DETECTION OF STOCHASTIC AND HETEROGENEOUS BEHAVIORS IN DNA NANODEVICES BY SUPER-RESOLUTION FLUORESCENCE MICROSCOPY

by

Alexander E. Johnson-Buck

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2013

Doctoral Committee:

Professor Nils G. Walter, Chair
Professor Hashim Al-Hashimi
Professor Raoul Kopelman
Associate Professor Jennifer Ogilvie
Assistant Professor Roger Sunahara

To my family.

ACKNOWLEDGMENTS

No body of work is the product of a solitary individual. I am grateful to the many people who have provided me with the support and mentorship needed to make this dissertation a reality.

I feel privileged to have worked with Dr. Nils Walter, whose dedication to scholarship and formidable work ethic set an inspiring example for his students. I am thankful for the considerable time he has devoted to overseeing my growth as an independent scientist and thinker.

I also owe thanks to the members and alumni of the Walter lab. The success of a research group rests heavily on the mutual support of its members. I am indebted to the many bright young scientists that have helped me to carve my own path through the open wilderness of discovery, especially Drs. Miguel Pereira, Anthony Manzo, and Chamaree de Silva.

Finally, I would like to acknowledge my family, without whose support I would never have reached this milestone. I want to thank my wonderful wife Lauren for her patience and unconditional love; you are my safe harbor in stormy seas. I am indebted to my parents for their enduring support, and for teaching me what really matters. I am also grateful to my sisters, in-laws, and other family and friends for their encouragement, companionship, and advice. This dissertation is, in a very real sense, a product of your efforts as well.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

Detection of Stochastic and Heterogeneous Behaviors in DNA Nanodevices by Super-Resolution Fluorescence Microscopy

by

Alexander E. Johnson-Buck

Chair: Nils G. Walter

Decades of advances in structural biology have inspired efforts to emulate and expand upon the functional capabilities of natural nanomachines. Recently, DNA nanotechnology has emerged as a promising route to realizing this ambition. With the help of advanced approaches like DNA origami, the structural and functional repertoire of this biopolymer has expanded far beyond the humble double helix, permitting designs of nearly arbitrary shape and growing sophistication.

The increased complexity and multifunctionality of synthetic devices constructed of DNA introduces more opportunities for erroneous assembly and otherwise heterogeneous performance. However, in most cases, functional characterization of DNA nanodevices is carried out in bulk or by techniques like atomic force microscopy (AFM) and transmission electron microscopy (TEM). Such approaches, while undeniably powerful, do not provide full access to the coupling between spatial, temporal, and chemical properties of individual nanodevices, rendering it difficult to understand how reproducibly they perform.

Recent progress in single-molecule fluorescence detection and super-resolution microscopy make it possible to address this gap in understanding. In this dissertation, I

report the development and application of several such techniques to elucidate spatiotemporal properties of single DNA nanodevices. First, a single-particle tracking assay is developed to characterize the complex movement of a synthetic DNA-based walker on prescriptive landscapes. The assay, combined with numerical modeling and population-level AFM results, confirms the designed path-following behavior of individual spiders and provides insights into their stochastic and heterogeneous behavior. Second, I develop a two-color version of the super-resolution technique PAINT (points accumulation for imaging in nanoscale topography) to interrogate the spatiotemporal dependence of oligonucleotide hybridization to arrays of dense targets on individual DNA origami, revealing surprisingly variable and persistent spatial patterns of binding kinetics. Finally, I examine the kinetics of oligonucleotide hybridization reactions on single DNA origami arrays bearing different densities of targets, showing evidence of at least two mechanisms by which hybridization kinetics differ from those in solution. Together, these results provide new insights into the degree of stochasticity and heterogeneity in the performance of DNA nanodevices and furnish new tools with which to design more sophisticated devices in the future.

CHAPTER 1:

PROGRESS TOWARDS A SINGLE-MOLECULE UNDERSTANDING OF BIOMOLECULAR NANOMACHINES

1.1 Introduction

Since Antony van Leeuwenhoek first peered into the world of microorganisms in the late seventeenth century, humans have striven to observe biological phenomena at ever smaller scales. In the intervening centuries, microscopy and other tools have transformed our understanding of biology, and indeed all matter. By measuring or inferring the microscopic spatial distribution of interactions between matter and radiation, we have identified the cell as the fundamental unit of life on Earth, elucidated the structures of molecules responsible for the genotype and phenotype of organisms, and begun to understand how these molecules function in concert to produce the fantastic diversity of behaviors we observe among living things.

Particularly enormous strides towards comprehending the chemical and physical underpinnings of life were made in the latter half of the twentieth century. The three-dimensional structure of the B-form helix of deoxyribonucleic acid (DNA) was solved in 1952, revealing for the first time the molecular basis of the genetic code[1]. Since then, techniques such as X-ray diffraction, electron microscopy (EM), nuclear magnetic resonance (NMR) spectroscopy, and fluorescence spectroscopy have borne fruit in the form of the now-mature field of structural biology. For instance, the structure of the ribosome, first discovered by electron microscopy in 1955[2], was elucidated over several decades, culminating in high-resolution structural data that revealed the ribosomal ribonucleic acid (rRNA) as the central catalyst of peptide synthesis[3–5]. Cryo-EM, X-ray

diffraction, and single-molecule fluorescence studies have also uncovered the presence of large-scale conformational changes and hybrid states during peptide elongation[6]. Out of several decades of similar work has crystallized an emerging picture of proteins and nucleic acids as nanometer-scale machines that perform the catalytic, mechanical, and regulatory functions that distinguish living from non-living matter.

