. Walker, A. K. Malik, S. Eisenberg, Analysis of the interactions of functional domains of a nuclear origin of replication from Saccharomyces cerevisiae. *Nucleic Acids Research* **19**, 6255-6262 (1991).

92. C. S. Newlon, J. F. Theis, The structure and function of yeast ARS elements. *Current Opinion in Genetics and Development* **3**, 752-758 (1993).

93. J. A. Belsky, H. K. Macalpine, Y. Lubelsky, A. J. Hartemink, D. M. Macalpine, Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. *Genes and Development* **29**, 212-224 (2015).

94. M. L. Visnapuu, E. C. Greene, Single-molecule imaging of DNA curtains reveals intrinsic energy landscapes for nucleosome deposition. *Nature Structural and Molecular Biology* **16**, 1056-1062 (2009).

95. D. A. Natale, C. J. Li, W. H. Sun, M. L. DePamphilis, Selective instability of Orc1 protein accounts for the absence of functional origin recognition complexes during the M-G1 transition in mammals. *EMBO Journal* **19**, 2728-2738 (2000).

96. M. L. DePamphilis, Cell cycle dependent regulation of the origin recognition complex. *Cell Cycle* **4**, 70-79 (2005).

97. M. L. DePamphilis, The 'ORC cycle': A novel pathway for regulating eukaryotic DNA replication. *Gene* **310**, 1-15 (2003).

98. J. F. X. Diffley, J. H. Cocker, S. J. Dowell, A. Rowley, Two steps in the assembly of complexes at yeast replication origins in vivo. *Cell* **78**, 303-316 (1994).

99. B. P. Duncker, I. N. Chesnokov, B. J. McConkey, The origin recognition complex protein family. *Genome biology* **10**, 214 (2009).

100. A. J. Kuo *et al.*, The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* **484**, 115-119 (2012).

101. P. Müller *et al.*, The conserved bromo-adjacent homology domain of yeast Orc1 functions in the selection of DNA replication origins within chromatin. *Genes and Development* **24**, 1418-1433 (2010).

102. M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner, H. L. Ploegh, Sortagging: A versatile method for protein labeling. *Nature Chemical Biology* **3**, 707-708 (2007).

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

104. D. Duzdevich, S. Redding, E. C. Greene, DNA dynamics and single-molecule biology. *Chemical Reviews* **114**, 3072-3086 (2014).

105. F. Wang *et al.*, The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion. *Nature Structural and Molecular Biology* **20**, 174-181 (2013).

106. J. Gros, S. Devbhandari, D. Remus, Origin plasticity during budding yeast DNA replication in vitro. *EMBO Journal* **33**, 621-636 (2014).

107. K. F. On *et al.*, Prereplicative complexes assembled in vitro support origin-dependent and independent DNA replication. *EMBO Journal* **33**, 605-620 (2014).

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

109. K. Van Eunen *et al.*, Measuring enzyme activities under standardized in vivo-like conditions for systems biology. *FEBS Journal* **277**, 749-760 (2010).

110. N. L. Bogenschutz, J. Rodriguez, T. Tsukiyama, Initiation of DNA replication from non-canonical sites on an origin-depleted chromosome. *PLoS ONE* **9**, (2014).

111. B. Wang *et al.*, The essential role of Saccharomyces cerevisiae CDC6 nucleotide-binding site in cell growth, DNA synthesis, and Orc1 association. *Journal of Biological Chemistry* **274**, 8291-8298 (1999).

112. M. Takehara, M. Makise, H. Takenaka, T. Asano, T. Mizushima, Analysis of mutant origin recognition complex with reduced ATPase activity in vivo and in vitro. *Biochemical Journal* **413**, 535-543 (2008).

113. N. Y. Yao, M. O'Donnell, The RFC clamp loader: structure and function. *Sub-cellular biochemistry* **62**, 259-279 (2012).

114. F. Bleichert, M. R. Botchan, J. M. Berger, Crystal structure of the eukaryotic origin recognition complex. *Nature* **519**, 321-326 (2015).

115. L. S. Drury, G. Perkins, J. F. X. Diffley, The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. *Current Biology* **10**, 231-240 (2000).

116. C. Liang, B. Stillman, Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in cdc6 mutants. *Genes and Development* **11**, 3375-3386 (1997).

117. S. Mimura, T. Seki, S. Tanaka, J. F. X. Diffley, Phosphorylation-dependent binding of mitotic cyclins to Cdc6 contributes to DNA replication control. *Nature* **431**, 1118-1123 (2004).

118. M. Oehlmann, A. J. Score, J. J. Blow, The role of Cdc6 in ensuring complete genome licensing and S phase checkpoint activation. *Journal of Cell Biology* **165**, 181-190 (2004).

119. S. Piatti, C. Lengauer, K. Nasmyth, Cdc6 is an unstable protein whose de novo synthesis in G1 is important for the onset of S phase and for preventing a 'reductional' anaphase in the budding yeast Saccharomyces cerevisiae. *EMBO Journal* **14**, 3788-3799 (1995).

120. A. Dillin, J. Rine, Separable functions of ORC5 in replication initiation and silencing in Saccharomyces cerevisiae. *Genetics* **147**, 1053-1062 (1997).

121. C. A. Fox, S. Loo, A. Dillin, J. Rine, The origin recognition complex has essential functions in transcriptional silencing and chromosomal replication. *Genes and Development* **9**, 911-924 (1995).

122. M. Foss, F. J. McNally, P. Laurenson, J. Rine, Origin recognition complex (ORC) in transcriptional silencing and DNA replication in S. cerevisiae. *Science* **262**, 1838-1844 (1993).

123. B. Özaydin, J. Rine, Expanded roles of the origin recognition complex in the architecture and function of silenced chromatin in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **30**, 626-639 (2010).

124. K. Shimada, S. M. Gasser, The Origin Recognition Complex Functions in Sister-Chromatid Cohesion in Saccharomyces cerevisiae. *Cell* **128**, 85-99 (2007).

125. J. Sun *et al.*, Structural and mechanistic insights into Mcm2-7 double-hexamer assembly and function. *Genes and Development* **28**, 2291-2303 (2014).

126. S. Vijayraghavan, A. Schwacha, The eukaryotic Mcm2-7 replicative helicase. *Sub-cellular biochemistry* **62**, 113-134 (2012).

127. J. M. Bailis, D. D. Luche, T. Hunter, S. L. Forsburg, Minichromosome maintenance proteins interact with checkpoint and recombination proteins to promote S-phase genome stability. *Molecular and Cellular Biology* **28**, 1724-1738 (2008).

128. F. L. Tsai *et al.*, Mcm2-7 is an active player in the DNA replication checkpoint signaling cascade via proposed modulation of its DNA gate. *Molecular and Cellular Biology* **35**, 2131-2143 (2015).

129. M. O'Donnell, H. Li, The eukaryotic replisome goes under the microscope. *Current Biology* **26**, R228 (2016).

130. C. Caillat, A. Perrakis, Cdt1 and geminin in DNA replication initiation. *Sub-cellular biochemistry* **62**, 71-87 (2012).

131. A. Costa *et al.*, DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. *eLife* **3**, 1-17 (2014).

132. Z. Yuan *et al.*, Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. *Nature Structural and Molecular Biology* **23**, 217-224 (2016).

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

134. C. S. Detweiler, J. J. Li, Cdc6p establishes and maintains a state of replication competence during G1 phase. *Journal of Cell Science* **110**, 753-763 (1997).

135. J. L. Bowers, J. C. W. Randell, S. Chen, S. P. Bell, ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. *Molecular Cell* **16**, 967-978 (2004).

136. S. Donovan, J. Harwood, L. S. Drury, J. F. X. Diffley, Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 5611-5616 (1997).

137. S. Tanaka *et al.*, CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature* **445**, 328-332 (2007).

138. R. E. Georgescu *et al.*, Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nature Structural and Molecular Biology* **21**, 664-670 (2014).

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

140. H. Masumoto, A. Sugino, H. Araki, Dpb11 controls the association between DNA polymerases α and ε and the autonomously replicating sequence region of budding yeast. *Molecular and Cellular Biology* **20**, 2809-2817 (2000).

141. A. C. Simon *et al.*, A Ctf4 trimer couples the CMG helicase to DNA polymerase ε in the eukaryotic replisome. *Nature* **510**, 293-297 (2014).

142. F. Villa *et al.*, Ctf4 Is a Hub in the Eukaryotic Replisome that Links Multiple CIP-Box Proteins to the CMG Helicase. *Molecular Cell* **63**, 385-396 (2016).

143. J. Walter, J. Newport, Initiation of eukaryotic DNA replication: Origin unwinding and sequential chromatin association of Cdc45, RPA, and DNA polymerase α. *Molecular Cell* **5**, 617-627 (2000).

