Single-molecule DNA rotation tracking using DNA origami rotors

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Altheimer, Benjamin D. 2019. Single-molecule DNA rotation tracking using DNA origami rotors. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:42013119">http://nrs.harvard.edu/urn-3:HUL.InstRepos:42013119</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Single-molecule DNA rotation tracking using DNA origami rotors

A DISSERTATION PRESENTED
BY
BENJAMIN D. ALTHEIMER
TO
THE COMMITTEE ON HIGHER DEGREES IN BIOPHYSICS
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
BIOPHYSICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
JULY 2019
Single-molecule DNA rotation tracking using DNA origami rotors

ABSTRACT

The rotation of DNA is an intrinsic part of many protein activities. Many DNA motor proteins, including helicases, translocases, and polymerases, rotate their DNA substrate due to DNA’s helical structure. Others induce DNA rotation by distorting the DNA upon binding or introduce turns as their primary enzymatic activity. Additionally, in the constrained environment of the cell, DNA rotation affects the buildup of torsional strain, which can influence the activity of many proteins. The direct measurement of single-molecule DNA rotation can serve as a readout of these protein activities, creating opportunities to study mechanism and regulation. However, existing methods are limited in resolution, throughput, and by certain experimental requirements. In this dissertation, I describe the development of Origami-Rotor-Based Imaging and Tracking (ORBIT), a new approach for direct, single-molecule measurement of DNA rotation at high resolution and throughput.

ORBIT uses structural amplification to turn the very small movement associated with DNA rotation into a signal measurable using fluorescence tracking. Specifically, we use fluorescently labeled DNA origami rotors to report on the rotation of double-stranded DNA that extends perpendicular to the structure. High resolution is achieved by minimizing Brownian noise with a small origami rotor and short double-stranded DNA. ORBIT does not require applying a stretching force or a biasing torque to the DNA. We demonstrate that ORBIT is capable of resolving rotation equivalent to the twist between adjacent base pairs with an integration time of only ~20 ms.
We have applied ORBIT for three biological applications. First, by measuring the Brownian dynamics of origami rotors tethered to a coverslip surface by double-stranded DNA, we determined the torsional rigidity of DNA in the absence of applied force. This value determines how twist and torque are transmitted through DNA and influences how the activity of one protein can affect another. Second, we applied ORBIT to the homologous recombination enzyme RecBCD during initiation and processive unwinding of DNA. We characterize the pausing and backtracking behaviors of the enzyme, finding evidence of two distinct pausing states during processive unwinding. Furthermore, with ORBIT, we are able to track activity starting from DNA binding, allowing us to observe the series of events during initiation on double-stranded breaks with variable end geometries. We find reversible, ATP-independent 5 bp transitions between wound and unwound states during initiation, and that the engagement of RecB with the 3’ strand is important for initiation. Third, we track the rotation of DNA induced by RNA polymerase (RNAP) during transcription, and directly detect single base-pair steps in the rotational movement of the DNA, indicating that RNAP tracks the double helix at this short length scale. We anticipate that ORBIT will be useful for studying DNA rotation in many contexts.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title Page</td>
<td>i</td>
</tr>
<tr>
<td>Copyright</td>
<td>ii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Table Of Contents</td>
<td>v</td>
</tr>
<tr>
<td>Citation to previously published work</td>
<td>vii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>viii</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td></td>
</tr>
<tr>
<td>1.1 Why study DNA rotation?</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Existing methods to track DNA rotation</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Desirable features for tracking DNA rotation</td>
<td>12</td>
</tr>
<tr>
<td>1.4 DNA vs protein rotation</td>
<td>20</td>
</tr>
<tr>
<td>1.5 DNA nanotechnology</td>
<td>21</td>
</tr>
<tr>
<td>1.6 Outline of the dissertation</td>
<td>30</td>
</tr>
<tr>
<td><strong>2 Design of the ORBIT assay</strong></td>
<td>31</td>
</tr>
<tr>
<td>2.1 Overview of the ORBIT assay</td>
<td>32</td>
</tr>
<tr>
<td>2.2 Localization precision and Brownian dynamics</td>
<td>35</td>
</tr>
<tr>
<td>2.3 Design of the origami rotor</td>
<td>39</td>
</tr>
<tr>
<td>2.4 Design of the rotor-dsDNA junction</td>
<td>43</td>
</tr>
<tr>
<td>2.5 Incorporation of the Cy3 staple strands</td>
<td>47</td>
</tr>
<tr>
<td>2.6 Expected hydrodynamic drag of the origami rotor</td>
<td>49</td>
</tr>
<tr>
<td>2.7 DNA origami anchor design</td>
<td>51</td>
</tr>
<tr>
<td><strong>3 Brownian dynamics and the stiffness of DNA</strong></td>
<td>57</td>
</tr>
<tr>
<td>3.1 Measuring Brownian dynamics</td>
<td>58</td>
</tr>
</tbody>
</table>
3.2 Analyzing the Brownian dynamics for real motion ........................................ 58
3.3 Brownian dynamics of the origami rotors ...................................................... 61
3.4 Determining the torsional rigidity of DNA .................................................... 67
3.5 Torque and angular lag due to the drag .......................................................... 69

4 INITIATION AND PROCESSIVE DNA UNWINDING DYNAMICS OF RecBCD 71
4.1 Introduction to RecBCD .............................................................................. 72
4.2 ORBIT assay for RecBCD ........................................................................ 78
4.3 Stopped-flow assay for RecBCD ............................................................... 81
4.4 Comparing ORBIT and ensemble RecBCD measurements ......................... 84
4.5 Characterization of the unwinding phase .................................................... 87
4.6 Initiation on blunt-end substrates .............................................................. 91
4.7 Initiation on DNA substrates with single-stranded overhangs ...................... 95
4.8 Stopped-flow validation of RecBCD initiation results .................................. 103
4.9 RecBCD conclusions .............................................................................. 106

5 TRANSCRIPTION BY RNA POLYMERASE .................................................. 108
5.1 Introduction to RNA Polymerase ............................................................... 109
5.2 Kinetics of elongation .............................................................................. 110
5.3 Single base pair stepping ......................................................................... 111

6 MATERIALS AND METHODS .................................................................... 115
6.1 DNA origami preparation and purification .................................................. 115
6.2 AFM and TEM imaging ........................................................................... 118
6.3 Single-molecule imaging of RecBCD-induced DNA unwinding using ORBIT 118
6.4 Single-molecule data analysis for RecBCD ................................................. 120
6.5 RecBCD pausing analysis ......................................................................... 122
6.6 RecBCD initiation analysis ..................................................................... 123
6.7 Stopped-flow ........................................................................................... 124
6.8 Single-molecule measurements of RNAP-induced DNA rotation .......... 125

7 CONCLUSION ............................................................................................. 127
7.1 Summary of this thesis ............................................................................ 128
7.2 Future directions ..................................................................................... 130

APPENDIX A DNA SEQUENCES .................................................................. 134

REFERENCES ................................................................................................. 157
Citation to previously published work

This thesis was adapted from the following publication:

I am deeply grateful to the many people who have supported and guided me throughout my time in graduate school and who made the work in this thesis possible. First, I would like to thank my advisor, Xiaowei Zhuang, for her support, guidance and feedback throughout this process. I have learned a great deal from her dedication, persistence, and commitment to tackling hard, ambitious scientific problems. I have also learned so much about doing science thanks to the incredible group of scientists she has brought together.

The work presented in this thesis was a collaboration and I am lucky to have had the opportunity to work with so many talented scientists. This work was done together with Pallav Kosuri, without whom this project would not have been possible. From Pallav, I have learned how to approach tough scientific problems and communicate the value of my ideas. I also thank our collaborators, Peng Yin and Mingjie Dai, who helped us formulate the project, introduced us to the exciting world of DNA nanotechnology, and made major contributions to the design of the assay described in this thesis.

I have been very lucky to be able to work with and learn from a talented group of scientists in the Zhuang lab during my PhD. First, I would like to thank Bryan Harada, who guided me as a rotation student and beyond, and trained me in single-molecule FRET, the molecular biology of nucleosomes and chromosome remodelers, and how to design and analyze single-molecule biology experiments. I also learned a great deal from the other single-molecule enzymology team members who were in the lab when I started, Will Hwang and
Sebastian Deindl, who were always happy to help out and discuss difficult problems and exciting directions.

There are many other current and former members of the Zhuang lab who played important roles during my PhD, including Yari Sigal, who provided training and support when our project’s microscope needs evolved, and Jeff Moffitt, who provided insights into how our new technology fit into the existing single-molecule technique landscape. This work also would not have been possible without help from our lab manager, Alec Goodman, and the lab administrators we’ve had over the years, Matt Kilroy, Danielle Reuter, Julie Zabbo, Rhonda Pautler, Teri Howard, and Ryan Hanna. I would also like to thank all the members of the Zhuang lab during my time here for their advice and feedback, and for always being willing to think deeply about someone else’s project. I would also like to thank them for their support, camaraderie, occasional volleyball wins, and many good conversations over the years.

I would like to thank Jim Hogle, Michele Jakolou, and Venky Murthy for all of their hard work making the Biophysics program a wonderful program and for their tireless support of us students. I also thank the faculty members who have served on my qualifying exam, dissertation advisory, and defense committees, Gerhard Wagner, Sun Hur, Andrew Kruse, Adam Cohen, Peng Yin, Joe Loparo, and Jeff Moffitt, for their time, feedback, and advice.

I also thank my previous research advisors. During my undergraduate studies at Oberlin College, Michael Nee and Manish Mehta introduced me to conducting original research and opened the door to all that has followed. I thank James Chou and Adam Cohen, as well as Jess Williamson and Min Ju Shon, who supervised my rotations in their labs, for guiding me through the early days of this journey.

I would like to thank my fellow students in the Biophysics program, especially my classmates, for introducing me to a wide range of scientific areas and for sharing this journey with
me. I also thank my friends, especially my former roommates, for their friendship, support, and many good times.

Finally, I would like to thank my family, without whom none of this would be possible. My parents have always supported and encouraged me in my endeavors, academic and otherwise, and helped nurture my interest in science. I thank my siblings, Andrew and Alyssa, from whose examples, advice, and questions I have learned so much. I also thank my in-laws, Nurhan, Ihsan, and Tuğçe, for welcoming me into their family and for always believing in me. Most importantly, I’d like to thank my wife, Gizem, for her constant love and support throughout this journey.
1.1 Why study DNA rotation?

The rotation of DNA is an intrinsic part of many protein activities. Many DNA motor proteins, including helicases, translocases, and polymerases, rotate their DNA substrate as they move along DNA’s helical pitch (Figure 1.1) [1–3]. Others induce DNA rotation by opening a bubble in the double-stranded DNA (dsDNA) or otherwise distorting the DNA upon binding
or activation [4–6]. The direct measurement of DNA rotation can thus serve as a readout of these protein activities, potentially providing a powerful tool to study their activity, mechanism, and regulation. Furthermore, tracking rotation at single-molecule resolution overcomes many challenges of ensemble enzymology, including ensemble averaging over asynchronous activity and heterogeneous populations [7–11]. With single-molecule tracking, it is often possible to directly observe the sequence of events during an activity of interest, including visits to short-lived or sparsely populated states and heterogeneous behaviors.

DNA rotation is especially important when considered in the constrained environment of the cell, where DNA rotation affects the buildup of torsional strain and supercoiling. This mechanical change has been shown to influence protein activity and chromatin structure [14–17]. For example, high positive supercoiling reduces the activity of RNA polymerase
(RNAP), as transcription on supercoiled DNA requires working against an opposing torque [15]. For this reason, cells have an entire class of enzymes, the topoisomerases, which induce or remove twist in DNA molecules [18] and at least one protein appears to have evolved to minimize its effect on DNA twist [19]. Cells may also buffer against torsional stress buildup through changes in protein interactions (e.g. through changes in DNA wrapping in tetrasomes in eukaryotes [20]). Although not the focus of this thesis, rotation measurements make it possible to determine the relationship between enzyme activity and DNA rotation, leading to a more complete understanding of how the activity of a given enzyme influences the overall state of the cell’s DNA and providing mechanistic insights. For example, the earliest measurement of DNA rotation induced by an enzyme, RNAP, was used to establish that RNAP tracked the double-stranded DNA helix during transcription [1], while later work (using both direct and indirect rotation tracking) has shown that certain other enzymes do not [3, 19]. Furthermore, the surprising finding that these enzymes (the bacterial DNA translocation enzyme FtsK and the ϕ29 viral DNA packaging motor) operate without DNA helix tracking provides information about their translocation mechanisms. Additionally, rotation measurements can be used to understand the physical properties (e.g. torsional stiffness) of DNA and DNA-protein complexes, which determine how DNA and chromatin respond to applied forces and torques [21–24].

Rotational measurements of DNA can be used as a probe of enzyme activity to address mechanistic questions. This approach is especially powerful if the relationship between protein activity and DNA rotation is understood (using single-molecule rotational measurements, ensemble approaches, or inferences from structural biology). Unwinding can then be understood as a direct proxy for the activity of interest. In practice, it has been possible to realize this potential in only a small number of cases. A series of rotation studies on E. coli gyrase have led to a sophisticated model of the series of protein and DNA conformational changes
during its activity introducing DNA turns [25–29]. Several rotational methods have also been applied to the homologous recombination enzyme Rad51, establishing that Rad51-DNA filaments nucleate cooperatively but grow via addition of Rad51 monomers [30] and measuring the physical properties of these filaments [31]. In other cases, including measurements of RuvAB, RecA, and reverse gyrase, rotational measurements have been limited to proof-of-principle measurements of protein-induced DNA rotation [32–34], likely due to the limitations of existing methods (discussed in the next two sections).

Thus, the direct measurement of DNA rotation allows us to (1) directly measure the activity of many important classes of enzymes and (2) better understand how their activity might influence the state of the DNA. Through single-molecule measurements of DNA rotation, we may be able to significantly improve our understanding of a wide range of protein-DNA interactions. However, in practice, the application of existing methods has been limited. In the remainder of this introduction, I will introduce existing approaches to measuring rotation, discuss features desirable for rotational measurement approaches, and describe DNA origami, the technology which enables the rotation tracking method presented in this thesis.

1.2 Existing methods to track DNA rotation

1.2.1 Supercoiling-based approaches

In relaxed DNA, each base pair is twisted ~35° from the previous pair (Figures 1.1C and 1.2A). Equivalently, the helical pitch of DNA is about 10.5 turns. DNA rotation causes a change in this twist. If the DNA is topologically constrained, this results in persistent over-winding or under-winding of the DNA (Figure 1.2B). This distortion can also be converted to DNA writhe, resulting in larger scale deformations known as supercoils (Figure 1.2C). For torsionally or topologically constrained DNA (such as closed circular DNA), the twist and
Figure 1.2: DNA twist and writhe (A) Relaxed DNA has about 10.5 bp per helical turn (corresponding to an average twist angle of ~35° between subsequent base pairs). (B) DNA can become over- (left) or under-wound (right) due to induced rotations if the DNA is torsionally constrained (represented here by the orange line; also note the DNA here is constrained on both edges). (C) The over- and under-winding can relax into DNA writhe, forming supercoils. Note the shape and effective size of the DNA are altered by the formation of supercoils. Adapted from reference [16] under a Creative Commons license.

writhe are constrained to have a constant sum, known as the linking number [35, 36]. Typically, a population of circular DNA molecules will contain a range of topoisomers, each with a specific, fixed linking number. Rotation introduced by enzyme activity can change the linking number, resulting in changes in twist and supercoiling. Changes in the amount of supercoiling change the physical shape and effective size of the DNA, which is detectable using ensemble approaches, most notably density gradient centrifugation and gel electrophoresis [37]. Similarly, local deformations due to protein or small molecule binding can significantly change the local twist, leading to additional supercoiling and/or twist in the remainder of the molecule to preserve the global linking number. Allowing the DNA to relax using topoisomerase then creates a permanent change in linking number, creating a record of the induced change which can be measured on a gel [37]. Additional information, for example the dependence of twist or structural transitions on buffer conditions, can be gathered by running 2-dimensional gels, where each dimension is run under different conditions [38].
This approach has been useful for characterizing several types of DNA rotation. For example, the twist induced by binding of a DNA intercalator, ethidium bromide, was determined through the number of intercalators required to remove all supercoils, as measured using cesium gradient centrifugation [39]. More significantly, the extent of DNA opening by RNAP during initiation and transcription can be indirectly measured through the effect of the local induced twist on the relaxed linking number of the DNA [40]. This approach also enabled the discovery of gyrase, which was found to processively introduce supercoils into closed circular DNA [41], as well as many additional studies of gyrase and other topoisomerases. Measurements of DNA topoisomer populations continue to prove useful for enzyme characterization [42], and new approaches in this area continue to be developed [43]. However, such methods are inherently limited to ensembles of DNA molecules, measured at a particular time, after any activity of interest. Furthermore, the measurement depends on supercoiling, not directly on the actual rotation introduced. This limits resolution because the introduced rotation is split between the twist and the writhe according to the stiffness of the DNA [35].

Changes in DNA supercoiling also formed the basis for the earliest single-molecule measurements of rotation (Figure 1.3A) [44–46]. In these magnetic-tweezer experiments, DNA is stretched between a surface and a magnetic bead. Both ends of the DNA are torsionally constrained via multiple attachment points, and a magnetic field creates a strongly preferred equilibrium angular position for the bead (with a spring constant of \(\sim 10^3 - 10^4\) pN nm rad\(^{-1}\) [47]), acting as a strong restoring torque and thus preventing rotation. Thus, as with closed DNA, the linking number is conserved and induced rotations are reflected in the changing number of supercoils in the DNA. These supercoils change the effective length of the DNA, bringing the magnetic bead closer to the surface. By monitoring this height, the induced rotation can be inferred. This approach was first applied to enzyme activity with topoisomerase
Figure 1.3: Single-molecule measurements of DNA rotation. (A) Measurement of DNA effective-length changes due to rotation of DNA. Supercoils sequester part of the DNA, changing the height of the bead, which is pulled upwards by a magnet. Adapted from reference [46] with permission. (B) Direct measurement of DNA rotation induced by RNAP. Despite residual restoring torque from the magnetic tweezer, the magnetic bead is able to rotate as the DNA is unwound by RNAP. Rotation can be tracked using the asymmetrically arranged small fluorescent beads (pink) on the magnetic bead (blue). During this process, RNA (red) is generated. Adapted from reference [1] with permission. (C) Rotor-bead assay for direct DNA rotation tracking. Because of the nick in the dsDNA backbone, the non-magnetic rotor bead (green) can rotate freely as the lower DNA twists spontaneously or due to protein activity (such as gyrase-induced rotation). Adapted from reference [25] with permission.

In these experiments, supercoils are first added to the DNA by rotating the magnet that constrains the magnetic bead. Topoisomerase is then added, relaxing the supercoils and effectively lengthening the DNA. The step size of this motion confirmed that topoisomerase relaxes DNA in increments of two supercoils, and the force-dependence implied that the enzyme shortens the DNA during the rate-limiting step.

As with the ensemble supercoiling-based approaches, this magnetic-tweezer based method has found widespread utility in measuring rotational changes caused by a wide range of proteins, including topoisomerases, ligase, and RNAP [19, 26, 48–55]. While, compared to gel-based approaches, it has the benefit of real-time measurement, it is an indirect, low resolution measurement of rotation, and can be confounded by other sources of length changes, such as DNA wrapping around a protein. The DNA is also held under both force and torque, features which will be discussed in more detail in Section 1.3.4.
1.2.2 Direct tracking of DNA rotation

The first direct, real-time measurement of single-molecule DNA rotation was published by the Kinosita group in 2001 [1]. This assay, shown in Figure 1.3B, built on the earlier supercoiling-based magnetic tweezer measurements, but instead of tracking the height of the magnetic bead, the DNA rotation was directly tracked by adding smaller fluorescent beads to the magnetic bead, creating an asymmetric object. This 850 nm multi-bead complex was attached to a 5 kb segment of DNA, onto which RNAP was initiated and stalled. Critically, the RNAP was then attached to the surface of a microscope flow chamber by non-specific adsorption, resulting in multiple contact points which create a torsionally constrained attachment. Thus, any rotational movement between the DNA and protein can only relax toward the magnetic bead, which then spun as a result of transcription by RNAP. This asymmetric bead was easily imaged in a light microscope, and, importantly, mechanically amplifies the size of the movements, from the rotation of a ~2 nm double helix to a 850 nm bead.

This experiment, which establishes that RNAP most likely faithfully tracks the dsDNA double helix during transcription was a major step forward in rotational measurements. However, it suffered from multiple drawbacks. First, to suppress lateral motions of the bead, which would obscure the rotation of interest, a small (~0.1 pN) stretching force had to be applied. This created an inherent requirement that the protein-DNA complex be formed prior to watching any activity: without pre-initiation of the enzyme on the dsDNA, the applied force, which is required to generate a meaningful signal, would physically separate the surface-attached enzyme and the bead-bound DNA substrate. Thus, this method cannot be used to track any initial binding/initiation processes. Second, due to both the magnetic field acting on the paramagnetic bead with a preferred magnetization axis and the hydrodynamic drag from this large bead, the system experienced a large torque, which the authors
believed contributed to extensive supercoiling of the DNA. (The torque from the drag was at least 5 pN nm, and there is an unknown contribution from the magnetic field. A later study using a similar approach found a potential well with torsional stiffness of $\sim 20$ pN nm rad$^{-1}$ from the magnetic field [30], although such a large torque would likely stall RNAP [56].) This supercoiling complicates the interpretation of the unwinding measurement, as discussed in the previous section. Third, even without the complication of supercoiling, the resolution of the method is inherently limited by the Brownian dynamics (discussed in more detail later), which are significant due to the size of the bead and the length of the DNA. Finally, the measurement was very low throughput, with less than one active complex per field of view.

Despite these drawbacks, this or very similar approaches, sometimes referred to as free rotation magnetic tweezers (FRMT) because the rotation is less constrained compared to traditional magnetic tweezers with their very high torques [30], have been used in a small number of cases: to study RuvAB kinetics during Holliday junction branch migration [32], Rad51 nucleation and extension [30], and reverse gyrase kinetics, and force, torque, and temperature sensitivity [34, 57]. Note that in these later cases the DNA was directly attached to the surface and the protein bound internally on the DNA, allowing the study of different types of enzymatic processes within a similar experimental framework.

1.2.3 LOWER TORQUE APPROACHES

Multiple different approaches to measure rotation without such large torques have since been developed. In one early study, the requirement for an applied force and torque was removed by using a large double-bead complex and a short dsDNA tether [58]. With this geometry, the lateral Brownian dynamics (see Section 1.3.1) were small enough compared to the size of the beads to allow rotational tracking. However, because of the size of the bead, the Brow-
nian dynamics occur on slow timescales (relaxation time of 1.6 seconds), giving very limited spatiotemporal resolution.

Several groups have built off the Kinosita group’s general approach for measuring DNA rotation while better controlling \([22, 47, 59, 60]\) or effectively eliminating \([33]\) the magnetic restoring torque by carefully controlling the magnetic field. In the latter case, called freely-orbiting magnetic tweezers (FOMT), the residual torque from the magnetic field acting on the bead was reduced to negligible levels by creating a magnetic field aligned vertically (along the same axis as the DNA) \([33]\), making it possible to measure the physical properties of DNA, DNA-protein complexes, and DNA-intercalator complexes, as well as filament formation by RecA and Rad51 without applied torque \([20, 31, 33, 61, 62]\). While these approaches can reduce or eliminate the applied torque, they still require the application of a stretching force, large, resolution-limiting beads, and, until very recently \([63]\), remained one-at-a-time measurements.

Alternatively, the torque from the magnetic tweezer can be fully decoupled from the substrate DNA by the use of two beads with specialized roles instead of a single magnetic bead. In rotor bead assays (Figure 1.3C), one bead is used to apply a stretching force to the molecule using a magnetic tweezer, optical trap, or pipette while a second bead, which is rotationally decoupled from the first via a nick on one DNA strand, is used to report rotation \([3, 21, 25, 64]\). Importantly, this bead specialization also opens the door to using very small beads to track the rotation \([27, 28]\). Using smaller beads moves the Brownian noise to higher frequencies where it is more rapidly averaged out (see Section 1.3.1). However, position measurements of smaller beads are also more easily corrupted by lateral movements, requiring higher stretching forces \([28]\). The rotor bead assay has been used to study DNA physics \([21, 64–67]\), and investigate the mechanisms of DNA gyrase \([25, 27–29]\), and a viral
DNA packaging motor \cite{3}, but requires complex assembly and a stretching force, and remains a low throughput approach.

Finally, rotational control and tracking can be achieved using a variant of the standard optical trapping technique which uses optically anisotropic particles and polarized light to both apply and detect torque, allowing indirect determination of particle rotation \cite{68, 69}. While this approach inherently applies a large torque to the DNA, active feedback or a rapidly oscillating polarization can be used to achieve an effectively low torque \cite{68, 70, 71}. Optical torque trapping has been used extensively to study DNA physics \cite{69, 72–74} as well as the effect of torque on transcription by RNAP \cite{56} and nucleosome stability \cite{14}, but has not been applied to directly track DNA rotation due to enzyme activity.

Collectively, these approaches have enabled an initial set of direct rotational measurements of DNA with applications to a small number of important enzymes (RNAP \cite{1}, gyrase \cite{25–29, 34, 57}, DNA recombination enzymes \cite{30–33}, and a viral DNA packaging motor \cite{3}). They operate on a similar principle (except for the angular optical trap): the small movement of DNA rotation about its own axis is mechanically amplified using a lever arm. Despite the simplicity of this core idea, these methods have been used in what is still a relatively small number of cases compared to the scope of protein-nucleic acid interactions which induce DNA rotation due to unwinding, tracking, or simply binding to DNA. The limited application of DNA rotation assays is a consequence of the limitations described above, most notably force/torque application, system complexity, low resolution, and low throughput. Thus, the full potential of DNA rotational measurements to investigate protein-DNA interactions has not been realized, and additional approaches are needed. In the next section, I discuss what features would be desirable for a DNA rotation tracking method based on lessons we can learn from the existing approaches.
1.3 Desirable features for tracking DNA rotation

In this section, I consider features of DNA rotation tracking methods which would make direct, real-time rotational tracking a more useful and widespread single-molecule strategy. These topics are closely inter-related and ultimately are derived from the need to faithfully track the motion of a rotational probe (such as a bead) attached to a relatively flexible DNA molecule. I also discuss how previous methods have improved these features, and the trade-offs that have been required.

1.3.1 Resolution

First, a rotational method should be able to achieve high spatiotemporal resolution. The length and timescales of biological processes set the required resolution: many biological processes occur on the millisecond timescale or even faster over length scales of ~1 base pair (bp) or ~35° [75]. For methods that rely on tracking the movement of a microscale probe (such as a bead) attached to a DNA tether (which is also the experimental substrate), the resolution is fundamentally limited by the Brownian dynamics of the probe [28, 76–78]. These motions are a consequence of random thermal fluctuations, which cause noise in both the lateral (side-to-side movements of the probe) and angular dimensions. For a rotational measurement, conceptually, the consequence of this Brownian noise in the angular dimension is clear - the angular noise is added to the angular position of interest to give the observed signal. The factors determining the magnitude of this noise will be discussed shortly.

Less intuitively, the lateral dimension of this noise needs to be considered, even for angular measurements. The lateral noise can cause the position of the probe to be a poor reporter of the angular position because it may not be possible to fully disambiguate the angular and lateral movements. This has necessitated the use of a stretching force [1, 25, 28] to
suppress the lateral movements in all but one direct rotational measurement. The exception, which used a large double bead complex and short DNA is instructive: the force was not required because the short DNA did not geometrically allow large enough lateral movements to affect the rotational measurement [58]. Note the converse is also true - as probes become smaller, they amplify the motion less, and a higher force is required to more completely suppress lateral Brownian motion [28]. However, using larger probes is not an ideal solution, as discussed in the next paragraph.

When no external force/torque is applied to the probe, the total magnitude of the Brownian fluctuations in the lateral and angular dimensions depends on the stiffness of the DNA tether, which in turn scales with the total length of DNA [78]. With an applied force or restoring torque, the relevant stiffness is the combined effect of the applied force/torque and the DNA tether, significantly reducing the movements but also, in the case of the restoring torque, biasing the angular measurement or even preventing rotation. In either case, the timescale of these fluctuations is affected by the hydrodynamic drag of the probe [79–81]. This drag causes the movements to be correlated over time, requiring additional time averaging to achieve a desired signal-to-noise ratio. With lower drag probes, the noise is effectively pushed to higher frequencies, improving the spatial resolution at any finite sampling rate. Thus, high spatiotemporal resolution requires the use of short DNA tethers and small probes [28, 76–78]. However, short tethers are often incompatible with the experimental geometry (due to bead size and surface attachment approaches), and smaller probes result in less amplification of the rotational motion, lowering the angular precision with which the probe can be localized. With small probes, lateral Brownian dynamics also more significantly influence angular measurements, as discussed in the previous paragraph. I will describe the Brownian dynamics, and their connection to the DNA stiffness and probe hydrodynamic drag, in more quantitative detail in Sections 2.2 and 3.2.
1.3.2 Throughput

Second, we would like to be able to achieve high throughput in order to be able to reliably statistically compare many conditions within a reasonable time frame. Existing methods have generally only been able to observe one molecule at a time due to several limitations. Several methods, including the angular optical trap [68, 69] and at least some implementations of the rotor bead assay [28], use a detection scheme which can only capture position information about a single molecule. Even if widefield detection is used, many complexes are not usable because of inactive enzymes or improperly assembled DNA-bead complexes, which require multiple components to be present, intact, and properly constrained [1, 25, 28, 63]. Careful alignment within the magnetic field is also an important consideration for achieving low torque in certain magnetic tweezer experiments [33]. Experiments which use optical traps, including early rotor bead implementations [21] and the angular optical trapping technique [68, 69] are limited to trapping single molecules within the optical trap laser spot. While the throughput of most methods are not explicitly discussed, rotation measurement studies typically include a small total number of complexes. Recently, higher throughput versions of magnetic torque tweezers (MTT) and freely orbiting magnetic tweezers (FOMT) were developed by improving the uniformity of the magnetic field [63, 82]. This allowed up to 80 complexes to be imaged in a field of view, although the actual samples sizes suggest that in practice many fewer usable complexes are imaged per experiment [63]. Additionally, these multiplexed approaches have not been used to study protein activity, an additional complication which would likely further reduce yield. Another difficulty is measuring a large field of view while maintaining the high frame rates required for high resolution imaging [28], although this problem has been significantly alleviated by recent developments in scientific CMOS camera technology [83]. Overall, throughput remains a significant challenge for
single-molecule rotation measurements. To address this issue, a method should use a wide-
field imaging approach covering a field of view with many DNA molecules, and as large a
fraction as possible of these molecules should be active, fully intact, and properly torsionally
constrained complexes.

1.3.3 EXPERIMENT DESIGN

Third, I consider several aspects of the experimental design and geometry which affect ro-
tational methods. It would be desirable for the components in a rotational method to have
a molecularly well-defined identity so that, for example, the exact length and geometry of
the DNA being studied is known. In practice, the requirement for rotationally constrained
attachment points on beads and coverslip surfaces has required the use of multiple protein-
substrate interactions (biotin/streptavidin or digoxigenin/anti-digoxigenin) over moderately
long stretches of DNA. For example, digoxigenin-based attachment has required multiple,
stochastic, incorporation of digoxigenin in a >300 bp segment, which inherently leads to some
uncertainty in the exact molecular state of the resulting complexes. This type of attach-
ment also introduces a bend into the DNA at the surface (and potentially at the bead, de-
pending on its orientation in the magnetic field) because the DNA must run parallel to the
surface over the length of this modified DNA segment. The uncertainty over precise length
and the presence of this bend likely make such an approach incompatible with short segments
of dsDNA. We would like to be able to better control the substrate geometry through more
precisely engineered attachment points.