1.2 Biological nanomachines behave stochastically and heterogeneously[1]

Despite the clockwork-like sophistication of biomolecular machines like the ribosome, the analogy to macroscopic machines is not perfect. The quantum-mechanical nature of molecules[8], combined with complex energetic landscapes punctuated by finite energy barriers, introduces a strong stochastic element to the behavior of individual molecules. For instance, while the catalysis mediated by a population of millions of enzymes is well approximated by deterministic equations[9], an individual enzyme catalyzes a reaction at random intervals with a certain *probability per unit time*. Using modern descendants of van Leeuwenhoek's microscope, this characteristic has been repeatedly verified for diverse protein and RNA molecules[10–12] *via* techniques like single-molecule fluorescence resonance energy transfer (smFRET)[13].

Intriguingly, for many systems studied *in vitro*, individual copies of a biomolecule undergoing a chemical reaction do not necessarily adhere to a single set of rate constants. Rather, molecules can often be divided into subpopulations whose fluctuations between states are apparently governed by different kinetic or thermodynamic parameters[7,14–16]. Such so-called heterogeneity of behavior has a variety of causes, from alternate folding to irreversible chemical damage[7]. The conformational energy landscape of RNA appears to be particularly rugged and therefore prone to kinetic folding traps leading to heterogeneity[7]. The capacity to misfold into kinetically trapped non-functional conformations seems to be the rule rather than the exception for RNA enzymes (ribozymes)[17–21]. In some cases, though, there appear

[1] Adapted in part from Marek, M.S., Johnson-Buck, A. and Walter, N.G. (2011) *Phys. Chem. Chem. Phys.* 13, p. 11524-11537. [7]

to be distinct populations of molecules – distinguished by the kinetics of conformational changes – that are all nevertheless catalytically active, suggesting the existence of multiple native states[14,16]. We have uncovered evidence for this genre of 'molecular schizophrenia' in a Varkud satellite ribozyme, as is detailed in Appendix 1. Intriguingly, in at least one case of an *in vitro* selected RNA, the presence of multiple distinct conformations of a single molecule appears to be not only tolerated, but functionally obligatory[22]. These studies show that, while the sequence of a peptide or nucleic acid predisposes the molecule to fold in particular ways, it does not in general define a unique three-dimensional structure. Each of the above classes of folding heterogeneity thus constitutes a kind of phenotypic diversity arising from a single genotype.

In the context of the living cell, there are many other stochastic sources of variation. While a great deal of phenotypic diversity arises from random genetic mutations culled by natural selection, many other molecular sources of phenotypic variation have been elucidated over the last several decades, including the action of transcription factors and repressors[23], covalent modification of histones and DNA[24], RNA interference[25], riboswitches[26–29], and alternative splicing[30,31]. These mechanisms make a large degree of phenotypic diversity possible even in populations of genetically identical cells. In some cases, phenotypic variation can arise stochastically in a population, such as through the translation of low-copy-number mRNAs in cells[32,33], or the presence of very low concentrations of transcriptional regulators[23]. Thus, stochastic molecular events have an important impact on the fate of an entire organism.

Such stochastic variation of phenotype may have important consequences for biological evolution as well. Genetic diversity within a population of organisms correlates with the fitness of that population[34], conferring upon the population variable resistance to parasites, toxins, and other environmental insults. Numerous examples for such effects have been observed, including genetic resistance to certain human diseases[35], resistance of insects towards pesticides[36], and the appearance of antibiotic-resistant strains of bacteria[37]. By analogy, the adoption of multiple non-interconverting functional conformations by a single molecule – conformational quasispecies – could allow an

organism to bypass evolutionary bottlenecks by, for instance, surviving a toxin that targets one conformer but not another[7] (Figure 1.1b), particularly in environments with fluctuating selection pressures[38].

The biological significance of stochastic variation of phenotype and the heterogeneous properties of single molecules remains unclear. For instance, recapitulating the natural co-transcriptional folding of RNAs has been shown to suppress alternative folding in some cases[39–41], suggesting that at least some kinds of folding heterogeneity observed *in vitro* may not reflect biologically important processes. Furtheremore, there exist cases in which robust, deterministic behavior of biological networks is preferred, and damping of stochastic variation appears to have been selected for[42–44]. Regardless of whether molecular heterogeneity is generally a nuisance or a boon to the organism, it is a very common feature of biopolymers, and therefore has strong implications for efforts to understand natural biochemical systems as well as to design artificial molecular systems inspired by biology.

1.3 Learning from nature: the construction of increasingly complex nanomachines from DNA[2]

A growing appreciation of the molecular details of biology has spurred efforts to adapt, mimic, and ultimately expand upon the functions of natural nanomachines. Within this nascent field, known as bionanotechnology, one of the most promising materials in the near term is DNA. Its well-defined double-helical structure and relatively simple, yet highly predictable organizational base pairing rules make DNA a strong candidate for the design of structures with nanometer precision[45].