144. A. K. Bielinsky, S. A. Gerbi, Chromosomal ARS1 has a single leading strand start site. *Molecular Cell* **3**, 477-486 (1999).

145. A. K. Bielinsky, S. A. Gerbi, Discrete start sites for DNA synthesis in the yeast ARS1 origin. *Science* **279**, 95-98 (1998).

146. A. B. Loveland, S. Habuchi, J. C. Walter, A. M. Van Oijen, A general approach to break the concentration barrier in single-molecule imaging. *Nature Methods* **9**, 987-992 (2012).

147. M. D. Sekedat *et al.*, GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Molecular Systems Biology* **6**, (2010).

148. D. S. Johnson, L. Bai, B. Y. Smith, S. S. Patel, M. D. Wang, Single-Molecule Studies Reveal Dynamics of DNA Unwinding by the Ring-Shaped T7 Helicase. *Cell* **129**, 1299-1309 (2007).

149. H. Yardimci, A. B. Loveland, S. Habuchi, A. M. Van Oijen, J. C. Walter, Uncoupling of Sister Replisomes during Eukaryotic DNA Replication. *Molecular Cell* **40**, 834-840 (2010).

150. R. C. Alver, G. S. Chadha, J. J. Blow, The contribution of dormant origins to genome stability: From cell biology to human genetics. *DNA Repair* **19**, 182-189 (2014).

151. B. D. Pope, D. M. Gilbert, The replication domain model: Regulating replicon firing in the context of large-scale chromosome architecture. *Journal of Molecular Biology* **425**, 4690-4695 (2013).

152. J. C. Rivera-Mulia, D. M. Gilbert, Replicating Large Genomes: Divide and Conquer. *Molecular Cell* **62**, 756-765 (2016).

153. C. Collart, G. E. Allen, C. R. Bradshaw, J. C. Smith, P. Zegerman, Titration of four replication factors is essential for the Xenopus laevis midblastula transition. *Science* **341**, 893-896 (2013).

154. C. Köhler *et al.*, Cdc45 is limiting for replication initiation in humans. *Cell Cycle* **15**, 974-985 (2016).

155. K. Yoshida, A. Poveda, P. Pasero, Time to be versatile: Regulation of the replication timing program in budding yeast. *Journal of Molecular Biology* **425**, 4696-4705 (2013).

156. A. M. Woodward *et al.*, Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. *Journal of Cell Biology* **173**, 673-683 (2006).

157. J. H. Taylor, Increase in DNA replication sites in cells held at the beginning of S phase. *Chromosoma* **62**, 291-300 (1977).

158. M. A. Kuipers *et al.*, Highly stable loading of Mcm proteins onto chromatin in living cells requires replication to unload. *Journal of Cell Biology* **192**, 29-41 (2011).

159. T. Pape *et al.*, Hexameric ring structure of the full-length archael MCM protein complex. *EMBO Reports* **4**, 1079-1083 (2003).

160. J. M. Dewar, M. Budzowska, J. C. Walter, The mechanism of DNA replication termination in vertebrates. *Nature* **525**, 345-350 (2015).

161. M. Maric, T. Maculins, G. De Piccoli, K. Labib, Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science* **346**, (2014).

162. S. P. Moreno, R. Bailey, N. Campion, S. Herron, A. Gambus, Polyubiquitylation drives replisome disassembly at the termination of DNA replication. *Science* **346**, 477-481 (2014).

163. T. Maculins, P. J. Nkosi, H. Nishikawa, K. Labib, Tethering of SCF/Dia2 to the Replisome Promotes Efficient Ubiquitylation and Disassembly of the CMG Helicase. *Current Biology* **25**, 2254-2259 (2015).

164. D. T. Long, V. Joukov, M. Budzowska, J. C. Walter, BRCA1 promotes unloading of the CMG Helicase from a stalled DNA replication fork. *Molecular Cell* **56**, 174-185 (2014).

165. C. Santocanale, K. Sharma, J. F. X. Diffley, Activation of dormant origins of DNA replication in budding yeast. *Genes and Development* **13**, 2360-2364 (1999).

166. J. B. Grimm *et al.*, A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nature Methods* **12**, 244-250 (2015).

167. J. Zhang *et al.*, DNA interstrand cross-link repair requires replication-fork convergence. *Nature Structural and Molecular Biology* **22**, 242-247 (2015).

168. S. Sofueva *et al.*, Ultrafine anaphase bridges, broken DNA and illegitimate recombination induced by a replication fork barrier. *Nucleic Acids Research* **39**, 6568-6584 (2011).

169. J. P. Duxin, J. M. Dewar, H. Yardimci, J. C. Walter, Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell* **159**, 349-357 (2014).

170. B. A. Berghuis *et al.*, Strand separation establishes a sustained lock at the Tus-Ter replication fork barrier. *Nature Chemical Biology* **11**, 579-585 (2015).

171. N. B. Larsen, E. Sass, C. Suski, H. W. Mankouri, I. D. Hickson, The Escherichia coli Tus-Ter replication fork barrier causes site-specific DNA replication perturbation in yeast. *Nature communications* **5**, 3574 (2014).

172. A. Brambati, A. Colosio, L. Zardoni, L. Galanti, G. Liberi, Replication and transcription on a collision course: Eukaryotic regulation mechanisms and implications for DNA stability. *Frontiers in Genetics* **6**, (2015).

173. J. Gros *et al.*, Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. *Molecular Cell* **60**, 797-807 (2015).

174. S. K. Powell *et al.*, Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. *EMBO Journal* **34**, 531-543 (2015).

175. B. D. Pope *et al.*, Replication-timing boundaries facilitate cell-type and species-specific regulation of a rearranged human chromosome in mouse. *Human Molecular Genetics* **21**, 4162-4170 (2012).

176. E. E. Arias, J. C. Walter, Replication-dependent destruction of Cdt1 limits DNA replication to a single round per cell cycle in Xenopus egg extracts. *Genes and Development* **19**, 114-126 (2005).

177. M. L. Hoang *et al.*, Structural changes in Mcm5 protein bypass Cdc7-Dbf4 function and reduce replication origin efficiency in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **27**, 7594-7602 (2007).

178. M. D. Ramer *et al.*, Dbf4 and Cdc7 proteins promote DNA replication through interactions with distinct Mcm2-7 protein subunits. *Journal of Biological Chemistry* **288**, 14926-14935 (2013).

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

180. S. Chen, S. P. Bell, CDK prevents Mcm2-7 helicase loading by inhibiting Cdt1 interaction with Orc6. *Genes and Development* **25**, 363-372 (2011).

181. N. Dhingra, I. Bruck, S. Smith, B. Ning, D. L. Kaplan, Dpb11 protein helps control assembly of the Cdc45•Mcm2-7•GINS replication fork helicase. *Journal of Biological Chemistry* **290**, 7586-7601 (2015).

182. S. MacNeill, Composition and dynamics of the eukaryotic replisome: a brief overview. *Sub-cellular biochemistry* **62**, 1-17 (2012).

183. S. Onesti, S. A. MacNeill, Structure and evolutionary origins of the CMG complex. *Chromosoma* **122**, 47-53 (2013).

184. L. Pellegrini, The Pol α-primase complex. *Sub-cellular biochemistry* **62**, 157-169 (2012).

185. R. Okazaki, T. Okazaki, K. Sakabe, K. Sugimoto, A. Sugino, Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proceedings of the National Academy of Sciences of the United States of America* **59**, 598-605 (1968).

186. K. Sugimoto, T. Okazaki, R. Okazaki, Mechanism of DNA chain growth, II. Accumulation of newly synthesized short chains in E. coli infected with ligase-defective T4 phages. *Proceedings of the National Academy of Sciences of the United States of America* **60**, 1356-1362 (1968).

187. R. L. Perera *et al.*, Mechanism for priming DNA synthesis by yeast DNA Polymerase α. *eLife* **2013**, (2013).

188. L. M. Dieckman, B. D. Freudenthal, M. T. Washington, PCNA structure and function: insights from structures of PCNA complexes and post-translationally modified PCNA. *Sub-cellular biochemistry* **62**, 281-299 (2012).

189. W. Du, M. E. Stauffer, B. F. Eichman, Structural biology of replication initiation factor Mcm10. *Sub-Cellular Biochemistry* **62**, 197-216 (2012).

190. A. A. Larrea *et al.*, Genome-wide model for the normal eukaryotic DNA replication fork. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 17674-17679 (2010).

191. T. A. Kunkel, P. M. Burgers, Dividing the workload at a eukaryotic replication fork. *Trends in Cell Biology* **18**, 521-527 (2008).