Further, a rotational method should be customizable and easily applied to investigations
of different types of protein behaviors: with minor adjustments, it should be possible to use
the same method to study proteins which cause processive rotation and those that distort the
DNA a smaller amount upon binding. Thus far, direct rotational measurements have been
used to study processive motion (RNAP, φ29, RuvAB, gyrase at low force), or repetitive, uni-
directional motions (gyrase at high force) and filament formation (RecA, Rad51), which give
signals similar to processive motions [1, 25, 30, 32, 33]. They have not been used to study
smaller changes in rotational state due to, for example, bubble formation during RNAP ini-
tiation [84] or Cas9 target search [85]. Such changes are likely difficult to study with the low
resolution provided by most existing methods, but at least in principle are observable using
existing experimental geometries.

The required experimental geometry is also constrained by whether the protein intro-
duces a net change in the DNA twist or transfers twist between the upstream and down-
stream DNA. In the former case, such as with gyrase activity [25], it is sufficient to know the
relative angle between the two ends of the dsDNA molecule to measure the DNA rotation.
Thus, the DNA can be directly tethered to the slide surface and the protein can bind from
solution during the experiment, without any additional torsional constraint on the protein
(as in Figure 1.3C). In the latter case, without net rotation the upstream increase (decrease)
in twist can be balanced by the downstream decrease (increase) in twist, resulting in no net
change, and no net rotation. This is the case for RNAP during elongation, which unwinds
upstream DNA and rewinds the downstream DNA as it exits the enzyme [1, 15]. If the rel-
ative angle between the two ends of the DNA were studied, only transient changes in angle
due to the drag of the RNAP as it spun around the DNA would be observed. Thus, to study
these changes, the enzyme itself has to be attached to the slide surface so that the relative
rotation between the enzyme and DNA is forced to relax through the rotational probe (as in
Figure 1.3B) [1].

Overall, the geometric and practical requirements of rotational methods introduce a va-
riety of constraints for rotation experiments. In developing a rotational method, we would
like to precisely control the experimental geometry and use a flexible design that is compatible with different types of required experimental geometries with minimal changes.

1.3.4 FORCE AND TORQUE

Finally, applying force and torque to the system during rotational measurements leads to a number of both benefits and drawbacks. As discussed Section 1.3.1, a stretching force on the DNA has traditionally been required in order to suppress lateral Brownian motions of the rotational probe due to bending of the DNA molecule. The necessary force depends on the size of the rotational probe [28], as the magnitude of the acceptable fluctuations drops as the bead size is reduced. This has presented a significant challenge in previous work that has improved the resolution of rotor-bead tracking: for tracking with the smallest demonstrated bead size (80 nm), a minimum stretching force of ~5 pN was required [28]. This limitation motivated the use of a larger bead, despite the correspondingly lower resolution, for the enzyme studies presented in that work. In the opposite extreme, it is possible to avoid the use of force by using a sufficiently large probe so that lateral motions do not significantly affect the angle determination. This approach has been demonstrated previously in just one case for DNA rotation, where a large dual-bead system was used to track DNA rotation induced by intercalating dyes [58]. Also, before a force can be applied both ends of the DNA must be tethered (e.g. to a bead and to a coverslip or enzyme attached to a coverslip). In particular, for experimental geometries with the protein itself attached to the surface (Section 1.3.3), the protein-DNA interaction is a structural component holding the DNA to the slide so the system has to be pre-assembled prior to the application of the required stretching force, making it impossible to study the initial binding and full initiation process of the protein on DNA.

Perhaps the most important consideration when applying a force is its effect on the biological system. Note first that the physical properties of DNA depend on applied force, in-
cluding the effective torsional stiffness [33, 47, 86] and twist angle between bases [64]. For many enzymes, an applied stretching force also affects activity. Force dependence is expected in any enzymatic step that involves a change in the length of the DNA, for example due to moving along the DNA or wrapping the DNA, because the force changes the free energy landscape of the step [87, 88]. The overall effect on enzyme kinetics depends on whether that step is rate limiting. This effect means that measurements under force may give different results than, for example, bulk biochemistry experiments, but also provide an additional control parameter which can be used to dissect the reaction mechanism. For example, the force dependence of gyrase was used to propose a mechanistic model explaining both activity initiation rates and processivity (both force dependent properties) and to conclude that the rate limiting step (which was not force-dependent) was not a DNA-wrapping step [25]. Overall, careful consideration of force-dependencies can provide meaningful additional information because force perturbs the system being studied.

The applied restoring torque presents a different kind of challenge in rotational experiments. Unlike the force, which can generally be held constant, and in the same direction (‘up’, stretching the DNA), the restoring torque in the approaches described in Section 1.2 results in a preferred angle in a potential energy well. In the most extreme case, the traditional magnetic tweezer, this torque completely prevents any observable rotation of the bead, obviously preventing direct rotational tracking. Reducing this torque made it possible to directly track rotation for a number of DNA biophysics and enzyme studies. In many cases, the magnitude of the residual torque and its potential effects on the observed behavior was unclear [1, 32, 34, 57]. Note that while even with significant applied torque (e.g. trap stiffness of ~20 pN nm rad$^{-1}$ [30]) it is possible to track enzyme-induced DNA rotation, this torque will significantly bias the position of the bead during each rotation.
Unlike the force, torque has not been required for rotational measurements and approaches such as the freely-orbiting magnetic tweezer (FOMT) [33] allow torque-free measurements of DNA rotation (see Section 1.2.3). Promisingly, the magnetic-torque tweezer [47], which allows the application of a moderate, well-defined restoring torque, can be altered to allow variable torque application in a single experiment, allowing a range of torque measurements to be taken on a single molecule [89], and likely can be taken to zero torque. As with force, torque can also change the structure of the DNA [21] and alter the energy landscape for enzyme activity. Simultaneously measuring the DNA rotational angle and applying a well-defined, position-independent torque will likely require methods with active feedback [68, 70] or rapidly oscillating fields [71] to apply a constant, position-independent torque.

Ultimately, whether the application of an external force/torque to the system should be considered a benefit or drawback of an approach depends on the question being asked. Given the current state of the field, it is currently more urgent to develop force- and torque-free approaches.

1.3.5 Features of an ideal approach

In this section, I have discussed a number of fundamental and practical features of single-molecule direct DNA rotation measurements. These features are interconnected in complex ways, and optimizing for one generally requires trade-offs in another (e.g. achieving higher resolution has required applying larger stretching forces). Taken together, though, the ideal approach would be (1) high resolution (requiring a short dsDNA molecule and small rotational probe), (2) high throughput (requiring widefield imaging and a large fraction of intact, fully assembled complexes), (3) compatible with different modes of enzyme action (e.g. binding to DNA ends or internally on the DNA) and (4) able to operate without an applied force or torque but compatible with their application. Additionally, ideally the method would be
relatively straightforward to implement experimentally and use minimal specialized equipment in order to facilitate more widespread adoption. Such an approach would make it possible to study a wide range of protein-DNA interactions in great detail.

1.4 DNA vs protein rotation

The focus of this thesis, and this introduction, is measuring the rotation of DNA. The rotation of protein domains has also been of great interest to the biophysics community and can be measured to address a distinct set of biological questions. While generally outside the scope of this thesis, a brief discussion of the differences in measuring DNA and protein rotation is helpful for understanding why the former has often been more challenging.

A key distinction between tracking protein and DNA rotation is the difference in structural flexibility. As discussed in detail above, the resolution of DNA rotation measurements is fundamentally limited by the flexibility of DNA. While protein molecules of course vary greatly in their physical properties, in general protein domains are smaller and reasonably rigid compared to typical DNA segments. Thus, rotational measurements using structural rotation amplification can sometimes be made with minimal Brownian motion. For example, for ATP synthase, rotation of the central domain relative to the rest of the protein can be measured with ~4° precision with 4 ms integration time without applying a force to the small rotor bead [90].

The rotation of protein domains can also be measured using polarization fluorescence microscopy [91, 92]. These assays take advantage of the relationship between the dipole moment of the fluorescent dye and the polarization of the excitation and/or emission light to measure the orientation of the dye. Assuming the dye is rigidly attached to a protein domain of interest, and the protein is constrained relative to the imaging system, this dependence
reports on angular changes. Such approaches can be used to measure protein domain orientational changes on the millisecond timescale [93]. However, such approaches have not been extended to study DNA rotation, likely due to the additional noise introduced by the DNA Brownian dynamics, the short lifetime of single fluorescent dyes under the high illumination intensities needed for these experiments, and difficulties in achieving highly orientationally constrained incorporation of the fluorescent dye in the DNA structure.

It is also worth noting that studies of protein domain rotation require directly labeling the protein, either with a structure to mechanically amplify the motion (e.g. a bead [90] or an actin filament [94]) or a rigidly incorporated fluorescent dye [93], which requires protein engineering for each protein of interest. By contrast, when studying DNA, the same type of DNA-probe attachment can be used for each protein, which should make it easier to generalize DNA rotation assays.

1.5 DNA nanotechnology

Most of the DNA rotation tracking methods described in sections 1.2.2 - 1.2.3 are based on a simple principle: a bead is used as a lever arm to mechanically amplify the small rotation of DNA into a larger signal. These structures are limited in terms of size and shape (and thus drag and experimental geometry), and cannot be optimally combined with the dsDNA of interest. We would like to instead be able to use a highly customizable, highly DNA-compatible material as the amplifying rotor. DNA nanotechnology, and in particular, the DNA origami technique, provide a way to generate a rotor with ideal and customizable properties.
1.5.1 Approaches for Building DNA Nanostructures

In many ways, DNA provides an ideal material for the construction of bottom-up nanofabricated objects [95, 96]. Segments of DNA can be brought together in a precisely controlled arrangement using the standard base pairing interactions of DNA, and the four nucleotide alphabet is sufficient to achieve very precise interactions with reasonable lengths of DNA but simple enough for structures to be designed relatively easily and predictably. Furthermore, DNA strands can be purchased at low cost, and can be functionalized (e.g. with fluorescent dyes, biotin, or chemically reactive groups) in a very well-defined manner. Thus, in contrast to, for example, protein structure design (which, while quickly improving, remains a challenging problem [97]), the construction of complex 3D structures using solely DNA has rapidly matured as a field. It is now possible to build DNA nanostructures with a wide range of shapes and functionalizations with high reproducibility and customizability.

The design of nucleic acid sequences for the construction of small nucleic-acid objects based on simple base-pairing rules was first proposed by Seeman in 1982 [98] and experimentally demonstrated shortly thereafter, with a simple stable DNA junction [99]. Over the next decade, the Seeman lab built increasingly complex small nucleic acid structures using junctions between dsDNA edges, including a quadrilateral [100], a small cube [101] and an octahedron [102]. However, folding these 3D structures required multi-step processes, limiting the potential for increasing the scale of DNA nanostructures. In parallel, the Seeman lab also developed the basic building block that would define most DNA nanotechnology moving forward, the multi-helix crossover structure [103, 104], which initially consisted of two parallel or anti-parallel DNA double helices with two crossover points to constrain the geometry and dynamics. This double-crossover design provides sufficient rigidity in the structure for the creation of large 2-D lattices, built hierarchically from small squares [105] or DNA junctions
using base-pairing interactions of short ‘sticky’ overhangs. This cross-over approach using only short DNA oligomers was also used to build the first six-helix DNA bundles [107], the key design component used for the experiments in this thesis. It was also demonstrated around this time that these structures can also be built from RNA [108], although DNA remains the dominant choice of building material due to its higher stability.

While these approaches can be used to build large lattice structures, building more highly customized, large aperiodic structures required an additional advance: the use of a relatively long DNA scaffolds to organize the structure and provide uniquely addressable sites. These were initially generated biochemically via PCR in order to fold a templated 3-D octahedron structure using a 1,669 nucleotide (nt) scaffold and five shorter oligomers [109]. The use of this scaffold made it possible to simplify the folding process in a one-pot reaction with a simple temperature ramp, the double cross-over design made the structure rigid, and, unlike for the previous 3D structures, no covalent linkage of the DNA oligomers (via DNA ligation) was required. However, in this structure, the scaffold was mostly base-paired to itself, not the additional oligomers, limiting the ease with which the shape can be customized as well as the achievable complexity and size.

The introduction of the DNA origami approach in 2006 increased the achievable size, customizability, and ease of preparation by using ~200 short ‘staple strands’ that base pair to a ~7 kb single-stranded scaffold [110]. Each staple strand is complementary to the scaffold at multiple separate regions, causing those regions to come together in a cross-over structure (Figure 1.4A). Using this approach, a large number of 2D structures on the 100 nm length scale could be easily designed and folded with high fidelity. Additionally, the scaffold chosen was single-stranded viral genomic DNA (M13mp18), which can be prepared in large quantities directly from the virus. This approach was extended to a simple 3-D structure, the six helix bundle, shortly after in order to create bundles of well-defined size for liquid crystal
Figure 1.4: 2D and 3D DNA origami. (A) Design of a 2D DNA origami structure. Non-contiguous segments of the template strand (black) are folded together using base pairing with shorter staple strands (assorted colors). Reprinted from reference [110] with permission. (B) Design (top two panels) and TEM images (bottom two panels) of several 3D DNA origami structures. Scale bar: 20 nm. Reprinted from reference [111] with permission.

A few years later, the same principles, combined with a more sophisticated design approach, were used to generate a diverse set of 3D DNA origami structures (Figure 1.4B) [111]. Around this time, software which greatly facilitates the design process was published [113], and DNA structural nanotechnology began to be more widely used for a variety of applications. A number of additional modalities for designing and folding DNA structures have been introduced, including single-stranded tiles [114], DNA bricks [115–117], single component RNA and DNA origami [118, 119], wireframe structures [120–124], and DNA-protein hybrids [125], adding to the diversity of DNA structures that can be created for various applications. These new technologies introduced new capabilities, but for the work described in this thesis, the DNA origami approach remains optimal due to the ease of design and preparation, achievable structural rigidity, and stability.
1.5.2 DNA Origami Design Principles

The development of the ORBIT assay described in this thesis required the design of two DNA origami structures. In this section, I introduce the principles that guide the design and preparation of DNA origami.

As described in the previous section, DNA origami consists of a long scaffold strand (typically ~7 kb) and ~200 short staple strands (usually, mostly 20-50 nt long [111]). The staples are complementary to two or more non-contiguous regions (domains) on the scaffold; forming connections between the staple and the multiple separate regions of the scaffold results in folding of the structure. To design such structures, generally the scaffold strand is first routed through a space of the desired shape, and then the staples are added following a number of rules which increase the likelihood of proper folding [110]. These rules have been developed using geometric and steric constraints, through studies of the nature of the folding process, and through experience. Nonetheless, many rules remain heuristics and not all can necessarily be satisfied simultaneously, so design can require an iterative, trial-and-error approach [126].

At the most basic level, the crossovers between separate domains of the scaffold must be designed such that the helix geometry is properly aligned. For example, a pair of crossovers between two helices should be separated by an integer number of turns of DNA to avoid straining the DNA [110]. Similarly, crossovers that lead to DNA on opposite sides of a helix should be separated by a half turn. Since a turn of DNA is about 10.5 bp, some strain in the structure is inevitable, which can be minimized by adjusting exact crossover positions [110]. In 3D origami, to minimize this strain, crossovers are only allowed between two helices at positions 21 bp apart [111, 113]. The crossover pattern between sets of neighboring helices is determined by the type of lattice used to organize the helices: in the honeycomb lat-
tice, the helices surrounding one DNA are 120° apart, and crossovers are allowed (to different neighboring helices) every 7 bp \cite{113}. These rules can be automatically applied to generate an initial origami design, which is typically then refined based on additional considerations \cite{113}. For example, short unpaired regions of the scaffold are generally needed at the end of each helix to avoid base stacking between origami complexes, which can lead to aggregation \cite{111}.

Even with a slow temperature ramp, the folding process does not necessarily occur as an equilibrium process. This can be observed, for example, from the hysteresis in the folding and unfolding reaction progress during temperature ramps \cite{127, 128} or from the differences in folding and unfolding pathways \cite{129}. Additionally, the folding path is highly influenced by the set of staple strand domains that bind earliest. Minor staple strand changes that affect which domains are favored to form first can alter the final shape of origami cleverly designed to fold into multiple, distinguishable structures, even though the staple changes should not thermodynamically favor one shape over another \cite{130, 131}. More generally, many origami structures and DNA model systems have been shown to fold or unfold starting from a particular region, which nucleates further folding \cite{129, 132–134}, and truncating a small number of early-binding staples can dramatically slow down the folding process \cite{126}. Non-equilibrium folding makes kinetic traps a significant concern in the design process. Understanding and controlling the folding pathway (in particular, the order in which different parts of the origami fold) has proven critical in the preparation of certain topologically-complex origami \cite{135}.

In addition to occurring out-of-equilibrium, folding is a highly cooperative process - as one staple strand brings together distant domains of the scaffold, the entropic costs of adding another neighboring strand decreases \cite{110, 126, 127, 132}. Additionally, newly formed double helices can base stack with existing helices in the structure, contributing further to cooper-
ativity [136]. Direct evidence of cooperativity comes from measuring the binding of specific staples: staples bind at lower temperatures (later in folding) if nearby staples are omitted [128]. Because of this cooperativity, the origami folding process occurs over a much narrower temperature range than would be predicted based the melting temperatures of the individual strands alone, making it possible, at least in some cases, to fold origami isothermally [127].

One potential problem during folding is the blocking of scaffold domains by competing staple strands with the same identity. Ideally, exactly one of each unique staple strand will bind to the scaffold, but, depending on the design, it is possible to have two bound (transiently, or not) to the separate scaffold domains which should eventually be brought together [137]. Blocking can be resolved by one strand unbinding, allowing the other to take its place, or perhaps by direct strand invasion. However, if the strands bind to the two domains too strongly, the blocked configuration may become a kinetic trap. For example, by the time (or equivalently, temperature) it is favorable for the two scaffold domains to come together, unbinding of either domain may be very slow, preventing formation of the proper structure. For this reason, it is beneficial to design origami staples with one domain, a 'seed' region, much more stable (i.e. longer) than the others [138, 139]. Blocking may also be more likely with high staple concentration, as at higher concentration there are more staples competing for the limited binding sites.

Another challenge is the imperfect incorporation of staple strands: one group found that, when using the typical 10-fold excess of staples over the scaffold, ~15% of staples may not be present in a structure [140]. An earlier study of another structure found only 1-4% of staples were missing from several 3-D structures, but used a much less direct approach to quantify defects [141]. Of course, the actual number of staples missing will depend on the structure and folding conditions. While this does not generally prevent the structure from
folding properly (defined as appearing to be the correct structure in TEM) [141], it does suggest that it may be desirable to use a larger excess of staples for critical components.

Collectively, this description of the folding process leads to a number of design guidelines which can increase the likelihood of an origami structure folding properly, including considering which region of the origami should ideally fold first and designing staples with strongly binding seed regions to control the folding pathway.

1.5.3 APPLICATIONS OF DNA NANOTECHNOLOGY

The ability to build customized DNA structures has proven useful for a wide range of applications. These applications are remarkably diverse, and include DNA computation [142], periodic [143] and non-periodic [144] scaffolding, liquid crystal formation for protein NMR [112], templating of polymer synthesis [145], drug delivery [146], and sensing [147–149]. Because of the scope and number of these applications, this is only a representative list; the interested reader is directed to a number of recent and comprehensive reviews on the applications of DNA origami [150–157].

Most importantly for this thesis, DNA nanostructures have been applied in a variety of innovative ways in single-molecule biophysics. Due to its well-defined structure and composition, DNA origami is a convenient, reliable standard for developing or calibrating new techniques (especially in fluorescence imaging) [158–164] or for structural or fluorescent bar-coding [165–167]. DNA origami has also been used to build a ‘frame’ to hold DNA with precise positioning to constrain the end-to-end distance [168, 169] or twist [170], or apply a force using single-stranded DNA as an entropic spring [171]. Similarly, researchers have created platforms for conveniently tracking enzyme motion along an incorporated double-stranded DNA molecule [172] or for arranging motor proteins to characterize their collective behavior [173, 174]. Origami can even be folded into a strained state to create a spring [175].
In this thesis, I will present work using DNA origami as a readout probe for single-molecule measurements. In addition to origami’s ability to be easily functionalized and its compatibility with DNA substrates, this will rely on three features of origami which have been increasingly recognized and used by the field over the last several years: (1) Origami can be used as a readout probe, (2) Origami can be made structurally rigid, and (3) Origami can be used as a lever arm. First, simple origami structures can be designed to report on biologically-relevant motions. For example, origami can be easily observed in AFM, and thus can report on the position of two ends of a single-stranded DNA molecule [176]. Small, fluorescently-labeled origami structures can be used in a variant of the classic tethered particle assay, reporting on the length of dsDNA in real-time through the Brownian dynamics of the origami [177]. Second, origami can be made much more rigid than regular dsDNA. For example, a six-helix bundle of DNA has a bending persistence length of ~2 μm, roughly 40-fold higher than dsDNA [178]. This property can be used to suppress noise in optical trapping experiments, which are sensitive to DNA stiffness [179], and generally can make origami a reliably rigid structural component for single-molecule experiments. Third, the rigidity of origami makes it possible to use origami as a lever arm to amplify signal amplitude. This was first demonstrated using a pliers-like DNA structure, which bind an analyte near the hinge, closing the pliers, and thus resulting in a structural change which could easily be observed by AFM [180]. Recently, an origami arm was also used to amplify the random rotational motion of an unconstrained origami rotary connection, albeit with insufficient time resolution for position tracking [181]. A hinged structure can also be used to amplify changes in the wrapping state of a nucleosome tethered between the arms [182] or the interaction between two nucleosomes [183]. Alternatively, the lever arms can be used in the opposite sense: a hinged origami structure can be used to very precisely control the distance between the arms near the hinge point by using less precise control over the distance near the distal ends [184].
Combining these features, origami can be used to structurally amplify motion for real-time single-molecule tracking of movements that would otherwise be too small.

1.6 Outline of the dissertation

This dissertation presents the development and application of a new method, Origami-Rotor-Based Imaging and Tracking (ORBIT) for tracking single-molecule DNA rotation in real time. This tracking is achieved using DNA origami rotors to structurally amplify the small rotation of double-stranded DNA into a much larger signal that can be readily tracked. I then present the development and characterization of ORBIT in Chapters 2 and 3. Next, I describe our application of ORBIT to study the torsional stiffness of DNA molecules by measuring the thermally-driven Brownian dynamics of the origami rotor in Chapter 3. In Chapter 4, I detail the use of ORBIT to study the unwinding of DNA by the homologous repair enzyme RecBCD during initiation and processive unwinding. I present the application of ORBIT to study transcription by RNA polymerase in Chapter 5. I will separately introduce these two enzymes in the relevant chapters. The methods used in each of these studies and the full set of DNA sequences used to generate the origami are given in Chapter 6 and Appendix A, respectively. Finally, I will summarize these developments and point to future directions for improvements of this method and the scope of potential applications in Chapter 7.
In Chapter 1, I described the motivation for developing improved methods for tracking the rotation of single molecules of double-stranded DNA (dsDNA), especially when being acted upon by a protein. I then discussed how the DNA origami technology is a natural tool for structural amplification of the rotation of DNA. In this chapter, I first describe, conceptually, the design of the ORBIT assay and the origami rotor. I then discuss in detail how the design parameters are expected to affect the performance of ORBIT, with a focus on the expected,
resolution-limiting, Brownian dynamics of the rotor. Next, the detailed origami design, con-
struction, and validation are given, as well as two key design considerations required to create
a robust assay. Finally, I discuss a second origami structure, the ‘anchor’, which can be used
for alternative attachment of the rotor to the microscope coverslip, generating a distinct ex-
perimental geometry that allows for a different set of experiments. These structures will be
applied in the following three chapters for three separate applications of ORBIT.

2.1 Overview of the ORBIT assay

Before considering the detailed design of the ORBIT assay, it is important to understand an
overview of how the assay works, as illustrated schematically in Figure 2.1. ORBIT was de-
designed to track the rotation of single dsDNA molecules relative to a motor protein acting on
the dsDNA (for example, as in Figure 2.1, bound to the end of the DNA). The protein gen-

![Figure 2.1: Schematic representation of the ORBIT assay with a DNA motor protein. The motor protein is attached to a microscope slide or coverslip in a torsionally constrained manner and acts upon a regular dsDNA substrate, causing it to rotate. This dsDNA is incorporated into the DNA origami rotor in a torsionally rigid way, such that rotation in the dsDNA must relax through rotation of the rotor. Multiple dyes are placed on one arm of the rotor to provide a fluorescence readout of the rotor rotation. This fluorescence can be captured on a standard fluorescence microscope (not shown).](image)
erates rotation of the dsDNA around the dsDNA helical axis (for example, by unwinding the
dsDNA helix), which results in small physical motions since the radius of the DNA is only
\(~1\) nm (small red arrow in Figure 2.1). To directly tracking this rotation, it is be mechani-
cally amplified by a rotor, which acts as a lever arm. In the ORBIT method, this rotor is a
DNA origami structure. The end of the dsDNA opposite the protein is seamlessly integrated
into the origami rotor structure, so that rotation of the dsDNA is faithfully output to the ro-
tor. The rotor amplifies the rotation (larger red arrows), making it easier to detect changes
in the position of the rotor arm(s). The rotor should be designed to structurally amplify the
rotational motion without adding excessive drag, but does not necessarily require the shape
shown in Figure 2.1.

To enable detection of the rotation of the origami, the rotor is labeled on one or more
arms with fluorescent dyes. The position of these dyes during each frame can be measured
using a fluorescence microscope and the center of the fluorescent spot can be very precisely
determined by fitting to a 2D Gaussian function, allowing the determination of the position
of the rotor in each frame at very high spatial resolution (~10-15 nm). This amplification
scheme relies on the transmission of DNA rotation from the protein to the rotor. Any rota-
tion of the protein itself (relative to the lab frame) would interfere with tracking, as it would
prevent the relative rotation between the protein and DNA from being faithfully transmit-
ted to the rotor. Thus, the protein must be torsionally constrained relative to the lab frame
of reference (or, equivalently, relative to the microscope coverslip). Otherwise, rotation of
the entire protein-DNA-rotor complex can interfere with DNA tracking. This is achieved
by attaching the protein to a stationary surface. Since tracking is based on fluorescence mi-
croscopy, the surface is either a microscope slide or coverslip that have been incorporated
into a flow chamber for controlled buffer exchange.
Importantly, because fluorescence measurements can be taken using a widefield total internal reflection fluorescence (TIRF) microscope, we can image many such protein-DNA-rotor complexes in parallel. Also, note that no external force or torque is applied to the origami. As a result, the protein-DNA complex does not necessarily have to be pre-assembled: the rotor-substrate complex’s initial binding to the motor protein can be observed (see Section 1.3.3). Additionally, the observed activity is not biased by applied force or torque (see Section 1.3.4).

To turn the schematic in Figure 2.1 into an actual design, there are several important design choices to consider, which I outline here, starting from the protein and moving toward the dyes. First, how will the motor protein be attached to the surface? To achieve torsional rigidity, the protein needs to be attached through multiple points - the single-point attachment typically used to tether proteins to surfaces allow free rotation. Throughout this thesis, multiple attachment is achieved using passive adsorption of the protein to the coverslip; since each point of contact is a weak, non-covalent attachment, it is highly unlikely for a protein to remain stably bound without multiple attachment points. Other options will be considered in Section 7.2. Second, how long will the dsDNA portion be (see Section 2.2)? Third, how is the dsDNA incorporated into the origami structure to ensure torsionally rigid incorporation (Section 2.4)? Fourth, what are the ideal shape and size for the origami structure (Section 2.3)? Fifth, how should fluorescent dyes be incorporated (Section 2.5)? The third and fifth points will be considered after presenting the final rotor structure in Section 2.3 to facilitate discussion of the detailed design points without the need to present intermediate full rotor structures. Finally, in Section 2.7, I present the design of a DNA origami ‘anchor’ which allows enzyme-free surface attachment.
2.2 Localization precision and Brownian dynamics

The resolution of ORBIT is fundamentally limited by two factors: (1) How well can we track the position of the labeled tip of the origami rotor? (2) How well does the position of the tip of the rotor track the underlying angle of interest, namely the angle of the DNA at the motor protein? In this section, I consider these two factors, and use them to motivate the design choices related to the dsDNA length and the origami rotor shape and size.

ORBIT is a particle-tracking assay, so it is critical to know very precisely where the particle (origami rotor) is during each frame. The position of the tip of the origami rotor is determined by fitting the fluorescence intensity point spread function to a Gaussian function \[185\]. The precision of this localization depends on a number of factors, but most importantly, on the intensity of the fluorescent spot. Thus, we can improve our ability to track the origami by using high laser powers or multiple fluorescent dyes. Because we are interested in tracking rotation, the Cartesian coordinate position of the origami must be transformed into polar coordinates. Since we are actually interested in how well we can track the angle, not the (x,y) position of the origami arm, this gives a second way to improve the tracking: a longer lever arm. The uncertainty in the angular position, \(\Delta \theta\), scales with the uncertainty in the Cartesian coordinates, \(\Delta x\), approximately as \(\Delta \theta \approx \Delta x/r\), where \(r\) is the radius of the circle traced by the origami. However, we will see shortly that increasing the size of the origami arm comes with significant drawbacks. In any case, we see there are three options for increasing the precision with which we can track the origami arm: increase the laser power, increase the number of fluorescent dyes, or increase the length of the origami arm.

Although it is of critical importance to know where the origami rotor is during each frame, it is imperfectly related to the actual quantity we are interested in: the angle of the DNA at the enzyme contact point, which reports on the activity of the enzyme. On average,
the origami arm angle tracks this position because the DNA has a preferred twist, but it is also affected by random thermal fluctuations in the twist angle of the DNA between the DNA and the rotor as well as movements of other components, such as the rotor-dsDNA junction. These fluctuations away from the equilibrium twist of the DNA incur an energetic penalty, which can be described as a simple harmonic potential over the thermally accessible range of angles (even under an applied torque, only slight deviations from the linear force response are observed at large DNA distortions) [21]. That is, the rotor feels a restoring force proportional to the angular displacement away from the equilibrium position. These Brownian fluctuations contribute to the noise and are a fundamental factor limiting the resolution for ORBIT (as well as previous particle-tracking assays such as rotor bead tracking) regardless of how well the position of the rotor itself can be tracked [28].

The Brownian dynamics of the protein-dsDNA-rotor system can be determined by considering the equation of motion for the rotor, which treats the potential due to the torsional stiffness of DNA as a simple harmonic well [79]

\[ m\ddot{\theta}(t) + \gamma \dot{\theta}(t) + \kappa \theta(t) = (2k_B T \gamma)^{1/2} \eta(t), \tag{2.1} \]

where \( m \) is the mass of the rotor, \( t \) is the time, \( \gamma \) is the hydrodynamic drag of the rotor moving through the solvent, \( \kappa \) is the spring constant of the system, \( \theta \) is the angular deviation from the equilibrium position, \( k_B T \) is the thermal energy, and \( \eta(t) \) is a stochastic, time-varying, thermal force due to collisions from solvent molecules. In aqueous solution, the drag dominates over the inertial contribution, and the mass term can be dropped. Because of the contribution of the hydrodynamic drag, the resulting fluctuations are time-correlated (even though the thermal forces are random and uncorrelated), and the resulting fluctuations can
be Fourier transformed to give the power spectrum $P(f)$, which provides the contribution of
the Brownian fluctuations to the noise in the angular position at each frequency:

$$P(f) = \frac{k_B T}{2\pi^2 \gamma (f^2 + f_c^2)}, \quad (2.2)$$

where $f$ is the frequency, and $f_c$ is the corner frequency,

$$f_c = \frac{\kappa}{2\pi \gamma}. \quad (2.3)$$

The corner frequency, or more conveniently, the relaxation time, $\tau = \gamma / \kappa$, indicates the
timescale over which the correlations in the fluctuations in the position of the rotor decay.