Early efforts to design structured scaffolds from DNA involved joining different short double-stranded DNA (dsDNA) domains in a programmable fashion using single-stranded DNA (ssDNA) overhangs known as 'sticky ends'[46]. While the stiffness of

[2] Adapted in part from Michelotti, N., Johnson-Buck, A., Manzo, A.J., M. and Walter, N.G. (2012) *WIREs: Nanomed Nanobiotechnol.* 4, p. 139-152.[45]

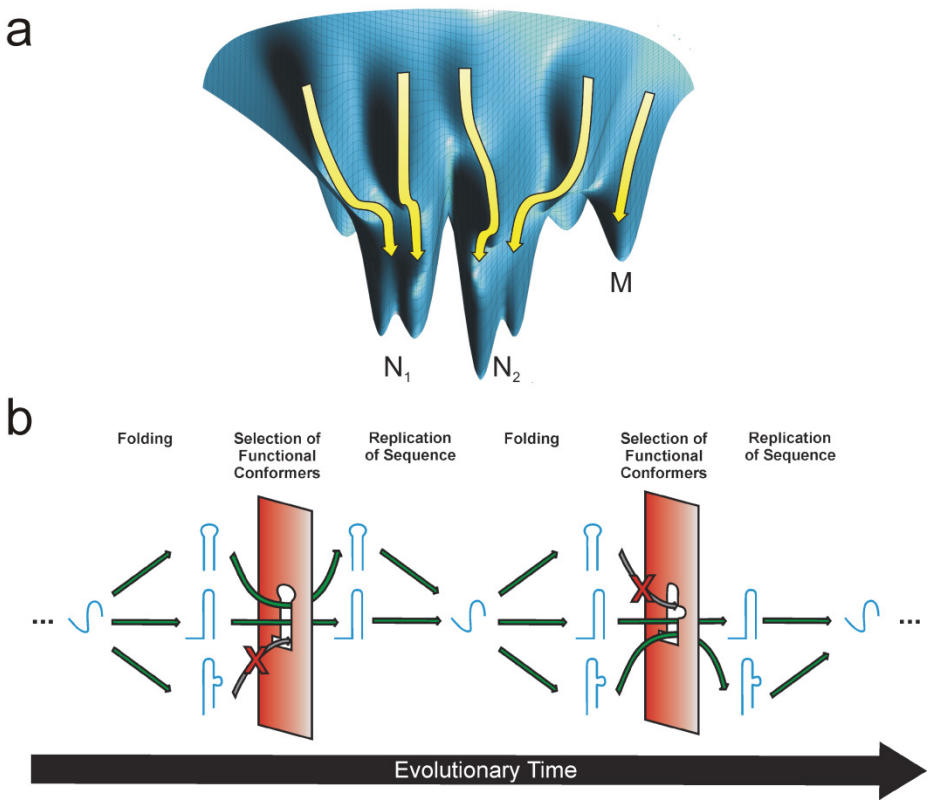


Figure 1.1. (a) Schematic representation of the rugged conformational free energy landscape (blue surface) of an RNA molecule exhibiting heterogeneity, such as the HpRz. An individual molecule folds along one of many possible pathways (yellow arrows) to one of multiple native states (N1, N2) separated by relatively large energetic barriers. These native states sample similar conformations, albeit with different kinetics, and thus possess similar activities. Alternatively, the molecule may enter a trapped misfolded state (M) that is non-functional. (b) Schematic representation of a possible adaptive role for conformational quasispecies of biomolecules. A single sequence (blue) may fold into several stable conformers, or native states, with varying functionality. Changing environmental conditions may impose certain restrictions (red) on the fitness of conformers, but the success of a subset of these conformers will enable the replication of the sequence and the survival of all stable (and kinetically accessible) conformers. Reproduced from reference[7] by permission of the PCCP Owner Societies.

dsDNA makes it a suitable raw material for the edges of stable two- and three-dimensional structures, the vertices of such structures remain flexible, resulting in a range of angles between domains[46]. Inspired by naturally occurring Holliday junctions[47,48], more rigid structures were accomplished using reciprocal exchange to generate double[49] and triple crossover motifs[50]. While such domain-combining approaches originally resulted in variable yields of assembled structures without defined boundaries, more recent approaches using single-stranded DNA tiles appear to have overcome these limitations[51,52]. As an alternative to combining multiple domains, increased yield and stability are accomplished using Rothemund's DNA origami method (Figure 1.2a-c), in which a circular single-stranded DNA molecule self-assembles into a predefined shape with the help of hundreds of shorter complementary ssDNA strands called staples[53]. DNA origami has been used to assemble sophisticated 2D and 3D structures with high yield[53–55] (Figure 1.2d,e). The ability to integrate hundreds of unique DNA strands into a single, well-defined structure makes it possible to precisely interact with and position other materials, including proteins[56,57], carbon nanotubes[58], and metal nanoparticles[59,60].

In addition to static structures, DNA has been used to design multi-component, dynamic devices with nanoscale precision. For example, it has been possible to design and execute molecular assembly lines[59,61], artificial cascades of enzymes separated by defined distances[56], and DNA computers that perform arithmetic operations[62] or play a game against a human opponent[63]. Molecular walkers analogous to the motor proteins myosin, kinesin, and dynein have been constructed completely or primarily from DNA and have attained increasing levels of autonomy as well as responsiveness to external instruction[59,61,64]. Coupling such devices with DNA computing[62,65] may further improve the complexity and range of responses to environmental stimuli or instructions.

Many of these devices incorporate dozens or hundreds of distinct DNA strands, raising an important question: how reliably and reproducibly does a given system perform? Given the diminutive alphabet of four natural nucleobases, it soon becomes difficult to exclude the possibility of unintended interactions and kinetics traps in a