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

193. Z. F. Pursell, I. Isoz, E. B. Lundström, E. Johansson, T. A. Kunkel, Yeast DNA polymerase ε participates in leading-strand DNA replication. *Science* **317**, 127-130 (2007).

194. R. E. Georgescu *et al.*, Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. *eLife* **2015**, (2015).

195. O. Yurieva, M. O'Donnell, Reconstitution of a eukaryotic replisome reveals the mechanism of asymmetric distribution of DNA polymerases. *Nucleus* **7**, 360-368 (2016).

196. I. Miyabe, T. A. Kunkel, A. M. Carr, The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved. *PLoS Genetics* **7**, (2011).

197. R. E. Johnson, R. Klassen, L. Prakash, S. Prakash, A Major Role of DNA Polymerase δ in Replication of Both the Leading and Lagging DNA Strands. *Molecular Cell* **59**, 163-175 (2015).

198. P. M. J. Burgers, D. Gordenin, T. A. Kunkel, Who Is Leading the Replication Fork, Pol ε or Pol δ? *Molecular Cell* **61**, 492-493 (2016).

199. J. L. Stodola, P. M. Burgers, Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale. *Nature Structural and Molecular Biology* **23**, 402-408 (2016).

200. J. A. St Charles, S. E. Liberti, J. S. Williams, S. A. Lujan, T. A. Kunkel, Quantifying the contributions of base selectivity, proofreading and mismatch repair to nuclear DNA replication in Saccharomyces cerevisiae. *DNA Repair* **31**, 41-51 (2015).

201. Y. Seol, K. C. Neuman, The dynamic interplay between DNA topoisomerases and DNA topology. *Biophysical Reviews* **8**, 221-231 (2016).

202. A. Prakash, G. E. O. Borgstahl, The structure and function of replication protein A in DNA replication. *Sub-Cellular Biochemistry* **62**, 171-196 (2012).

203. J. Fan, N. P. Pavletich, Structure and conformational change of a replication protein A heterotrimer bound to ssDNA. *Genes and Development* **26**, 2337-2347 (2012).

204. T. J. Lockett, A bacteriophage λ DNA purification procedure suitable for the analysis of DNA from either large of multiple small lysates. *Analytical Biochemistry* **185**, 230-234 (1990).

205. J. Schindelin *et al.*, Fiji: An open-source platform for biological-image analysis. *Nature Methods* **9**, 676-682 (2012).

206. C. E. Aitken, R. A. Marshall, J. D. Puglisi, An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical Journal* **94**, 1826-1835 (2008).

207. A. Mehanna, J. F. X. Diffley, Pre-replicative complex assembly with purified proteins. *Methods* **57**, 222-226 (2012).

APPENDIX I

An Overview of Eukaryotic DNA Replication

Every time a cell divides it must generate a copy of its genomic DNA. This process of DNA replication is fundamentally the same in all living things (**Fig. AI.1**). Replication initiates at specific loci on the DNA called origins of replication, or simply origins. Origins are defined when protein initiation factors bind to them. For replication to begin and the DNA sequence information to be copied, the two complementary strands that make up double stranded DNA (dsDNA) must separate locally to expose the sequences that will serve as templates for the synthesis of new DNA. Many proteins are required for progressively moving along and separating the parental DNA while generating new daughter strands; the collective assembly of these proteins is termed a replisome. Some proteins become incorporated into a replisome at replication initiation and remain as components of that replisome until replication termination. Other components are in equilibrium with replisomes, with individual proteins turning over. Typically, during the early stages of a replication initiation pathway, two replisomes are assembled at an origin, facing in opposite directions along the DNA axis. This pair of sister replisomes separates as replication begins. The two replisomes move away from each other and from the site of replication initiation along the substrate parental DNA. The region in between the diverging replisomes consists of two

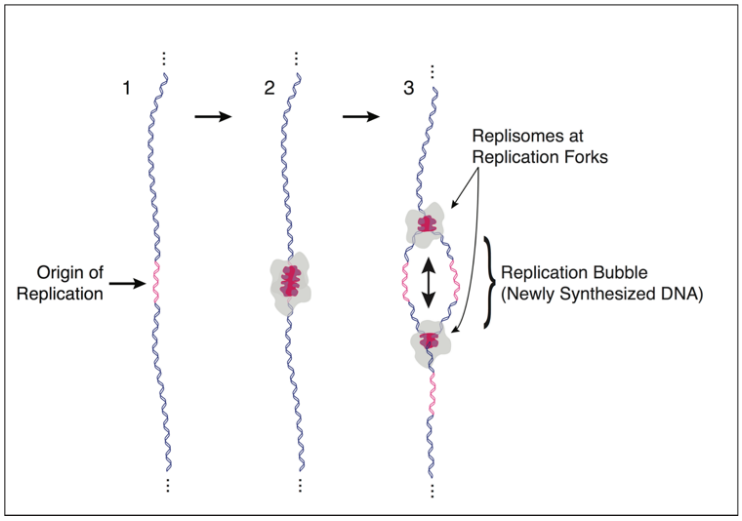

Figure AI.1 Generalized Schematic of DNA Replication
(1) An origin site is the location on the DNA where replication will begin. It may or may not be sequence-defined.
(2) Initiation factors and then the replication machinery assemble at the origin.
(3) Typically, a pair of replisomes is assembled at an origin and replication proceeds bidirectionally (that is, the replisomes move away from each other, in opposite directions). Each replisome consists of the machinery needed to replicate both strands of DNA and occupies a progressing replication fork, the local site of DNA synthesis.

identical double helices of dsDNA, each harboring one newly synthesized strand and one strand that served as the template for synthesis. These daughter strands of dsDNA form the so-called replication bubble, with two diverging replication forks occupied by active replisomes at either end. DNA is synthesized in the 5' to 3' direction on any given template strand. However, dsDNA is antiparallel, with complementary strands running in opposite directions. This results in continuous leading strand synthesis and discontinuous lagging strand synthesis at each replication fork.

Although this overall picture of replication applies very broadly, there are differences in the proteins involved and how they operate from species to species and especially domain to domain (*73*). The work presented here is concerned with eukaryotic genomic DNA replication, with the budding yeast *Saccharomyces cerevisiae* serving as a model system. Though the focus is on research performed

using budding yeast, replication proteins and processes are very highly conserved among eukaryotes (*23, 24*).

A distinctive feature of eukaryotic replication is that the initiation pathway is intimately coupled with the cell cycle. Many of the genes required for replication were first identified in screens for temperature sensitive yeast mutants that arrest at defined points in the cell cycle, as characterized by microscopic observation (*19-21*). For example, mutations that arrest cells prior to the separation of chromosomes and their partitioning to daughter cells were characterized by imaging cells that had arrested with their otherwise duplicated chromosomes visibly unseparated. Inhibition of such core processes is lethal, but by using temperature sensitive mutants the researchers were able to grow the cells at a permissive temperature and subsequently switch colonies to a non-permissive temperature and look for cell cycle arrest. That many of the genes identified in this way were subsequently found to code for replication proteins immediately shows that the replication program communicates with the cell cycle program: inhibition of replication also inhibits the cell cycle. Origins are usually defined by initiation proteins during the G1 phase of the cell cycle and the chromosomes are duplicated during S phase. By G2 phase, and prior to mitosis, the genomic DNA will have been copied once and only once. Eukaryotic cells employ a variety of mechanisms to ensure that chromosome copy number is correctly maintained,

with the chromosomes duplicated only once every cell cycle so that each daughter cell receives one copy of the genome.

The budding yeast genome harbors many hundreds of replication origins (*29*). Atypically among characterized eukaryotes, budding yeast origins can generally be identified by sequence. This was discovered during attempts to propagate extrachromosomal plasmids in budding yeast. Successful plasmid propagation was found to require the presence of specific sequences, termed autonomously replicating sequences, or ARS sites (*74, 75*). An ARS site was deemed essential for copying a plasmid and the several isolated ARS loci were identified as origins of replication (*29*). Without an ARS site, a plasmid effectively lacks an origin of replication and cannot be copied and therefore cannot propagate. However, origin sites are ultimately defined by replication initiation proteins that bind to them: if initiation proteins bind a locus and recruit the factors needed for replication to start at that site, then the site is an origin in that instance. Budding yeast ARS loci are targets for the binding of initiation factors, as discussed below and in Chapter 2. In that sense, "ARS" and "origin" are synonymous in the case of budding yeast. However, the important point is the binding of initiation factors, and in many other eukaryotes this happens largely independently of sequence (*67, 71*) and even budding yeast can use non-sequence defined origins, with local chromatin features thought to be the primary determinant other than sequence (*63, 101, 110*).