More intuitively, the noise resulting from the Brownian dynamics when data is integrated for
a duration $t$ can be determined by integrating the power spectrum, giving [28, 80]

$$\sigma(t) = \sqrt{\frac{2k_B T}{\kappa} \left( \frac{\tau}{t} - \frac{\tau^2}{t^2} \left( 1 - \exp \left( -\frac{t}{\tau} \right) \right) \right)} \approx \sqrt{\frac{2k_B T \gamma}{\kappa^2 t}}, \quad (2.4)$$

where the approximation is valid for $t >> \tau$. The approximation reveals more clearly the
important factors affecting resolution: for a given integration time, the Brownian noise depends
on the drag, $\gamma$, and the spring constant of the DNA, $\kappa$. Physically, these dependencies can be
understood as follows. A weaker spring constant allows the rotor to sample a larger range of
angles, resulting in more noise. From the equipartition theorem

$$\text{var}(\theta) = k_B T / \kappa, \quad (2.5)$$

where $\text{var}(\theta)$ is the variance in the angular position, we can also see that the total range of
angles explored by the rotor depends only on $\kappa$ (for a given experimental temperature). The
drag influences the noise by affecting the relaxation time of the motion: since the position of
the origami during each frame is correlated, additional time-averaging is required to average out these fluctuations. Thus, due to the Brownian fluctuations, high precision requires a large spring constant and low drag.

The spring constant of the system is determined by the torsional stiffness of each component. Most notably, as the DNA becomes longer, the spring constant decreases as the inverse of the length of the DNA \[78\]. Additionally, there may be additional flexibility at the junction between the origami and the dsDNA, which ideally should be minimized. Finally, the origami arm itself could contribute flexibility, depending on the stiffness of the design. However, in practice this should be negligible when using even a simple six-helix bundle, which has a persistence length of \(-2 \mu m\) \[178\].

The hydrodynamic drag is determined by the size and shape of the rotor, suggesting that the origami should be as small as possible. Of course, the role of the rotor is to amplify the motion, so in practice there is a trade-off between the signal amplification and drag. Thus, it may be beneficial to use different amplification factors for different experiments, depending on the timescales of interest and the required tracking circle radius. At a given amplification factor, we can also consider the shape of the rotor. Typically, motions have been amplified previously using approximately spherical beads \[78\] for practical reasons. However, a spherical structure introduces additional drag at a given size (see Section 2.6, where the drag of the origami rotor and a spherical bead are estimated). Thus, as implied in Figure 2.1, the rotor will be designed as a planar structure.

In summary, achieving high resolution will require high angular localization precision (high laser intensity, multiple fluorescent dyes and/or a long rotor arm) and minimal Brownian fluctuations on the timescale of interest (high spring constant, from a short dsDNA segment between the motor protein and rotor, and low drag, from a small origami structure).
2.3 Design of the origami rotor

The full origami routing diagram for the origami rotor, designed using CADNano [113], and several 3-D renderings of the folded structure are shown in Figure 2.2. The staple strand sequences used to fold the structure are given in Table A.1. The structure consists of two arms. The arms are constructed as six-helix bundles, which are a standard origami component with very high rigidity [107, 178]. One of the arms is a fully intact bundle, while the second has an opening for the first to pass through (Figure 2.2B). The area around this junction point is reinforced with additional helices to reduce flexibility. This design is reminiscent of the previously published ‘slotted cross’ origami, but with much thinner arms [111]. Each arm is ~160 nm in length, which should result in the ends of the arms tracing out a circle of ~80 nm radius when the rotor is rotated around its center, on an axis perpendicular to both arms. With an experimentally reasonable localization uncertainty of 15 nm, this arm length should result in an angular uncertainty due to localization precision of about 10°. This length is also compatible with the standard single-stranded DNA scaffold used for origami folding, the M13mp18 viral plasmid [110]. To reduce the change of topological and kinetic traps, the structure was designed with primarily local folding (most staple strands are restricted to a small stretch of a single six-helix bundle).

At the end of one arm, six of the staple strands (indicated in green in Figure 2.2A) were directly labeled with the Cy3 fluorescent dye. This direct attachment (as opposed to labeled secondary DNA oligomers) was an important design choice, as will be described in Section 2.5. The dyes are spaced out around the helix and over the last 10 nm of the arm to reduce photophysical effects such as quenching from nearby dyes [186, 187]. The opposite end of the arm is designed in the same way, and can easily be labeled as well if additional
Figure 2.2 (following page): Origami rotor design. (A) Routing diagram. The origami rotor structure consists of two 160-nm arms. The intact arm (a six-helix bundle) passes through a break in the orthogonal arm (two half-length six-helix bundles). Additional helices stabilize the junction. Two of these helices contain staple strands (black, near center of diagram) that are extended beyond the origami structure (extension not shown; see (C) and Table A.1). The portion of these strands extending from the origami consist of 14 bp of complementary DNA and a single 12-nt overhang on one staple for ligation. Six staples within 14 nm of the end of the intact arm (light green, on bottom-left) are labeled with Cy3 at their 3’ ends. (B) 3D rendering of the rotor design. In (A) and (B), the scaffold is in blue and the staples are in the remaining colors. (C) Two staple strands (red) are extended from the center of the structure, forming a 14-bp double stranded region and a 12-nt overhang on one strand for ligation. (D) The overhang is ligated to a longer piece of dsDNA, which serves as the substrate of the DNA motor protein.
Figure 2.2: (continued)
information about the movement of the rotor is required (e.g. to separate rotational and lateral motion).

The rotor was additionally designed with two complementary strands that extend from the structure (shown in red in Figure 2.2C). One of these strands has a 12-nt single-stranded overhang for convenient and direct addition of a longer dsDNA extension, which serves as the substrate for the motor protein. The precise design of this junction point was an additional critical design consideration, considered in more detail in Section 2.4. A rendering of the final complex, after ligation of a dsDNA extension which serves as the substrate for the motor protein, is shown in Figure 2.2D. Note that the four-arm design was chosen to encourage binding of this structure in an ‘upright’ configuration: the angle of the dsDNA makes with the surface-bound motor protein is limited to be close to the vertical due to steric constraints from the four arms. This should encourage binding to enzymes which are oriented with such that they bind DNA pointing directly away from the surface, making it more likely the rotors can rotate unimpeded. However, this design feature may not be necessary for rotation tracking.

The required staple strands can be purchased at modest overall cost; only staples which are modified (e.g. with a dye or 5’ phosphorylation) require purification. The rotor was assembled by mixing together the scaffold strand and staple strands, heating the mixture to 80°C and slowly cooling to 4°C over three days, as described in methods Section 6.1. The resulting structure was characterized using agarose gel electrophoresis (Figure 2.3A), which shows a single predominant product and a minor larger structure, likely a dimer or other higher-order product, as is typically observed in preparing origami [111]. The purified monomer band was imaged using atomic force microscopy, which showed the expected rotor structures (Figure 2.3B, C).
Figure 2.3: Origami rotor characterization. (A) Agarose gel purification of the origami rotor after folding, showing a Cy3 fluorescence scan. The rotor is in each lane. The bright bottom band consists of the Cy3-labeled free staple strands, which are present in excess during folding. The major band above the free staples consists of the rotor, which is cut out and extracted, and the faint higher band is likely a rotor dimer or other multimer. (B) AFM images of DNA origami rotors. Scale bar: 100 nm. (C) A large field of view AFM image of origami rotors, showing a high yield of properly folded origami. Scale bar: 1 μm.

2.4 Design of the rotor-dsDNA junction

A critical consideration in designing the origami rotor is the torsional rigidity of the rotor - dsDNA extension connection. The short dsDNA that extends from the center of the origami can be ligated to an additional DNA extension, such as those given in Table A.3. This dsDNA acts as the substrate for the motor protein, and it is the rotation of this DNA that we wish to measure. Thus, to create any meaningful signal, the connection must be topologically constrained. This is achieved by the design shown in Figure 2.4A: both strands of the extension are incorporated into the helices of the rotor. When properly assembled, the dsDNA and
Figure 2.4: Origami rotor-dsDNA junction. (A) Simplified routing diagram of the rotor-dsDNA junction. Two helices of the origami are shown. The two extension strands (black) are base paired to the scaffold (blue) and each other, and surrounded by regular staple strands (green). The two extension strands come out of the structure adjacent to each other in 3D space. One extension strand has an overhang, for ligation to a longer extension strand. The extensions have $s$ nt base paired to the scaffold and $t$ nt base paired to each other. (B) Simplified routing diagram of the junction, in the problematic case where two dsDNA extensions are present. The structure contains two extension dimers, and each dimer has only a single scaffold-pairing domain inserted properly into the origami. For clarity, one of the extensions is shown above the helices; in the actual structure the two double helices would likely run nearly parallel, adjacent to each other. (C) Experimental design for determining number of linker strands extending from each origami rotor. The rotor (blue) is attached to a streptavidin-coated coverslip using twelve biotin secondaries on one arm (red; 'b' indicates biotin; only two are shown for clarity). Another arm contains twelve Cy3 dyes (indicated by the green star), which are used to find the origami. One of the extension strands is labeled with a single Cy5, and the number of Cy5 bleaching steps is used to estimate the number of extension strands present in the rotor. (D) Optimization of design and linker strand concentration for single-linker incorporation. The number of bp the linker strands make inside the structure ($s$), with each other ($t$), and their fold-excess over the scaffold concentration during folding are shown. For each origami, identified using the Cy3 dyes, the photobleaching on the red channel was categorized as 0, 1, 2 or 'other' steps.
rotor cannot rotate relative to one another; however, if only one extension strand is properly base paired, torsional rigidity cannot be achieved. Thus, optimizing this connection point is critical.

First, formation of the intact structure needs to be favorable. Toward this end, the junction is designed such that the two extension strands enter the origami structure on adjacent helices, where the helices are properly oriented to have a strand extending out from the structure. Second, the structure and folding conditions need to be designed such that there is a single extension DNA on each rotor. Of particular concern is the case where two extension dsDNA are present, each with one strand incorporated into the origami and a second remaining as single-stranded DNA as shown in Figure 2.4B. In this situation, each staple blocks the other from one of its two binding domains, preventing the origami from reaching the desired structure. With this undesirable product, neither dsDNA can report rotation and the rotor can become attached to multiple motor proteins. Experimentally, improper incorporation results in freely rotating rotor, randomly sampling the circumference of a circle.

In our initial design, each incorporated staple strand had significant base pairing (25 bp) to the rotor scaffold strand, as well as 20 bp to each other, in the external section (shown in red in Figure 2.2C). This was designed to make these connections highly favorable and stable. Additionally, we used a large excess of the extension strands, to ensure that they would be present in each origami. However, this design results in a large fraction of rotors which are not torsionally constrained, so we optimized this junction by using a direct readout of the number of dsDNA extensions present in each rotor. To determine the number of extensions present, we added a Cy5 fluorescent dye to one of the extension DNA strands and a biotin to the rotor itself (Figure 2.4C). The rotors were directly tethered to a streptavidin coated surface and identified using the Cy3 on each rotor. The Cy5 was imaged until bleaching, and the number of photobleaching steps was determined as a measure of the number of dsDNA
extensions present. As can be seen in Figure 2.4D, in the original design, a large minority of rotors contained two extensions, limiting the number of functional rotors.

The junction point is fundamentally different from the rest of the origami structure in that the two staple ‘extension’ strands anneal directly to each other. In all other parts of the structure, the staples are complementary only to the scaffold. To understand this result, we need to consider an alternative DNA nanotechnology design approaches, DNA tiles [188] or DNA bricks [115], which are scaffold-free design strategies. In these approaches, the entire structure consists of short strands, which base pair to each other to form a 2D or 3D structure. The extension junction can be viewed as an example of this architecture, with the two extension strands and the scaffold each binding in a pairwise interaction. Generally, for such structures, a wide range of undesired structures can form [189, 190]; clearly, the simplicity of our junction restricts the space of possible outcomes. Nonetheless, we can take two ideas from this field. First, folding is improved by reducing the concentration of components which are over-represented in the undesired structures (here, the extensions) [189]. Second, nucleation behavior, which determines when the structure begins to form, how quickly it forms, and the likelihood of reaching a kinetic trap, depends on the strength of the binding interaction [190]. In our simple case, we can equivalently consider whether the strength of the interactions prevents improperly forming structures from recovering.

In order to reduce the number of structures with two extensions, we redesigned the structure with weaker base pairing. This should encourage exchange of the staple strand during folding, perhaps giving structures which have two extensions at some point during folding a chance to find the desired structure. It should also result in the extensions binding later in the folding process, when the two binding sites are more likely to be physically adjacent. This should increase the chance that both binding sites will be bound around the same time, cooperatively, instead of early in the binding process when the two extension binding domains
are likely very far apart. Additionally, we reduced the concentration of the extension strands in solution, reducing the likelihood of having a second strand bind before the first finds its target. As can be seen in Figure 2.4D, these changes significantly reduced the fraction of rotors with two extensions. The design in the rightmost panel (15 bp inside the rotor, 14 bp outside) was used for all experiments in this thesis to minimize the number of origami with multiple, unconstrained dsDNA extensions and is given in Section 2.3 and Appendix A.1.

2.5 Incorporation of the Cy3 staple strands

A second subtle design consideration is the attachment of the fluorescent dyes to the rotor arm. The number and spacing of dyes is important, as multiple dyes provide much brighter signal and better localizations, but if dyes are too close together they will quench [191]. In our design, the dyes are all ~5 nm apart, spaced around the helix. There are two straightforward routes to incorporate the six desired Cy3 dyes in the rotor structure. The dyes can be directly incorporated into six distinct staple strands near the end of the rotor arm (Figure 2.5A). This is conceptually simple but requires purchasing six dye-labeled oligomers with distinct sequences, which makes the origami more expensive. These oligomers also are not re-useable between different origami structures, as the sequence must be custom designed for the particular structure based on the scaffold routing. Alternatively, the six strands can be extended with primary binding sites for a Cy3-labeled secondary oligomer (Figure 2.5B). The dye can be placed on the end of the secondary near the helix, resulting in a similarly tight cluster of dyes. The extension can also be extended so that multiple secondaries can bind to each. A single primary extension sequence can be used for each extension. Thus, only a single dye-labeled oligomer is needed, and it can be re-used across origami (there are no distinct
Figure 2.5: Two designs for dye incorporation. (A) Cy3 (green star) is directly incorporated into six distinct staple strands (red) near the end of the rotor arm. Between the staple strand’s scaffold complementary sequence and the dye is a single dT base (not shown). Because of the 3D structure of the arm, each dye is attached to DNA where the staple is on the exterior of the six-helix bundle. Unlabeled strands are indicated in black and the scaffold in blue. (B) For each dye position, a secondary strand (purple) is added as a direct continuation of the staple strand. Each purple segment has the same sequence. A single DNA strand (green) is labeled with Cy3 (green star), which binds to the purple single-stranded extension. Longer extensions can also be used, allowing two dye-labeled strands to bind. In (A) and (B), the schematic of the entire end of the arm is shown on the left, a section showing only a small section of the arm with a single label is shown top-right, and a head-on view of the origami arms (gray) with the position of the dyes on the exterior of the bundle is shown bottom-right. (C) Fluorescence localizations from a rotor with Cy3 and Cy5 on opposite arms, labeled using the secondary strategy shown in (B), except with two secondaries bound per extension. The rotor is introduced to a flow chamber after a DNA motor protein was adsorbed onto the surface. The rotor shows motion consistent with binding of the DNA motor protein to the Cy3-labeled arm (giving tight localizations, shown in green) while the other end of the arm explores a large space (larger than the diameter of the arm, shown in red). This is likely a consequence of the dsDNA regions created by the secondary labeling strategy, favoring the use of the direct labeling strategy shown in (A). Scale bar: 100 nm.
sequence requirements for the primary and secondary strands across structures). For this rea-
son, we initially favored the second design.

However, this design is problematic for some applications: each dye-labeled oligomer
duplex becomes an additional dsDNA helix terminating in a blunt end. When conducting ex-
periments with DNA motor proteins that bind to dsDNA ends, such as RecBCD (the topic of
Chapter 4), this creates additional, undesired, binding sites for the motor protein at the end
of the labeled arm. For rotors labeled on one arm only, this should result in tightly clustered
localizations as the dyes are not able to move. This problem can be more clearly identified
with rotors labeled on opposite arms with Cy3 and Cy5. For such structures, many origami
show random motions along the circumference of a circle with diameter up to approximately
twice the size of the origami (Figure 2.5C) when a DNA motor protein is attached to the
surface. This outcome is most easily explained as binding of the protein to one rotor arm
while the origami is otherwise free to move. Nearly 40% of rotors showed motion of this type,
with one arm highly constrained while the other moved randomly. For this reason, the de-
sign shown in Figure 2.5A was used in this work and is described in Section 2.3. For other
applications, the primary-labeled secondary design in Figure 2.5B may be preferable due to
its lower cost.

2.6 Expected Hydrodynamic Drag of the Origami Rotor

In Section 2.2, I motivated the choice of an effectively 2D origami rotor design for signal am-
plification on the basis of the lower drag, compared to, for example, a spherical bead provid-
ing the same amplification. In this section, I present an estimate of the drag for the structure
described in 2.3 and for a comparable spherical bead.
We estimate the hydrodynamic drag of the origami rotor by treating it as a pair of cylinders of length 160 nm and radius 3 nm rotating about the middle point along the length of the cylinder. For a single cylinder rotating about its middle point along the long axis, the drag $\gamma$ can be estimated as \[ \gamma = \frac{\pi \eta L^3}{3(\ln(p) + \delta_T)} \] (2.6) where $\eta$ is the solvent viscosity, $L$ is the cylinder length, $\delta_T$ is an end-effect correction, and $p = L/(2R)$, where $R$ is the cylinder radius. Note $L$ is the length of the entire bundle (160 nm) and the cylinder radius is the size of the cross-section of the bundle (not to be confused with with the radius of the circle traced out by the rotor). The radius $R$ is expected to be $\approx 3$ nm for a six-helix bundle [107]. The end-correction for the associated size ($p \approx 25$) is given by reference [192] as $\delta_T = -0.616$. Using this approach, and taking into account both arms, we estimate the drag of the origami rotor in water to be $\approx 3$ fN nm s. For experiments done in 10% glycerol, the viscosity is 30% higher than water, giving an expected drag of $\approx 4$ fN nm s. Note that this estimate does not take into account the increase in the drag of an object moving near a surface [193, 194], which will be revisited in Section 3.3.

For comparison, for a sphere of diameter $D = L$ rotating about its center, which would give the same amplification of DNA rotation as our origami rotor, the hydrodynamic drag would be $\gamma = \pi \eta D^3$ [78], giving $\gamma = 12$ fN nm s at 0% glycerol, or $\approx 4$ times larger than the drag of our origami rotor. This would lead to a $\approx 2$-fold lower time resolution in rotation measurements while achieving similar signal amplification. (In reality, the signal amplification may be lower, depending on what feature of the bead is tracked.)
2.7 DNA origami anchor design

The rotor complex shown in Figure 2.2D requires another component to rigidly tether the dsDNA to the surface. When linked to a dsDNA that ends in a double-strand break, this connection can be achieved using a protein adsorbed onto a surface (as in Figure 2.1). While this is sufficient for the enzyme studies in this thesis, it is highly desirable to be able to attach the rotor-dsDNA complex directly to the surface, without a mediating DNA-interacting protein for three reasons. First, this is required in order to characterize the inherent (Brownian) noise in the measurement, in the absence of any enzyme dynamics which might affect the measurement. Second, this also would make it possible to use ORBIT to study the physical properties of DNA. Third, although outside the scope of this thesis, protein-free surface attachment would be useful for studying enzyme activity that involves DNA distortion or opening but not processive motion along DNA.

Previous work has used a single dsDNA with a large number of biotin or digoxigenin groups over a few hundred base pairs for torsionally constrained surface attachment (e.g. [25, 28, 30, 32, 47, 69]). Attachment of two or more gives a rotational constraint. However, such an approach is not compatible with short dsDNA sections and gives slightly different constructs for each single molecule. This approach also requires that the dsDNA run parallel to the microscope coverslip surface over the length of the attached section, requiring a force to align the dsDNA upward and introducing a bend into the DNA being studied (Figure 2.6A). Thus, we designed an anchor structure to tether a rotor-dsDNA system to the microscope coverslip surface in a well-defined manner.

Perhaps the most obvious approach is to use the previously described origami rotor (with added biotin for attachment to a streptavidin-coated surface) as an anchor. However, counterintuitively, this approach led to dramatically more flexibility and thus noise, com-
Figure 2.6: Rejected surface attachment designs. (A) Direct attachment of dsDNA to coverslip using biotin-streptavidin. Torsional rigidity can be obtained with a simple dsDNA segment with multiple biotin modifications attached to multiple streptavidin (red circles) on a coverslip. However, this design requires that the biotinylated dsDNA segment runs parallel to the surface. In order to generate detectable rotation of the rotor (blue), a bend would be needed in the dsDNA, which requires an applied force. This general approach, except with a rotor bead instead of the origami structure, has been used in previous work (e.g. [25]). (B) Attachment of dsDNA to coverslip using a second origami rotor structure as an anchor. At the junction points, the dsDNA is routed directly into the origami, with a 90° bend (Figure 2.2C). As described in the text, this design suffers from additional noise due to the flexibility of the junction point. Because the dsDNA is between the lower junction and the readout dyes, it acts as a lever arm for tilting motions at the anchor-dsDNA junction.

pared to using this rotor to track rotation when attached to an enzyme. The difficulty derives from the flexibility of the origami-dsDNA junction to tilting motions (that is, changes in the angle between the rotor arm and the dsDNA). Such motions are likely highly accessible for this structure. The effect of such motions at the rotor junction on tracking is minimal: the projection of the tilting in the image plane is small over the accessible region. In practice, these tilting motions should slightly decrease the observed radius of the rotation. However, when such tilting motions are allowed at the anchor-dsDNA junction, the dsDNA itself acts as a lever arm, amplifying the motion and allowing the rotor arm to explore a large space (Figure 2.6B). Thus, this approach for the anchor will not be described further. Note that this flexibility could also be removed from a future version of the origami rotor: creating a 90° bend in the helical direction without introducing significant tilting flexibility is possible but likely requires the careful design of multiple helices before and after the bend [195].
In order to build an anchor structure which is rigid toward these tilting motions, we
designed an origami structure that incorporated a dsDNA extension without breaking its
double-helical structure at the junction point (Figure 2.7A and Table A.2). The dsDNA he-
lix extension is incorporated directly into the origami structure, with one strand continuing
straight into the structure and the other forming a nearby, parallel helix. Due to the base-
stacking interactions, such a structure should strongly favor remaining upright. The remain-
der of the structure consists of a series of short six-helix bundles, forming arms that extend
out in three directions. At the end of each arm several staple strands are extended with bind-
ing sites for secondary strands, labeled with biotin. Two 3D renderings of the structure are
shown in Figure 2.7B. This base can be ligated to a rotor structure, either directly or indi-
rectly via a regular dsDNA molecule with single-stranded overhangs (Section 6.1), to create
the anchor-dsDNA-rotor complex shown in Figure 2.7C. This complex can then be attached
to a streptavidin-coated coverslip using the biotin groups on the anchor. Note this structure
may also be useful in other types of single-molecule experiments, where it is desirable to have
a constrained dsDNA molecule with defined orientation.

This origami was folded as described in Section 6.1 and purified using agarose gel electrophoresis (Figure 2.8A), which predominately showed a single product. TEM imaging showed
the expected anchor structure (Figure 2.8B). After ligating to the origami rotor, we observe
both rotor and rotor-anchor complexes via agarose gel purification, and extract the slower
electrophoresing (top) band (Figure 2.8C).
Figure 2.7 (following page): Origami anchor design. (A) Routing diagram. The origami anchor structure consists of three 20-nm arms, each made of a short six-helix bundle motif. Several staple strands were extended with binding sites for biotin-labeled secondary oligomers for surface attachment (Table A.2). From the center of the structure, three strands (black) were used to make an adaptor to allow ligation to additional DNA. Following the final strand crossover, the adaptor consists of 26 bp of dsDNA followed by a 12-nt ssDNA overhang. (B) 3D renderings of the origami anchor structure. In (A) and (B), the scaffold is in blue and the staples are in the remaining colors. (C) Surface anchored origami structure used for characterizing the Brownian dynamics. The origami rotor, anchor, and a dsDNA linker (as needed), were ligated together. The origami anchor is attached to the microscope surface using multiple biotin tags (not shown).
Figure 2.7: (continued)
Figure 2.8: Origami anchor and anchor-rotor complex characterization. (A) Agarose gel purification of the origami anchor, showing predominantly a single origami band (second brightest band). The bright bottom band is from the excess staple strands. The faint top band is likely dimer or other multimer. The DNA was visualized using the fluorescence from the Sybr Safe DNA binding dye. Each lane contains the same rotor sample. (B) TEM images of the origami anchors. (C) Agarose gel purification of three origami anchor-rotor complexes. The DNA was visualized using the fluorescence of the Cy3 dye on the rotor. The top dark band was cut out and the origami extracted. The bottom band is unligated Cy3-labeled rotor; the unligated anchor origami is not visible.
Brownian dynamics and the stiffness of DNA

The resolution of ORBIT is fundamentally limited by the random, thermally-driven Brownian dynamics of the origami rotor, as introduced in Section 2.2. In this chapter, I present the characterization of these Brownian dynamics for rotor-anchor complexes, including how the Brownian dynamics are measured and how the key parameters that describe the Brownian motion, the spring constant and hydrodynamic drag, are determined. Then I will present results from a series of rotor-anchor complexes which yield these parameters. These results are
further used to estimate the torsional rigidity of DNA. Finally, I will estimate the drag-induced torque and angular lag during rotor motion, showing that these quantities can be neglected in this work.

3.1 Measuring Brownian dynamics

The Brownian dynamics of the origami rotor were measured using the rotor-anchor complex (Figure 3.1A and Section 2.7). The anchor was attached to a microscope coverslip coated with biotin-BSA and streptavidin, and the location of the Cy3 dyes on the end of the labeled origami arm were tracked on a fluorescence microscope at up to 3 kHz frame rates. The location of the origami arm was determined by fitting the fluorescent spot to a Gaussian function, which showed motion along an arc (Figure 3.1B). The center of this arc was determined by fitting the localizations to a circle, and the Cartesian positions were converted to polar coordinates, giving angular trajectories such as the example shown in Figure 3.1C. The distribution of angular positions is approximately Gaussian, consistent with motion in a simple harmonic well (Figure 3.1C). Additionally, as expected (Section 2.2), these motions are not equally distributed over all frequencies but rather fall off at high frequencies due to the hydrodynamic drag (Figure 3.1D).

3.2 Analyzing the Brownian dynamics for real motion

From power spectra such as the example shown in Figure 3.1D we wish to determine the spring constant and drag of the anchor-rotor complex. These can be extracted by fitting to a Lorentzian function, as described in Section 2.2. However, Equation 2.2 neglects several features of the real data which need to be considered for accurate determination of these parameters. First, each camera frame integrates over a finite time, $f_s^{-1}$, where $f_s$ is the sampling
Figure 3.1: Measuring Brownian motion using the rotor-anchor complex. (A) Rotor-anchor complex used to characterize Brownian dynamics. $L$ indicates the length of the single helix of dsDNA between the two origami structures. (B) Example of localization positions of the origami arm during random Brownian motion. Scale bar: 100 nm. (C) Example angular trajectory for a rotor-anchor complex with $L = 52$ bp at 3 kHz (black) and down-sampled to 50 Hz (red) from a rotor-anchor complex, showing Brownian fluctuations away from the equilibrium position (set to 0°). Right box: Histogram of the raw angular positions. The Gaussian distribution of angles is consistent with the expectation for random motion in a simple harmonic well. (D) Power spectrum of the angular trajectory in (C), showing the expected Lorentzian shape for Brownian motion in a simple harmonic well.
frequency. Because the rotor moves during this time, the observed signal is affected by motion blur \([80, 81]\). Second, due to the finite sampling frequency, the experimental power spectrum only has defined values over a finite bandwidth, from 0 Hz to \(f_s/2\). Noise at frequencies below this cutoff is folded back into the spectrum (aliased) to \(f_{in} = |f - nf_s|\), where \(n\) is an integer that brings the frequency \(f_{in}\) into the range of the power spectrum \([81]\). Third, we have to consider the contribution of localization uncertainty. Localization errors are uncorrelated in time, and thus give a white noise (frequency-independent) contribution to the power spectrum \([80, 196]\). Note also that, unlike the spring constant and drag, we expect the localization uncertainty to be dependent on the fluorescence intensity for each complex, which in turn depends on the local light intensity in the TIRF field as well as the number of dyes incorporated into the structure. There should also be Brownian dynamics in the lateral directions (contributing to both radial and angular uncertainty), but such motions are not amplified by the lever arm and thus are likely small. We validate the treatment of localization uncertainty and/or lateral motions with a frequency-independent term in the following section. Taking these experimental contributions into account gives an improved expression for an experimental power spectrum \([80, 81, 196]\),

\[
P(f) = \frac{2k_B T \gamma}{\kappa^3} \left( \kappa + \frac{2\gamma f_s \sin^2 \left( \frac{\pi f}{f_s} \right) \sinh \left( \frac{\kappa}{f_s} \right)}{\cos \left( \frac{2\pi f}{f_s} \right) - \cosh \left( \frac{\kappa}{f_s} \right)} \right) + \epsilon \quad (3.1)
\]

where \(f_s\) is the camera frame rate, and \(\epsilon\) is the contribution from the localization uncertainty, and the other variables are as defined in Section 2.2. The radial position variance was converted to an angular equivalent \(\sigma_L^2\) using the radius of the arc of localizations. To determine \(\epsilon\) from the radial variance, the variance \(\sigma_L^2\) was divided by the size of the frequency range of the power spectrum \((f_s/2)\), giving
Additionally, an unweighted fit was performed, as a weighted fit, while improving the precision, biases the result due to correlations between the experimental value of the power spectrum at each frequency and the weight [197].

After determining the spring constant and drag for a rotor-anchor complex, as well as the localization precision for a particular complex, we can estimate the uncertainty (standard deviation) in the measurement of the angle as a function of the integration time $t$ by adding an addition term to Equation 2.4 to account for the localization uncertainty [80]:

$$
\sigma(t) = \sqrt{\frac{2k_B T}{\kappa} \left( \frac{\tau}{t} - \frac{\tau^2}{t^2} \left( 1 - \exp \left( -\frac{t}{\tau} \right) \right) + \frac{\sigma_L^2}{f_s t} \right)}.
$$

(3.3)

Note that unlike the Brownian noise, the localization uncertainty $\sigma_L^2$ per camera frame is uncorrelated from one frame to the next and the variance due to localization uncertainty falls as the inverse of the square root of the integration time $t$. Also, if the integration time is one frame, then $f_s t = 1$, and the variance from the Brownian dynamics and localization uncertainty are simply added. Finally, note this expression for contribution of the localization precision is not meaningful below the exposure time, $f_s^{-1}$, because the localization precision is not well-defined.

### 3.3 Brownian dynamics of the origami rotors

In order to characterize the Brownian dynamics of the origami rotor, we measured the position of the origami arm in the rotor-anchor complex for three complexes with different lengths of dsDNA, $L$, between the rotor and anchor. We determined the angular position
of the dye-labeled arm and used this measurement to determine the parameters determining the Brownian motion, the spring constant and drag, as described in Section 3.2.