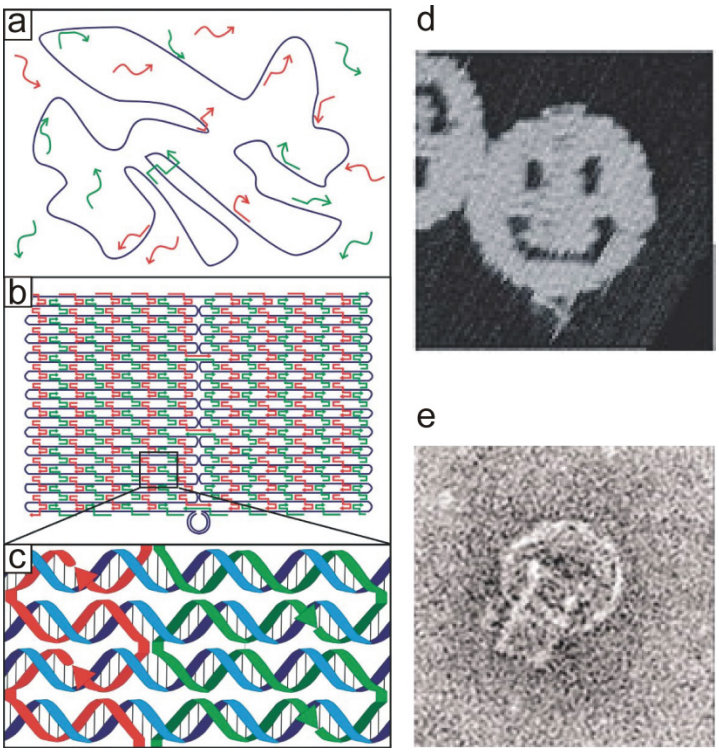


Figure 1.2. The DNA origami method. (a) The ssDNA 'scaffold' strand (purple) is folded into a particular shape by hybridization with hundreds of short complementary ssDNA 'staple' strands (red and green). (b) Resultant rectangular origami tile after all the staples have bound to the scaffold. (c) Triple crossover motifs formed between the staples and the scaffold. As each staple is unique and therefore specifies an 'address' on the completed structure, the 5′ ends of the staples can be extended to create overhangs for site-specific decoration with other components. (d) Atomic force micrograph of a two-dimensional DNA origami pattern[53]. (e) Transmission electron micrograph of a three-dimensional vase-shaped DNA origami structure[55]. Panels (a)-(c) reprinted from reference[45] with permission from John Wiley & Sons, Inc. Panel (d) Reprinted by permission from Macmillan Publishers Ltd: Nature[53], copyright 2006. Panel (e) reprinted with permission from the American Association for the Advancement of Science[55].

DNA-based system. The yield of properly folded DNA origami can be as high as 90%, but may be significantly lower depending on the particular design target[53]. If an application is very sensitive to the performance of individual devices, these execution errors may be significant or even detrimental.

Despite these considerations, most functional characterization of dynamic nanodevices is carried out in bulk[56,61,63,66,67], even if the structural integrity of individual complexes is verified by atomic force microscopy (AFM) or transmission electron microscopy (TEM). There thus exists a gap between the single-molecule characterization of structures and the typically ensemble-level assessment of function. An accurate functional characterization of these devices at the single molecule level would provide important feedback in the design process and contribute to more precise and reproducible behavior.

1.4 Super-resolution fluorescence microscopy: bridging the structure-function gap in characterizing DNA nanodevices

Assessment of structural integrity and yield of DNA nanodevices is typically carried out with AFM or TEM[53–56,58–60,68–70]. With sub-nanometer spatial resolution and the capacity for three-dimensional topological imaging[71,72], they are currently the most powerful techniques for visualizing the structure of nanoscale DNA structures. However, these techniques have drawbacks with respect to real-time imaging of functional DNA nanomachines. AFM, which depends on physical contact between a metallic cantilever tip and the sample, can mechanically perturb delicate biological structures[73], making repeated time-lapse imaging of a sample difficult to interpret. AFM is also rather insensitive to flexible components such as single-stranded DNA (ssDNA)[69], a common feature of functional DNA devices, and cannot reliably resolve even chemically distinct features that are in physical contact with one another[56]. TEM can damage organic samples by bombarding them with energetic electrons, also rendering time-lapse imaging difficult[72]. Although TEM can potentially provide information about elemental composition, it cannot readily distinguish, for example, two nucleic acids with different

sequences[72]. These limitations render it challenging to use AFM or TEM for longitudinal imaging of single complexes or spatiotemporal detection of specific chemical moieties such as flexible biopolymers.

On the other hand, visible light microscopy enables prolonged non-perturbative imaging of living samples, as van Leeuwenhoek demonstrated in the early 1700s with the first observations of microorganisms. In particular, fluorescence microscopy[74] has been an enormously useful tool for interrogating biological samples with molecular precision, owing to the fact that specific molecules can often be stained or site-specifically labeled with small organic fluorophores[75], fluorophore-labeled antibodies[76], or fluorescent proteins[77]. Fluorescence provides enhanced sensitivity and contrast, permitting the detection of even single fluorescent molecules[78]. The main drawback of fluorescence microscopy is resolution.

Due to the diffraction of light, the maximum possible resolution attainable on a wide-field light microscope is the Abbe limit[79],

$$d = \frac{\lambda}{2N} \qquad (1)$$

where *d* is the resolution, λ is the wavelength of light and *N* is the numerical aperture of the lens. As *N* is typically equal to 1.2-1.4 for modern objective lenses, the theoretical maximum resolution is on the order of one-half to one-third the wavelength of light. When a single fluorophore is detected by light microscopy, then, it appears not as an infinitesimal speck but as a diffraction-limited pattern known as an Airy disc[79]. At first sight, this poses a problem for imaging nanoscale structures and devices, since the Abbe limit corresponds to approximately 150-300 nm for visible light. However, the diffraction-limited intensity pattern can be numerically modeled, most commonly with a two-dimensional Gaussian function, in order to locate its centroid with high precision. The localization accuracy is limited only by the number of photons collected, the amount of background noise in the measurement, and the effective pixel size of the detector[80],

but as a practical matter typically falls in the range of 1-25 nm. Using this so-called super-resolution strategy, individual fluorescent particles have been successfully tracked over time with nanometer precision, for example revealing the hand-over-hand walking behavior of the protein motors myosin V, myosin VI, and kinesin[81]. Clearly, this provides an avenue for characterizing synthetic locomotive nanomachines as well[82], including DNA-based walkers.