The Initiation Pathway

Origin sites across the genome are bound by the heterohexameric Origin Recognition Complex (ORC) during G1. ORC was isolated through its ATP-dependent binding to a cloned ARS sequence (*49*). The ATPase activities of ORC subunits may play important roles in how the complex binds DNA and recruits other factors (*36, 38, 39, 60*). ORC subunits harbor domains needed for ATP binding and hydrolysis, DNA binding, and chromatin interactions (*99*).

ORC bound to DNA on its own has no replication initiation activity: it must interact with the key cofactor Cdc6 (*33, 35-39, 44, 60, 118, 136*). Cdc6 is also an ATPase and its levels fluctuate throughout the cell cycle, peaking in late G1 (*36, 52, 60, 61, 111, 115, 117-119, 134*). Together, DNA-bound ORC and Cdc6 sequentially recruit two heteroheptameric Mcm2-7/Cdt1 complexes to the origin. Cdt1 associates only with correctly formed Mcm2-7 complexes in the cytoplasm, and Cdt1 harbors a nuclear localization signal that targets it for import into the nucleus (*44, 130, 176*). This ensures that only Mcm2-7 can enter the nucleus and interact with DNA-ORC-Cdc6, rather than any hexamers with incorrect stoichiometry. Cdt1 is released from Mcm2-7 after the complex is recruited to DNA by DNA-ORC-Cdc6 (*38-40, 42, 44, 60, 125*). Mcm2-7 has an opened ring-like structure in solution, and DNA-ORC-Cdc6 is thought to clamp the opened ring down onto dsDNA (*38-44, 50, 51, 125*). By unknown mechanisms, a second Mcm2-7/Cdt1 is then loaded by DNA-ORC-Cdc6, whereupon the second Cdt1,

ORC, and Cdc6 are released from the DNA and leave behind an Mcm2-7 double hexamer (*40, 42, 125*). The two hexamers face in opposite directions, away from each other along the DNA axis. The double hexamer is inactive in this state, during G1, but in downstream steps, each Mcm2-7 hexamer forms the core of the replicative helicase, the molecular machine responsible for separating dsDNA within the replisome.

Replication

At the G1-to-S transition, two important cell cycle regulated kinase activities direct the assembly of replisomes around the initially inactive Mcm2-7 double hexamers: Dbf4-dependent kinase (DDK) and the S-phase Cyclin-dependent Kinase S-CDK (*28*). DDK phosphorylates several sites on Mcm2-7. Phosphorylated Mcm2-7 is a target for binding by the accessory factors Cdc45 and Sld3 (see below) (*177-179*). It is also possible that phosphorylation partially remodels the Mcm2-7 hexamer, priming it for the significant conformational changes it must undergo during its conversion to the core component of the replicative helicase. S-CDK has many targets, though only some are directly important for the replication pathway. These include ORC, Cdc6, Sld3/7, and Sld2. When ORC is phosphorylated, it becomes inactive for Mcm2-7/Cdt1 recruitment and double hexamer loading (*180*). This ensures that once S-phase and DNA replication begin, any ORC molecules bound to newly replicated origins cannot load additional double hexamers, which could

lead to re-replication at that site. Similarly, Cdc6 is also phosphorylated and thereby targeted for degradation: this is another mechanism to ensure that double hexamers are only loading during G1 but not S phase (*115, 134*). An additional fail-safe is the presence of the S-CDK inhibitor Sic1 specifically during G1 (*28*). S-CDK activity and changes in S-CDK activity therefore ensure that Mcm2-7 double hexamers are only loaded during G1 phase, in inactive form.

S-CDK also phosphorylates the replication cofactors Sld2 and Sld3 (*137*). Phosphorylated Sld2 and Sld3 bind Dpb11 to form an Sld2-Dpb11-Sld3/7 complex (*181*). (Sld3 and Sld7 form a single complex.) It is unclear whether *in vivo*, the Sld2-Dpb11-Sld3/7 complex forms in solution or only in the context of an assembling replisome, but in either case its primary function appears to be recruitment of the leading strand polymerase, Polε, through an Sld2-Polε interaction (*140*). Sld2, Dpb11, and Sld3/7 are accessory factors required only during replisome assembly, and they are not part of the active replisome (*182*).

DDK-phosphorylated Mcm2-7 is bound by Cdc45 and the heterotetrameric GINS complex. Each Mcm2-7 hexamer in the double hexamer associates with one Cdc45 and one GINS to form a Cdc45-Mcm2-7-GINS (CMG) complex. The CMG is the eukaryotic replicative helicase (*46, 131, 132, 183*). At an unknown point on the pathway to replication initiation, through unknown mechanisms, the Mcm2-7 ring is opened at the Mcm2-Mcm5 interface (the Mcm2-Mcm5 gate) and the ring is assumed to close back down around single-stranded DNA (ssDNA). Additional

channels in the CMG helicase created by the presence of Cdc45 and GINS may serve to handle the other, excluded ssDNA of the parental DNA. The active CMG moves in a 3' to 5' direction encircling the parental ssDNA that is the template for leading strand synthesis (*48*).

Each stretch of DNA synthesis begins from a short RNA primer, generated by the Polα-primase complex. Polα-primase synthesizes RNA, then a small amount of DNA before a primary polymerase takes over (*184*). The primary polymerase on the leading strand is Polε. Because leading strand synthesis is continuous, Polα-primase is only needed once, at the initiation of leading strand synthesis. The lagging strand polymerase is Polδ, and the dynamics are more complicated. Lagging strand synthesis is discontinuous, resulting in a series of synthesized stretches called okazaki fragments (*185, 186*). Therefore, Polα-primase is loaded for each stretch, and then locally displaced by Polδ, which synthesizes the rest of the okazaki fragment. This happens reiteratively for each okazaki fragment (*184, 187*). Another replisome component, the Ctf4 homotrimer, couples the CMG to Polε and Polα (*141, 142*). The ring-like PCNA complex is a processivity factor that encircles DNA and interacts with the polymerases (among many other factors). This helps keep the polymerases at the site of DNA synthesis (*188*). The RFC "clamp-loader" is responsible for loading PCNA onto DNA, and in the case of lagging strand synthesis, this may be required for each okazaki fragment (*113*). MCM10 may also link the Mcm2-7 hexamer to Polα, and it has

other poorly-characterized roles in stabilizing the replisome (*189*). This may be key for the reiterative use of Polα/Primase on the lagging strand.

The leading and lagging strand designations of the polymerases were assigned by culturing cells harboring error-prone mutants of one or the other polymerase and looking for higher mutation rates on nascent DNA known to have originated from leading or lagging strand synthesis, based on directionality relative to an origin site (*190-193*). Additional structural and genetic work agrees with the designations (*46, 138, 194-196*), however there remains some controversy over the potential role of Polδ in leading strand synthesis (*197, 198*). The polymerases consist of multiple subunits and multiple activities. Polα/Primase can synthesize RNA on a DNA template without a primer and DNA on an DNA template, but it is relatively error-prone. However, the region initially synthesized by Polα/Primase is subsequently excised by Dna2/Fen1 and resynthesized by a primary polymerase to consist entirely of DNA, and a ligase seals the nicks to yield a continuous DNA backbone (*199*). Polε and Polδ synthesize DNA on a DNA template, require a primer, and have proofreading functionality. Combined with the DNA mismatch repair machinery, thought to trail the replisome, the DNA replication error rate is on the order of one in a billion (*200*).

As the two strands of dsDNA are pried apart by the CMG in a progressing replisome, the DNA ahead of the replication fork becomes stressed because unwinding of the double helix forces the as-yet-unwound region to rotate. This

topological stress would inevitably build up in the context of a chromosome and stall the replisome. The problem is solved by topoisomerases that cleave the DNA in a controlled manner either on one backbone, thereby allowing free rotation about a single bond, or both backbones, to relieve supercoils (*201*). The action of CMG also generates regions of transient ssDNA, especially on the lagging strand. The single-stranded DNA binding heterotrimeric protein Replication Factor A (RPA) is essential for protecting this exposed ssDNA. Many RPAs will bind along a region of exposed ssDNA and protect it from damage; RPA also disrupts or otherwise prevents the formation of secondary ssDNA structures that could potentially disrupt the intricate processes at an active replisome (*143, 202, 203*). This brief overview only covers the main steps of the eukaryotic replication pathway and the core components needed to assemble and activate replisomes for DNA replication. Even in this very limited format, however, the complexity of replication is apparent, and much remains to be explored and understood.