As discussed in the previous section, localization uncertainty and potentially lateral Brownian motion can contribute to the angular uncertainty. We can evaluate these contribution using the radial dimension of the position signal. If the localization uncertainty is the dominant factor in the radial noise, the standard deviation in the radial position should be strongly affected by the fluorescent intensity of the origami, as is observed in Figure 3.2A. Additionally, the power spectrum of the radial position are frequency independent (Figure 3.2B), consistent with the uncorrelated nature of localization errors in subsequent frames. Although they would be expected to be frequency-dependent, lateral Brownian motions might also contribute to this signal, either as a minor component of the noise which is buried in the localization error, or if the corner frequency of the motion is too high to observe in 3.2B given our finite sampling frequency. This contribution can either be neglected or treated as

![Figure 3.2](image.png)

**Figure 3.2:** Characterization of the radial noise for the rotor-anchor complex. (A) Dependence of the radial noise on the fluorescence intensity. The standard deviation in the radial positions is strongly affected by the fluorescence intensity (in arbitrary units), consistent with the radial uncertainty being dominated by the localization precision. (B) Example power spectra of the radial noise for the three dsDNA lengths (red: 52 bp, green: 92 bp, blue: 163 bp). The power spectra are independent of frequency, consistent with the expected uncorrelated nature of noise in the localizations.
Figure 3.3: Power spectrum showing the Brownian noise in the angular position of the rotor attached to the anchor by 52 bp of dsDNA imaged at 3 kHz. Red line shows the modified Lorentzian fit, as described in Section 3.2 (Equation 3.1). This fit yields the torsional stiffness, $\kappa$, and hydrodynamic drag, $\gamma$.

After validating the use of a single, frequency-independent term to describe the contributions to the noise other than the angular Brownian noise, we fit the power spectra (such as the example shown in Figure 3.3) derived from many rotor-anchor complexes observed at 1500 Hz to Equation 3.1. We found that the inverse spring constant of the complex depended on the length of dsDNA between the rotor and anchor (Figure 3.4A), showing that the dynamics are primarily a result of the inherent flexibility of the dsDNA, not the junctions between the dsDNA and the origami or the origami and the surface. The hydrodynamic drag (Figure 3.4B) was similar to the value predicted in Section 2.6, $\sim 3 \text{ fN nm s}$. We also found that the observed hydrodynamic drag of the rotor structure depended on the length of
Figure 3.4: Characterization of the Brownian noise of the origami rotor. 

(A) Dependence of the inverse of the torsional stiffness $\kappa$ on the length of DNA between the rotor and origami. Using longer DNA in the complex results in higher flexibility. 

(B) Dependence of the drag on the DNA length. As discussed in the main text, the drag is expected to depend on the height above the coverslip surface due to hydrodynamic effects. 

(C) Dependence of the relaxation time, $\tau = \gamma/\kappa = 1/(2\pi f_c)$, on the length of dsDNA between the rotor and anchor origami calculated using the drag and the torsional stiffness. $f_c$ is the corner frequency of the power spectrum. 

(D) Dependence of the hydrodynamic drag $\gamma$ of the origami rotor, derived from the power spectrum fit, on the viscosity of the buffer. The drag of the origami rotor connected by a 92-bp dsDNA to the anchor was determined in 0%, 10% and 25% glycerol. Data points are mean ± s.e.m, generated from 133-210 single molecule trajectories each. Except for (D), measurements of the anchored origami parameters are for buffer containing 0% glycerol. Data were collected at 1500 Hz.
the DNA. This is due to the hydrodynamic effects of motion near a surface, which result in a higher, height-dependent, effective drag, as has been described previously [193, 198, 199]. We determined the relaxation time $\tau = \gamma/\kappa$ for each linker length (Figure 3.4C). This time constant determines how quickly correlations between angular fluctuations decay. Finally, to provide additional validation, we altered the viscosity of the solution by varying the glycerol concentration of the buffer (0% - 25%) [200]. The drag shows the expected linear dependence on the viscosity, with a nearly zero intercept (Figure 3.4D; note that at zero viscosity, there would be no drag; see Equation 2.6).

Next, using the measured hydrodynamic drag and torsional stiffness and Equation 3.3, we estimated the angular uncertainty as a function of integration time for a DNA rotor connected to a 52-bp dsDNA (Figure 3.5). We made this estimate with (red curve) and without (blue curve) considering the contribution of the localization uncertainty ($\epsilon$ in Equation 3.3). This estimate agreed quantitatively with the experimentally measured angular uncertainty, determined by downsampling the experimental data to the indicated integration time. These results indicate that we can detect an angular change due to a single base-pair (bp) rotation (34.6°) with a signal-to-noise ratio of 3 using an integration time of only 20 ms without the need to apply an external force or torque. The required integration time increased with the length of the dsDNA linker (for example, to ~30 ms for 92 bp dsDNA). For comparison, other torque-free DNA rotation tracking methods would require an integration time of 80 ms to over an hour to achieve the same angular precision [28, 33] and additionally require the application of a stretching force to the DNA. Furthermore, the highest resolution rotational measurement actually applied to study protein-induced DNA rotation, using a 140 nm bead, would require ~300 ms for 1 bp resolution [28]. ORBIT also compares favorably to linear optical trap measurements, whose resolution depends on DNA length, applied stretching force, bead size, and setup stability. From a recently published high resolution optical trapping
power spectrum measurement, using 5 pN force, a 3.5 kb dsDNA molecule, and 1 μm beads, we can estimate that ~200 ms would be required for 1 bp resolution [201], and under slightly different conditions, ~1 s required integration time has also been estimated [77].

**Figure 3.5:** The standard deviation of the angular positions of the rotor as a function of integration time. Black line shows the standard deviation in the position of a single rotor connected to the anchor with a 52-bp dsDNA tracked at 3 kHz after down-sampling the measured angle to the indicated integration time. Red and blue curves show predicted precision with and without taking into account the contribution of localization uncertainty (Equations 3.3 and 2.4, respectively). (Parameters: $\kappa = 7.8$ pN nm rad$^{-1}$; $\gamma = 5.0$ fN nm s; localization uncertainty per frame $\sigma_L^2 = 0.038$ rad$^2$) The torsional stiffness $\kappa$ and drag $\gamma$ were derived from the measurements of multiple rotors with 52-bp dsDNA connecting the rotor to the surface anchor as described above. The per-frame localization uncertainty $\sigma_L^2$ was estimated for the trajectory used to generate the black curve using the radial uncertainty in position, converted to an angular value using the radius of the circular trajectory. The upper and lower dashed lines correspond to the single base-pair rotation angle (34.6°) and 1/3 of the single base-pair rotation angle, respectively, and the crossing points of these lines with the standard deviation vs. integration time curve give the integration times required for detection of single base-pair rotation with a signal-to-noise ratio of 1 and 3, respectively.
3.4 Determining the torsional rigidity of DNA

The response of DNA to physical deformations such as the twisting induced by many genome-processing enzymes has important implications for enzyme activity and DNA organization, as described in Section 1.1. Random thermal fluctuations of the structure of DNA also influence its interaction with binding partners [202]. In both cases, the rigidity of DNA is a critical parameter for understanding protein-DNA interactions. When referring to this property of DNA, I will refer to the stiffness of a particular molecular construct as the spring constant or stiffness and the intrinsic (per length) quantity as the rigidity. Note that in the literature these terms are not used in a consistent fashion, and both quantities are often referred to without distinction as the torsional stiffness, so care should be taken in comparing references.

The twisting and bending of DNA have been the subject of numerous experimental ensemble and single-molecule experiments (for example, [33, 44, 47, 74, 203]; for reviews, see [24, 204, 205]). The observed mechanical properties can be reasonably well described using the twistable worm-like chain model (TWLC), which includes energy penalties for bending and twisting [206]. Although this model does not include any coupling between the twisting and bending motions, through the contribution of writhe, the bending modes decrease the effective torsional rigidity. This reduction depends on the applied stretching force, as this force (if applied) reduces the contribution of DNA bending to the observed torsional rigidity. The experimental results can be more accurately modeled using a model from Marko and Siggio (the MS model), which includes the contribution of a twist-bend coupling term to the energy [86, 207]. Physically, this coupling term is a consequence of the asymmetry of DNA (with its major and minor grooves). Like the TWLC, the MS model predicts that the effective torsional rigidity is reduced with applied force, and it predicts the rigidity-force relation more accurately than the TWLC [86].
The torsional spring constant $\kappa$ derived from the power spectrum fit (Equation 3.1) of the anchored origami complex can be used to estimate the effective torsional rigidity constant of dsDNA. The value of $\kappa$ contains contributions from the dsDNA linker between the anchor and the rotor as well as contributions from the flexibility of the attachment points to the origami and between the anchor and surface. These components contribute to the flexibility as springs in series, giving

$$\frac{1}{\kappa} = \frac{1}{\kappa_{DNA}} + \frac{1}{\kappa_{other}} = \frac{L}{C} + \frac{1}{\kappa_{other}},$$

(3.4)

where $\kappa_{DNA}$ is the contribution of the dsDNA linker, $\kappa_{other}$ is the other contributions from the attachment points, $L$ is the length of the dsDNA linker, and $C = \kappa_{DNA} * L$ is the effective torsional rigidity constant of DNA with an applied stretching force.

The additional term, $\kappa_{other}$, present due to flexibility in the system from sources other than the dsDNA of interest, prevents the determination of $C$ from a single rotor-anchor complex. However, when measuring $\kappa$ of the anchored origami complexes with several dsDNA linker lengths but otherwise identical design (i.e. identical $\kappa_{other}$), the $\kappa_{other}$ term acts as an offset and does not prevent determination of $C$. Using this approach, we fit the $\kappa^{-1}$ values for the three complexes with different lengths of dsDNA between the rotor and anchor (Figure 3.6), and found $C = 200 \pm 10$ pN nm$^2$ rad$^{-1}$, which is within the range of previously measured results under zero force [86]. Note that due to the contribution of the bending modes at zero force, this value is an effective torsional rigidity, which is smaller than the stiffness of DNA under an applied force as described above [86]. We additionally found $\kappa_{other} = 30 \pm 8$ pN nm rad$^{-1}$, which is equivalent to about 20 bp of dsDNA. Thus, the Brownian noise is primarily the consequence of the inherent flexibility of the dsDNA in the complex, not the attachment points.
Figure 3.6: Determination of the torsional rigidity constant of DNA. The dependence of the inverse torsional stiffness on the dsDNA length (data from Figure 3.4A) was fit to a line. The slope of this linear fit yields the torsional rigidity per unit length of the dsDNA in the absence of an applied force (Equation 3.4), \( C = 200 \pm 10 \text{ pN nm}^2 \text{ rad}^{-1} \), which is within the range of previous measurements under zero force [86]. The inverse of the y-offset from this fit, \( \kappa_{\text{other}} = 30 \pm 8 \text{ pN nm rad}^{-1} \), is the torsional stiffness of the remainder of the structure, which is the equivalent of ~20 bp of dsDNA.

3.5 Torque and Angular Lag Due to the Drag

In addition to affecting the Brownian noise, the hydrodynamic drag of the origami also causes a torque and angular lag in position. In this section, I estimate the effect sizes for reasonable enzyme-induced rotation rates and show that these effects can be neglected for ORBIT measurements.

The motion of the DNA origami rotor through solvent causes a resisting torque due to the origami’s hydrodynamic drag. This torque due to the drag is given by \( \tau_d = -\gamma \omega \), where \( \omega \) the angular velocity. At a rotation rate of 215 bp s\(^{-1}\) (130 rad s\(^{-1}\); this is the fastest average enzyme rate in this thesis - see the 300 \( \mu \text{M} \) ATP rate for RecBCD in Section 4.2). Using the rotor’s drag of 4.9 fN nm s in 10% glycerol (Section 3.3), we estimate the typical torque under these conditions as -0.6 pN nm rad\(^{-1}\), which is much smaller than the contribution from the thermal energy, ~4 pN nm, and thus inconsequential in rotation measurements.
The torque is smaller during RNAP experiments as the enzyme is slower. Further, note that RecBCD is a very fast enzyme, and thus this torque will generally be small for most studies.

The drag will also cause the origami rotor to lag behind the true angular change of DNA at the enzyme. Since this lag is resisted by the stiffness of the DNA, we can estimate the angular lag as

\[ \theta_{\text{lag}} = \frac{\tau_d}{\kappa} = -\frac{\gamma \omega}{\kappa}. \quad (3.5) \]

At the beginning of RecBCD translocation when the dsDNA linker is 80 bp long, we use the relationship between the torsional stiffness and dsDNA length in 3.6 to find \( \kappa \approx 6 \text{ pN nm rad}^{-1} \). At this stiffness, we estimate \( \omega_{\text{lag}} \) to be -0.10 rad (or -6°) at 10% glycerol and 130 rad s\(^{-1}\). Thus, the angular lag is small compared to our measured angular noise (~35°) due to Brownian motion and can be neglected. Note that both the torque and lag would be larger if the drag were larger (i.e. if the rotor were larger). In fact, previous rotor bead tracking experiments have corrected for the drag-induced angular lag, and even noted the torque sometimes reached levels sufficient to unwind the DNA [3].
Initiation and processive DNA unwinding dynamics of RecBCD

In this chapter, I describe how we applied ORBIT to study the activity of the DNA repair enzyme RecBCD, a processive helicase that detects double-stranded breaks and initiates homologous recombination on DNA [208]. First, I will introduce RecBCD, with an emphasis its initiation and DNA unwinding activities. Then I will describe the ORBIT assay for RecBCD
and the data that is generated. Next, I describe how we can validate the results from our ORBIT assay using an ensemble approach. After these sections, I will detail two investigations of the biology of RecBCD: the pausing and backtracking behaviors of RecBCD during processive unwinding and the initiation mechanism of RecBCD on different types of DNA substrates.

4.1 Introduction to RecBCD

Double-stranded breaks in DNA occur due to chemical and radiation damage and frequently during replication due to nicked or damaged DNA [208–210]. This type of damage is extremely harmful, and must be repaired for continued cell viability and to complete replication. Multiple pathways exist for such repair, including non-homologous end joining, which generally leads to DNA sequence alterations, and homologous recombination, in which the DNA is repaired using a second copy of the damaged DNA [208, 211]. The homologous recombination pathway in E. coli begins with the unwinding and partial degradation of the damaged DNA by the heterotrimer RecBCD.

While unwinding and degrading E. coli DNA, every ~5 kb RecBCD will reach a properly oriented regulatory sequence, the crossover hotspot instigator (‘Chi’), which can prompt the enzyme to prepare the DNA with a long 3’ single-stranded DNA by nicking the 3’ strand at the Chi site and subsequently cutting only the 5’ strand [208, 211–216]. This switch in state involves a several second pause in activity, followed by resumed unwinding at a slower rate [217]. The Chi regulatory sequences are over-represented in the genome and tend to be aligned toward replication forks, highlighting the important role of recognition of this sequence by RecBCD for rescuing problems during replication [218]. After Chi recognition, RecBCD helps to coat the single-stranded DNA with RecA [219]. These RecA molecules
then facilitate the invasion of the single-stranded DNA into the undamaged DNA copy for homologous recombination [220]. While RecBCD’s role can be partially compensated by the activity of the RecF pathway (which typically fills single-stranded DNA gaps), due to the frequency and severity of double-stranded breaks, deletion of RecBCD leads to reduced viability [208, 220]. Beyond its role in DNA repair, RecBCD also plays an important role in viral defense, degrading linear viral DNA (or viral DNA linearized by restriction enzymes) that has infected the cell [208].

RecBCD is a complex molecular machine with multiple enzymatic activities. It contains two separate SF1 family single-stranded DNA translocation domains: a 3'-5' motor in RecB [221, 222] and a 5'-3' motor in RecD [223–225]. RecB and RecD each contain helicase activity in isolation, but are much faster and more processive in the context of the RecBCD complex, where both motors actively translocate on separated DNA strands [221–225]. Additionally, mutations that prevent DNA hydrolysis in either motor reduce the complex’s unwinding rate and processivity, suggesting the two motors may operate cooperatively [226–228]. Any cooperativity likely cannot involve concerted motion between the two motors as the motors operate at different rates, forming a single-stranded DNA loop that has been observed between the upstream dsDNA and the lagging motor [225, 229]. RecB also contains an extension, described as an ‘arm’, which binds to the dsDNA as it approaches the remainder of the enzyme [230]. At the opposite end of the complex, the nuclease domain of RecB provides the regulated cutting activity essential for degrading viral DNA and preparing long 3’ single-stranded substrates for recombination [230, 231]. RecBCD is generally thought to cut the DNA frequently during unwinding [208, 211, 232, 233], although this cutting frequency is condition dependent and the physiological nuclease behavior remains controversial [216].

The RecC domain forms part of the channels for each DNA strand and contains the recognition domain for the Chi sequence [230]. RecC also contains a ‘pin’ region, which sits
at the junction between the upstream dsDNA and the separated single-stranded regions [230]. When in complex with RecB (forming RecBC), RecC significantly increases the rate and processivity of the RecB motor [222], likely due to the contributions of the pin and the additional contacts with DNA [230]. Adding RecD to the RecBC complex further increases the rate and processivity. Perhaps because of the dual-motor nature of this enzyme, RecBCD has an exceptionally high processivity (~30 kb) [234, 235]. Note that the smaller RecD subunit was not discovered until 1986 [236]. Before this discovery, the protein is referred to in the literature as RecBC (which is also a stable, active heterodimer) or exonuclease V; RecD is likely also present in the complex in these earlier papers. Finally, although not mechanistically understood, the RecBC and RecBCD complexes also appear to load RecA onto this 3’ strand after Chi recognition [219].

The initiation of RecBCD activity requires a nearly blunt end DNA; up to ~25 nt overhanges on either strand are allowed during initiation [237]. This ability to initiate on a relatively wide range of substrates (blunt-end, 3’ overhang, 5’ overhang) is important for RecBCD’s function, as the wide range of sources of DNA break can lead to these different substrates [208–210]. Importantly, longer overhangs, which RecBCD cannot directly process, can be preprocessed by other enzymes such as RecJ [238–240]; however, note that even in these cases, substrates with a range of short overhangs (on both the 3’ and 5’ strands) can result [241].

The mechanism of RecBCD binding and initiation has long been of interest. RecBCD binds very tightly to blunt ends, with nanomolar affinity [242, 243]. Early DNA protection experiments suggest that, in the presence of Mg$^{2+}$ and absence of ATP, RecBCD opens the terminal 5-6 bp of DNA [244]. This opening, which brings the single-stranded DNA toward the motor domains, can also be observed in the crystal structures [230, 245]. In these structures, the RecB and RecD motor domains are about 5 and 10 nt from the pin, respectively, suggesting that this initial opening may serve to bring the DNA to the RecB motor. In the
absence of magnesium (preventing DNA unwinding), RecBCD binds more strongly to substrates with 6 nt 3’ strand overhangs and 10-20 nt 5’ strand overhang than blunt end DNA; the favorable free energy change associated with the additional contacts between the overhangs and the DNA were associated with RecBCD’s ability to unwind the first few bases of DNA without ATP [246]. Additionally, the RecBC complex shows increased binding affinity when magnesium is added, but only if there is not a 3’ strand overhang of 6 nt or more, suggesting that the increased affinity is from DNA unwinding, and thus that RecBC also can unwind about 6 nt [247].

Kinetic modeling of ensemble quenched-flow and stopped-flow measurements has suggested that several kinetic steps are required before processive DNA unwinding [248, 249], and that these steps are single-stranded overhang dependent [250]. In particular, dual-overhang substrates (with both 3’ and 5’ dT single-stranded overhangs long enough to reach RecB and RecD, respectively) were found to skip some initiation steps compared to blunt end substrates or substrates with the same 3’ overhang but a shorter 5’ overhang [250]. This was interpreted as indicating a critical role for RecD during initiation. However, this kinetic modeling requires the assumption of a homogeneous enzyme population; the highly heterogenous RecBCD population [251] may not be well described by these models. In support of this model, optical trapping experiments showed that adding a 5’ overhang to the DNA substrate increases the observed noise, which was interpreted as increased enzyme position fluctuations [252]. Because this noise profile was more similar to that observed during a translocation stall and during translocation than the noise with a blunt-end DNA substrate after binding, it was suggested that filling the RecD motor with the 5’ strand results in a translocation-like state, providing additional indirect evidence of a role for the 5’ strand and RecD motor during initiation. On the other hand, a recent cryo-EM structure of RecBCD bound to DNA found
evidence of ATP-analog binding to RecB but not RecD, suggesting that RecB might start processive unwinding [253].

After initiation, RecBCD transitions into a processive unwinding phase. During this phase, RecBCD can unwind DNA for tens of kilobases at hundreds to over 1000 bp/s [208]. Single molecule experiments have revealed that RecBCD molecules display a wide range of unwinding rates [217, 235, 251, 254, 255]. The heterogeneity has been attributed to long-lived conformational differences [251]. Despite significant interest, the step size of RecBCD during this unwinding phase remains unknown. Bulk studies have suggested a step size of ~4 bp [248–250], but bulk analysis of step sizes is confounded by the inherent heterogeneity of RecBCD, likely resulting in an inflated apparent step size [256]. Separately, a single-molecule study of RecBCD at very low ATP concentrations (2-4 μM) and under a 6 pN opposing force found rapid ~1-1.5 nm (~4 bp) fluctuations in position. This higher than expected noise profile was interpreted as an inherent part of the translocation cycle and evidence against a simple single base pair stepping model [252]. On the other hand, the RecB and RecD motors are SF1A and SF1B family enzymes, which generally are thought to move using a 1 bp ‘inchworm’ mechanism [257]. They also hydrolyze 1 – 1.5 ATP per bp per motor domain, suggestive of a single bp step mechanism [258, 259].

RecBCD is a remarkably powerful enzyme, and is able to push or displace other tightly bound proteins including RNA polymerase and nucleosomes [260–262]. However, it is strongly affected by moderate applied forces (~6 pN), which can induce large backtracking events [254]. Most studies of RecBCD have lacked the resolution required to characterize features of the unwinding phase of activity, except for the relatively long (seconds) pause at the Chi sequence [217]. Optical trap experiments tracking the motion of RecBCD on DNA have shown that, under an applied force, RecBCD can pause and backtrack during processive motion [254]. Follow-up experiments revealed higher a larger than expected noise profile, which was
interpreted as ~4 bp fluctuations in the RecBCD position on the DNA [252], perhaps due to frequent, rapid backtracking. Whether these activities have functional roles remains unclear. Surprisingly, while RecBCD generally requires dsDNA, it is able to bypass short (~20 nt) single-stranded gaps during its unwinding activity [263, 264]. Several recent papers from the Lohman lab have additionally focused on a putative ‘secondary translocase’ ability driven by the RecB motor and carried out by the RecB arm, perhaps facilitating movement along both double-stranded and single-stranded DNA [265–268].

In addition to its importance for DNA repair, RecBCD is a common model system for developing new ensemble and single-molecule techniques. Some of the earliest ensemble helicase assays were developed to track RecBCD’s helicase activity. These measure DNA unwinding using the quenching of the intrinsic fluorescence of the single-stranded DNA binding protein when it binds to the unwound DNA product [269] or using the displacement of fluorescent intercalating dyes from the dsDNA substrate [270]. The latter approach, combined with optical trapping, formed the basis of the earliest real-time single molecule measurements of RecBCD activity [235], which was used repeatedly to study the mechanism of RecBCD [217, 251, 271] and ultimately was multiplexed in the ‘DNA curtains’ technique [261]. RecBCD was also one of the first enzymes studied using the tethered particle assay [272], and was used in the first application of a variant of the tethered particle assay which uses a small force to improve resolution [273]. Separately, a stopped-flow assay for measuring helicase activity was developed for RecBCD [249]. In this approach, an initially quenched Cy3 labeled oligomer is displaced after a certain number of base pairs are unwound, resulting in a fluorescence signal of unwinding.
4.2 ORBIT assay for RecBCD

As RecBCD unwinds dsDNA, its two motors, RecB and RecD, should each track along one of the DNA strands [222, 224, 225, 230], which is expected to generate a rotation of the DNA with respect to the enzyme of ~34.6° per unwound base pair. Due to the fast DNA unwinding rate of the enzyme [235, 251, 254, 261, 269, 272], RecBCD-induced DNA rotation has not been observed previously and such detection would require methods with very high spatiotemporal resolution. To measure RecBCD-induced DNA rotation in real time, we attached RecBCD onto the coverslip surface of a microscope flow chamber using passive adsorption, and then used a flow system to introduce ATP along with dsDNA substrates of RecBCD, each 80 bp long and attached to an origami rotor (Figure 4.1A). Critically, with ORBIT, we observe the rotor-dsDNA substrate spontaneously bind to RecBCD on the surface during the observation time. This is only possible here because we do not have to pre-assemble the complex before tracking rotation, as would be required for this type of experiment if a stretching force was applied (see Section 1.3.4).

When RecBCD binds to the double-stranded break at the end of the dsDNA substrate and begins to unwind the DNA duplex, the resulting DNA rotation should be amplified by the origami rotor and result in processive motion of the fluorescent dyes along a circular path. Using a total internal reflection fluorescence (TIRF) microscope, we tracked the dye movements at sampling rates up to 1 kHz. The fluorescent dyes on the origami rotor arm displayed unidirectional movements along a circle with a diameter similar to the diameter of the rotor (Figure 4.1B). The circle diameter is slightly smaller than the origami arm length because (1) the dyes are not all at the very end of the arm, (2) Brownian motion along the arc during a single frame contributes to motion blur which brings the localizations slightly closer
Figure 4.1: Measuring single-molecule DNA rotation by RecBCD using ORBIT. (A) Schematic depiction of the ORBIT method with RecBCD. Rotation of a dsDNA segment attached to RecBCD is amplified by a DNA origami rotor and detected by tracking the position of fluorescent dyes attached to the tip of a rotor blade. To measure enzyme-induced DNA rotation, the RecBCD molecules are attached to the surface of a microscope slide. DNA substrates with attached origami rotors are then added. Substrate binding and subsequent DNA rotation is captured using a total internal reflection fluorescence (TIRF) microscope. (B) Localization trajectory of the fluorescent dyes from a single origami rotor connected to a dsDNA substrate being unwound by the RecBCD helicase. The rotation angle $\theta$ was measured from the position of dyes along a circular path. Time is indicated by the color bar. Scale bar: 100 nm. (C) Radial localization uncertainty during processive unwinding by RecBCD. The uncertainty in the radial position of the origami arm is limited by the average fluorescence intensity during unwinding. For the highest intensity fluorescent spots, the standard deviation of the radial position is ~10 nm. We apply a localization uncertainty threshold of 16 nm (0.1 pixel), shown in red, to select only trajectories with relatively high localization precision for further analysis.
to the center and (3) the origami may sample tilted states (tilting of the rotor relative to the dsDNA extension) which bring the dyes slightly closer to the center position.

The wide-field imaging nature of ORBIT allowed us to collect, in parallel, many single-molecule trajectories per experiment. The localizations are fit to a circle and converted from Cartesian to polar coordinates using the center position of the fit circle. Then, the angular measurement can be converted into the position of RecBCD along the dsDNA substrate by using the average DNA twist of 34.6° per bp. We use the radial standard deviation to estimate the localization precision (see Section 3.2), and find we can track the position of the arm with 10-15 nm precision (Figure 4.1C). To ensure reliable tracking, we restrict further analysis to trajectories with better than 16 nm (0.1 pixel) localization precision. A small fraction of trajectories (~10%) showed signs of interactions between the rotor arms and the coverslip surface (angular noise dropping below the expected Brownian noise) and were excluded as described in Section 6.4.

We measured the unwinding rate under a wide range of ATP concentrations (15 – 300 μM) at room temperature and with a solution pH of 8. Examples of trajectories acquired at different ATP concentrations are shown in Figure 4.2. The average rates of individual RecBCD molecules exhibited a broad distribution even at the same ATP concentration (Figure 4.3). This heterogeneity is consistent with previous single-molecule studies of RecBCD [217, 235, 251, 254], and is thought to be due to long-lived conformational differences between enzymes [251]. Despite the heterogeneity, the ensemble-averaged unwinding rate derived from many molecules showed a clear ATP dependence (Figure 4.4), which could be fit to a Michaelis-Menten relation, yielding $v_{max} = 304 \pm 13$ bp/s and $K_M = 124 \pm 12$ μM at solution pH 8, which did not vary with the solution pH (at solution pH 6, $v_{max} = 290 \pm 10$ bp/s and $K_M = 130 \pm 10$ μM). These values for $v_{max}$ and $K_M$ are consistent with previously measured values from both bulk and single-molecule experiments [235, 254, 269, 272].
4.3 Stopped-flow assay for RecBCD

We next carried out an ensemble measurement for RecBCD helicase activity. Because the previously measured unwinding rates for RecBCD span over a range, we used these experiments to facilitate better comparison of our results of surface-bound enzymes from ORBIT measurements with ensemble measurement results of freely diffusing enzymes in solution. Even more importantly, in the following sections, this assay will be used to validate new findings about RecBCD.

We adapted a stopped-flow fluorescence assay previously designed for measuring the unwinding of DNA by RecBCD [249, 250]. In the previously described assay, a DNA hairpin substrate is first formed from three DNA strands (Figure 4.5A). Two of these strands are dye labeled with Cy3 and Cy5 such that they are in close proximity, quenching the Cy3 via FRET with the Cy5 dye. The RecBCD is prebound to the enzyme, and the RecBCD-DNA
complex is then quickly mixed with ATP and heparin. Heparin, a competitive inhibitor of DNA binding, is used to achieve single-turnover conditions. The Cy3 strand is displaced by the RecBCD helicase activity, resulting in unquenching of the dye and an increase in green fluorescence. For a closer comparison to our ORBIT assays, which do not involve pre-binding of the RecBCD to the DNA substrate, we instead used a two-stage mixing procedure (Figure 4.5B). The RecBCD and DNA were first mixed for a short time (200 ms) and then combined with the ATP and heparin. This shorter pre-binding time reduces the likelihood of missing kinetic steps before the ATP is added. The resulting fluorescence measurements are shown in Figure 4.5C.

Figure 4.3: Histograms of the average unwinding rate of individual molecules at various ATP concentrations (solution pH 8.0): (A) 15 µM ATP, (B) 25 µM ATP, (C) 50 µM ATP, (D) 75 µM ATP, (E) 150 µM ATP, and (F) 300 µM ATP.
Figure 4.4: Average unwinding rate as a function of ATP concentrations at room temperature. (A) Unwinding rate at solution pH 8. The average rates were fit to Michaelis-Menten kinetics ($v_{max} = 304 \pm 13$ bp s$^{-1}$; $K_M = 124 \pm 12$ μM). Data are generated from ~30-90 single molecule trajectories from at least three experiments at each ATP concentration. (B) Unwinding rate at solution pH 6 ($v_{max} = 290 \pm 10$ bp s$^{-1}$, $K_M = 130 \pm 10$ μM). Data are generated from ~40 – 90 single molecule trajectories from at least three experiments at each ATP concentration. For both (A) and (B), data points are mean ± standard error of the mean (s.e.m.).

We collected stopped-flow fluorescence measurements using DNA substrates with two different DNA lengths before the Cy3 dye (Table A.4), and used the difference in half-rise times to estimate the unwinding rate of the RecBCD at 50 μM ATP. Because the initial DNA was identical in both cases, this difference should be the time required to translocate by the difference in DNA length between the two substrates (26 bp). This bulk measurement yielded an estimate of unwinding rate of ~100 bp/s at 50 μM ATP (Figure 4.5C), which is comparable to the value (85 bp/s) measured by our single-molecule experiments at the same ATP concentration, suggesting that the surface attachment did not significantly perturb the RecBCD unwinding activity.
Figure 4.5: RecBCD unwinding measured using ensemble stopped-flow assay. (A) Design of the DNA substrate for the stopped-flow experiments [249, 250]. The Cy3 on strand A is initially quenched by the Cy5 on strand B. RecBCD activity causes the dissociation of strand A, resulting in an increase in fluorescence. The hairpin on the left side (strand C) ensures that activity can only begin from the right side. (B) Dual mixing stopped-flow experiment design. The RecBCD and DNA were mixed together for 200 ms in the delay loop before mixing with ATP and heparin, which prevents additional RecBCD-DNA binding after single turnover. The unquenching of the Cy3 was measured after the RecBCD-DNA and ATP were mixed. (C) Stopped-flow fluorescence measurements on blunt-end substrates with strand A either having 26 (black) or 52 bp (green), at 50 μM ATP and pH 8. The ratio of the difference in strand A lengths for the two samples to the difference in measured half-rise times of the two samples allows the unwinding rate to be estimated as ~100 bp/s, assuming that the initiation kinetics are not different for these two substrates because they have the same geometry and sequence at the double stranded break.