Additional challenges arise when multiple targets must be localized within a diffraction-limited region, where Gaussian fitting of the overlapping intensity distributions of single molecules may no longer converge to the correct particle positions. This is crucial in the case of many DNA origami-scaffolded devices, which often bear patterns of multiple interacting components within the confines of a sub-100-nm structure[56,59,67,83]. An effective solution to this problem has been the family of stochastic reconstruction microscopies[84–87], in which a sparse subset of features in a crowded field are sampled at random and localized individually to reconstruct a super-resolution image of the sample. This can be achieved by controlled, stochastic activation of organic fluorophores[84,87] or fluorescent proteins[85], or by reversible binding of luorescently labeled probes to specific targets[86]. This latter approach, known by the acronym PAINT (points accumulation in nanoscale topography), has distinct advantages for imaging DNA nanostructures. First, since it relies on exchangeable rather than static probes, it is less sensitive to loss of probes from photobleaching, rendering time-lapse imaging more feasible. Second, it permit the facile exchange of one set of probes for another, enabling consecutive assays of a single sample for different sets of chemical features. It has recently been applied to super-resolution imaging of sparse features displayed on DNA origami[88], but its potential for multicolor and time-lapse imaging, as well as the resolution of very dense features, has yet to be realized in the context of nanotechnology.

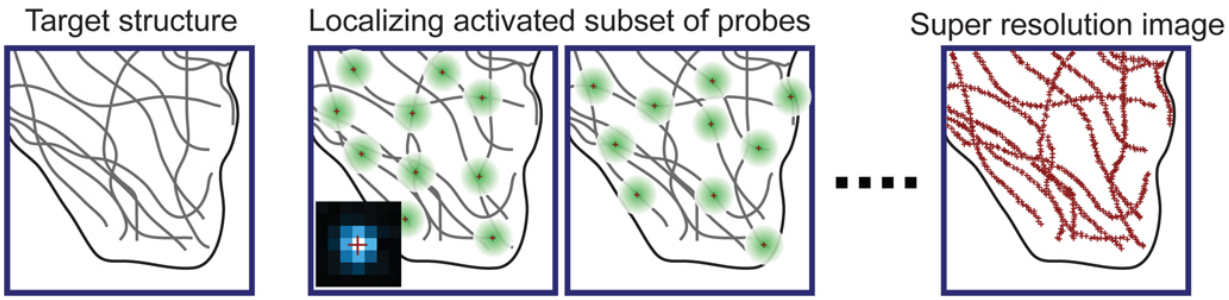


Figure 1.3. Principle behind stochastic reconstruction microscopy. A target structure (left) with features smaller than the Abbe diffraction limit is labeled with static or exchangeable fluorescent probes. In a given interval of observation, only a sparse subset of these probes is activated or bound and localized to nanometer precision (middle). After many cycles of activation or binding, a super-resolution image is reconstructed from the accumulated localizations of probes (right). Reproduced from reference[89] with permission from Annual Reviews.

1.5 Overview of the dissertation

In the following chapters, we describe several applications of single-molecule and super-resolution fluorescence microscopies to the study of individual nanoscale devices and structures composed of DNA. In Chapter 2, we describe single-particle tracking of individual DNA walkers known as molecular spiders, which use multivalent binding and catalysis to advance on tracks of oligonucleotide substrate positioned in space by DNA origami. We also show that their patterns of motion are qualitatively consistent and in relatively good quantitative agreement with a simple Monte Carlo model of their walking behavior based on bulk solution kinetics of isolated spider legs.

In Chapter 3, we extend the powerful PAINT technique to multicolor imaging of dense patterns of oligonucleotide features on DNA origami. We achieve sufficient resolution to distinguish different sub-100 nm patterns of features, and show that it is possible to quantitatively follow chemical reactions over time using this technique. We also report an unexpected spatial heterogeneity of target accessibility over the surface of origami. These patterns, which we refer to as “fingerprints,” differ substantially between individual origami, in analogy to the heterogeneous behavior of simpler nucleic acids described in Section 1.2. We propose that these fingerprints are due to variable spacing of weakly interacting features over the origami surface, an interpretation supported by both kinetic evidence and coarse-grained conformational modeling of the DNA origami tile.

Finally, in Chapter 4, we develop a FRET-based kinetic assay of hybridization kinetics on single origami. Using this assay, we show that the kinetics of hybridization and dissociation of oligonucleotides on a DNA origami tile differ systematically from the corresponding reactions in solution. Depending on the nature of the binding partners, dissociation from the tile can apparently be slowed by two mechanisms: (1) direct passing of oligonucleotide probes between their origami-bound targets, and/or (2) nonspecific interactions with the origami tile itself. Finally, we show that origami with sufficient numbers of binding targets behave very reproducibly in terms of their overall

kinetic behavior, while there is more stochastic variation between origami bearing fewer targets, and that this behavior is consistent with Monte Carlo predictions.

Together, the work presented in this dissertation contributes to the analytical toolkit of nanotechnology. We aimed to provide the means to better “micro-manage” individual DNA nanodevices by understanding the degree to which their behavior can be monitored, predicted and controlled. We also promote the notion that simultaneous acquisition of spatial, temporal, and chemical information about DNA nanodevices provides a rich foundation for their further improvement. Ultimately, we harbor the hubristic hope of enabling the design of nanomachines approaching the beauty, complexity, and utility of those that constitute the living world around us.