APPENDIX II

Experimental Procedures

Nanofabrication

Pairs of precisely-positioned 1.4mm holes are drilled into fused silicon dioxide microscope slides (G. Finkenbeiner, Inc.) using diamond-coated drill bits (McMaster-Carr) fitted to a drill-press (Servo). The slides are immersed in water during drilling to wash away dust and minimize glass breakage. Drilled slides are cleaned with piranha solution (3:1 sulfuric acid:30% hydrogen peroxide) (Fisher Scientific) and thoroughly washed with deionized and purified water (Biocel- or MilliQ-grade, EMD MilliPore, used throughout). Each slide is sequentially layered with 3% w/v in anisole 25k or similar average molecular weight PMMA (Polymer Source), 1.5% w/v 495k average molecular weight PMMA (Microchem), and aquaSAVE 53za (Mitsubishi Rayon) in a spin coater (Laurell). The aquaSave forms a dry conducting layer needed for electron-beam lithography and can conveniently be washed away with water. Alternatively, a thin (tens of nanometers) layer of aluminum sputtered onto the PMMA works equally well but requires several extra steps prior to chromium deposition (see below).

The features to be nanopatterned are created in DesignCAD “write files” and loaded into the NPGS software system (JC Nabity Lithography Systems) used to direct the electron microscope (FEI). The software executes the desired pattern

in the layered PMMA; appropriate electron dosages are determined beforehand with dose-tests on each of the desired features. PMMA is developed in a 1:3 solution of isopropanol:MIBK (MicroChem). Electron-beam evaporation (Semicore) of a chromium slug deposits a 250nm layer of metal onto the slide. PMMA liftoff is in boiling acetone for 10min.

Flowcell Assembly and Usage

A narrow channel is excised in double-sided tape (3M) and the tape is placed over the slide surface, the channel connecting the two drilled holes and encompassing the region with the nanopatterns. A coverslip (Fisher Scientific) is layered onto the tape and firmly compressed by hand. The slide and coverslip are then sandwiched between standard glass microscope slides and clipped down on all four edges with binder clips. The flowcell is cured in a vacuum oven (VWR) at 175^{o}C for ~30min. to seal the tape. The drilled holes are fitted with nanoports to connect to downstream microfluidics using standard fittings and tubing (IDEX).

Lipids (Avanti Polar Lipids) are dissolved into chloroform (ethanol-stabilized, EMD MilliPore) to make a master mix (MM) of 100mg/ml DOPC (1,2-dioleoyl-*sn*-glycero-3-phosphocholine), 10mg/ml PEG-2000 DOPE (1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-2000]), and 0.5mg/ml biotinyl cap PE (1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-N-(cap biotinyl)). Note that the PEGylated lipids are only essential for experiments with QDs. 200µl

of MM is measured into a thoroughly cleaned 2ml glass vial (Thermo Scientific) and the chloroform gently evaporated with a stream of purified anhydrous nitrogen. The lipid film is placed under vacuum at least overnight and rehydrated in 2ml lipid buffer (10mM Tris-HCl, pH 7.6 or 7.8, 100mM NaCl) for at least 30min. The rehydrated lipids are sonicated with a microtip in 20s bursts (Misonix) on ice until the solution goes clear, or, in the absence of PEGylated lipids, for the same amount of time it would take the PEGylated mix to go clear. The solution is filtered with a syringe-driven 2µm PES filter (Millex) and stored at 4ºC for up to two weeks.

Flowcells are prepared for experimentation by depositing a lipid bilayer with a 20min. incubation of a 5% solution of lipids, as prepared above, in lipid buffer. Excess lipids are washed out and a second lipid incubation can be included for a high-quality surface. A 20% solution of a 1mg/ml PBS stock of anti-DIG F_{ab} fragments (Roche) is incubated in the flowcell for 30min. for experiments requiring double-tethered curtains. Excess material is flushed out and the flowcell is equilibrated with BSA buffer (25mM Tris-HCl, pH 7.6, 1mM $MgCl_2$, 1mM DTT [Sigma-Aldrich], and 0.8mg/ml BSA [Sigma-Aldrich]). A 0.5% solution of a 1mg/ml PBS stock of streptavidin (Sigma-Aldrich) in BSA buffer is then incubated for 30min. Excess material is flushed out in BSA buffer and the desired concentration of DNA is introduced. Varying DNA concentration and incubation time can be used to control the density of DNA curtains.

Slides can be salvaged by soaking in alcohol, removing the coverslip and tape, followed by cleaning in 3% Hellma II cleaning solution (Hellmanex), 1M NaOH, and three thorough rinses in water.

DNA Substrates

λ_{ARS1}[‡]

A 231bp DNA fragment harboring the 193bp *ARS1* sequence was PCR amplified from the pKS-*ARS1* plasmid and cloned into the unique Xho1 and Nhe1 restriction sites in the λ phage genome (New England Biolabs). The cloned product was then packaged into phage particles using phage extract (MaxPlax, Epicentre). Plaques were generated on LE392 *E. coli* bacterial lawns (Epicentre), and screened for the *ARS1* insert. A screened plaque was used as a phage source to purify λ_{ARS1} DNA by lytic growth (see below). The final λ_{ARS1} DNA substrate is 47,551bp in length and the center of the *ARS1* site is at 33,647bp from the left end of the phage genome. DNA substrates were end-labeled by annealing and ligating biotinylated and DIG-tagged ssDNA handles (IDT) at 42°C to the naturally-occurring 12bp overhangs of λ at a 1:14 molar ratio of λ_{ARS1} DNA to each handle. Excess oligonucleotide was separated out by gel filtration (Sephacryl S-1000 or S-200, GE Healthcare).

[‡] All paragraphs or sections marked with "‡" are excerpts reprinted from reference (*133*) with permission from Elsevier.

$\lambda_{2X\mathit{ARS1}}$

The λ_{ARS1}-DNA-harboring phage generated in (*133*) was used to grow plaques on LE392 *E. coli* bacterial lawns. The presence of the *ARS1* insert was confirmed by PCR and sequencing. A screened plaque was used to purify λ_{ARS1} by lytic growth (see below). Another copy of the *ARS1* sequence was PCR amplified from the pKS-*ARS1* plasmid with SpeI-cleavable ends and cloned into the unique XbaI site of λ_{ARS1} (restriction enzymes from New England Biolabs). The cloned $\lambda_{2X\mathit{ARS1}}$ product was then packaged into phage particles using phage extract, and the phage were again plated, screened, and purified. The product was end-labeled as above. Excess handles were separated out by treating the reaction mix with 30% w/v 5,000 average molecular weight polyethylene glycol (PEG_{5000}) (Sigma-Aldrich) at 4°C overnight on a rotator. Centrifugation at 17,000xg on a tabletop centrifuge was used to precipitate the large $\lambda_{2X\mathit{ARS1}}$ DNA, which was washed with 70% ethanol and resuspended in TE150.

Lytic Growth of Phage to Purify DNA

This protocol was derived from (*204*) with significant unique modifications specifically developed for these experiments. 0.4ml of an overnight culture of LE392 is added to 0.4ml of (10mM $CaCl_2$, 10mM $MgCl_2$) and inoculated with a single screened plaque. Extracting the phage beforehand yields the same results as using a plug directly. The inoculate is added to 200ml of NZCYM broth in a 2l

flask; vigorous aeration is crucial. The culture is grown at 37°C at 125rpm for 4hrs., at which point it should be cloudy. OD_{600} rises above 1 and then drops rapidly. When it begins to climb again, 0.5ml chloroform is added and the mix is transferred to a 200ml glass bottle with dry NaCl to 1M and incubated on ice for 10min. The mix is centrifuged at 12,000xg for 10min. to remove cell debris. Supernatant is collected and centrifuged again at 12,000xg for 8min. to remove remaining cell debris. PEG_{8000} is added to 10% w/v, mixed thoroughly, and incubated on ice for 30min. The mix is centrifuged at 12,000xg for 15min. to pellet the phage, which is then resuspended in 10ml phage dilution buffer (10mM Tris-HCl, pH 8.0, 100mM NaCl, and 10mM $MgCl_2$). RNase and DNase (Sigma-Aldrich) are added to 20µg/ml and 5µg/ml final concentrations and the mix is incubated at 37°C for 30min. The λ DNA remains protected at this stage in phage capsids. 10ml of 0.3M Tris-HCl, pH 7.6, EDTA to 100mM (6ml), SDS to 1.25% (3.7ml), and proteinase K (New England Biolabs) to 50µg/ml are added and the mix is incubated for 10min. at 65°C. 10ml of ice-cold 3M KAc is added and the mix is incubated on ice for 10min. The KAc forces the SDS and associated proteins out of solution, the potassium salt of SDS being insoluble, forming a thick white gel. The gel is centrifuged out at 8,000xg for 10min. Isopropanol is used to precipitate the DNA, which is washed with 70% ethanol, resuspended, phenol:chlorofom extracted, reprecipitated, and finally resuspended in ligation buffer in preparation for handle ligation as described above.