4.4 Comparing ORBIT and ensemble RecBCD measurements

In comparing across methods, it is important to take into account local differences in conditions that may be present. Of particular note in comparing ORBIT measurements to typical ensemble biochemistry measurements, such as stopped-flow, is the passive adsorption of the enzyme in the RecBCD assay to the microscope coverslip. This approach is used here
to achieve torsional rigidity: multiple attachment points are required to prevent the enzyme from freely swiveling with respect to the surface. However, this is not the only attachment approach that can achieve torsional rigidity; alternative approaches will be discussed in Section 7.2. Note that in all cases, including ensemble biochemical measurements, the local conditions are quite different from those in vivo [274] and thus it is generally advisable to confirm that key results are not highly condition-dependent.

The effect of surface binding (both passively and by chemical crosslinking) on enzymes has been of interest for a long time, primarily for industrial applications [275–277]. There are several potential effects to consider when designing ORBIT experiments. First, in some cases, surface adsorption can lead to protein unfolding, resulting in inactive enzyme [278]. Note that such an effect will be surface-chemistry specific; for example, with RecBCD we have observed that treatment of the surface with the hydrophobic silanization chemical dichlorodimethylsilane results in loss of RecBCD activity in ORBIT experiments, which could be due to inactivation or a loss of binding. For enzymes that respond in this manner, no productive enzyme activity is expected using this strategy, and the approaches outlined in Section 7.2 should be pursued. Slightly counter-intuitively, surface binding can also help stabilize an enzyme in harsh conditions as the multiple attachment points between the native folded structure and the surface can favor that structure over the unfolded state, and because immobilization can prevent aggregation [276, 277, 279, 280]. This approach has been used for applications where maintaining the protein stability over time and in harsh conditions is important, and may allow ORBIT to be used to study a wider range of conditions than is possible using bulk biochemistry. Second, near the surface (within a few nm) the solvent is more ordered than in bulk solution, leading to a higher local viscosity [281–283]. A similar effect is also seen near protein surfaces, membranes, and other biomolecules [283–285], contributing to higher viscosity in the cellular context. Although enzymes are sometimes slower at higher viscosity
we have already seen in Figure 4.5 that RecBCD’s processive unwinding occurs at nearly the same rate while on the surface or freely diffusing in solution.

Third, the silicate glass coverslips surfaces used here are negatively charged due to the SiO$^-$ groups on the surface. This creates an electrostatic potential at and near the surface, leading to local accumulation of H+ ions and thus a local pH shift [290–292]. This pH shift is related to the potential according to [292]

$$pH_s = pH_b + \frac{F\psi}{2.3RT} \quad (4.1)$$

where $pH_s$ is the effective pH at the surface, $pH_b$ is the bulk solution pH, $F$ is Faraday’s constant, $\psi$ is the electric potential due to the surface charge, and $RT$ is the thermal energy.

Measurements of the potential at the surface at a bulk pH of 8 give $\psi$ value of 170 mV, which corresponds to a pH shift of ~2.9 pH units right at the surface interface [291]. Because of the finite size of the enzyme attached to the surface, and the charge screening effect, the effective pH shift should be smaller than ~2.9 pH units. An experimental measurement of this pH shift for an enzyme acting at a silicate surface previously gave the value of ~2 pH units [290] when the solution pH is around 8-9. Because the concentration of SiO$^-$ groups on the surface depends on the pH [291], this effect size should be reduced at lower pH. As expected, in another case, with enzyme adsorbed onto a silicate surface, a pH shift of around 1 unit was observed around a solution pH of 5 [293]. Note this effect is not specific to surface adsorption and occurs when the enzyme is tethered near a surface or charged structure, such as DNA origami, or within charged polymers [292, 294–296] as well as when the substrate, not enzyme is attached to the surface [290].

Since RecBCD is very near the surface during ORBIT measurements, this effect is expected to reduce the local pH experienced by the enzyme. For this reason, throughout the
RecBCD section, we will carefully consider whether the observed results are pH dependent. Note we have already seen that the unwinding rate (Figure 4.4) is pH independent, and is approximately the same when measured using ORBIT or stopped-flow (Figure 4.5). Additionally, this pH offset will be taken into account when comparing ORBIT and stopped-flow measurements quantitatively in Section 4.8.

4.5 CHARACTERIZATION OF THE UNWINDING PHASE

We observed three distinct features in the processive DNA unwinding phase in our single-molecule trajectories: unwinding, pausing, and backtracking (Figure 4.6). RecBCD was previously observed to pause and backtrack under an opposing force [254]. Our results showed that pausing and backtracking of RecBCD also occurred in the absence of an opposing force, but the pause durations and backtracking distances that we observed were substantially smaller than those observed under opposing forces [254]. Furthermore, the pause frequency decreased with increasing ATP concentration (Figures 4.7A and 4.8A), suggesting that pause entry preferentially occurs when the enzyme is in the apo (non-ATP bound) state, poten-

![Figure 4.6: Example single-molecule unwinding trajectories at 25 μM ATP showing (A) pausing and (B) backtracking recorded at 500 Hz. For display purposes, the initiation phase, discussed in Section 4.6, is not shown here.](image)

1.87
tially because ATP binding causes a conformational change that disfavors pause entry. On the other hand, the average pause duration remained largely constant across all tested ATP concentrations (Figures 4.7B and 4.8B), suggesting that pause exit occurs through an ATP-independent process.

Pauses were followed either by resumed unwinding or by a backtracking event (i.e. rotation of DNA in the opposite direction), presumably reflecting rewinding of the substrate’s two DNA strands. The backtracking distance was exponentially distributed with an average around 6 bp (Figures 4.7C and 4.8C). Previous optical trap experiments have shown rapid RecBCD fluctuations on the DNA during the processive DNA unwinding phase \[252\], which may be related to the backtracking events that we observed here. The backtracking events were typically followed by a pause, which we refer to as the recovery pause, before DNA unwinding was resumed. Interestingly, and in contrast to the above-described pauses that occurred during forward unwinding, the recovery pause duration showed a dependence on the ATP concentration (Figures 4.7D and 4.8D), suggesting that these two types of pauses represent distinct enzymatic states and that backtracking is associated with a conformational change that outlasts the actual backtracking movement.

While it remains unclear if backtracking by RecBCD has a functional role, one possibility is that backtracking is necessary to resolve situations where the enzyme has stalled, for instance due to DNA damage-induced crosslinking of the two DNA strands. In such cases, to avoid covering the damage site (thus blocking repair), the enzyme would need to shift its position or dissociate, both of which could be accomplished through backtracking. A similar scenario has been described for RNA polymerase, which can backtrack when encountering impassable obstacles such as pyrimidine dimers \[297, 298\]

Based on these results, we suggest the following kinetic model of DNA unwinding by RecBCD (Figure 4.9): During DNA unwinding, pausing occurs frequently and some of the
Figure 4.7: Dependence of pausing and backtracking on ATP at pH 8. (A) The dependence of the mean pause frequency per molecule on the ATP concentration. The error bars are s.e.m. (B) The ATP dependence for the median duration of pauses that are not associated with backtracking. (C) The mean backtracking distance at various ATP concentrations. The error bars are s.e.m. (D) The ATP dependence for the median recovery pause duration after a backtracking event. The error bars in (B) and (D) are the standard deviation of the median derived from resampling. For the recovery pause durations, the differences between the 25 μM ATP data point and the 50, 75, and 300 μM ATP data points are statistically significant. (p-values: 0.03, 0.003, and 0.008, respectively, derived from Kolmogorov-Smirnov test of the distributions of the recovery pause durations; n = 31, 15, 25, and 13 backtracking events for the 25, 50, 75, and 300 μM points, respectively. Note that not all trajectories contain a backtracking event. Data are generated from ~30-150 single molecule trajectories from at least three experiments at solution pH 8 at each ATP concentration.)
Figure 4.8: Dependence of pausing and backtracking on ATP at pH 6. (A) The dependence of the pause frequency on the ATP concentration. The pause frequency is determined as the number of pauses per second for each single-molecule trajectory, and the error bars are s.e.m. (B) The ATP dependence for the median duration of pauses during forward unwinding. (C) The mean backtracking distance at various ATP concentrations, and the error bars are s.e.m. (D) The ATP dependence for the median recovery pause duration after a backtracking event. The error bars in (B) and (D) are the standard deviation of the median derived from resampling. For the recovery pause durations, the p-values for the differences between the 25 μM ATP data point and the 50, 75, and 300 μM ATP data points are 0.05, 0.004, and 0.09, respectively, derived from Kolmogorov-Smirnov test of the distributions of the pause durations (n = 20, 22, 16, and 10 backtracking events for the 25, 50, 75, and 300 μM points, respectively. Note that not all trajectories contain a backtracking event.) Data are generated from ~40 – 90 single molecule trajectories.
Figure 4.9: Schematic of a kinetic model of RecBCD-induced DNA unwinding. RecBCD enters the paused state during processive unwinding, with a likelihood that decreases with [ATP]. The paused state can recover to the processive-unwinding state, with a characteristic time independent of ATP, or it can transition to the backtracking state. After backtracking, the complex enters the recovery pause state, which recovers to the processive-unwinding state through an ATP-dependent process.

pauses lead to enzyme backtracking; the enzyme can then exit the backtracking state and resume DNA unwinding through a recovery pause intermediate, which is distinct from the pause state entered by the enzyme during forward unwinding.

4.6 INITIATION ON BLUNT-END SUBSTRATES

The initiation of RecBCD activity at a double-stranded break constitutes the first step of the homologous recombination pathway but the precise sequence of events during initiation remains unclear. During processive activity, RecBCD unwinds dsDNA by pulling the two DNA strands in different directions across a pin using its two motor subunits, RecB and RecD (Figure 4.10) [224, 225, 230]. Because the active sites of these two motor domains are positioned several nanometers from the pin where the two DNA strands separate [230], an additional mechanism may be required to initially unwind the dsDNA ends and bring them into contact with the motors. Biochemical and structural studies have shown that RecBCD can unwind the first several base pairs of DNA in an ATP-independent manner [230, 244–246],
Figure 4.10: Relative positions of the pin and motor domains. (A) Crystal structure of the RecBCD-DNA complex (black: RecBCD; red: DNA), showing partial unwinding of the DNA helix, which positions the ssDNA near the RecB motor domain. Adapted with permission from [230]. (B) Schematic showing the separation between the RecBCD motor active sites and the pin position where the two strands of the duplex DNA are separated.

potentially providing a mechanism. Kinetic modeling of ensemble data has suggested that initiation requires multiple steps, perhaps related to this unwinding [248–250]. However, note this modeling assumes a homogeneous population of enzymes, while RecBCD exhibits highly heterogeneous rates [251]. Direct evidence for the role of this ATP-independent unwinding mechanism in RecBCD initiation is still lacking.

Strikingly, our single-molecule ORBIT trajectories revealed an initiation phase exhibiting repetitive rotational transitions of the DNA between two well-defined states (Figure 4.11). The average magnitude of these transitions was $169^\circ \pm 5^\circ$, which corresponds to $\approx 5$ bp of DNA unwinding. The processive unwinding phase consistently started from the higher rotational angle, unwound state (green state, Figure 4.11), suggesting that this step is an on-pathway intermediate. We also observed these transitions in the absence of ATP. Upon addition of ATP to the RecBCD-DNA complexes undergoing these two-state transitions in the
absence of ATP, the complexes eventually transitioned into the processive unwinding phase, which again started from the higher rotational angle, unwound state (Figure 4.12). The magnitude of the step sizes was the same with and without ATP (Figure 4.13A).

We also compared the transition kinetics between the wound and unwound states as a function of ATP concentration. We determined the time constants for the unwinding (magenta to green) and rewinding (green to magenta) at three ATP concentrations (0, 50, and 300 μM ATP). These rates are constant across the ATP range (Figure 4.13B), including in the absence of ATP. To calculate the rate constants, we assumed single-exponential kinetics. Thus, the unwinding rate was simply the inverse of the average dwell time in the magenta state. The dwell time in the green state is influenced by two rate constants [299] - the rewinding rate, and the transition rate to the processive unwinding phase. We accounted
Figure 4.12: Reversible unwinding transitions in the absence of ATP followed by addition of 50 μM ATP to start the processive unwinding phase. Note that the recording in this experiment started after the DNA substrate was added to the sample and hence the initial binding of the substrate to the enzyme was not observed here. ATP was added using a syringe pump, activated after 6 seconds, as indicated by the dashed line. Arbitrary vertical offsets are applied to different trajectories for display purposes.

for this competition in calculating the time constant for the rewinding rate when ATP was present (see Section 6.6 for details).

Taken together, our results suggest that the ATP-independent, transient unwound state is an obligatory intermediate during RecBCD initiation on blunt-end DNA. We note that in free dsDNA, the final base pair is dynamic and frequently open [202], and thus the wound state could be fully base-paired or contain ~1-2 unpaired bases prior to binding. In either case, the ssDNA length in the unwound state (~5-7 nt) is comparable to the amount of ss-DNA required to engage with the RecB motor (~6 nt) inferred from the crystal structure
Figure 4.13: Quantification of initiation step size and rate. (A) Rotation angle distributions for the reversible unwinding transitions during the initiation phase for blunt-end DNA bound to RecBCD with 50 μM ATP (top; mean = 169° ± 5°) and without ATP (bottom; mean = 172° ± 3°). (B) Transition rates from the wound to unwound state (magenta) and from the unwound to the wound state (green) at three different ATP concentrations. Green to magenta rate constants in the presence of ATP take into account the competition with the alternative pathway (to processive unwinding; see text or Section 6.6). Error bars are standard error. Data are generated from ~30-100 single-molecule trajectories from three experiments at each of the ATP concentrations at solution pH 8.

[245], providing a mechanistic explanation of why this transient state is an obligatory, on-pathway intermediate.

4.7 Initiation on DNA substrates with single-stranded overhangs

As described in Section 4.1, natural substrates for RecBCD include dsDNA with blunt ends as well as 3’ or 5’ ssDNA overhangs of various lengths [208]. How RecBCD is able to initiate on this diverse range of substrates remains unclear. We hypothesized that the transient, ATP-independent unwinding transitions observed on blunt-end dsDNA may not be required for initiation on substrates with ssDNA overhangs that are long enough to reach the cor-
Figure 4.14: RecBCD initiation on double-stranded breaks with different break geometries. (A) Single-molecule trajectories recorded at 500 Hz showing RecBCD initiation and processive unwinding on a dsDNA substrate with a 6-nt 3' overhang and (B) three different dsDNA substrates with 5' overhangs: 10-nt 5' overhang, 15-nt 5' overhang, and a 10-nt 5' overhang without G-C pairs in the first 5 bp ("-GC"). Transient two-state transitions are observed for the 5'-overhang substrate, and the green/magenta color coding, and as well as the ‘Δ’ and ‘ΔF’ notations, are as described in Figure 4.11.

responding motor domain. To test this hypothesis, we first designed a substrate with a 6-nucleotide (nt) 3’ overhang, which should be sufficiently long to engage the RecB motor [245]. We observed RecBCD-mediated DNA unwinding on this substrate but did not observe the two-state transitions during the initiation phase (Figure 4.14A). The overall mean initiation time at 50 μM ATP was similar with and without the 3’ overhang, likely because the wound (magenta) state makes only a small contribution to the total initiation time (Figure 4.15A).
Figure 4.15: Characterization of initiation phase for variable overhang substrates. (A) Mean cumulative dwell times of the green and magenta states prior to processive unwinding on the blunt-end (blunt), 3’-overhang (3’ oh) and the 5’-overhang (5’ oh, 10 nt, 15 nt, and 10 nt -GC) DNA substrates. The green and magenta portions indicate the cumulative dwell times in the green and magenta states, respectively. Error bars are s.e.m. P-values (two-sided t-test): 0.36 (blunt-end vs. 3’ oh), 0.002 (blunt vs. 10-nt 5’ oh), 0.005 (blunt vs. 15nt 5’ oh), 0.77 (10-nt vs. 15-nt 5’ oh), and 0.0002 (10-nt 5’ oh vs. 10-nt -GC 5’ oh). n = 104, 90, 51, 46, and 137 trajectories for the blunt-end, 3’ oh, 10 nt 5’ oh, 15 nt 5’ oh, and 10 nt -GC 5’ oh, respectively. (B) Mean rotation angle for the transient two-state transitions prior to processive DNA unwinding (as indicated in Figures 4.11 and 4.14), observed for the blunt-end and 5’ overhang substrates. Error bars are s.e.m. The data are generated from 50-100 single-molecule trajectories from at least three experiments for each substrate geometry at 50 μM ATP, solution pH 8.

These results support the hypothesis outlined above, where an ATP-independent unwinding step is required to engage the RecB motor, thereby initiating the processive, ATP-dependent unwinding on blunt-end dsDNA, and demonstrate that having a ssDNA overhang engaging the RecB motor can bypass this ATP-independent unwinding step.

We next tested whether a 5’ overhang would have the same effect. An earlier study using ensemble stopped-flow measurements has suggested that a 10-nt 5’ overhang can allow RecBCD to bypass an initiation step [250], but the nature of the initiation step observed in
these ensemble measurements is unclear. Moreover, the substrates used in that study con-
tains not only a 5’ overhang but also an additional 3’ overhang, which complicates interpreta-
tion. To test whether the 5’ overhang allows bypassing the ATP-independent unwinding step
that we observed during initiation, we designed substrates without a 3’ overhang but with ei-
ther a 10-nt or a 15-nt 5’ overhang, both of which should be sufficiently long to engage RecD
[245, 250].

Notably, in contrast to the 3’-overhang substrate, the 5’-overhang substrates still exhib-
ited the transient, two-state transitions in the initiation phase with a magnitude similar to
that observed for the blunt-end dsDNA (Figure 4.14B; Figure 4.15B). Like for the blunt-end
substrate, processive unwinding of the 5’-overhang substrate started from the state where ~5
bp were unwound (green state, Figure 4.14B). Interestingly, compared to the blunt-end sub-
strate, the 5’-overhang substrate had a substantially longer initiation phase due to a much
longer dwell time in the wound state, but spent less total time in the unwound state be-
fore processive unwinding started (Figure 4.15A). Because of the increased dwell time in the
wound state, initiation on these 5’-overhang substrates was strongly rate-limited by unwind-
ing of the first ~5 bp. We thus hypothesized that it might be possible to speed up initiation
by weakening these base pairs. Indeed, conversion of the G-C pairs in the initial 5 bp to A-T
(‘-GC’) substantially reduced the average initiation time of the 5’-overhang substrate (Figure
4.15A) while still exhibiting the 5 bp initiation step (Figure 4.15B).

Finally, we characterized the dependence of the RecBCD initiation pathway and kinet-
ics on the pH of the buffer solution and the ATP concentration. We found that for all tested
pH conditions, the same qualitative behavior was observed: the blunt-end and 5’-overhang
substrates showed 5 bp initiation steps, while the 3’-overhang substrate did not (Figure 4.16).
The initiation kinetics, by contrast, showed a kinetic dependence on pH, with faster kinetics
at higher solution pH over the range tested (solution pH 6 - 8; Figure 4.17). Although the ab-
Figure 4.16: RecBCD initiation trajectories at solution pH 6. (A) Single-molecule trajectories showing RecBCD initiation on blunt end substrates at 50 μM ATP, solution pH 6, recorded at 500 Hz. Arbitrary vertical offsets are applied to different trajectories for display purposes. Repetitive two-state transitions are evident during the initiation phase prior to processive DNA unwinding. The dash-boxed region is magnified and shown in (B), in which the lower and higher rotational angle states are marked in magenta and green, respectively. (C) Single-molecule substrates showing RecBCD initiation on 6-nt 3’ overhang (D) and 5’ overhang (10-nt) substrates under the same conditions. The dash-boxed region is magnified and shown in the inset.
Figure 4.17: RecBCD initiation times at different pH values. Mean initiation state dwell time, mean cumulative green state dwell time and mean cumulative magenta state dwell time as described in the main text, at (A) solution pH 6, (B) solution pH 7 and (C) solution pH 8. Note the different scaling on the time axes. Data are mean ± s.e.m. Data are generated from ~30 – 130 single molecule trajectories.

Solute rates depended on pH, the relative rates (with the 5’ overhang slower to initiate than the others) remained the same at lower pH. Similarly, the same dependence on overhang geometry was observed in experiments conducted at 6-fold higher (300 μM) ATP (Figure 4.18). Thus, RecBCD follows the same initiation mechanism across a range of conditions.

Taken together, our results suggest that engagement of the 3’ DNA strand with the RecB motor plays an important role in the initiation of DNA unwinding and that for DNA substrates that lack a 3’ overhang, an ATP-independent unwinding transition is used to engage the 3’ DNA strand with RecB (Figure 4.19). Interestingly, a recent cryo-EM structure detected nucleotide analog binding to the RecB motor, but not RecD, perhaps suggesting that RecB more strongly binds ATP before processive unwinding begins [253]. Furthermore, it was suggested that a helix from the RecB domain in this structure might stabilize the closed conformation of the RecD motor, perhaps preventing RecD activity, although it is un-
clear if this contact would be sufficient to have such an effect, or whether it might be affected by the amount of DNA contacting the RecB motor [253].

The slower transition to the unwound state in the case of the 5'-overhang substrate may be due to a higher activation barrier created by the additional contacts that are present between the 5' overhang and RecBCD [246], which likely need to be broken as part of the unwinding step. Supporting the idea that the presence of the overhang might disfavor further unwinding, a cryo-EM structure of RecBCD with a DNA substrate with a 3-nt 3' overhang and a 12-nt 5' overhang did not show unwinding [253].

We note that processive, ATP-dependent DNA unwinding does not start immediately after the ATP-independent unwinding of the blunt-end and 5'-overhang substrate or immediately after RecBCD binding to the substrate with a sufficiently long 3' overhang to engage RecB. Additional waiting time was observed before processive unwinding in both cases, indicating the presence of another rate-limiting step. This particular waiting time was shorter for substrates with a 5' overhang (Figure 4.15A), suggesting a role of the engagement of the
5’ overhang with RecD in initiation once the 3’ strand has reached RecB, consistent with the previous suggestion of the involvement of RecD during initiation from both single-molecule and stopped-flow experiments [250, 252]. The rate enhancement in this step, the transition between the unwound initiation state and the processive unwinding state, may be a sign of cooperativity between the two motors in RecBCD when both are in contact with DNA. Comparison of RecBCD structures with different 5’ strand lengths showed that having more DNA in contact with the RecD domain leads to a change in the RecD-RecB contacts, potentially providing a structural explanation for this effect [253]. Note also that cooperativity during the processive unwinding phase has been previously suggested based on the decrease in unwinding rate when either motor is mutated [208, 228].
It is also interesting to note that RecBCD with an ATPase-deficient RecB mutant can initiate processive unwinding on dsDNA with a 5’ overhang but not on a blunt-end dsDNA [224, 225, 228], suggesting that RecD can partially compensate the role of RecB in initiation when RecB is defective or that the role played by RecB in this case does not necessarily rely on ATP hydrolysis.

4.8 Stopped-flow validation of RecBCD initiation results

To further test our single-molecule results, we used an ensemble stopped-flow assay to measure the pre-steady state kinetics of RecBCD-mediated DNA unwinding. These experiments are similar to those described in Section 4.3, except dT overhangs are added to the dsDNA end as indicated (Figure 4.20A). Using this approach, we found that processive unwinding was indeed delayed on a 5'-overhang substrate as compared to a blunt-end substrate, and that G-C to A-T conversion in the first 5 bp increased the rate of strand displacement (Figure 4.20B-C), corroborating our single-molecule results.

We further generated predictions of ensemble time courses using the initiation and unwinding rates determined from single-molecule experiments and compared them with the stopped-flow measurement results. For these predictions, unwinding rates and initiation dwell times were drawn from the experimental distribution repeatedly to generate an expected ensemble result. For this simulation, unwinding was assumed to occur as a series of 1 bp steps. The Cy3 unquenching was assumed to occur in a single step, when the RecBCD had unwound all of the DNA between the dsDNA end and the dyes. (Note that during unwinding on the 3’ overhang the unwinding phase is 5 bp longer than for the other substrates, as no initial unwinding step occurs.)
Figure 4.20: RecBCD unwinding kinetics measured using ensemble stopped-flow assay. (A) Design of the DNA substrate for the stopped-flow experiments, as described in Figure 4.5, except that single stranded overhangs were added as indicated (dashed lines) to prepare the overhang substrates. (B) Stopped-flow fluorescence measurements for blunt end, 6-nt 3’ overhang (3' oh), and 10-nt 5’ overhang (5' oh) substrates, showing slower kinetics when the 5’ overhang is added. (C) Stopped-flow fluorescence measurements for 10-nt 5’ overhang and an additional 10-nt 5’ overhang substrate with G-C pairs in the initial 5 bp converted to A-T (5’ oh -GC). Stopped-flow measurements were conducted at 50 μM ATP, solution pH 6.

Recall from Section 4.4 that the proximity of RecBCD to the silica coverslip surface is expected to lead to a pH shift of about 2 units. This effect should be explicitly taken into account in the quantitative comparison because the initiation kinetics of RecBCD depend on pH (Figure 4.17). We additionally observe this pH dependence in the stopped-flow measurements, and find excellent agreement with the simulation based on the blunt-end ORBIT kinetic data with the expected pH offset of 2 pH units (Figure 4.21A). Note that, due to the
Figure 4.21: Comparison of simulated and experimental stopped-flow time courses. (A) Comparison of predicted ensemble time course (red line) based on the initiation and unwinding rates derived from single molecule data (collected at a solution pH of 8) to measured stopped-flow time courses (dots) on the 26 bp blunt end substrate at several pH values. As expected, due to the shift in local pH near the surface (see text), the predicted curve from the initiation and unwinding rates measured by single-molecule experiments at solution pH of 8 matches the stopped-flow data measured at pH 6. (B) Comparison of stopped-flow data (red symbols) for blunt end substrates at pH 6 to the time courses predicted from both initiation and unwinding rates (red) or unwinding rates alone (black line), derived from single molecule data obtained at a solution pH of 8 (surface pH of 6). Inclusion of the initiation phase in the simulation is required to match the experimental results. (C)-(D) Comparison of stopped-flow data at pH 6 (color symbols) shown in Figure 4.20 to predicted ensemble time courses (lines with matched color) using initiation and unwinding rates derived from the single-molecule data at a solution pH of 8 (surface pH of 6) as described for (A). All stopped-flow experiments were conducted at 50 μM ATP, pH 6 except for (A), for which pH is indicated in the legend.
short distance between the dsDNA end and the Cy3 dye (and thus short unwinding phase time), the initiation dwell time contributes significantly to the observed signal. This can be seen by omitting the initiation phase from the simulation, which leads to a much faster rise in fluorescence than that observed in the stopped-flow data (Figure 4.21B). Importantly, after accounting for this local pH shift, we observed excellent agreement between single-molecule and stopped-flow measurements for all blunt-end, 3’-overhang and 5’-overhang substrates tested without any fitting parameters (Figure 4.21C-D). This provides further validation that the RecBCD mechanism observed in the ORBIT experiments while the enzymes are adsorbed onto the surface is also operative in freely diffusing enzymes.

As single-molecule experiments are often carried out using surface-anchored molecules, note that it may be generally important to characterize the effect of surface-induced pH shift on enzyme behavior if the enzyme is near a charged surface. As a further technical note, although passive surface adsorption to silica coverslips did not perturb the activity of RecBCD except for this observed pH shift, such attachment could potentially perturb some other enzymes, in which case specific linkage at multiple sites on the enzyme [300] or surface modification could be used to mitigate such perturbation yet still achieve torsionally constrained enzyme immobilization. This point will be addressed in more detail in Section 7.2.

4.9 RecBCD conclusions

In summary, we have applied the ORBIT assay to study the helicase activity of the DNA repair enzyme RecBCD. ORBIT can be used to track the rotation of DNA induced by RecBCD’s DNA unwinding activity during initiation and processive unwinding. This study was made possible by the very high spatiotemporal resolution of ORBIT, required to track the rapid ro-
tation of DNA induced by RecBCD, and the capability of ORBIT to track activity starting from initial binding of the DNA substrate.

We first studied the processive unwinding phase of RecBCD activity, finding ATP-dependent and heterogeneous unwinding kinetics. Additionally, we directly observed two types of pauses (one independent of backtracking, and the other occurring as part of recovery after backtracking) with different dependencies on ATP, as well as a distinct backtracking phase. During initiation on blunt-end substrates, we observed reversible 5 bp unwinding events, which likely bring the DNA in contact with the RecB motor. Adding a 3' overhang to the DNA substrate allowed the enzyme to bypass this unwinding step because the DNA could directly contact the motor. Surprisingly, extending the 5' DNA such that it can reach the RecD motor did not have a similar effect: initial 5 bp unwinding was still required. Thus, these measurements of initiation suggest the model of initiation described in Figure 4.19, where contact between the RecB motor and the 3' DNA strand occurs prior to initiation, even if the RecD motor can already contact DNA. These findings were further validated using an ensemble, stopped-flow assay.

In addition, the results found here may have implications for related enzymes. RecBCD is widely conserved in Gram-negative bacteria [211]. Gram-positive bacteria and mycobacteria instead contain the AddAB and AdnAB, respectively, families of enzymes, which are partial RecBCD homologs that fulfill a similar role but are less well understood [208, 233]. While RecBCD does not have direct homologs in eukaryotes, the same role in recombination is played by a collection of other enzymes, including ExoI, Dna2, and Sgs1 [211]. Additionally, the dual motor architecture (3'-5' and 5'-3' DNA translocation domains acting in tandem) has recently been recognized in other complexes. Thus, our results might shed light on other systems with two motors, such as the Rep-DnaB complex [301], and the human complexes hDNA2-WRN and hDNA2-BLM [302].
In this chapter, I describe a third application of the ORBIT method: measuring the rotation of DNA induced by *E. coli* RNA polymerase (RNAP) during transcription. After introducing RNAP in Section 5.1, I describe the use of ORBIT to measure the kinetics of elongation during transcription in Section 5.2. I then present results at low [NTP] that show the single-base pair stepping behavior of RNAP using ORBIT.
5.1 Introduction to RNA Polymerase

The transcription of RNA from DNA is a key step in the production of gene products. Transcription leads to the production of messenger RNA, which are further processed by the ribosome to produce proteins, as well as a number of other RNA molecules with enzymatic and regulatory roles [36]. In *E. coli*, transcription is carried out by a single enzyme, RNAP; by contrast, eukaryotes contain multiple, specialized types of RNAP [36]. *E. coli* RNAP has significant sequence homology and shares a similar structure with other bacterial and eukaryotic RNAP enzymes [303, 304].