CHAPTER 2:

SINGLE-PARTICLE TRACKING AND MONTE CARLO MODELING OF WALKING MOLECULAR SPIDERS[3,4]

2.1 Introduction

Nature employs a variety of processive motor proteins for coordinated processes such as directed superdiffusive transport[90] and cell division[91]. Their common feature is the transformation of chemical energy, usually in the form of ATP, into work. For instance, the molecular motor myosin V employs a cycle of ATP binding, hydrolysis, and ADP release to coordinate a series of conformational changes in its two head domains, giving rise to a highly processive hand-over-hand gait[92,93].

Over the past decade, efforts have been made to emulate certain characteristics of these sophisticated natural nanomachines using DNA. In analogy to the binding arms of a protein motor, ssDNA can bind site-specifically to a track or landscape. Furthermore, strand-displacement reactions[94–96] or catalysis[97] can provide the free energy to bias the motion of a walker on a track. DNA walkers originally depended on real-time direct input from the experimenter to achieve directionality, but have recently achieved a degree of autonomy by the incorporation of catalytic legs[66] or thermodynamic engineering of strand displacement reactions[98]. Still, no previously reported walker has

[3] Reproduced in part from Lund, K.; Manzo, A. J.; Dabby, N.; Michelotti, N.; Johnson-Buck, A.; Nangreave, J.; Taylor, S.; Pei, R.; Stojanovic, M. N.; Walter, N. G.; Winfree, E.; Yan, H. Nature 465, 206–210. Copyright Nature Publishing Group, 2010.

[4] Alexander Johnson-Buck performed displacement and velocity analysis of single-particle tracking experiments, fluorescence assays of spider leg cleavage, and Monte Carlo simulations of spider walking. Anthony J. Manzo and Nicole Michelotti performed single-particle tracking experiments and RMSD analysis. Anthony J. Manzo, Nicole Michelotti, and Alexander Johnson-Buck collaborated in devising single-particle tracking experiments and trajectory classification criteria. AFM experiments were conducted by Kyle Lund, Jeanette Nangreave, and Nadine Dabby. Analysis of AFM experiments was performed by Nadine Dabby, Kyle Lund, Jeanette Nangreave, and Steven Taylor. Spiders were synthesized, purified, and their integrity confirmed and monitored by Steven Taylor. SPR experiments were performed by Renjun Pei. The contents of the manuscript on which this chapter is based were produced with collaboration from all the above individuals as well as Nils G. Walter, Hao Yan, Milan N. Stojanovic, and Erik Winfree.

achieved more than a few processive steps, nor has their behavior – and any degree of stochasticity or heterogeneity in it – been characterized at the level of single molecules.

Here, we present the single-molecule characterization of a class of DNA walker called the molecular spider. Composed of a streptavidin body attached to four biotinylated DNA enzyme (DNAzyme) legs, it has previously been demonstrated to undergo autonomous random walks in fields of DNAzyme substrate[99]. We convert this to a processive, directed walk by providing spiders with a highly structured environment: a track of dozens of ssDNA substrate molecules assembled on an addressable DNA origami tile[53]. When exposed to different track geometries constituting alternative 'programs' of movement, the spiders follow instructions with high efficiency as measured by AFM. We also examined the walks of individual spiders in a time-resolved fashion using single-particle fluorescent tracking, revealing a variety of both designed and non-designed behaviors. We further used a Monte Carlo model of spider walking to show that the aggregate motion of the majority of moving spiders is qualitatively consistent with expectations based on solution binding, dissociation, and cleavage kinetics of isolated DNAzyme legs in solution. These results constitute initial steps towards the development of intelligent 'molecular robots' that autonomously sense and react to cues in their environment in a programmable fashion.

2.2 Materials and Methods

ABBREVIATIONS

iSp18 is a hexa-ethyleneglycol internal spacer; Bio is biotin; and BioTEG is biotin-tetra-ethyleneglycol.

PREPARATION OF SPIDERS

Materials and Instrumentation for the Preparation and Characterization of NICK3.4A+1 and NICK$_{3.4A+1}$• (Cy3)$_3$. Synthesis and purification of the modified DNA strands used to construct **NICK3.4A+1** and **NICK$_{3.4A+1}$• (Cy3)$_3$** were carried out by Integrated DNA Technologies (Coralville, IA) and used as received. Streptavidin was obtained from

Pierce, product number 21125 (Rockford, IL). IE-HPLC purification was performed using a Shimadzu LC-6AD pump equipped with a Shimadzu SPD-M10A PDA detector, with separation carried out on an anion exchange TSKgel DEAE-NPR column, 4.6x50 mm (IDxL) (Tosoh Biosciences). Concentrations of oligonucleotides were determined on an Amersham Biosciences Ultrospec 3300 pro UV/visible spectrophotometer.

*Assembly of NICK3.4A+1***.** *Part A*; capture leg [5′ - GCC GAG AAC CTG ACG CAA GT/iSp18//iSp18//3Bio/ - 3′] (**C**) (47 nmoles in 10 mL of 10 mM HEPES,150 mM NaCl, pH 7.4) was added drop-wise to a stirred solution of streptavidin (STV) (5 mg, 94 nmoles in 1 mL of 10 mM K3PO4,pH6.5). The desired one-to-one conjugate product ("STV-(**C**)1") was purified by ion exchange (IE) HPLC (see Figure A1.2 for details of purification conditions). *Part B*; deoxyribozyme leg [5′ - /5BioTEG//iSp18//iSp18/TCT CTT CTC CGA GCC GGT CGA AAT AGT GAA AA - 3′] (**L**) (100 μM, in water) was titrated into the isolated 1:1 conjugate HPLC fraction from 'Part A' above, until all three remaining biotin binding sites of the 1:1 conjugate "STV-(**C**)1" were occupied by **L** to give the final desired product "STV-(**C**)1(**L**)3" i.e. **NICK3.4A+1**. The titration was monitored by IE-HPLC, and was deemed complete when a slight excess of **L** was observed with no intermediate species, i.e. no "STV-(**C**)1(**L**)1" or "STV-(**C**)1(**L**)2", present, see Figure A1.3. The assembly was purified by IE-HPLC (see Figure A1.3 for details of purification conditions) and the volume of the elutent reduced (by centrifugation) to give a final concentration of 2.3 μM, as determined by absorbance at 260 nm. Characterization of the assembly was carried out by IE-HPLC and PAGE (Figures A1.2-A1.5). The assembly was stable at -20 °C for at least six months.