Single-molecule Data

Images were acquired using NIS-Elements software and a TE2000-U inverted microscope (Nikon) fitted with a 100x plan-APO oil immersion objective (Nikon) and a Photometrics Cascade II CCD camera (Photometrics). TIRFM lasers included a 200mW 488nm blue line (Coherent), 200mW 532nm green line (Coherent Sapphire), and 100mW 640nm red line (CrystaLaser). Laser intensities at the prism face were calibrated for each experiment and generally fell in the tens of mWs range.

Images were acquired at 5Hz with 200ms integration time unless otherwise noted. Specifically, Cdc6 lifetime experiments were performed at 12.5Hz with 80ms integration time, and replication experiments were performed at 1 frame per 60s or 80s with 200ms integration time. Raw TIFF images were imported as image stacks into Fiji (*205*). Kymograms were generated directly from image stacks by defining a 1 pixel (px) wide sub-stack and lining up the individual component images along the time axis. To localize fluorescently labeled and DNA-bound proteins along the λ DNA substrate, a series of 11 images centered about a given time point was averaged for intensity.[‡]

Replication experiments with Mcm2-7$^{Mcm4\text{-}SNAP\text{-}JF646}$ were performed at 1 frame per 100s or 110s with 667ms integration time using ~30mW laser power at the prism (CrystaLaser 640nm, 100mW).

Bulk Pre-RC Assembly Assays‡

A total of 480fmol ORC, 960fmol Cdc6, and 1.44pmol Mcm2-7/Cdt1 were sequentially added to an 80µl reaction containing 240fmol bead-coupled (Dynabeads M-280 Streptavidin, Life Technologies) 7.4kbp linear p*ARS1*-Nco-Nco plasmid DNA in helicase loading buffer (25mM HEPES-KOH, pH 7.6, 12mM MgAc, 50µM ZnAc, 300mM KGlut, 3mM ATP, 1mM DTT, 4mg/ml BSA, and 0.03mg/ml biotin). In reactions involving QD or streptavidin labeling, the target protein was incubated with a four times molar excess of QD streptavidin conjugate (Life Technologies) for a minimum of 30min. on ice prior to the loading reaction. The reaction mix was incubated at 25°C at 1,200rpm for 30min. in a Thermomixer (Eppendorf). Beads were washed three times with buffer H + 300mM KGlut and 0.02% NP-40, and DNA-bound proteins were eluted from the beads by incubation with 1U of DNase I per reaction for 10min. at 25°C and 1,200rpm. Eluted proteins were separated by SDS-PAGE and stained with fluorescent protein stain (Krypton, Thermo Scientific). For high-salt wash experiments, helicase loading buffer + 0.5M NaCl (instead of KGlut) was used for the second wash step.

Single-molecule Pre-RC Experiments‡

Single-molecule pre-RC experiments were conducted at room temperature in reaction buffer containing 25mM HEPES, pH 7.6, 12mM MgAc, 50µM ZnAc, 1mM DTT, 225mM KGlut, 3mM ATP, 20mM creatine phosphate (Roche), 0.04mg/ml

creatine kinase (Roche), 4mg/ml BSA (Sigma-Aldrich), 0.03mg/ml biotin (Sigma-Aldrich) (in all experiments involving biotinylated protein), and 1mM Trolox (Sigma-Aldrich) (in all experiments involving organic dyes). For salt stability measurements the standard reaction buffer was diluted by half with 1M NaCl, yielding a final concentration of 0.5M NaCl.

Prior to use, ORC-Sort-bio was incubated with a four times molar excess of QD 605 or 705 streptavidin conjugate (Life Technologies) for 1 hour on ice. In all cases, the pre-incubated samples were brought to the indicated final working concentrations of protein immediately prior to experimentation. Control experiments with ORC lacking the biotin tag show that it cannot be labeled with QD streptavidin conjugate, indicating that the labeling was specific to the biotin at the N-terminus of Orc1 (see Appendix III). In experiments where ORC and Cdc6 were incubated on a DNA curtain simultaneously, concentrated Cdc6 was added after dilution of ORC to the working volume to a final concentration four times greater than the final ORC concentration. The flowcell was pre-equilibrated with reaction buffer, and protein was injected into the reaction chamber at 100µl/min. The protein was allowed to equilibrate with the DNA for 2min., unless otherwise noted, in the absence of buffer flow, and excess protein was then flushed out at 800µl/min. In the case of double-tethered curtain experiments, buffer flow was then terminated.

Mcm2-7/Cdt1 complex made with MCM4-fSNAP-bio was prepared for microscopy as described for ORC above. After dilution to the working concentration of 10nM, ORC and Cdc6 were added to 1nM and 4nM, respectively, unless otherwise noted. For all Mcm2-7 constructs, association with DNA was observed for 15min. before excess protein was flushed out. Binding distributions and photobleaching analyses were performed after this loading reaction. No high salt wash was used in these experiments.

For experiments involving fluorescently labeled Cdc6, Cdc6-bio was prepared for microscopy as described for ORC above with the exception that an 8-times molar excess of Alexa Fluor 532 or 647 streptavidin conjugate (Life Technologies) was used.

Bulk Replication Assay[‡]

~0.1nM λ_{ARS1} coupled to magnetic beads (Dynabeads M-280 Streptavidin, Life Technologies) was incubated with 0.5nM ORC, 4nM Cdc6, and 15nM Mcm2-7/Cdt1 at room temperature for 35min. at a final volume of 200µl in reaction buffer containing 25mM HEPES, pH 7.6, 12mM MgAc, 50µM ZnAc, 1mM DTT, 300mM KGlut, 3mM ATP, 20mM creatine phosphate (Roche), 0.04mg/ml creatine kinase (Roche), 4mg/ml BSA (Sigma-Aldrich), and 0.03 mg/ml biotin (Sigma-Aldrich) (in all experiments involving biotinylated protein). 200µl water was added along with DDK to a final concentration of .0075µg/µl and incubated for 15min. Beads were

magnetized and resuspended for 1 hour in 100μl reaction buffer + 12.5µg/µl S-phase extract, 200µM of rUTP, rCTP, and rGTP, 40µM each dNTP, 23µCi dCTP [α-P32], and, as indicated, 0.05µg/µl competitor native λ (New England Biolabs). Beads were serially magnetized and washed with Wash Buffer (WB) (25mM HEPES, pH 7.6, 1mM MgAc, 300mM KGlut, 10% glycerol, and 0.02% NP-40), WB + 2 M NaCl, and WB. The beads were again magnetized and resuspended in 10µl 25mM HEPES, pH 7.6, and 150mM KGlut. This was added to 5ml scintillation fluid (ScintiVerseTM BD, Fisher Scientific) and vortexed. Scintillation readings (LS 6500, Beckman) were made for 3 x 3 minutes for each vial and background subtracted based on a blank. Results were reported as the ratio of experimental signal to a minus DDK control (generated for each experimental run).

Single-molecule Replication Experiments[‡]

A double-tethered DNA curtain was incubated with 0.5nM ORC, 4nM Cdc6, and 10nM Mcm2-7$^{Mcm4\text{-}DY549}$/Cdt1 at room temperature for 20min. in reaction buffer containing 25mM HEPES, pH 7.6, 12mM MgAc, 50µM ZnAc, 1mM DTT, 300mM KGlut, 3mM ATP, 20mM creatine phosphate, 0.04mg/ml creatine kinase, 4mg/ml BSA, and 1mM Trolox. Excess protein was flushed out and followed by DDK at a final concentration of .001µg/µl in 50% reaction buffer and incubated for 15min. Excess protein was flushed out, followed by reaction buffer with 0.1mM DTT and 1.5mM Trolox (Sigma-Aldrich) + 12.5µg/µl or 6.25µg/µl S-phase extract, 25µM

protochatechuate (Sigma-Aldrich), 50nM protochatechuate dioxygenase (Sigma-Aldrich) (*206*), 200μM of rUTP, rCTP, and rGTP (Invitrogen), 40μM each dNTP (Invitrogen), and 0.225μg/μl competitor salmon sperm DNA (Sigma-Aldrich). In general we observed a significant drop in the signal-to-noise ratio on the introduction of S-phase extract, which has a high autofluorescence and otherwise scatters photons.

In experiments with fluorescent Cdc45, WT Mcm2-7/Cdt1 was used instead of Mcm2-7$^{Mcm4\text{-}DY549}$/Cdt1. The standard S-phase extract (*28*) contains Cdc45-3xHA; Alexa Fluor 532 streptavidin conjugate was pre-incubated with biotinylated anti-HA F_{ab} fragments (Roche) for 1 hour, and then incubated with the S-phase extract for 1 hour on a rotator at room temperature. In these experiments, the replication step buffer was supplemented with 0.03mg/ml biotin.