Transcription by RNAP begins with the recognition of a promoter sequence and the opening of a transcription bubble in a cleft in the RNAP structure [305–307]. The complex begins to add ribonucleotides and, possibly after multiple rounds of futile cycling, can transition into elongation, the processive transcription state [84, 308]. During elongation, the DNA upstream of the enzyme is unwound, giving the ribonucleotides access to the template DNA strand [84, 308]. Unwinding requires rotation of DNA relative to the RNAP, which has been directly observed at low resolution using a magnetic tweezer [1]. At the back end of the enzyme, the DNA is re-wound. If the RNAP rotates around the DNA, this can lead to no net rotation, but in practice movement is constrained by the drag of the complex and the torsional stress can both cause rotation of the RNAP and the introduction of positive (on upstream DNA) and negative (on downstream DNA) twist into the DNA [15, 309]. Furthermore, torsional strain in the DNA affects the rate of elongation [56]. The torsional strain induced by the enzyme itself is sufficient to slow elongation unless the strain is actively removed [15].

RNAP activity is highly regulated, with many of the regulatory steps reflected in pausing and backtracking behaviors observed by ensemble and single-molecule approaches [84,
These behaviors also play important roles in proof-reading and error correction [84]. As with RecBCD, RNAP is known to exhibit significant heterogeneity between molecules during elongation [311, 312]. Elongation depends on applied force, and RNAP can transcribe against 25 pN of opposing load [87, 313]. Single-molecule measurements have revealed that RNAP takes 1 bp steps along DNA [201, 313]; these experiments will be described in more detail in Section 5.3. Single-molecule measurements of RNAP activity have also been applied for biotechnology applications, such as sequencing of DNA [314, 315]. Because of both the physiological and technological importance of RNAP, new information about RNAP derived from applying new technologies to monitor its activity can have widespread impact.

5.2 Kinetics of elongation

To further demonstrate the general utility of our method, we used ORBIT to probe transcription by RNAP. We ligated dsDNA transcription templates with the T7A1 promoter [316] followed by a region depleted of dT bases (Figure 5.1A) to the origami rotors. We first assembled stalled transcription elongation complexes by initiating transcription in the presence of three of the four NTPs (excluding UTP to cause the complex to stall) [317], with each complex comprising an RNAP enzyme bound to a dsDNA template (with 143 bp between the stall site and origami rotor) attached to an origami rotor. We then deposited these complexes onto a microscope coverslip through passive surface adsorption of RNAP and resumed transcription elongation by adding all four NTPs.

The ORBIT trajectories revealed processive rotational motion punctuated by long pauses (~1-25 s; Figure 5.1B), reminiscent of the dynamics of the linear movements of RNAP translocation observed in previous single-molecule experiments [87, 318–320]. We measured the dependence of the elongation rate on NTP concentration and found that it was well described
by Michaelis-Menten kinetics, with $K_M = 240 \pm 30 \mu M$ and $v_{max} = 17.5 \pm 0.8 \text{ bp/s}$ (Figure 5.2A), consistent with previous results [1, 318–320]. As with RecBCD, the observed average rates for individual RNAP molecules were heterogeneous (Figure 5.2B-D), consistent with previous single molecule work on RNAP [311, 312, 320]. We additionally note that previous work has shown that RNAP activity is not perturbed when passively attached to a coverslip [1].

5.3 Single base pair stepping

Optical trap studies have detected single-base-pair translocation steps of RNAP by measuring linear movements of RNAP during transcription under an applied force [201, 313]. In
Figure 5.2: Rotation kinetics of RNAP during transcription detected by ORBIT. (A) Dependence of elongation rate on NTP concentration at room temperature. The average rates were fit to Michaelis-Menten kinetics ($v_{max} = 14 \pm 1$ bp s$^{-1}$; $K_M = 240 \pm 30$ μM). Error bars indicate s.e.m. Data are generated from ~30-40 individual molecules from at least three separate experiments at each ATP concentration. (B) Distribution of single-molecule transcription rates at 0.1 mM NTP. (C) Distribution of single-molecule transcription rates at 0.5 mM NTP. (D) Distribution of single-molecule transcription rates at 1 mM NTP.

2005, Abbondanzieri et al. [313] observed elongation by RNAP under low NTP concentrations using an optical trap. They then calculated the autocorrelation function of the position trajectory and analyzed the autocorrelation function’s periodicity to determine an underlying step size of 0.37 nm, or about 1 bp. However, only a small fraction of the elongation data was of sufficient quality to show stepping, and the analysis was only applied to this small
subset. Later, Righini et al. [201] used an improved optical trapping setup to record higher resolution stepping data, allowing quantitative analysis of most of the elongation data. By analyzing the resulting data using a Hidden Markov model, they also found evidence of single base-pair stepping, with a step size of 0.32 nm. In both cases, these measurements required the application of a large applied force to increase the stiffness of the DNA and thus improve the resolution of the measurement. This large assisting force may rectify or otherwise influence the motion of the RNAP along DNA. Furthermore, while RNAP generally rotates along the DNA helix [1], rotational steps have not been previously observed and hence the extent to which the rotational motion of RNAP follows the helix on the scale of single base pairs remains unclear.

We thus sought to determine whether RNAP moves in single base-pair steps in the angular dimension and in the absence of an applied force. We recorded ORBIT data on RNAP at low NTP concentrations. Following Abbondanzieri et al. [313], we used unequal concentrations of the four NTPs ([GTP] = 5 μM, [UTP] = 5 μM, [ATP] = 2.5 μM and [CTP] = 1.25 μM) due to their different incorporation rates. In many segments of the trajectories, steps consistent with single base pair motion (~35°) were visually apparent (Figure 5.3A). We systematically characterized this stepping behavior by using a variable step-size hidden Markov model [321] (Figure 5.3B) to find the most probable step size in each trajectory (Figure 5.3C) and the probability distribution for all forward steps (Figure 5.3D). These distributions showed a single, clear peak at ~35° with most steps distributed between 25-40°, consistent with the 27-40° range of twist angles between subsequent base pairs in B-DNA [322]. These results thus revealed RNAP rotational movement steps corresponding to single-base-pair translocation, suggesting a close coupling between the rotational motion of the enzyme and the DNA helical structure down to the scale of a single base.
Figure 5.3: Single base-pair stepping of RNAP during elongation. (A) Examples of single-base-pair sized rotational steps of RNAP at low NTP concentration (gray: raw 50 Hz data; black: 3-point boxcar filter). Horizontal dashed lines are spaced according to the average single-base-pair twist angle (34.6°) of DNA. (B) An example single-molecule trajectory of RNAP-driven DNA rotation (gray: raw 50 Hz data, black: 3-point boxcar filter) shown together with the hidden Markov model (HMM) fit (red). Bottom right inset: High resolution measurement of a single step of RNAP-induced DNA rotation. Gray: raw 200 Hz data, red: trajectory showing means before and after the step. Scale bar: 1 s. Data were acquired with [GTP] = 5 μM, [UTP] = 5 μM, [ATP] = 2.5 μM and [CTP] = 1.25 μM. (C) Distribution of most probable step sizes >5° from the HMM-derived step size distribution of individual single-molecule trajectory. For each measured single-molecule trajectory, a most probably step size is determined and the distribution is constructed from many such trajectories (N = 31). (D) Full probability distribution of forward step sizes (>5°) from the HMM analysis of the single-molecule trajectories of RNAP-induced DNA rotation. Here all detected step probabilities in individual single-molecule trajectories are used to construct the histogram, instead of using the most probable step size of each trajectory as shown in (C).
6.1 DNA origami preparation and purification

DNA origami rotors and anchors were designed using CADNano [113] and prepared as described previously [111], with modifications. All DNA oligomers, including origami staple strands and additional DNA linkers, were ordered from Integrated DNA Technologies (IDT). DNA oligomers for the rotor and anchor are given in Tables A.1 and A.2, respectively. All
origami structures contained extension strands with single-stranded DNA (ssDNA) overhangs for ligation to additional DNA. Oligomers containing dye, phosphorylation for ligation, or biotinylation modifications and any strands being ligated were ordered with HPLC or PAGE purification. DNA modifications, namely Cy3 fluorescent dyes, biotinylation, and 5’ phosphorylation, are indicated using their IDT codes (see Tables A.1-A.4). To attach the DNA origami anchor to the streptavidin-coated coverslip, it contains six strands (labeled ‘TTHr21’ in Table A.2) with binding sites for the biotinylated oligomer (‘Hr21_5Bio’).

DNA staple strands (Tables A.1, A.2) and the single-stranded M13mp18 viral DNA (New England Biolabs) used as the scaffold [110] were mixed in folding buffer: 10 mM Tris, pH 8.0, 1 mM EDTA, and 18 or 9 mM MgCl₂ for the origami rotor and for the anchor, respectively. The concentrations of the DNA components were 10 nM for the scaffold strand and 100 nM for the unlabeled staple strands, and 0.5 - 1 μM for the Cy3-labeled staple strands. The origami mixtures were incubated and annealed using a thermocycler. For the origami rotor, these mixtures were held at 80° C for 5 minutes, and annealed by cooling, first to 65° C in 1° C steps every 5 minutes, then to 25° C in 1° C steps every 105 minutes. The origami anchor was folded by heating to 80° C for 5 minutes, and annealed by cooling, first to 65° C in 1° C steps every 1 minute, then to 25° C in 1° C steps every 20 minutes.

For RecBCD and RNAP experiments, the folded origami rotors were purified, ligated to double-stranded DNA (dsDNA) segments serving as RecBCD substrates, and then again purified. We first PEG precipitated the origami to remove most of the free staple strands [323]. The origami sample was mixed 1:1 with 2x PEG precipitation buffer (15% PEG-8000, 5mM Tris, pH 8.0, 1 mM EDTA, 500 mM NaCl), incubated 30 minutes at 4° C, and centrifuged at 8000g for 30 minutes. The pellet was washed with 1x PEG wash buffer (7.5% PEG, 10 mM Tris, pH 8.0, 1 mM EDTA, 18 mM MgCl₂). After being resuspended in T4 ligation buffer, the pre-annealed short extension strands were ligated to longer DNA oligomers (Table A.3;
note the sequence descriptions in Table A.3 indicate final extension DNA length after ligation) using T4 DNA ligase (New England Biolabs) for 2 hours at room temperature. The reaction mixture was treated with Proteinase K (New England Biolabs) for 1 hour at room temperature to degrade the ligase. The origami sample was then purified by agarose gel electrophoresis. Electrophoresis was performed with a 2% agarose gel in an ice bath in running buffer containing 89 mM Tris, 89 mM borate, 2 mM EDTA, and 10 mM MgCl₂. The origami band was excised from the gel and the origami extracted using a freeze ‘n’ squeeze spin column (Bio-Rad) by spinning at 1000 g for 60 minutes. The sample was concentrated by PEG precipitation as described above.

To characterize the Brownian-limited angular resolution of the rotor, we prepared a complex with a DNA origami rotor attached to an origami anchor via dsDNA linkers of various lengths (Figure 3.1A). To remove excess origami extension strands, the origami structures were gel purified before ligation. The two origami structures were ligated together as described above. The length of the linker DNA between the two origami structures included the 14 base pairs (bp) extending from the origami rotor (Figure 2.2), 26 bp on the origami anchor (Figure 2.7), 12 nucleotides (nt) of ssDNA overhang on both structures, and any additional DNA added between the two origami. For the shortest DNA linker length, 52 bp, the linker consisted entirely of DNA present on the rotor and anchor structures. Note that for this sample, to create the direct connection, the DNA oligomer Anchor_Ext0_oh (Table A.2) was replaced with Anchor_Ext0_oh_direct (Table A.3). The two longer lengths used additional dsDNA linkers in the ligation reaction (Table A.3). These were either purchased from IDT (92 bp) and annealed prior to ligation or prepared by PCR and dU excision (163 bp). In the latter case, the DNA was prepared using PCR with PfuTurbo Cx Hotstart DNA Polymerase (Agilent) and primers with a dU base 12 nt from their 5’ ends (Table A.3). Following purification on a column (Zymo DCC-100), the product’s dU bases were excised with
the USER enzyme system (New England Biolabs) to create 12 nt overhangs for ligation with the rotor and anchor structures. Finally, the ligation products were again purified using electrophoresis.

6.2 AFM AND TEM IMAGING

AFM images were obtained using an Asylum MFP-3D system (Asylum Research) at the Center for Nanoscale Systems at Harvard University. A 2 μL droplet of purified sample (low nM concentration) and then a 20 μL drop of buffer containing 5 mM Tris, 0.5 mM EDTA, 10 mM MgCl₂, 10 mM NiCl₂ were applied to a freshly cleaved mica surface and left for approximately 2 minutes. The images were taken under liquid tapping mode, with C-type triangular tips (resonant frequency, $f_0 = 40-75$ kHz; spring constant, $k = 0.24$ N m⁻¹) from the SNL-10 silicon nitride cantilever chip (Bruker Corporation).

For TEM imaging, sample was adsorbed onto glow discharged carbon-coated TEM grids for 2 minutes and then stained for a few seconds using a 2% aqueous uranyl formate solution containing 25 mM NaOH. Imaging was performed using a JEOL JEM-1400 TEM operated at 80 kV.

6.3 SINGLE-MOLECULE IMAGING OF RecBCD-INDUCED DNA UNWINDING USING OR-BIT

Single-molecule fluorescence imaging was conducted using a Nikon Eclipse Ti inverted microscopy body with a 60x 1.4 NA oil objective (Nikon) and a high speed scientific CMOS camera (Hamamatsu Orca-Flash 4.0 v2). The camera field of view was cropped as needed to achieve high frame rates. The sample was illuminated using objective-type total internal reflection with a 1 W 532 nm laser (CrystaLaser). Laser intensity was controlled using
an acousto-optical tunable filter (Crystal Technologies). The microscope filter cube con-
tained a dichroic mirror (Chroma Technology Corp ZT532/640rpc-UF3) and an emission fil-
ter (Chroma Technology Corp ZET532/640m-TRF). The focus was maintained with an IR laser reflection focus lock system. Each camera pixel corresponded to 160 nm in the sample plane. The microscope hardware was controlled with custom software written in Python.

Origami were imaged in a flow chamber consisting of a glass coverslip (VWR, No. 1.5) attached to a microscope slide with double sided tape. Two holes were drilled in the slides to facilitate buffer exchange. The coverslips were cleaned by sonication in 95% ethanol, rinsing in water, drying thoroughly with compressed nitrogen, and plasma cleaning under argon atmosphere (Harrick Plasma PDC-32G). Between uses, slides were soaked in acetone and water to facilitate flow chamber disassembly, then scrubbed with alconox, sonicated in 1 M KOH, rinsed in water, briefly flamed with a propane torch, and plasma cleaned. After assembling the flow chambers with double sided tape and sealing the sides with epoxy, they were used the same day or vacuum sealed until use to prevent dust accumulation. Tubing was inserted into the slide holes and epoxied in place. A syringe pump (KD Scientific KDS210) was used to pull solution into the chamber.

Standard reaction buffer contained 50 mM Tris, 2 mM Trolox, 50 M of the Trolox quinone (prepared using UV irradiation and quantified by UV-Vis absorbance spectroscopy before adding to the buffer [324]), 5 mM protocatechuic acid (PCA), 10% glycerol, and 10 mM MgCl₂, adjusted to pH 8.0 unless otherwise indicated. Before imaging, 0.25 U/mL protocatechuate-dioxygenase (PCD; sold by OYC Americas as rPCO) was added. The PCA/PCD system acts as an oxygen scavenger and Trolox suppresses dye blinking [325, 326]. ATP (Affymetrix or ThermoFisher Scientific) was added when indicated in the text. PCD was added ~10 minutes before imaging to allow the PCA/PCD system to remove oxygen. All experiments were done at room temperature (~23° C).
Experiments were conducted by first flowing 300 units/mL RecBCD (New England Bio-labs) in reaction buffer into the chamber to passively adsorb RecBCD molecules to the slide surface. After ~1 minute, unbound RecBCD was washed out with ~100 μL of reaction buffer twice. The second wash contained PCD and the desired ATP concentration. Finally, the origami-dsDNA substrate complex was added to the chamber in reaction buffer with PCD and the desired ATP concentration. Data were typically acquired for 3-4 minutes at 500 Hz – 1 kHz frame rates.

6.4 Single-molecule data analysis for RecBCD

Single-molecule imaging data were analyzed using custom Python code. Briefly, to identify RecBCD-bound rotor complexes, the movie was first divided into 100-frame segments. Fluorescent spots were selected for further analysis by applying a median filter in the time dimension to each segment and then identifying local intensity maxima in the resulting median images. Using a median filter allowed us to ignore rotors that entered the evanescent field transiently without successful binding to RecBCD. Single molecule trajectories were determined by fitting each selected fluorescent spot to a 2D Gaussian in each frame. The Python code used for this analysis is available at https://github.com/altheimerb/python-sma/.

Single molecule trajectories were further analyzed using custom code in Igor Pro. Among the binding events of the rotor-dsDNA substrate to the surface (fluorescent spots appearing at the surface for at least 0.2 s; typically, ~250 spots per field of view), ~50% of these spots showed localizations moving along the full circumference of a circle and were selected for further analysis. The remaining 50% showed either no apparent motion (~30%), likely corresponding to rotors bound directly to the surface (i.e. not through the dsDNA linker which would have led to Brownian motion of rotors), or showed constrained movement along a par-
tial circular arc (~20%). The latter could be due to rotors bound in a non-canonical manner with some degree of flexibility, for instance via the dsDNA linker to an incorrect part of an enzyme or directly to the coverslip surface, or from canonical attachment of dsDNA linker to inactive enzymes (which would be expected due to the fraction of inactive enzymes present in typical preparations of RecBCD solutions).

For the spots displaying motion along a complete circle, we fit each set of localizations to a circle, the center of which was used to convert (x,y) positions to polar coordinates (see Figure 4.1B). ~50% of the circular trajectories showed largely unidirectional rotation, due to processive RecBCD activity, while the remainder showed random fluctuations in the angular direction. This latter fraction, likely due to mis-incorporated DNA linkers with only a single DNA strand in the origami-linker DNA complex, was not considered for subsequent analysis. Furthermore, high localization precision is required for high-accuracy rotational tracking. We used the radial variance (localization variance in the radial direction orthogonal to the circular path), which depended on the photon number detected from the fluorescent spot in each frame (Figure 4.1C), as an approximate measure of the localization precision and included only trajectories with a localization precision better than 16 nm (0.1 pixel) in our analysis of DNA unwinding by RecBCD. We expect the Brownian dynamics in the radial direction to be small because (1) unlike the angular fluctuations, the radial motions due to lateral motions of the origami (resulting from DNA bending) are not amplified by a lever arm effect and (2) tilting motion of the arms relative to the surface, although amplified by the lever arm, has only a very small lateral projection on the imaging plane. Typically, among the trajectories that displayed processive RecBCD activity, ~30% of them were removed due to this localization precision cut. In these remaining trajectories, the angular noise was dominated by the Brownian dynamics, which was substantially larger than the localization precision measured by the radial variance. However, a small fraction of traces (~10% of the remaining trajecto-
ries) showed periods when the angular motion of the origami was comparable to the localization precision, likely due to interactions with the surface or steric hindrance by the surface if the rotor is tilted, and were excluded from further analysis. The surface sticking of origami rotor was minimal potentially because the negative charge on the glass (silica) surface tends to repel the negatively charged DNA. After the above three filters, a total of ~30% (~50% x 70% x 90% ) of circular trajectories remained and were considered in our RecBCD activity analysis. Angular changes were converted to base pairs unwound using the average DNA angular twist of 34.6° per bp.

6.5 RecBCD pausing analysis

We used an automated pause-finding algorithm to identify pauses in the single-molecule unwinding trajectories. Briefly, the trajectories were subjected to a 20 Hz half-transmission frequency binomial smoothing filter, and time derivatives of the trajectory were then used to determine the instantaneous velocity. Frames showing a velocity below a threshold (1° per frame at 500 Hz), were identified as pause frames and frames moving backwards faster than this rate were identified as potential backtracking frames. Because long pauses tended to get broken up due to short fluctuations, another binomial smoothing (8.4 Hz) was applied and additional frames were called as pauses here using the same threshold. Since smoothing tends to blur the edges of the pausing phase, pauses were extended forward and backward until the angle in the raw data moved outside of a 1 bp window from the pause location. Adjacent pauses at the same angular location were merged. Because of signal fluctuations, we only consider pauses that lasted at least 100 ms during forward unwinding to avoid false positive detection of pauses. Likewise, we only considered backtracks that lasted at least 100 ms (including pre-backtracking pause and recovery pause) and exhibited a minimum of 100° (~3 bp)
backward motion as real backtracks. For pauses that occurred between backtracking and re-
covered forward unwinding, we did not set a threshold on pause duration because of the low
probability of false positive detection of such events.

6.6 RecBCD initiation analysis

The RecBCD initiation phase was defined as the period of the trajectory between binding
and the start of processive unwinding. Reversible unwinding transitions were detected during
this initiation phase. To determine the transition rates between the wound (magenta) and
unwound (green) states, the initiation phase was segmented according to the angle, and the
dwell time in each state was determined. In the presence of ATP, exit from the green state is
a competitive process, with transitions occurring both back to the magenta state or forward
to processive unwinding. This reduces the dwell time in the green state, increasing the appar-
ent transition rate between the green state and the magenta state. Assuming both processes
are first order transitions, the apparent transition rate, measured as the average dwell time
of individual green states, should be equal to the sum of the real transition rate from green
to magenta state and the transition rate from green state to the processive unwinding state.
Mathematically, the real transition rate from the green state to magenta state, $k_{g-m}$, is re-
related to the average dwell time in the green state, $<t_g>$, and the probability of transitioning
from the green state to the magenta state, $P_{g-m}$ according to [299]

$$k_{g-m} = \frac{P_{g-m}}{<t_g>}.$$  \hspace{1cm} (6.1)

$P_{g-m}$ is related to the average number of times each trajectory visits the green state, $c_g$,
before unwinding by considering all possible number of visits and their probability [327]
\[ <c_g> = \sum_{n=1}^{\infty} (1 - P_{g-m}) \times (P_{g-m})^{n-1} \times n = \frac{1}{1 - P_{g-m}} \] (6.2)

Based on our measured distribution of the number of times each trajectory visited the green state, we derived obtained \( P_{g-m} = 0.52 \pm 0.06 \) (50 \( \mu \)M ATP) and \( P_{g-m} = 0.29 \pm 0.03 \) (300 \( \mu \)M ATP). We then used this probability and the average dwell time of each green states to determine \( k_{g-m} \), using Equation 6.1. For the 50 \( \mu \)M condition, we excluded a single outlier with >6x more transitions than the trajectory with the next largest number of counts in calculating \( P_{g-m} \); including this point does not significantly affect the result (\( P_{g-m} = 0.65 \pm 0.10 \)). Since the transition rate from the magenta to green state, \( k_{m-g} \), is not affected by the transition from the green to processive unwinding state, \( k_{m-g} \) was simply determined from the average dwell times of the magenta states. To determine these transition rates in the absence of ATP, experiments in which ATP was initially absent and then added after 6 seconds were similarly segmented into states. In these experiments, dwell times beginning after the addition of ATP were excluded. Only trajectories which showed processive unwinding after adding ATP were considered.

6.7 Stopped-flow

DNA unwinding by RecBCD was measured in a pre-steady state stopped-flow assay similar to that of reference [249] using a stopped-flow spectrophotometer (KinTek SF-2004) to measure the unwinding of fluorescently labeled DNA substrates (Figure 4.5; Table A.4). In this substrate design, each sample forms a hairpin consisting of a “C” strand annealed to a pair of “A” and “B” strands, as shown in Figure 4.5A. The emission intensity from a Cy3 dye was initially quenched by energy transfer to a Cy5 dye in each substrate. RecBCD-catalyzed
DNA substrate unwinding dissociated the Cy3 labeled DNA strand, resulting in reduced quenching of Cy3, which was detected as an increase in Cy3 fluorescence.

We used a 2-stage mixing protocol in order to measure single-turnover unwinding reactions shortly after enzyme-substrate binding (Figure 4.5B). In the first stage, we mixed RecBCD with DNA for 200 ms, allowing the enzyme-substrate complex to form. In the second stage, we mixed the enzyme-substrate complexes with a solution containing ATP and heparin while recording the resulting fluorescence. Heparin inhibits RecBCD-DNA binding and thus prevents enzymes from engaging new DNA substrates after the first turnover [249]. The use of heparin to achieve single-turnover conditions requires that the RecBCD-DNA complex is formed before the ATP and heparin are added. Although this can be achieved by pre-mixing the RecBCD and DNA prior to loading the mixture into the stopped-flow instrument, we chose to instead briefly mix RecBCD and DNA for 200 ms in the stopped-flow apparatus before the addition of ATP to trigger unwinding in order to prevent long equilibration of the RecBCD-DNA complex in the absence of ATP which could potentially change initiation phase characteristics and to more closely match our single-molecule experimental conditions, where the substrate and ATP were simultaneously added to the surface-bound enzyme. The 200 ms mixing time was the shortest time that allowed enough RecBCD-DNA complexes to form to create a reliably detectable fluorescence signal after adding ATP. Each experimental fluorescence time course shown in the figures represents an average of $n > 10$ individual measurements.

6.8 Single-molecule measurements of RNAP-induced DNA rotation

RNAP experiments were conducted similarly to the RecBCD experiments described above. The origami complexes were ligated to dsDNA extensions (see Table A.3) which contained
the T7A1 promoter [316]. The reaction buffer for these experiments contained 50 mM Tris, 2 mM Trolox, 50 μM of the Trolox quinone, 5 mM protocatechuic acid (PCA), 20 mM NaCl, and 10 mM MgCl₂, adjusted to pH 8.0. Stalled ternary (RNAP-DNA-RNA) complexes were formed by adding RNAP and origami complexes to 10 μM ATP, GTP, and CTP, 250 μM of the dinucleotide ApU, and 5 mM DTT in reaction buffer. The dsDNA extensions (see Table A.3) contained a 20 bp sequence without T’s after the T7A1 promoter, causing the complexes to become stalled in the absence of UTP [1]. After ~20 minutes at room temperature, 200 μM sheared salmon sperm DNA (Invitrogen) was added to the solution containing stalled ternary complexes. The complex was then incubated at least 5 minutes at room temperature and added to microscope flow chambers. After washing out excess complexes, reaction buffer with all four NTPs, 100 μM salmon sperm DNA, and 0.25 U/mL PCD was added to start the reaction while imaging. RNAP experiments were recorded at 200 Hz, except for those at low [NTP] for single base pair stepping detection, which were recorded at 50 Hz or 100 Hz and down-sampled to 50 Hz. Following previous work [313], experiments for base-pair stepping analysis were conducted with a 4:4:2:1 ratio of GTP:UTP:ATP:CTP (5, 5, 2.5, and 1.25 M, respectively). All other experiments were done with equimolar concentrations of NTPs.

The RNAP data was processed similarly to the RecBCD data. Single base pair stepping was analyzed using a previously described hidden Markov model (HMM) [321] on the elongation phase of each trajectory with high localization precision. We used a 5° cutoff to reject spurious steps smaller than the noise.
In this thesis, I have described the development of a new method, Origami-Rotor-Based Imaging and Tracking (ORBIT), for tracking single-molecule rotational motion with high resolution and throughput. I have further described three applications of this approach, to study the biophysical properties of DNA, the unwinding of DNA by RecBCD, and transcription by RNAP. In this final chapter, I summarize these results and then discuss potential future methodological developments and applications.
7.1 Summary of this thesis

In Chapter 1, I motivated why measurements of DNA rotation are important for investigating protein function and mechanism. In particular, these measurements provide a direct readout of one of the primary outcomes of many types of protein activities: DNA unwinding, DNA rotation due to helix-tracking, DNA bubble formation, and DNA distortion. Although a number of methods have been developed for studying these activities, they are limited by resolution and throughput, among other challenges, and have not seen widespread adoption. I then described DNA nanotechnology, and in particular, DNA origami, which provides a way to generate custom structures for use in single-molecule biophysics as well as many other applications.

In Chapter 2, I described the application of DNA origami to the problem of tracking DNA rotation. Using origami as a lever arm, we structurally amplify the small motion of DNA rotation into a larger motion. By adding fluorescent dyes to the end of the lever arm, we can track the rotation using a standard fluorescent microscope. Because of the simplicity of the method, and the customizability and flexibility of the DNA origami design, we anticipate our method will be useful for a range of applications requiring precision measurement of DNA rotation. I discussed in detail several key design considerations, which will remain important to consider while further customizing or improving the structure, including the origami-dsDNA junction.

In Chapter 3, I described the characterization of the thermally-driven Brownian dynamics of the origami rotor. By fitting the power spectrum of this motion, we determined the torsional stiffness and hydrodynamic drag of this complex, which set the spatiotemporal resolution. We found that ORBIT achieves the highest demonstrated resolution for tracking rotation of DNA, with only ~20 ms of integration time required to resolve single base-pair steps
when the dsDNA substrate is 52 bp long. Further, we used the dsDNA length dependence of the torsional stiffness to estimate the effective torsional rigidity constant of DNA under no applied stretching force, and found good agreement with the range of published values.

In Chapter 4, I presented the first application of ORBIT to study DNA rotation induced by enzyme activity. We used ORBIT to track the unwinding of DNA by the helicase RecBCD, a repair enzyme which prepares DNA for homologous recombination. Applying ORBIT to track DNA rotation allowed us to observe distinct phases of initiation, unwinding, pausing and backtracking during RecBCD-mediated DNA unwinding. In particular, we focused on the mechanism of initiation, finding that engagement of the 3’ DNA strand with the RecB motor plays an important role during for the transition to the ATP-driven processive unwinding phase, and that ATP-independent 5 bp unwinding steps can allow the 3’ strand to contact the RecB motor.

In Chapter 5, I described the application of ORBIT to study transcription by RNAP. In addition to characterizing the elongation rate using ORBIT, we directly observed single base pair rotational steps of RNAP. This represents the first observation of these steps for RNAP both in the absence of an applied assisting load and in the angular dimension, and suggest that RNAP motion is closely coupled to the DNA double-helical structure down to the base-pair scale. Furthermore, this work establishes that ORBIT can be applied to study a range of protein-DNA interactions, including two enzymes (RecBCD and RNAP) with very different mechanisms.

Finally, in Chapter 6, I give the methods used to make and apply the origami rotors and to analyze the resulting data. Additionally, in Appendix A, I list the DNA sequences used to make the origami structures. These resources should be useful for future applications of ORBIT.
Collectively, the studies presented in this thesis demonstrate the power of DNA nanotechnology to amplify biomolecular movements for mechanistic studies. In the final section, I discuss future developments and applications of the ORBIT technology.

7.2 Future directions

Looking forward, there are a number of improvements and adaptations of the ORBIT method which will increase its utility for addressing biological questions. In this section, I first outline potential improvements and technical developments and then discuss the types of additional problems that ORBIT may be applied to in the future.

While the resolution already achieved using ORBIT is an improvement over existing rotational bead tracking assays, several changes to the origami design will further increase the resolution. First, reducing the drag by decreasing the size of the structure would further shift the Brownian dynamics to higher frequencies. The drag depends strongly on the length of the origami arms [192] as well as the number of arms. An origami rotor with a single, slightly shorter arm will have several-fold lower drag. Design of these smaller origami will be facilitated by recent developments in the production of shorter template strands [328, 329]. Second, the relationship between the system’s stiffness and DNA linker length shows an offset in the y-intercept, indicating that some of the observed flexibility is due to the origami-duplex connection point. Redesigning this connection to reduce flexibility should reduce the compliance of the system, reducing the overall magnitude of the Brownian dynamics.

For many applications, notably on proteins that cause slow DNA rotation (such as RNAP or the Holliday junction migration complex RuvAB [32]), the high temporal resolution achieved in this work may be unnecessary. For these experiments, it would be desirable to trade resolution for other benefits. In particular, increasing the arm length would reduce the contribu-
tion of fluorescence fitting error to the angular uncertainty. Alternatively, it would allow the use of lower laser power densities, increasing photostability and making it possible to image either over a larger field of view to further improve the method’s throughput or with inexpensive, lower power lasers. This expanded set of tools, made possible by the customizability of DNA origami design, will make these rotation tracking methods applicable to questions on a wide range of timescales.