Assembly of $NICK_{3.4A+1}\bullet (Cy3)_3$[5]**.** *Part A and part B* were carried out in identical fashion to the assembly of **NICK3.4A+1** above, except "(C)" was [5′ – /5Cy3/GCC GAG AAC CTG ACG CAA GT/iSp18//iSp18//3Bio/ - 3′] and triethanolamine (20 mM) was used in place of HEPES and TRIS for the assembly and HPLC purification respectively. *Part C;* the volume of $\mathbf{NICK_{3.4A+1}\bullet (Cy3)_1}$, fraction isolated by HPLC, was concentrated to 1 mL

[5] The number of Cy3 dyes per spider is an average. This particular protocol sometimes produced an average of four Cy3 dyes per spider molecule, hence such spiders will be notated in the text as $\mathbf{NICK_{3.4A+1}\bullet (Cy3)_4}$.

(0.834 nmoles) and Cy3 Mono NHS ester (20 nmoles) (PA13101, Lot number 359269, GE Healthcare) dissolved in DMSO added to the solution containing the assembly (giving a total DMSO concentration of 10%). The resulting mixture was incubated at room temperature overnight, protected from light. Excess dye was separated from the **$NICK_{3.4A+1}$• $(Cy3)_3$** product by gel filtration (PD-10 column, 17-0851-01, lot 367770, GE Healthcare). Ratio of dye to streptavidin-DNA assembly was obtained by determining concentrations at 550 nm (ε_{max} 150,000 $M^{-1}cm^{-1}$) and 260 nm (ε_{max} 1,220,000 $M^{-1}cm^{-1}$) respectively.

SURFACE PLASMON RESONANCE (SPR)

Materials and Instrumentation for SPR Experiments. Immunopure avidin was purchased from Pierce (Rockford, USA). We used a Biacore X system and commercially available Biacore SA sensor chips, and Biacore C1 sensor chips, from GE Healthcare (Piscataway, USA). 1× HBS buffer (10 mM HEPES, pH7.4 with 150 mM NaCl) was employed as running buffer.

Preparation of Substrates on pseudo-2D Hydrogel Matrix Surfaces for SPR. A 20 μM solution of cleavable substrates (5′-BioTEG-TTTTTTTTCACTATrAGGAAGAG, "r" precedes a ribonucleotide) was applied to both channels of the SA sensor chip (carboxymethylated dextran matrix pre-immobilized with streptavidin) for 16 min at 5 μL/min, followed by a 60 s wash with 4 M urea and 15 mM EDTA in both channels to remove any nonspecifically adsorbed materials. The quantity of substrates adsorbed was calculated by the change in measured mass as described[99].

Preparation of Substrates on 2D Monolayer Surfaces for SPR. Avidin was covalently bound to the C1 sensor chip surface (a carboxymethylated monolayer) via amino groups using the following protocol. The carboxymethylated surface was first activated at a flow rate of 5 μL/min by using a 7 min injection pulse of an aqueous solution containing N-hydroxysuccinimide (NHS, 0.05 M) and N-ethyl-N`-(dimethylaminopropyl) carbodiimide (EDC, 0.2 M). Next, an 80 μL injection of 1 mg/mL avidin (in 1× HBS) was flowed over the activated surfaces of both channels for 40 min at 2 μL/min. The

remaining activated sites on the chip surfaces were blocked with a 35 µL injection of an ethanolamine hydrochloride solution (1 M, pH 8.5). Then, a 20 µM solution of cleavable substrate was applied to both channels of C1 sensor chip for 20 min at 4 µL/min, followed by a 60 s wash with 4 M urea and 15 mM EDTA. Based on the average SPR responses for avidin (~2,010 RU, 0.03 pmole/mm^2) and substrate (450 RU, 0.056 pmole/mm^2), there are two substrates bound for each avidin molecule. The average intersubstrate distance is 5.5 nm.

SPR Monitoring of Dissociation of NICK$_{3.4A+1}$ Spider on Non-cleavable Substrate and Product Surfaces. The non-cleavable substrate analog (substrate in which rA was substituted with A) or product surfaces were prepared in a similar manner to the preparation of substrate on 2D monolayer surfaces. The spider was loaded to channel 2, with channel 1 serving as a negative control. We calculated the ratio of spider to non-cleavable substrate or product by measuring the change in SPR response units (RU) after the spider was flowed onto the chip, then used the equation: ratio (spider/S or P)= M_w (S or P) × RU(spider)/[M_w (spider) × RU(S or P)] (Figure A1.21). Monitoring the dissociation of the spider was performed in 1× TA-Mg buffer (40 mM Tris, 20 mM acetic acid, 12.5 mM Magnesium acetate) with 1 mM $ZnCl_2$.