In control experiments, aphidicolin was added to the replication mix at a final concentration of 300μM. For Sic1 inhibition, Sic1 was added to the replication mix at a final concentration of 1μM.

Single-molecule Replication Experiments with Mcm2-7$^{Mcm4\text{-}SNAP\text{-}JF646}$

For standard loading reactions, a double-tethered DNA curtain was incubated with 1nM ORC, 12nM Cdc6, and 15nM Mcm2-7$^{Mcm4\text{-}SNAP\text{-}JF646}$/Cdt1 at room temperature (23 ± 1°C) for 20min. in reaction buffer containing 40mM HEPES, pH 7.6, 12mM MgAc, 50μM ZnAc, 0.5mM DTT, 225mM KGlut, 3mM ATP, 20mM

creatine phosphate (Roche), 0.04mg/ml creatine kinase (Roche), 4mg/ml BSA (Sigma-Aldrich), and 1mM Trolox (Sigma-Aldrich). Excess loading reactions used 5nM ORC, 67nM Cdc6, and 30nM Mcm2-7$^{Mcm4\text{-}SNAP\text{-}JF646}$/Cdt1. To initiate replication, excess protein was flushed out and followed by 7nM DDK at and 400 units of T4 DNA ligase (New England Biolabs) and incubated for 20min. Excess protein was flushed out followed by a reaction mix with final concentrations of 40mM HEPES, pH 7.6, 12mM MgAc, 50μM ZnAc, 300mM KGlut, 50mM NaCl, 3mM ATP, 20mM creatine phosphate, 0.04mg/ml creatine kinase, 4mg/ml BSA, 0.5mM Trolox, 12.5μg/μl S-phase extract, 24nM DDK, 25μM protochatechuate (Sigma-Aldrich), 50nM protochatechuate dioxygenase (Sigma-Aldrich) (*206*), 200μM of rUTP, rCTP, and rGTP (Invitrogen), 40μM each dNTP (Invitrogen), and 0.225μg/μl competitor salmon sperm DNA (Sigma-Aldrich).

Position Distribution Histograms[‡]

DNA-bound proteins were identified by observing characteristic lateral fluctuations. Intensity centroids were selected manually and pixel (px) values relative to the chromium barriers were converted to base pair values. 2-D Gaussian refinement was used where possible, on low occupancy curtains. Post steady-state position distribution histograms were generated with a custom Python script that bins the localization data to 1px, or ~1000bp, and defines an error based on 300 bootstrap samples and an 85% confidence interval.

Position Distribution Histograms for Mcm2-7$^{Mcm4\text{-}SNAP\text{-}JF646}$ Data

Particles were detected and tracked using a custom Python script. Particles were filtered using the following criteria: a particle had to last for more than 2 frames, exhibit an average radius of less than 2px, be present in the first frame of the data recorded, be more than 4px away from the edges of the image, and show at least one detection every three images (to accommodate blinking and tracking errors). Data were binned to 1px. Error bars were obtained by 1000 bootstrap iterations with a confidence interval of 95%.

Lifetime Analysis[‡]

Direct measurement of kymogram traces was used to generate dwell time values as the raw data for survival probability plots. Time values were either normalized to a defined time-point after the flow-out of excess protein (for ORC lifetime data), or both binding and dissociation points were captured for each binding event (for Cdc6 lifetime data). Dwell times were processed with a custom Python script that generates survival probabilities with 300 bootstrap samples of the input data binned to the acquisition framerate. Single exponential fits were used to extract lifetime variables.

The distinct lifetime subpopulations of Cdc6 were determined by maximizing the F-value in two-lifetime model F-tests in which each pair of

subpopulations was fitted by either two single exponential decay functions or one single exponential decay function. The highest F-value is given by treating bins 32, 33, and 34 as distinct ($F = 5024$, $P < .0001$). *ARS1* is located within bin 33.

Cdc6 On-rate Ratio Calculation‡

We begin with the steady-state equations for Cdc6 on-rates at *ARS1* (1) and non-*ARS1* (2). Dividing these two (3) eliminates the free Cdc6 concentration term and yields a series of three experimentally definable ratios: the ratio of the measured Cdc6 off-rates at *ARS1* and non-*ARS1*, the ratio of *ARS1* DNA to non-*ARS1* DNA, and the ratio of the effective Cdc6 occupancy at *ARS1* and non-*ARS1*. These values are used in (3) to give an estimate of the ratio of the Cdc6 on-rate at *ARS1* to non-*ARS1* sites.

(1) $k_{on,ARS1} = k_{off,ARS1}$[Cdc6-*ARS1*]/[*ARS1* DNA][$Cdc6_{free}$]

(2) $k_{on,non\text{-}ARS1} = k_{off,non\text{-}ARS1}$[Cdc6-non-*ARS1*]/[non-*ARS1* DNA][$Cdc6_{free}$]

(3) $k_{on,ARS1}/k_{on,non\text{-}ARS1} = (k_{off,ARS1}/k_{off,non\text{-}ARS1})$([non-*ARS1* DNA]/[*ARS1* DNA])([Cdc6-*ARS1*]/[Cdc6-non-*ARS1*])

$\Rightarrow (k_{off,ARS1}/k_{off,non\text{-}ARS1}) = .0719\ s^{-1}/.1255\ s^{-1}$

⇒ [non-*ARS1* DNA]/[*ARS1* DNA] = (non-*ARS1* DNA region)/(*ARS1* DNA region) = 47,358bp/193bp

$\Rightarrow$ [Cdc6-*ARS1*]/[Cdc6-non-*ARS1*] = ($N_{ARS1}\tau_{ARS1}/N_{non\text{-}ARS1}\tau_{non\text{-}ARS1}$), where

N = number of Cdc6 binding events in a given time-interval

$\Rightarrow(N_{ARS1}\tau_{ARS1}/N_{non\text{-}ARS1}\tau_{non\text{-}ARS1}) = (113)(13.9s)/(214)(7.97s)$

$\therefore k_{on,ARS1}/k_{on,non\text{-}ARS1} \approx 130$

Mcm2-7$^{Mcm4\text{-}DY549}$ Photobleaching Analysis‡

The fluorescent intensity of each particle was tracked by first defining a 3x3px region of interest (roi) about each bound Mcm2-7. The average intensity of the roi was calculated for each frame and plotted against time. A background intensity value of an immediately adjacent 3x3px region with no fluorescent dye was subtracted from each frame to account for lipid autofluorescence. Nearest neighbor averaging was used to smooth the final traces.

APPENDIX III

Protein Constructs and Preparations

Untagged ORC, Cdc6, and Mcm2-7/Cdt1

WT ORC and Mcm2-7/Cdt1 complexes were purified as in (*56*). Briefly, G1-phase arrested yeast were induced to overexpress FLAG-tagged Cdt1 and all six MCM subcomponents. Anti-FLAG resin was used to capture and wash Mcm2-7/Cdt1 complexes which were then eluted with FLAG peptide. The protein was concentrated and free peptide washed away in a spin concentrator (Vivaspin). WT Cdc6 was purified as described in (*55*).

Sortase-tagged ORC[‡]

Ub-Sort-ORC (containing ubiquitin, followed by a triglycine Sortase recognition tag, FLAG-tag, and Orc1 gene) was expressed in *S. cerevisiae* (W303) grown to OD_{600} = 1.2 in YEP supplemented with 2% w/v glycerol and induced with 2% w/v galactose for 3 hours at 30˚C. The cells were α-factor arrested for an additional 3.5 hours, then harvested and washed once with 50ml ice-cold 0.2mM PMSF, and once with 100ml buffer A (50mM HEPES, pH 7.6, 10% glycerol, 5mM MgAc, 1mM ZnAc) + 0.1mM EDTA, 0.1mM EGTA, and 1M sorbitol. The cells were

[‡] All paragraphs or sections marked with "‡" are excerpts reprinted from reference (*133*), with permission from Elsevier.

resuspended in approximately 1/3 of settled cell volume in buffer A + 500mM KCl, 0.1mM EDTA, 0.1mM EGTA, 0.01% NP-40, and protease inhibitors (Roche cOmplete tablets), and frozen dropwise in liquid nitrogen. The cells were lysed with a SamplePrep Freezer/Mill (SPEX) and the lysate was clarified by ultracentrifugation at 120,000xg for 90min. at 4°C. The supernatant was applied to 2ml anti-M2 FLAG resin (Sigma) pre-equilibrated in buffer A + 0.1mM EDTA, 0.1mM EGTA, 500mM KCl, and 0.01% NP-40 and incubated with rotation for 3 hours at 4˚C. The flow-through was discarded and the resin was washed with 40ml of buffer A + 0.1mM EDTA, 0.1mM EGTA, 200mM KCl, and 0.01% NP-40. Sort-ORC was eluted with buffer A + 0.1mM EDTA, 0.1mM EGTA, 200mM KCl, 0.01% NP-40, and 0.15mg/ml FLAG peptide. Note that N-terminal ubiquitin is cleaved off in the cells resulting in an N-terminal Sortase recognition tag on Orc1. The peak fractions were pooled and applied to 0.5ml of SP resin (GE Healthcare) pre-equilibrated with buffer A + 0.1mM EDTA, 0.1mM EGTA, 200mM KCl, and 0.01% NP-40. The resin was washed with 5ml of buffer A + 200mM KCl, and 0.01% NP-40, and Sort-ORC was eluted with buffer A + 500mM KCl, and 0.01% NP-40. The final concentration of KCl in the buffer was lowered to 200mM with the addition of buffer A + 0.01% NP-40.