Unlike earlier methods, ORBIT does not require the application of a stretching force to the DNA. This simplifies the experiments and avoids the potentially perturbative effects of applied force. However, in some cases, it can be informative to use applied force or torque as control parameters when studying enzyme mechanism. With simple changes, ORBIT is fully compatible with the application of force and/or torque. Most simply, the rotor structure could be adapted to have a second dsDNA extension on the ‘top’ surface, opposite the existing dsDNA extension. This extension could be attached to a magnetic bead or optical trap to apply force or torque [78]. Alternatively, the recently demonstrated ability to manipulate DNA origami with an external electric field [330] could be applied to achieve the same end with the existing structure by adapting the flow chamber to accommodate the application of electric fields.

In this thesis, the RecBCD and RNAP enzymes were attached to a cleaned silica coverslip using passive adsorption. This provided multiple attachment points between the enzyme and surface, which is critical for achieving the required torsional constraint for rotation measurements. While this approach was sufficient to study these two enzymes without significantly perturbing their activity, in other cases different approaches may be required if the enzyme either does not bind well to the surface or does not retain its activity. In such cases, there are several additional approaches to consider. First, the silica coverslip can be modified chemically, for example using the wide range of commercially-available silanization
reagents. Such treatment will provide different surface chemistries, affecting how the enzyme
binds to the surface. Second, the enzyme can be attached directly to a biotin-streptavidin
coated surface after specifically or non-specifically labeling the enzyme at multiple sites. This
approach has been demonstrated experimentally for kinesin experiments [300], although it
is unclear whether both attachment points in these experiments were via the streptavidin
or whether one attachment point might be non-specific. Third, to increase the likelihood of
successful attachment, an enzyme could be multiply attached to an origami frame like the an-
chor structure used in this work. Because of the control over shape and the addressability of
linker groups, the origami could be designed to optimally position the linker groups, instead
of the random placement of these groups on a coverslip such as with biotin-BSA/streptavidin
coated coverslips. The attachment of DNA nanostructures to a single enzyme via multiple
attachment points was recently demonstrated [331, 332], and single attachment of proteins
to origami at well-defined positions has long been of interest [144, 173, 292, 333–337]. These
origami adapters could then be attached to a biotin-BSA coated slide using streptavidin, as
described in Section 2.7 and 3.1 for the origami anchor.

Future applications of ORBIT fall broadly into three categories. First, ORBIT can be
used to characterize the physical properties of any structure that can be rigidly linked to the
origami anchor and rotor. The simplest version of this application is described in Section 3.4,
where the torsional rigidity of DNA is calculated. This could be directly extended to study
how this important physical parameter depends on sequence, buffer conditions, DNA mod-
ifications, DNA damage, and protein filament formation. These changes in rigidity could in
turn affect how twist is partitioned between different parts of the DNA. Since torsional strain
induces structural changes and local variations in DNA structure influence protein-DNA in-
teractions [338], changes in rigidity could also modulate how torsional strain affects DNA-
protein interactions. Additionally, the DNA could be replaced with double-stranded RNA.
or even other biomolecules, such as collagen [339], to study their torsional properties as well. Second, ORBIT can be applied to study other enzymes that processively move along DNA, like RecBCD and RNAP. This includes possible applications to several large and important classes of enzymes, including helicases, translocases (such as chromatin remodeling enzymes and packaging enzymes), and polymerases [310, 340–343]. Third, the rotor-dsDNA-anchor complex used to measure Brownian motion can also be used to study proteins and small molecules that bind to DNA without processive motion. Many DNA binding proteins, such as transcription factors, distort DNA [4, 338], resulting in twist; others, such as RNAP during initiation [84, 310] and Cas9 during sequence recognition [344, 345], open bubbles in the DNA. Small molecules, most notably intercalators (many of which have therapeutic applications), are also known to cause DNA underwinding [39, 58, 346]. The binding and activity of all of these should be directly observable with ORBIT. These three sets of applications highlight the potential versatility of ORBIT and the wide range of problems that can be approached with improved tools for tracking DNA rotation.
DNA sequences
Table A.1: DNA oligomers for origami rotor. DNA modifications (Cy3 dye labels and phosphorylation) are included using their IDT codes. The final two strands extend outside of the origami. They have 14 nt of complementarity (green) followed on one strand by a 12-nt single stranded overhang (purple) for ligation to additional DNA.

<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0[139]</td>
<td>GCAAGCGAAGCGAAAACGCGCGGGGAGATTCTTTTCACCAGT</td>
</tr>
<tr>
<td>2[293]</td>
<td>GATTGTATAAAGCACCACCCTCGTTGATAAAGCATGTTGGGAGCT</td>
</tr>
<tr>
<td>2[461]</td>
<td>AATTCATCAATAATCGGAATTATCATCAAGAAACCGAGGTAGA</td>
</tr>
<tr>
<td>4[531]</td>
<td>CGACCCTCAATCAATTGCAACAGGAAAAGAGAGCCAGCAGTT</td>
</tr>
<tr>
<td>7[112]</td>
<td>CTCAGTTGAGATTAAACGCGCTTTAAGCCAAAAGGAAAAACGAGGAAAAATC</td>
</tr>
<tr>
<td>8[125]</td>
<td>AGAAAGATTCATATTTAATAAAAACGAATCGGGGAAAGATGACC</td>
</tr>
<tr>
<td>8[503]</td>
<td>CATTTTCGCTACGTTTGCCTTTAGCTGATAGCGTAAGCAG</td>
</tr>
<tr>
<td>9[476]</td>
<td>CCCCTCAGAGCCCAAGCCACCCCTCAGAACCAGCCTCCCTCAGA</td>
</tr>
<tr>
<td>10[237]</td>
<td>TTGAGGCAAAGAAACGCGTTTTAATTCGTTCCATTAAACGTC</td>
</tr>
<tr>
<td>10[447]</td>
<td>CGACGCGGCTCAGTGTATAAGAGCAAGACCAGTAAGCGTCAG</td>
</tr>
<tr>
<td>13[257]</td>
<td>AAAGGAGCCAGGACTAGGAAGCAATATTTCGTCGCTG</td>
</tr>
<tr>
<td>1[154]</td>
<td>TTAGCGGAAGCATGCTAATTCACATTTAGGCGGCTCATGGTC</td>
</tr>
<tr>
<td>1[280]</td>
<td>GAAAATATTTAAATTTTTTTTGTATACGAGGCAAATCAT</td>
</tr>
<tr>
<td>1[448]</td>
<td>AATAATCTTGAGAGACCTACCATTACATCTTTGGCCCTACGAA</td>
</tr>
<tr>
<td>3[308]</td>
<td>AGCAAAACAGAGTTCATTGAGGAGATTACATTGACATTACAAAGGCTG</td>
</tr>
<tr>
<td>4[195]</td>
<td>CGAGGATCCCGGGGGCGCGTAACCACGACGCTTGAG</td>
</tr>
<tr>
<td>6[293]</td>
<td>CACTTTCCGAGGTGAAGTTATAGCTATTTCGCA</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>10[405]13[398]</td>
<td>TTGTACCGACAAAAATAATATCAGAGAGCGGATAAGTGCGCGTTTTGT</td>
</tr>
<tr>
<td>10[531]6[518]</td>
<td>AGAAACCATGATATTACCAGAGAAGGGACGCAAAATCTTT</td>
</tr>
<tr>
<td>21[275]16[280]</td>
<td>GCAAAATTAAGCAATAAGAAAACATTATGAGTTCTATTGACGCTGGT</td>
</tr>
<tr>
<td>2[251]5[251]</td>
<td>GTAATGGGATAGCACCAGTGGATATGGAGTGAGTACGAG</td>
</tr>
<tr>
<td>4[321]0[308]</td>
<td>CTAATCGATGAACGTTAAAGGGATTATTTATGAGGATCCTT</td>
</tr>
<tr>
<td>4[363]0[350]</td>
<td>CTCTGATGCAAATCAAGTGGTTTTTATAATGAGAGACTACCAG</td>
</tr>
<tr>
<td>7[341]19[363]</td>
<td>TTTACGGTTACCAACGCCTAACGAAATTTTATTGAGAC</td>
</tr>
<tr>
<td>7[448]10[448]</td>
<td>TGAAACAAATAATCTGAAATTTACGTTAAATGAGGACCAG</td>
</tr>
<tr>
<td>8[461]11[461]</td>
<td>CTTGATATTCACATAGCATTGGACAGGAGACCACACAAATAGC</td>
</tr>
<tr>
<td>9[273]18[273]</td>
<td>GCCCATTGGCGCGACAGTATTGTGGTGATACCGAATTTCAC</td>
</tr>
<tr>
<td>11[126]8[126]</td>
<td>ATAAATCGATACATAGGAATACCACATTTTAATTACAGGT</td>
</tr>
<tr>
<td>11[252]8[252]</td>
<td>CCAGACCAAGACTTACGGCTACAGAGCCATCAGCAAGC</td>
</tr>
<tr>
<td>11[546]8[546]</td>
<td>TAATACTTGGAGCAAGGTTGAAATTATCAAAGGGAGGAAGGT</td>
</tr>
<tr>
<td>16[377]11[384]</td>
<td>TTAGTACGGCATGTAGTGCCAGATGTCAGAGGGTAA</td>
</tr>
<tr>
<td>16[384]15[398]</td>
<td>AGGAGGTACAACGCGCAACATGTTATTTA</td>
</tr>
<tr>
<td>22[300]6[308]</td>
<td>ATGCTGTTTTAATCACCAGTTAGTAAAGTTCAAAAAGGGCGGAGAC</td>
</tr>
<tr>
<td>1[238]4[238]</td>
<td>CGGTACGTTTGGTGGTTGGGAGGGACGAGGTGTACGCTTGAGTAA</td>
</tr>
<tr>
<td>Label</td>
<td>Sequence</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2[503]5[503]</td>
<td>CTAACAACTAATTTTGGAATTGGAAGGCGAAATCGAAACAAT</td>
</tr>
<tr>
<td>4[489]0[476]</td>
<td>GAAGTAACATTATCCGGCCTTTGCTGTAACAATTTCGACAAGA</td>
</tr>
<tr>
<td>5[210]2[210]</td>
<td>AATGCGCTAACGCCTGTGCTGCAAGGCGCAGGCTCTTTCGCTA</td>
</tr>
<tr>
<td>6[139]9[139]</td>
<td>TAATGCCAAAAATCAACCAGTGCAAGCTTTAACGAACACGTAATAGGGAGACCTTCCGGCAGGGTT</td>
</tr>
<tr>
<td>7[196]10[196]</td>
<td>CTATAATTTGTGTCGAACGCGCAGGCTACTAATAGGGAGACCTTCCGGCAGGGTT</td>
</tr>
<tr>
<td>7[490]10[490]</td>
<td>GCTAGCCCGCTTATCACCGGAAACCAGGCCAGAAAAAGACGAAAAAGACGAAAAAGACGAAAAAGACGAAA</td>
</tr>
<tr>
<td>10[391]16[385]</td>
<td>GGTAAGTAAATTCTGAACCGTACTAATAGGGAGACCTTCCGGCAGGGTT</td>
</tr>
<tr>
<td>11[504]8[504]</td>
<td>ATAGCCCGAAATCATAACCTGTTGCCATCGGCGTCTTTCATCGG</td>
</tr>
<tr>
<td>16[398]7[405]</td>
<td>TTTCGAAGGAATATATAAAAAATAAGTATAGGCGGAGGAAAAAGACGAAAAAGACGAAAAAGACGAAAAAG</td>
</tr>
<tr>
<td>20[293]7[293]</td>
<td>CAGGCAACCTACGCAGGAGAAGGCCCCGGTACGATTCGAGGAGACCTTCCGGCAGGGTT</td>
</tr>
<tr>
<td>23[275]24[273]</td>
<td>GCTCCTTTTGATACGAGTAGATTTTAG</td>
</tr>
<tr>
<td>0[307]3[307]</td>
<td>GTAAAAAAGGCGAGTAAATCGTAAACTTCGAGAAAAAGGCCCCGG</td>
</tr>
<tr>
<td>0[433]3[433]</td>
<td>GCAAAAGCTTTGATACAGTAAACAGTACCCCTGATGCTTTTT</td>
</tr>
<tr>
<td>0[517]3[517]</td>
<td>GAGGATCCGAGCATATCTGCGAGTCTGTTATCTAAAATAAC</td>
</tr>
<tr>
<td>3[392]1[405]</td>
<td>TTTTATGGAATTATTTCTTTGGAATTATGAAACTTACATTTTA</td>
</tr>
<tr>
<td>4[405]0[392]</td>
<td>GACAGTACATAATGTCCATCAGCAGCCAGGATTGAAACTTACATTTTA</td>
</tr>
<tr>
<td>6[356]13[342]</td>
<td>TAAGAACGCTTACCTTACACAGTGGTGCTGTTATCTAAAATAAC</td>
</tr>
<tr>
<td>6[559]9[559]</td>
<td>CACCGACGGAATACGAAAATTATCATATGGCAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG</td>
</tr>
<tr>
<td>8[209]11[209]</td>
<td>TATCATCGCTTGGAAATTATACCAAGCAGCCTTTTTTAGCGGAGG</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>8[377]16[364]</td>
<td>CCAATTATTTTCGTTTTACGACGCCACC</td>
</tr>
<tr>
<td>10[312]20[294]</td>
<td>AGTACGGTGTTGCAAGTAATGCTTACACCTCA-GAAATCATGA</td>
</tr>
<tr>
<td>10[573]6[560]</td>
<td>TGTCACAAATCAATACCAAAGAAGACTGGCTCCTTATTACGGGT</td>
</tr>
<tr>
<td>19[301]21[316]</td>
<td>AATGCAATCCATAGCATAAGCAGTAGTA</td>
</tr>
<tr>
<td>22[370]25[370]</td>
<td>AATAGCATTAAGTACAGAAGCGCAGCGCCAAAAATAATATCAAC</td>
</tr>
<tr>
<td>25[287]23[300]</td>
<td>AATGGTCTCTGCGAAGAGGTCTTTTTT</td>
</tr>
<tr>
<td>25[371]22[371]</td>
<td>GACAATATCAGCTAAACAACCTGTTTTTTTTTAACAAAGAA</td>
</tr>
<tr>
<td>0[181]3[181]</td>
<td>TGCCTAATCAGCCTACCGAGCTGAAATAATTTAATCCCA</td>
</tr>
<tr>
<td>0[475]3[475]</td>
<td>ATATGGAAACTATATTTTGCGGAACAATATATATCAGGA</td>
</tr>
<tr>
<td>1[196]4[196]</td>
<td>GTGCGAAAGGGGAAGGTTTTTTCCACTACCCACC-CGCCAGGCTG</td>
</tr>
<tr>
<td>3[140]1[153]</td>
<td>GCCAGCTGCAATTAGAGTCCGGGAAACCTGTGACTGCCAGCCT</td>
</tr>
<tr>
<td>3[266]1[279]</td>
<td>CGTCTGGCCTTCGTGCCATCAAAATAAGTTTTAACCAATAG</td>
</tr>
<tr>
<td>3[350]1[363]</td>
<td>TATATGGAAATGAGGTTGTTGAGGTTATATATATATTTAACCCTCC</td>
</tr>
<tr>
<td>3[434]1[447]</td>
<td>TTAGGTTTTTACAGAAAGAATTTGCTACTCAGTAACGAAACAG</td>
</tr>
<tr>
<td>5[336]2[336]</td>
<td>GCCAGAAACCAGTAATGTACCCTAAACTACCTACATTTTCAG</td>
</tr>
<tr>
<td>6[181]9[181]</td>
<td>AAACACCCGCGATGCTAGGTGGCTGGATATTACATTACAC</td>
</tr>
<tr>
<td>6[223]9[223]</td>
<td>CATGTTACAAATATTACACTAAACACTGAAACAAGTACAA</td>
</tr>
<tr>
<td>6[475]9[475]</td>
<td>GGAAAGCATCTTACCAGCCACCAGAACAGTTGAGGCAAGTCA</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>9[518]7[531]</td>
<td>ACGTCACCAATGACCCATTAGCAAGGCTTTACCAGTAGCACCC</td>
</tr>
<tr>
<td>14[398]11[391]</td>
<td>GACGTAGTAAATGAATTTCGTAACTGATCTAAAAAGCAAAATGAGC</td>
</tr>
<tr>
<td>20[349]9[349]</td>
<td>GGTATTACCGACTCATCGAGAGGCTAATTTGCTAAACAG</td>
</tr>
<tr>
<td>20[356]18[343]</td>
<td>AAGACCGGAGGTTCCTGAATCACCT</td>
</tr>
<tr>
<td>20[377]7[370]</td>
<td>GCTGTCCTAGATTGCACCCAGCTACGCATCGT</td>
</tr>
<tr>
<td>13[323]14[326]</td>
<td>TAATGCGGGAAGTGAGAATAG</td>
</tr>
<tr>
<td>13[343]23[349]</td>
<td>GCCTGTAGCGAGGCTTAACGATTTTTGAGAGAATAACA-GCAAGC</td>
</tr>
<tr>
<td>0[391]3[391]</td>
<td>AATAGTGAGAGTGCTCAATATATGAGTAATTAAATTTCCTT</td>
</tr>
<tr>
<td>3[182]1[195]</td>
<td>GGTCGACTCTAGCAACAGGGCTTGCTGACCGGAACGACCGC</td>
</tr>
<tr>
<td>5[168]2[168]</td>
<td>GTACGCGGTGAGTGAAGAGGTAAAGCGCTTGGCTACAAATTCC</td>
</tr>
<tr>
<td>5[504]2[504]</td>
<td>ATTACCGTAGAAGTTCAATAGATATACGATCTTTAGGAGCA</td>
</tr>
<tr>
<td>10[153]6[140]</td>
<td>TACTGGGCTACTTAGATTGCTTTATCCCCTGATGAGATGGTTAC</td>
</tr>
<tr>
<td>11[210]8[210]</td>
<td>AAGACTTCTTAGCGGAAATCCGGGACCTTTACCAACGAGATTGT</td>
</tr>
<tr>
<td>14[310]11[312]</td>
<td>GGAATTGCGAATAAATGATATTCAACCGAGTAAAATATGCAACTAA</td>
</tr>
<tr>
<td>15[301]9[314]</td>
<td>CATATGCGTTTACCGCCTAACACATCAAT</td>
</tr>
<tr>
<td>Label</td>
<td>Sequence</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
</tr>
<tr>
<td>15[350]20[357]</td>
<td>AGGGCTTCCGCCACCCATATTATTTAAGCTTTCCAACAAGCATCATTCC</td>
</tr>
<tr>
<td>16[349]15[349]</td>
<td>CCTCAGAAACGCCACACGCTCAACAGT</td>
</tr>
<tr>
<td>23[357]10[350]</td>
<td>GGGAGAAGGCTTTATTTAAGCTCAAAAGCCATCCTTCCAATCAGA</td>
</tr>
<tr>
<td>0[223]3[223]</td>
<td>AGTTGGGCGCTACAGGCAAGCGCCATTCTGATCGGTGCGGGG</td>
</tr>
<tr>
<td>1[322]4[322]</td>
<td>AGCAATAGGAAACCCCAAGGCTTACGTCATTGACAGGAGAGAAAA</td>
</tr>
<tr>
<td>2[419]5[419]</td>
<td>ACAAAATCGCGCACAATAACGGATTCGCTTTTACTTTGAGC</td>
</tr>
<tr>
<td>3[518]1[531]</td>
<td>CTCAAAATATCCAAAGGCATCACCCTGCTATCAAATGAAAAAT</td>
</tr>
<tr>
<td>4[153]0[140]</td>
<td>ATAATGAATCGGCCAGGAGCGGGCTAATTGCGTTGCGCCA</td>
</tr>
<tr>
<td>4[237]0[224]</td>
<td>CATGCCGGAAACCAGGGCGCGTACTATGCGACAGTATCGGTA</td>
</tr>
<tr>
<td>5[126]2[126]</td>
<td>GGGAGAGGTCACCGGCCCTGAGAGATTTGCTTGAAGAGACGGGCA</td>
</tr>
<tr>
<td>7[154]10[154]</td>
<td>AACATTCAGTGAATGTAAATTTGGGCAGTTAAATCAGAGAG</td>
</tr>
<tr>
<td>7[371]14[363]</td>
<td>AGGAATCTGGACGAGCTGCTTACGATCGGTGCGGG</td>
</tr>
<tr>
<td>8[167]11[167]</td>
<td>AACAAAGCTGCTTGACAAGAAGCAGCTTTACATTGTACAG</td>
</tr>
<tr>
<td>9[140]7[153]</td>
<td>CGATTTTAAGAAATATTGTGAAATTACCTCAAAATTTCACCTTT</td>
</tr>
<tr>
<td>9[224]7[237]</td>
<td>CCTAAAAACGAAAAACCTACGAAGGCAGCGCCTGAAAATACGT</td>
</tr>
<tr>
<td>9[252]16[257]</td>
<td>GGATCGTCACCCACTAATAAGA</td>
</tr>
<tr>
<td>9[434]7[447]</td>
<td>TAATAAGTTTTACCGATACAGGAGGTATTATACATCGCTTT</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>11[462]8[462]</td>
<td>AATAGCTGCAGTCTCCTCATTAAAGCCAGCCAGACGATTGGC</td>
</tr>
<tr>
<td>14[362]6[357]</td>
<td>TAAACAACAGACAGCCCTCCTATTATCCGATATTC</td>
</tr>
<tr>
<td>15[257]9[272]</td>
<td>ATAAACACCGGAATCAAAGGCGTAACCATC</td>
</tr>
<tr>
<td>15[326]15[325]</td>
<td>AGTATAAAAGCCCTTAGTTAATTTATCTTTCTAAATTCCTTACC</td>
</tr>
<tr>
<td>16[279]9[293]</td>
<td>GATAAATTAATTACTAGAAAAAGCCTGAATACCGGGGCGCG</td>
</tr>
<tr>
<td>0[265]3[265]</td>
<td>CCGTGCAAACGTGCTTCATCAACATTAAGTGGAACAAACCAG</td>
</tr>
<tr>
<td>0[349]3[349]</td>
<td>TTTCTGTCTCTGTGACGAAACGAAAGCACAGAGACCCACAC</td>
</tr>
<tr>
<td>1[406]4[406]</td>
<td>ACAGAGCGAATTATGAAACAAACATCATTAACCGATCGGGA</td>
</tr>
<tr>
<td>1[490]4[490]</td>
<td>CCAGATTAGAGCCCGATTAGACTTTTACAATAATCCACACAGTT</td>
</tr>
<tr>
<td>2[335]5[335]</td>
<td>GGATAGCAAAGCGCAGTTCAATATATATTATGAGCGACGATAC</td>
</tr>
<tr>
<td>3[476]1[489]</td>
<td>TTTAAAAGTGGTGAATGCCCAGACGGGTATCTCTCGTATTTAAAT</td>
</tr>
<tr>
<td>4[279]0[266]</td>
<td>ATCTGTAGGCAGCTTTTCCTCGTTAGAAATCAAGCTCTTAAA</td>
</tr>
<tr>
<td>5[252]2[252]</td>
<td>CACGTATTCTGCCCATAGATGGGCAGTCCTTGGCGGATTTGACC</td>
</tr>
<tr>
<td>6[384]8[378]</td>
<td>TCAGATAATTACGCAGCCCAAAATAGGTGATCAA</td>
</tr>
<tr>
<td>6[517]9[517]</td>
<td>CATAAATCCAAAGGGGACCGTCATCCAGACTGTAGCGAA</td>
</tr>
<tr>
<td>8[251]11[251]</td>
<td>GACGACATCGGATGGGCGACTTTGGCGGAATTGAGCGGA</td>
</tr>
<tr>
<td>8[545]11[545]</td>
<td>AATATTTGACGGATAAAAAGGGGACATTTTTACCAACAGC</td>
</tr>
<tr>
<td>9[98]7[111]</td>
<td>GACGACGATAAACGCGATAACCCTCGTTTCCGTAAGAGCAACA</td>
</tr>
<tr>
<td>9[560]7[573]</td>
<td>AAACGCAAAGACAGGGGTGCAAACATATACCCAGTATGTAGGC</td>
</tr>
</tbody>
</table>

Continued on next page
Table A.1 (Continued)

<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>10[489]6[476]</td>
<td>CACCACCCCTCAGAGCGAAGCCCTTTTTAACCACCGGAAACCAT</td>
</tr>
<tr>
<td>11[84]8[84]</td>
<td>CCCCCCTCGAATCGTATGACTGGATAGCGTACTAATAGTAAAAT</td>
</tr>
<tr>
<td>13[287]13[286]</td>
<td>GCTTGATTTTTTTCACTGGAAAATTCGATTCGCGTTTATCA</td>
</tr>
<tr>
<td>18[309]19[300]</td>
<td>CCTGAGTAATGTGGTATTTTTA</td>
</tr>
<tr>
<td>2[209]5[209]</td>
<td>TTACGCGCAGCTGGCACTTTGGGAAGGCGCCATTCGCGCTT</td>
</tr>
<tr>
<td>3[224]1[237]</td>
<td>CACCGCTTCTGGACCTCCAGCCACCTTTATCTCAGGAAGAT</td>
</tr>
<tr>
<td>5[294]2[294]</td>
<td>AAACAGTTTCGCAATTTGAAACGTTAATACCTAAAAACAGGAA</td>
</tr>
<tr>
<td>5[462]2[462]</td>
<td>AGAAGTTACTCAAGGGTTTTTTTGGATTATATTTCATCTGATGAGC</td>
</tr>
<tr>
<td>6[97]9[97]</td>
<td>TACTGCGAATGCTGAGGCTTTTGCATGCCAGGAACTCAA</td>
</tr>
<tr>
<td>6[272]14[257]</td>
<td>AAACATGTTTAATTCAAAAAAAGGCTCCA</td>
</tr>
<tr>
<td>7[238]10[238]</td>
<td>AAACGAGGGTAGCTATTTCATGAGGAAGTACCTTCACAGGAG</td>
</tr>
<tr>
<td>7[532]10[532]</td>
<td>ATAATTATTCATTAAATTTGGGAATTAGACCCAGGAGCAGCAA</td>
</tr>
<tr>
<td>8[314]15[300]</td>
<td>CTAATTTAATCGGTAAAGGTTAATTTGAGTATTAGTATTAT</td>
</tr>
<tr>
<td>8[419]11[419]</td>
<td>TGAATGAATTAATATGCCCCCCTGCTATGCCCGTCACAGAAT</td>
</tr>
<tr>
<td>10[111]6[98]</td>
<td>TAAACAAAAATAGCTTAAACGTTTACATAGGAGGCATAAAA</td>
</tr>
<tr>
<td>10[195]6[182]</td>
<td>AGTACAGACCGGCGCCTCAAAAAGATTACGGTAGCTACAGTAG</td>
</tr>
<tr>
<td>10[293]22[275]</td>
<td>ACCTGTTGGAACATAGCTCGAGTACCTTTTAATT</td>
</tr>
<tr>
<td>11[420]8[420]</td>
<td>TGAGTTAGGGTTTTCTCAAGAGAAGGAATTTATTCTGAAACA</td>
</tr>
<tr>
<td>16[363]15[377]</td>
<td>CTCAGAAAATTTGAGAATCCGCCCATAATT</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>19[364]21[379]</td>
<td>CTTAATTTTCCTTAAGCCGTTTTTCAATAAT</td>
</tr>
<tr>
<td>1[364]4[364]</td>
<td>GGATAGCTTGGATTTTCAATAATCAGGTTCTGACTTTGCTT</td>
</tr>
<tr>
<td>2[167]5[167]</td>
<td>ACACAACATACGCGCTAAATCGTGCTATTGAATAACCGGCCACC</td>
</tr>
<tr>
<td>2[377]5[377]</td>
<td>GAAACATAGCGCTAAATCGTGCTATTGAATAACCGGCCACC</td>
</tr>
<tr>
<td>4[447]0[434]</td>
<td>GGTCAGATGAAATATAGTAATAAACATCACAATACCCATATTATTTGGA</td>
</tr>
<tr>
<td>5[378]2[378]</td>
<td>GAGTAAAAATTATGAGTGGCGCTGACGAGCCCTTTAGATAATCCTT</td>
</tr>
<tr>
<td>5[420]2[420]</td>
<td>AATACTTTAGATGATTTCTATTCAATTACGAGAATACCCATTTG</td>
</tr>
<tr>
<td>6[433]9[433]</td>
<td>GATTAGCAGCCCAACCTGAGTAAACAGGTTCTTCCGAACCTATGG</td>
</tr>
<tr>
<td>7[406]10[406]</td>
<td>GAGAGGCTGAGACTTTGCTAGTACAGGCTAGGCAACCATAAAACAG</td>
</tr>
<tr>
<td>9[182]7[195]</td>
<td>AGATGAGAAGGCTGCGCAAACCTTTGAGGACGAGGGAACCGGAAGGACAGAAGGCA</td>
</tr>
<tr>
<td>11[168]8[168]</td>
<td>AAGCAGAAGCAGAAGGCTGAGCAGCCAAATCAACG</td>
</tr>
<tr>
<td>19[273]20[275]</td>
<td>AAAATTTTTAGAAACGGCGAAAGATTTTGAATGCTATACATCGGCATACAGAAGGCTGAGCTGTACGCTGAGCAGAT</td>
</tr>
<tr>
<td>0[559]3[559]</td>
<td>TTAACACTTTTTGACTTTTTGATGCTATACATCGCCATACAGAAGGCTGAGCTGTACGCTGAGCAGAT</td>
</tr>
<tr>
<td>1[532]4[532]</td>
<td>CTGAAGATAAAAACCAACAGTGCGACGCTAGCTACGCTACGCTAATGCG</td>
</tr>
<tr>
<td>2[545]5[545]</td>
<td>GAACCACCACAAACTGATACGCCCCTAAATAGTCTTTGGAAAT</td>
</tr>
<tr>
<td>3[560]1[573]</td>
<td>AACCCCTTTGACCTTCTGGCCAAACAGAGGAGATTCACCAGTCA</td>
</tr>
<tr>
<td>4[573]0[560]</td>
<td>GGCACAGACAATACGTCAATTAGCTAATTGAGGACAGAGGAGATTCACCAGTCA</td>
</tr>
<tr>
<td>5[546]2[546]</td>
<td>ACCTACAGCCCTGCGAGGGTGAGGGCGGTCATAAAATACGGAAC</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0[97]3[97]TCy3</td>
<td>AGGGCGAGACGGGGAGATAGGGTTGAGTTTTGGAACAA-GACCT/3Cy3Sp/</td>
</tr>
<tr>
<td>1[112]4[112]TCy3</td>
<td>GGCTTCACCGCCTCTGGTTTGCCCCAGTGGCAGTGCG-TATT/3Cy3Sp/</td>
</tr>
<tr>
<td>2[125]5[125]TCy3</td>
<td>ACAGCTGATTGCTTCGCCAGGTTGTTTGGCGGTTAAAG-GAAT/3Cy3Sp/</td>
</tr>
<tr>
<td>3[98]1[111]TCy3</td>
<td>CTTATAATCAAGGGGGAATCGGCAAATCCTGTTTGATG-GTT/3Cy3Sp/</td>
</tr>
<tr>
<td>4[111]0[98]TCy3</td>
<td>TGAAGAATAGCCCGAAGGCCGGCAGCCAGGCGAAAAT-CAAT/3Cy3Sp/</td>
</tr>
<tr>
<td>5[84]2[84]TCy3</td>
<td>AGAGCTTAAAACCGTGGACTCACAACGATGTCCACTAT-TAAT/3Cy3Sp/</td>
</tr>
<tr>
<td>Rotor_Ext1</td>
<td>/5Phos/ACACAACCTGAGGAAAAAGGAACAACCTAAA</td>
</tr>
<tr>
<td>Rotor_Ext2_oh</td>
<td>TTCTAGCTGATAAATTTCCCTCAGTGTGTCTCGGACTAGT</td>
</tr>
</tbody>
</table>
Table A.2: DNA oligomers for origami base. DNA modifications (biotinylation and phosphorylation) are included using their IDT codes. The six strands ending in ‘TThr21’ contain a 21-nt binding site (red) for the Hr21_5Bio biotinylated secondary strand. The final three strands form the adaptor for ligating additional DNA to the anchor using the 12-nt overhang (purple). Two regions of complementarity between these adaptor strands are indicated in green and orange. Note Anchor_Ext0_oh is replaced by Anchor_Ext0_oh_direct (Table A.3) for the anchor-rotor complex linked by a 52-bp dsDNA.

<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1[70]4[77]</td>
<td>AAAGCGTAAGAATAATAGCCCCACCAGCCCACGCT</td>
</tr>
<tr>
<td>1[105]26[112]</td>
<td>GAATGCCATTGCAATCACCAGGTGCTGCTTACGCC</td>
</tr>
<tr>
<td>1[119]4[112]</td>
<td>CAGGAAAACAATATATTAGTATAACCGTTGTAGCA</td>
</tr>
<tr>
<td>1[133]26[126]</td>
<td>ATGGAAAGGATTATGGGTAACTCGGTGCGGGGCTTC</td>
</tr>
<tr>
<td>2[104]23[118]</td>
<td>TAGTCTTAGGTAGGCGGTACTTCTTTGTACCGCTCTAGAG</td>
</tr>
<tr>
<td>3[70]24[63]</td>
<td>AACGAACTAAAACAATAACTATTTTTCAATATATT</td>
</tr>
<tr>
<td>3[126]24[133]</td>
<td>ATAACATCACTTGTGGTAATGAGCTCGTTGTGAAA</td>
</tr>
<tr>
<td>4[76]22[63]</td>
<td>GAGAGCCATTAAGACCTTTTTAACCTCC</td>
</tr>
<tr>
<td>4[104]5[118]</td>
<td>TTAACACATCAAAACCCTCAGGTACGCCCA</td>
</tr>
<tr>
<td>5[70]8[77]</td>
<td>TAAAGCATCACCCTAACCAGTTTTTAGGATAGAAGT</td>
</tr>
<tr>
<td>5[119]8[112]</td>
<td>GAATCTGATTTTACGGACACAGGGCGGTACTAT</td>
</tr>
<tr>
<td>6[104]9[118]</td>
<td>TATCTGGTAGATTAGATAATATATATTTACACCGGGAAC</td>
</tr>
<tr>
<td>6[118]4[105]</td>
<td>GACAGGACTGTGTACTGCCCCGCTTTCAACATACATCAGTA</td>
</tr>
<tr>
<td>7[70]20[63]</td>
<td>AAATATCGAAAGGAAATATATTAGAATCCTTGAAG</td>
</tr>
<tr>
<td>7[91]17[104]</td>
<td>CAACTAATCAGTGGTGCTTTGTAATCAATCTAATTAGT</td>
</tr>
<tr>
<td>9[70]18[63]</td>
<td>ACTCGTAAATTAGACGATGTACCTTTTTAATG</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>9[77]12[63]</td>
<td>TTAAATCGGAAACAATATTTCCCTCTTCTGAATAATGGGAAGGGTT</td>
</tr>
<tr>
<td>10[90]18[91]</td>
<td>ATTTTGGCTTTGCGCCGAACGTTACATTAAAGCAAAATTTAAC</td>
</tr>
<tr>
<td>10[132]18[133]</td>
<td>GATTAGGAAAGGAAGGAACATATGCACCGCGGCGCGCTGTGCGTAT</td>
</tr>
<tr>
<td>10[139]16[126]</td>
<td>AGCCCCCCTTATAAGTCCACGCTGGTTT</td>
</tr>
<tr>
<td>11[70]16[63]</td>
<td>ATCATCAAGAAACCACAATAAATTATTCATTTCAGA</td>
</tr>
<tr>
<td>12[90]15[97]</td>
<td>GATTATAGATTATCGAGATGATCATTATCTGCTTTG</td>
</tr>
<tr>
<td>12[97]15[90]</td>
<td>TTGTTTGAAATTGCAGGTTTAAACGTGAGGCTGAT</td>
</tr>
<tr>
<td>12[111]16[112]</td>
<td>AAATATACACATCGTGTGTGTTGCGGGGAAGTTCCGACAGGCGA</td>
</tr>
<tr>
<td>13[70]14[63]</td>
<td>GCACGTTCAGTACAGTAGTAC</td>
</tr>
<tr>
<td>14[111]12[98]</td>
<td>AGTCCACGATGGGTGAAGTTTGAGTAAGGCAATTATCCTGAT</td>
</tr>
<tr>
<td>16[83]13[90]</td>
<td>GAGCGCAGGATTCATGAATAAAACAGAAATAAG</td>
</tr>
<tr>
<td>16[97]7[90]</td>
<td>TACAAAAATCGCGCACATCAAGGAGGATTGCACTAA</td>
</tr>
<tr>
<td>16[104]13[111]</td>
<td>GTCAAGTAATAACTTTATTTTGTCGATGATTTAAGAAC</td>
</tr>
<tr>
<td>16[111]7[118]</td>
<td>AAATCCTGAGACGGGGGCTCAAGAGCCTCGTGGTTG</td>
</tr>
<tr>
<td>16[125]13[132]</td>
<td>GCCCCAGAATCGGCGCAGTTGTGCCAAAGCGTCAAAGGG</td>
</tr>
<tr>
<td>17[112]6[105]</td>
<td>GCAACAGGTTTTTCTTTTTCACGGGAAACACATCAA</td>
</tr>
<tr>
<td>18[90]4[91]</td>
<td>AATTGCAAATAAGCCAAATCGCTGAACCCTCAAATCGGCCTGC</td>
</tr>
<tr>
<td>18[132]4[133]</td>
<td>TGGGCGCGCATTAATTAAGGAGAGTTGTTTATCCATCA</td>
</tr>
<tr>
<td>20[90]17[83]</td>
<td>AATTAATTTTCCCTGAGTGTTTGAATAAACAAA</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>21[84]23[97]</td>
<td>AAGAGTCAACAGTGAGAAGATAAAACAGTAATGCGTGCAAAT</td>
</tr>
<tr>
<td>21[91]1[90]</td>
<td>AATAGTGAATCATAGGTCTGAATGCTGACGAAACTGCGTGGA</td>
</tr>
<tr>
<td>21[112]2[105]</td>
<td>GAGCCCGGCGCTCAACAATTCGGTGACAGGCTAT</td>
</tr>
<tr>
<td>21[133]1[132]</td>
<td>AAGCCTGTTCCTGTGTGAAAGGTACCATCAGAAACGCTC</td>
</tr>
<tr>
<td>23[98]20[91]</td>
<td>CCAATCAACATCAAAATTTACAGATCGTGCTATT</td>
</tr>
<tr>
<td>24[97]21[83]</td>
<td>GACAAAGAAGCGAGAGAAAACCTTATCTGAGAGACTACGCTGAG</td>
</tr>
<tr>
<td>24[118]0[105]</td>
<td>CAAGCTTGCATGCCGGGGGATTCACACG</td>
</tr>
<tr>
<td>24[125]4[126]</td>
<td>CCAGTGCGATCCCCCTTGTATAAGCATAAAGGCTACGCAAAT</td>
</tr>
<tr>
<td>24[132]52[133]</td>
<td>ACGACGTGATTATTTACATTATAATTGATCGTGTAAGAG</td>
</tr>
<tr>
<td>25[84]50[91]</td>
<td>TGATAAAGGACATTTCACGGTGAAATCAAAGGCATATGGTCAACATA</td>
</tr>
<tr>
<td>26[97]51[111]</td>
<td>ATAAACCAACGCTCAACAGTAACCAACAGCTGCAATGGGCTCCGATATC</td>
</tr>
<tr>
<td>27[63]30[70]</td>
<td>TATAAAATTCTTTATAAAACAAATTTGTAAGTA</td>
</tr>
<tr>
<td>27[77]53[90]</td>
<td>CCACTTAAGCCACCGGAATAAAATTCGACATTCGAATT</td>
</tr>
<tr>
<td>27[112]47[111]</td>
<td>AAAGCGCCCGCTTCTGCCAGTGATGGGCACGATCGTGATCATT</td>
</tr>
<tr>
<td>28[97]49[111]</td>
<td>TTAATTGGAGCGGCTTAACCGCTGTGCATCTGGTTCAGCTC</td>
</tr>
<tr>
<td>28[139]30[126]</td>
<td>ACTCCAGACAGTATGGGATAGGTCACGT</td>
</tr>
<tr>
<td>30[97]31[111]</td>
<td>ATATAAACGTGGTTATCAACCGAGTAAC</td>
</tr>
<tr>
<td>31[63]34[70]</td>
<td>CAACATGGTCAGCTTAATCTTGACGAGGTATTA</td>
</tr>
<tr>
<td>31[112]43[111]</td>
<td>AACCCTTGTCGACATCATTTTGCTAATTTGGTTCCAAA</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>31[126]34[133]</td>
<td>TCCGTGGGAACAAACCTTCCCTGCCATCAGTTAATA</td>
</tr>
<tr>
<td>32[97]45[111]</td>
<td>ATAAAGTCTCAATAATCGGCAAATCAGCTTTAATGTGCATTTTT</td>
</tr>
<tr>
<td>34[97]35[111]</td>
<td>TTCCTTAAGCCGTTTTTATTCAGAAAAAG</td>
</tr>
<tr>
<td>34[125]46[126]</td>
<td>AAAATTTCTAAACCAATAGGAACGTAGCCAGCTGAAAAAATGGT</td>
</tr>
<tr>
<td>34[125]46[126]</td>
<td>AAAATTTCTAAACCAATAGGAACGTAGCCAGCTGAAAAAATGGT</td>
</tr>
<tr>
<td>35[70]38[63]</td>
<td>CGCACTCCCAATAGAGATATACCTTGTGGGAGGGTTT</td>
</tr>
<tr>
<td>35[112]39[111]</td>
<td>CCCAAAGTACCCCTGAGAATCTACAAAGGTCTACGTCA</td>
</tr>
<tr>
<td>36[97]41[111]</td>
<td>TCGTAGGTATTCTAAGACACGTCATTGGGTTGTAGACGCA</td>
</tr>
<tr>
<td>37[84]39[97]</td>
<td>TATCCGGAATCATTTTTATTTATCCCAATCCCAAGCGAGCTACA</td>
</tr>
<tr>
<td>37[126]43[125]</td>
<td>GCAAACACATGTGCTAAGAAACCCTTTTGCGGAGAAATAAAGC</td>
</tr>
<tr>
<td>38[139]40[126]</td>
<td>GAGGCTAGCTATTTTTGATATTTTCAAAAG</td>
</tr>
<tr>
<td>39[70]37[83]</td>
<td>ATTAGGGTCTATTTCCTCCGAGAAAGGCT</td>
</tr>
<tr>
<td>39[98]40[84]</td>
<td>ATTTTTATCAGCGAGGCTTTTTCAGCAACTGACCTAAGAGG</td>
</tr>
<tr>
<td>39[112]37[125]</td>
<td>AATCACCATCAATATTGAGGAGGTCTGGA</td>
</tr>
<tr>
<td>40[83]34[84]</td>
<td>CGTCTTTAGCCATAACCGCGCATCAGAAAACAGCATCATTCC</td>
</tr>
<tr>
<td>40[125]34[126]</td>
<td>GGTGAGAAAAATTTTTACATATAACAGGAAGATTGTTTTTGT</td>
</tr>
<tr>
<td>41[112]36[98]</td>
<td>AGGAATAAAAAGCGCGGAGACTGAAATCTTACAAATAAATTCCA</td>
</tr>
<tr>
<td>43[70]40[63]</td>
<td>AACATAAAAATAGCAATAAACCCAGAGCCTAATT</td>
</tr>
<tr>
<td>43[84]31[83]</td>
<td>GAAGCGCCAAAGGTCAGAGGTTAAGAATAGAAAAAAATGCAG</td>
</tr>
<tr>
<td>43[112]34[98]</td>
<td>TTAAGCAGCCCTTTATTTCAAAAACGATTTGAGGAGATTTGTCT</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>43[126]41[139]</td>
<td>CTCAGAGCATAAAGTGTAATAACTCATAT</td>
</tr>
<tr>
<td>44[132]31[125]</td>
<td>ATCATAAGGCAAGGGCGAGCTTTTCACGGATTC</td>
</tr>
<tr>
<td>45[91]42[84]</td>
<td>GCCCAATCCCTGAATTTAGACTTTTGTTAACGTCA</td>
</tr>
<tr>
<td>45[112]32[98]</td>
<td>CATTGGGCAAAGAATTAGAACTGAAACAATAAGAGAATAG</td>
</tr>
<tr>
<td>46[83]28[84]</td>
<td>CTATCTTAGCCGAACAAAAGGGCAGAGGCCATTTTCAGAATCG</td>
</tr>
<tr>
<td>46[125]28[126]</td>
<td>CAATAACACAGTGTGTTGTTGTTGATTGGGACGACGCCAGCTT</td>
</tr>
<tr>
<td>47[70]44[63]</td>
<td>AGCAGATAACCGAAGAACCCACAATTGAGCGCTAAT</td>
</tr>
<tr>
<td>47[112]30[98]</td>
<td>CCATATACTGTTTAGCTATAAGAAACAAGGAAAACCAAGAGA</td>
</tr>
<tr>
<td>47[126]44[133]</td>
<td>ATTCACATCTGCATTTCGCGAGGGCCAAATA</td>
</tr>
<tr>
<td>48[125]27[139]</td>
<td>AGTACGGCGGATGGTCGGCAATTCGCCATTCAGGCTGCGC</td>
</tr>
<tr>
<td>49[84]25[83]</td>
<td>AGCAAACCTAAAAGAAACGCAATCAATAGCATAATTGACCTG</td>
</tr>
<tr>
<td>49[91]46[84]</td>
<td>GTAGAAAATAATAAACAGAAATGAAATGAAATAG</td>
</tr>
<tr>
<td>49[112]28[98]</td>
<td>ATTTTTGTCTGTGAAGTGGAAACGCAATACATAGGAGGGC</td>
</tr>
<tr>
<td>49[126]25[125]</td>
<td>CTTAGAGTACCTTTAATTTGCTAATTTCGATTGGCTAAGGCGA</td>
</tr>
<tr>
<td>51[70]48[63]</td>
<td>GTCACAAGACACCCTACCGCAGAACTGGCCATGATT</td>
</tr>
<tr>
<td>51[112]26[98]</td>
<td>GCCGTTTTCTTTTTGATAAGTAAAGGGTGTTACCAGGATAAG</td>
</tr>
<tr>
<td>51[126]53[139]</td>
<td>GCTTCAATTAAAGAGCCCTGACTATTATA</td>
</tr>
<tr>
<td>51[133]48[126]</td>
<td>AGCGAATAGAGAGCCTAATATAATATGCAACTAA</td>
</tr>
<tr>
<td>53[91]1[104]</td>
<td>CACCGAATGACCATAAAGACATCCAAAACCGATATTATTTT</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>54[90]75[104]</td>
<td>GACTTGACAAGGCCGGAAGCCAGAGGGGAAAACGGCGCTACTACGTT</td>
</tr>
<tr>
<td>54[118]80[119]</td>
<td>CTTTAAAAAATCAGGTCTTTTGAAGGCCGGAGCCTGGCATGAT</td>
</tr>
<tr>
<td>55[105]59[118]</td>
<td>GGTATATTGCAAAAAAAGGATCAACTAGAAAAATCTACGTT</td>
</tr>
<tr>
<td>55[126]57[139]</td>
<td>CTGGATAAACCAATAAGAGCAACACTA</td>
</tr>
<tr>
<td>56[90]73[104]</td>
<td>CCAATGAAGCGTCAGACTGCATAACGCGAGAAGTCCTTTCCACA</td>
</tr>
<tr>
<td>56[118]75[125]</td>
<td>GAGGCTTGTAAAAATGTTTAGATCAAATGGACAATG</td>
</tr>
<tr>
<td>57[119]75[118]</td>
<td>GCCATAGATAGCGATGCGGGACTTGCAGGGAGTTATTGCGC</td>
</tr>
<tr>
<td>58[90]71[104]</td>
<td>CGTTTTCTCAAAATCCACCGCGTTGGGAATGCGAGCACCCTCAT</td>
</tr>
<tr>
<td>59[126]61[139]</td>
<td>CGAACTATCGGATTCCTTGAGATGGTTTA</td>
</tr>
<tr>
<td>60[90]69[104]</td>
<td>AGAGCCAAGCCACCACCTGAGAAACACCGACTCGCAGCCGAA</td>
</tr>
<tr>
<td>60[118]64[105]</td>
<td>CTGGGTCAATTATACCAAGAAGTAAGGCTACTGGCTTAGGCAA</td>
</tr>
<tr>
<td>62[97]64[84]</td>
<td>ACCAGAGCCGCCACAGACGATATTAAG</td>
</tr>
<tr>
<td>63[70]68[77]</td>
<td>GGTTGAGCAGGTCCAGAAACATGATTGCTCA</td>
</tr>
<tr>
<td>63[98]65[111]</td>
<td>GATATTCTGCCCCCTGTGAGAAGCTCCCTACAGAAGCGGKTGTTTC</td>
</tr>
<tr>
<td>63[126]66[119]</td>
<td>GTAATCTACCGGTGTACAGACCTCAATCACTACGGCC</td>
</tr>
<tr>
<td>64[83]66[63]</td>
<td>CCAGAATTGATGATTGCCGCTATAACAGTGAATG</td>
</tr>
<tr>
<td>64[104]67[97]</td>
<td>ACAAATAGTAATAAGGTCAGTGCCCTTGAAGAGGC</td>
</tr>
<tr>
<td>65[70]64[63]</td>
<td>TGGCTTTGGGAAACGCAGCTCT</td>
</tr>
<tr>
<td>65[112]62[119]</td>
<td>CAGACGGAGGCGACCTTCATCAAGAGGCTCATT</td>
</tr>
<tr>
<td>Start</td>
<td>End</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>66[118]60[119]</td>
<td></td>
</tr>
<tr>
<td>66[139]62[126]</td>
<td></td>
</tr>
<tr>
<td>67[70]60[77]</td>
<td></td>
</tr>
<tr>
<td>67[119]69[118]</td>
<td></td>
</tr>
<tr>
<td>68[76]58[77]</td>
<td></td>
</tr>
<tr>
<td>68[90]63[97]</td>
<td></td>
</tr>
<tr>
<td>68[139]66[126]</td>
<td></td>
</tr>
<tr>
<td>69[105]68[91]</td>
<td></td>
</tr>
<tr>
<td>69[119]57[118]</td>
<td></td>
</tr>
<tr>
<td>69[126]59[139]</td>
<td></td>
</tr>
<tr>
<td>70[139]58[126]</td>
<td></td>
</tr>
<tr>
<td>71[70]56[77]</td>
<td></td>
</tr>
<tr>
<td>71[105]58[91]</td>
<td></td>
</tr>
<tr>
<td>72[90]60[91]</td>
<td></td>
</tr>
<tr>
<td>72[139]70[126]</td>
<td></td>
</tr>
<tr>
<td>73[70]54[77]</td>
<td></td>
</tr>
<tr>
<td>73[105]72[91]</td>
<td></td>
</tr>
<tr>
<td>74[139]54[126]</td>
<td></td>
</tr>
<tr>
<td>75[70]52[77]</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>75[105]54[91]</td>
<td>AACTTAGAACCCTCATAAGTTAGCCGTAACGGATTTTAGTCACC</td>
</tr>
<tr>
<td>75[126]56[133]</td>
<td>ACAAACAGCTGAGGTCGCCCTCAATCAGCGAGTAA</td>
</tr>
<tr>
<td>76[90]56[91]</td>
<td>GAGAATACTGTATGGATCTAAACACGCCTGATAGCATTTCGTCAC</td>
</tr>
<tr>
<td>76[118]54[105]</td>
<td>GATACCGTCCAAAAGAAAGACTTCAAACCAATCAACAGTTCA</td>
</tr>
<tr>
<td>76[139]76[119]</td>
<td>GGTTGAAATTTCTTAAACAGCTT</td>
</tr>
<tr>
<td>77[70]52[63]</td>
<td>GAATAATTGAGGAGGAGGAG</td>
</tr>
<tr>
<td>77[77]26[70]</td>
<td>AATTTTTCTGGCACAACAGAGAAAATACCACCAAGAA</td>
</tr>
<tr>
<td>77[105]76[91]</td>
<td>AAAAGGCATTCAACAGTTTCAGCGGAGT</td>
</tr>
<tr>
<td>1[91]24[98]</td>
<td>CAGACAAATAAAAAGTAAGGCGTTAAAAATACGCA</td>
</tr>
<tr>
<td>7[126]20[130]</td>
<td>GTATAACGTGCCCTTAGCCGCGATGAATCAGCATTATAAT</td>
</tr>
<tr>
<td>9[119]14[112]</td>
<td>GTGGCGAAGCTGGAGGCGAGCCATCAACCCAAATCGTG-GACTGAAACAG</td>
</tr>
<tr>
<td>14[139]15[132]</td>
<td>GTTGAGTGTGTTTCAAAATCC</td>
</tr>
<tr>
<td>20[129]8[126]</td>
<td>GCGTTGCCTCGCCAGCTCAGCGTGCGATGCTGGCCCTTCACAC-CGCTAC</td>
</tr>
<tr>
<td>31[84]28[77]</td>
<td>AACGCCGCTACCGACAAAGGTTGAGGATCCAAAATGATAT-GTTCCATATT</td>
</tr>
<tr>
<td>42[83]45[90]</td>
<td>AAAATGAAAACAGGAAGACGATGTAAGAAAACCT-GAACGAGTAA</td>
</tr>
<tr>
<td>59[119]73[118]</td>
<td>AATAAAATAGGAATAGACTTTCCGGCACAGGAGGCTCGCGTTTT</td>
</tr>
<tr>
<td>69[77]70[70]</td>
<td>AGAGGGTACCGCCACCCCTCAG</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>71[77]72[70]</td>
<td>CCGCCACTGTACCGTAACACT</td>
</tr>
<tr>
<td>73[77]74[70]</td>
<td>CAACTAAGTTTTTGTCGTCCTT</td>
</tr>
<tr>
<td>75[77]76[70]</td>
<td>GAATTTTGAAAGGAACAACT</td>
</tr>
<tr>
<td>12[132]11[146]TTHr21</td>
<td>ACGTGAAGTGCCGTAAAGCACTAAATCGTTTACACATTCC-TAACCTTCTCATA</td>
</tr>
<tr>
<td>12[139]13[146]TTHr21</td>
<td>GCCCACTCGAAAAACCGTCTAATTACACATTCTAACTTCTCATA</td>
</tr>
<tr>
<td>35[133]37[146]TTHr21</td>
<td>ATAAGCAAAACTAGAGAGAATCGATGAATTTAACACATTCC-TAACCTTCTCATA</td>
</tr>
<tr>
<td>40[139]39[147]TTHr21</td>
<td>TAAAGATCAAACGTTCTAGCTGTTTAACACCTAACTTCTCATA</td>
</tr>
<tr>
<td>64[139]63[146]TTHr21</td>
<td>CAGATGATGACAAGAACCACAATTTTAACACCTAACTTCTCATA</td>
</tr>
<tr>
<td>65[126]65[146]TTHr21</td>
<td>TAAGGGAACCGAACTGACCAATTTTAACACCTAACTTCTCATA</td>
</tr>
<tr>
<td>Hr21_5Bio</td>
<td>/5BiotinTEG/TATGAGAAGTTAGGAATGTA</td>
</tr>
<tr>
<td>Anchor_Ext0_oh</td>
<td>TTAGTTAATTTCCATCTTTATGTATTGCTCATTATGCTGCAACCTCCGT</td>
</tr>
<tr>
<td>Anchor_Ext1</td>
<td>/5Phos/GCAATGATGAGTGAGCAATAAACATAACATCCACAGTGAA</td>
</tr>
<tr>
<td>Anchor_Ext2</td>
<td>GTGGATGGTCTGACCTAATTTAATGGTTTG</td>
</tr>
</tbody>
</table>

153
Table A.3: Additional DNA oligomers. DNA modifications (phosphorylation) are included using their IDT codes. Sample descriptions indicate the final linker duplex DNA length after ligation. This length includes contributions from one or both origami structures (14 bp for the origami rotor, 26 bp for the origami anchor, and a 12-nt overhang on each). Substrates with 3’ or 5’ overhang for RecBCD experiments are designed based on the 80-bp blunt-end DNA with additional 5’ or 3’ dT nucleotides as indicated in the main text. The strands for 80- and 92-bp duplexes are annealed prior to ligation to the origami structure(s). The primers used to generate the 163-bp DNA linker between the origami rotor and anchor contained dU bases. One strand of the PCR product for the 163-bp linker is shown prior to dU excision of the nucleotides shown in red, which created the 12-nt ssDNA overhang required for ligation to the origami structures. The other strand also contains a dU base to create ssDNA overhangs on each end. To generate the origami anchor-rotor complex with a 52-bp linker, strand Anchor_Ext0_oh (Table A.2) in the anchor folding reaction was replaced with Anchor_Ext0_oh_direct. For this sample, no additional DNA was required to connect the rotor and anchor. Overhangs for ligation to the origami rotor or anchor are shown in purple and bases complementary to Anchor_Ext1 (Table A.2) are in green.

<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>80-bp blunt-end DNA for RecBCD experiments</td>
<td>TCGAATTTCGCCCTATAGTGACGTGATTACAATTCACTGGCGAGCTGTTTTACAAAACTAGTCCGAGA</td>
</tr>
<tr>
<td>/5Phos/TTGTAAAAACGACGCCAGTGAATTTTGAATACGACT-CACTATAGGCCAGTGAGTAG</td>
<td></td>
</tr>
<tr>
<td>80-bp 10nt 5’ overhang DNA without GC in initial 5 bp (“-GC”)</td>
<td>TTTTTTTTTTATTATATCGCCTATAGTGACGTGATTACAATTCACTGGCCGGTCTTTTACAAAACTAGTCCGAGA</td>
</tr>
<tr>
<td>/5Phos/TTGTAAAAACGACGCCAGTGAATTTTGAATACGACT-CACTATAGGCCAGTGAGTAG</td>
<td></td>
</tr>
<tr>
<td>Adaptor for 92-bp linker between rotor and anchor</td>
<td>/5Phos/CCTCTAGAGTGGGAGCTCCGAACACTATACGGAGTTTCAGCA</td>
</tr>
<tr>
<td>/5Phos/ATAGTGTTTCCGAGCTCCCACTCTAGAGGACTAGTCCAAGG</td>
<td></td>
</tr>
<tr>
<td>Primers used for 163-bp linker production</td>
<td>TCTCGGACTAGUUTTGATAAAGCGACGGCAGTGAATTTG</td>
</tr>
<tr>
<td>TGCGAACTCCGUATAGTGTTTCCGAGCTCCCACTC</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
Table A.3 (Continued)

<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptor for 163-bp linker prior to dU excision</td>
<td>TGCGAACTCCGUATAGTGTTCCGAGCTCCCACTCTTAGAG-GATCCCCCGGTTACCCGAGCTCCGAATTCTCAGCTATAGTAGAGTCCG- TATTTACAATTCACGCGGTCTGTTTTACAAACTAGTCCGAGA</td>
</tr>
<tr>
<td>Replacement strand for 52-bp linker: Anchor_Ext0_oh_direct</td>
<td>TTAGTTAATTTTATATGGTGTATGCTACTATCATGCACTAGTGCAAAGGGTAGGTGCTGATCTTTCCTATAGGATGAGCTGTTTTACAACTAGTCCGAGA</td>
</tr>
<tr>
<td>Extension for variable [NTP] RNAP experiments</td>
<td>TTATCAAAAAAGAGTATTGACTTTAAGGT- TAACCTATAGGATACCTACGAAGCCTCAGA- GAGGGGACACGGGGATCCCTCTAGACTGCAGAAAGCGTGTGCTGCTACATGACTGACTGGTTGTAAAACGAGTGCCAGTGAATTGTAGCCACCAACCTTTCTGCAGTCTAGAGGATCCCCCGTGTCCCTC- GATGGCTGTAAGTATCCTATAGGTTAGACTTTAAGTCAATACTCTTTTTGATAA</td>
</tr>
<tr>
<td>Extension for low [NTP] RNAP experiments</td>
<td>TTATCAAAAAAGAGTATTGACTTTAAGGT- TAACCTATAGGATACCTACGAAGCCTCAGA- GAGGGGACACGGGGATCCCTCTAGACTGCAGAAAGGTGTTGGTGCTACAATTCACTGGCCGTCGTTTTACAACTAGTCCGAGA</td>
</tr>
</tbody>
</table>

/5Phos/ACAGTGCGATTTTAGCGAGCCCTCTAATCTTGTTAGCTTAGCTGATCTCTGACTGACTGACTGCTGATCTTTCCTGACGTCTAGAGGATCCCCCGTGTCCCTCTGATGGCTGTAAGTATCCTATAGGTTAGACTTTAAGTCAATACTCTTTTTGATAA

/5Phos/TTGTAAAAACGACGCGCAGTGATTTTGAGCCAC- CAACCTTTTCTGACGTTAGGATCCCGCGTGCTCCCTTCTC- GATGGCTGTAAGTATCCTATAGGTTAGACTTTAAGTCAATACTCTTTTTGATAA
Table A.4: DNA oligomers for hairpin substrates used in stopped-flow fluorescence experiments. Each hairpin consists of a “C” strand annealed to a pair of “A” and “B” strands, as shown in Figure 4.5. DNA modifications are included using their IDT codes.

<table>
<thead>
<tr>
<th>Label</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBE26._C</td>
<td>AGATCCTAGTGCAAGGTTTTCTGCACCTAGGATCTTTCTCTCAGTTGTTTTCTCGGACTAGTTTGTAAACGACGGCCAGTGAAATTGTAATTCGACTCACTATAGGGCGAATTCGA</td>
</tr>
<tr>
<td>HBE26._A</td>
<td>TCGAATTCGCCCTATAGTGAGTCGA/iCy3/A</td>
</tr>
<tr>
<td>HBE26._B</td>
<td>/5Cy5/TTACAATTCACTGCCCGTCGTTTTACAAAATAGTCGCCGA-GAACACAACAGGAGAA</td>
</tr>
<tr>
<td>HBE26._A_5oh10</td>
<td>TTTTTTTTTTGGCAATTCGCCCTATAGTGAGTCGA/iCy3/A</td>
</tr>
<tr>
<td>HBE26._C_3oh6</td>
<td>AGATCCTAGTGCAAGGTTTTCTGCACCTAGGATCTTTCTCTCAGTTGTTTTCTCGGACTAGTTTGTAAACGACGGCCAGTGAAATTGTAATTCGACTCACTATAGGGCGAATTCGATTTTTT</td>
</tr>
<tr>
<td>HBE52._C</td>
<td>AGATCCTAGTGCAAGGTTTTCTGCACCTAGGATCTTTCTCTCAGTTGTTTTCTCGGACTAGTTTGTAAACGACGGCCAGTGAAATTGTAATTCGACTCACTATAGGGCGAATTCGA</td>
</tr>
<tr>
<td>HBE52._A</td>
<td>TCGAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTGGTCGTTTTA/iCy3/A</td>
</tr>
<tr>
<td>HBE52._B</td>
<td>/5Cy5/TAACATAGTCCGAGAACAACACTGAGGAA</td>
</tr>
</tbody>
</table>
References


161


165


169


175