Equimolar amounts of Sortase and Sort-ORC were combined with a 200-times molar excess of biotinylated peptide (biotin-LPETGG) and $CaCl_2$ was added to a final concentration of 5mM. The reaction was incubated at room temperature

for 15min., and then quenched with 20mM EDTA. The reaction mix was applied to 0.5ml of SP resin (GE Healthcare) pre-equilibrated with buffer A + 0.1mM EDTA, 0.1mM EGTA, 200mM KCl, and 0.01% NP-40 and washed with 10ml of buffer A + 0.1mM EDTA, 0.1mM EGTA, 200mM KCl, and 0.01% NP-40. Biotin-Sort-ORC was eluted with buffer A + 0.1mM EDTA, 0.1mM EGTA, 500mM KCl, and 0.01% NP-40. Peak fractions were pooled and stored at -80˚C.

Sortase-tagged Cdc6[‡]

This protocol is based on (*207*). GST-SUMO-Sort-Cdc6 (pET23b-GST-SUMO-Sort-Cdc6) was expressed in Rosetta 2(DE3) pLysS cells in LB supplemented with 100µg/ml ampicillin and 25µg/ml chloramphenicol grown to OD_{600} = 0.6 and induced with 0.5mM IPTG for 5 hours at 18ºC. Cells were harvested by centrifugation, and washed once with 50ml of ice-cold 0.2mM PMSF. The pellet was resuspended in 50ml of buffer B (50mM K_2HPO_4/KH_2PO_4, pH 7.5, 5mM $MgCl_2$, 1mM DTT, 1% Triton X-100) + 2mM ATP, 0.15M KAc, protease inhibitors (Roche cOmplete tablets), and 100µg/ml lysozyme, and incubated for 15min. on ice. Cells were lysed by sonication and the lysate was clarified by ultracentrifugation at 120,000xg for 40min. at 4°C. The clarified lysate was incubated with 2ml of pre-equilibrated Glutathione Sepharose 4 Fast Flow resin (GE Healthcare) for 3 hours. The flow-through was discarded and the resin was washed with 40ml of buffer B + 2mM ATP, 0.15M KAc, and 1% Triton X-100. The

column flow was stopped and 50% slurry was made with buffer B + 2mM ATP, 0.15M KAc, and 300µg of Ulp1 protease. Note that Ulp1 protease cleaves immediately after SUMO, resulting in an N-terminal Sortase recognition tag on Cdc6. The mixture was incubated at 4˚C for 10min. with occasional swirling by hand to prevent resin settling, and the flow-through was recovered along with two 2ml washes with buffer B + 2mM ATP and 0.15M KAc. All the fractions were pooled and the KAc concentration was adjusted to 75mM by adding buffer B + 2mM ATP. The mixture was applied to 1ml of pre-equilibrated hydroxyapatite ceramic (BioRad, 80µm particle size) and washed with 5ml of buffer B + 2mM ATP and 0.15M KAc, followed by 5ml of buffer C (50mM Tris-HCl, pH 7.5, 5mM $MgCl_2$, 1mM DTT, 15% glycerol, 1% Triton X-100), and 5ml of buffer C + 0.15KAc. Sort-Cdc6 was eluted with buffer C + 0.4mM KAc.

Equimolar amounts of Sortase and Sort-Cdc6 were combined along with a 200-times molar excess of biotinylated peptide (biotin-H_{10}LPETGG), plus 5mM $CaCl_2$. The reaction was incubated at room temperature for 3min., and then quenched with 20mM EDTA. The reaction mix was applied to a Superdex 75 10/300 gel filtration column equilibrated with buffer D (50mM HEPES-KOH, pH 7.6, 5mM $MgCl_2$, 1mM DTT, 0.1mM EDTA, 1% Triton X-100, 15% glycerol) + 0.3M KGlut and 10mM imidazole. Peak fractions were pooled and incubated with 0.5ml of pre-equilibrated cOmplete-His-Tag Purification Resin (Roche) for 2 hours with rotation at 4˚C. The flow-through was discarded and the resin was washed with

5ml of buffer D + 0.3M KGlut and 10mM imidazole. Biotin-H_{10}LPETGGG-Cdc6 was eluted with buffer D + 0.3M KGlut and 0.3M imidazole. Peak fractions were pooled and stored at -80°C.

fSNAP-tagged Mcm2-7/Cdt1[‡]

S. cerevisiae cultures were grown to OD_{600} = 0.8-1.0 and α-factor arrested in G1 phase by the addition of α-factor (200ng/ml). Overexpression of Mcm2-7/Cdt1 containing a 1x FLAG epitope on Mcm3 and fSNAP on Mcm4 was induced by addition of 2% w/v galactose for 4 hours prior to harvesting. Harvested cells were resuspended in a 1/3 pellet volume of cell lysis buffer (100mM HEPES-KOH, pH 7.6, 0.8M sorbitol, 10mM MgAc, 2mM EDTA, 300mM KGlut) containing 1X protease inhibitor cocktail (Roche) and frozen dropwise into liquid nitrogen. The frozen cell droplets were milled with a SamplePrep Freezer/Mill (SPEX). After thawing, the cell lysate was centrifuged at 45,000rpm for 60min. (Ti70 Rotor, Beckman) and the supernatant was mixed with 0.4ml anti-M2 FLAG resin (Sigma) equilibrated with buffer H (25mM HEPES-KOH, pH 7.6, 1mM EDTA, 1mM EGTA, 5mM MgAc, 10% glycerol) + 300mM KGlut and 1mM ATP. The volume of the lysate was increased to 10ml with buffer H + 300mM KGlut and 1mM ATP and the mixture incubated for 4 hours at 4°C with rotation. The resin was washed two times with 10ml of buffer H + 300mM KGlut and 1mM ATP and eluted with buffer H + 300mM KGlut and 1mM ATP containing 0.1mg/ml 3xFLAG peptide. The protein-

containing fractions were pooled, DTT added to 1mM, and SNAP-biotin[9] (New England Biolabs) was added to a final concentration of 10μM. The mixture was incubated at room temperature for 1.5 hours, after which the SNAP-labeled Mcm2-7/Cdt1 was applied to a Superdex 200 gel filtration column equilibrated in buffer H plus 300mM KGlut and 1mM ATP and peak fractions were pooled and stored at -80°C.

Mcm2-7$^{Mcm4\text{-}DY549}$/Cdt1‡

Mcm2-7 $^{Mcm4\text{-}Sort}$ /Cdt1 was prepared as described for Mcm2-7/Cdt1 constructs described above, but prior to gel filtration was incubated with equimolar amounts of Sortase, and $CaCl_2$ was added to a final concentration of 5mM. This was mixed with 200μl of peptide carrying a Sort-tag and labeled with DY549-P1 (Dyomics). The reaction was incubated at room temperature for 15min., and then quenched with 20mM EDTA.

DDK and S-phase Extract

DDK was purified as in (*28*). Briefly, yeast containing Cdc7-FLAG and Dbf4 under the GAL1/10 promoter were grown to overexpress the protein. Cells were frozen and lysed, and anti-FLAG resin was used to capture and wash the complexes

[9] Or SNAP-JF646, as indicated.

which were then eluted with FLAG peptide. Protein-containing fractions were collected and snap-frozen in liquid nitrogen.

Whole-cell S-phase extract was also prepared as in (*28*). Briefly, yeast containing a temperature sensitive allele of Cdc7 (*cdc7-1*) and with Cdc45-3xHA, Sld2, Sld3, and Cdc28 (yMW004) under the GAL1/10 promoter were grown with protein overexpression and then arrested at S phase. Cells were frozen, lysed, resuspended, and dialyzed against storage buffer. The protein concentration and conductivity of the extract were measured, then the resulting extract was flash frozen in liquid nitrogen.