We could not directly measure the dissociation rate of spiders from cleavable substrate because 1) dissociation of the cleavage product from the surface accounts for the vast majority of the SPR response, and 2) the ratio of substrate to cleavage product changes with time, so the dissociation rate of spiders is not constant. Therefore, we instead monitored the SPR response to obtain the dissociation rate of spider on non-cleavable substrate, and product. We observed that over the course of 30 min >92% of spiders remained on a product covered surface and over the course of 60 min 86% remained bound (Figure A1.21). These percentages represent an upper-bound on spider dissociation from our tracks (which will be a mixture of substrates and products as the spider walks over it). So we estimate an upper-bound for the dissociation rate as less than 8-14 % over the time scale of our experiments on AFM and fluorescence microscopy.

SPR Monitoring of Cleavage of Substrates by NICK$_{3.4A+1}$ Spider. Spiders (0.8~6.3 nM in 1× HBS buffer) were loaded only on channel 2 at 5 µL/min, with channel 1 used as a negative control. The amount of spider applied was controlled by adjusting concentrations and the reaction times of spiders in the loading solution. Monitoring the cleavage of the substrate was initiated by switching to 1× TA-Mg buffer with 1 mM $ZnCl_2$ or 1× HBS buffer with 1 mM $ZnCl_2$ with the Biacore X system 'Working Tools Wash'. Product formation in real time was measured through the decrease in mass, using the formula 1,000 RU = 1 ng·mm^{-2}. Rates of cleavage were determined from the approximately linear region of the product release curves during the initial 10% of substrates cleaved. On the 2D monolayer surface, real-time processivity of spiders was measured to be ~79% (percentage of total substrate cleaved over the course of the experiment) at a 1:291 ratio of spider (17.8 RU) to substrate (448.4 RU) with a cleavage rate of 1.42 min^{-1} per spider. On the pseudo-2D matrix surface, spiders showed a real-time processivity ~86% of total substrate cleaved at a 1:990 ratio of spider (26 RU) to substrate (2,222 RU) with a cleavage rate of 2.81 min^{-1} per spider (Figure A1.10).

PREPARATION OF SPIDER-ORIGAMI ARRAYS

Assembly of Spider-Origami Arrays for Atomic Force Microscopy (AFM). The spider arrays consist of M13mp18 viral DNA (New England Biolabs) and 202 ssDNA staples (Integrated DNA Technologies, see Figure A1.1 for DNA sequences). The arrays were annealed in 1× TA-Mg Buffer (40 mM Tris, 20 mM acetic acid, 12.5 mM Mg^{2+}, pH 7.6) using a 1:3 ratio of M13 to staple strands and a final concentration of 10 nM (M13). The arrays were annealed in two hours from 94 ^{o}C to 25 ^{o}C using an Eppendorf PCR machine (Eppendorf). The **NICK3.4A+1 or NICK$_{3.4A+1}$• (Cy3)$_3$** were then added to the arrays at a 1:1 ratio of START strand to spider and left at room temp overnight. Because origami folding is sensitive to stoichiometry, we expect that some fraction of origami are missing the START strand and are thus unable to position a spider before the TRACK is deposited. The substrate strand and CONTROL strand were then added at a 1:1 (for initial ABD, EABC and Before EABD samples) or 1:3 (for 15, 30 and 60 minute EABD samples) ratio of staple probes to substrate or CONTROL and allowed to

bind overnight at room temperature (20 ^{o}C to 24 ^{o}C). We observed (by AFM) a larger percentage of apparently unbroken TRACKS when excess substrate was added. In the presence of excess substrate there is a low probability that a spider leg may bind to a free floating substrate or STOP strand that would deter or inhibit interactions with the TRACK. Note that the 8-17 deoxyribozyme has reduced but non-negligible activity in TA-Mg buffer (relative to maximal activity with Zn^{2+}; see PAGE Activity Assays, below), suggesting that spiders bound at START may cleave immediately neighboring substrates during the overnight incubation. Since spiders undergo (unbiased) walks on product tracks with little dissociation, this possibility is not a concern. To minimize stacking interactions that can cause aggregation of origami, the staples on the left and right edges of the origami were removed. Schematics of the assembled origami landscapes are shown in Figures A1.6-A1.9.

Modification of Spider-Origami Arrays for Fluorescence Microscopy. To make the origami arrays compatible with fluorescence microscopy, we returned 4 of the removed staples to the corners of the origami. In order to affix the origami to slides for analysis, we divided the corresponding staples into two strands so that we could affix biotin labels onto the 5′ end that is antiparallel to staple probes (as in Figures A1.7b, A1.8b and A1.9b). We modified the CONTROL strand by adding a Cy5 fluorophore to its 3' end, which resulted in 6 Cy5 fluorophores labeling the STOP position. On all landscapes, CONTROL staples were replaced with staples lacking the non-cleave-able substrate probes. The EAC landscape used in both fluorescence microscopy and AFM experiments lacked a CONTROL site. In addition, the EAC arrays for fluorescence microscopy were annealed in 5× SSC buffer (75 mM sodium citrate, pH 7.0, 750 mM NaCl), and the EABC and EABD arrays in 1× TA-Mg buffer. Fluorescence microscopy was also performed for origami arrays containing a truncated substrate TRACK, or product TRACK. The product strand is 8 nucleotides shorter than the full length substrate and includes only the sequence 5′ of the RNA base. The resulting 31 oligonucleotides have the same sequence as the corresponding portion of the full length cleavable substrate. All other assembly details for origami arrays for fluorescence microscopy including DNA concentrations, relative strand ratios, and binding conditions

were unchanged.

ATOMIC FORCE MICROSCOPY

AFM Imaging. "Before" samples were deposited on mica without the addition of TRIGGER or $ZnCl_2$. "After" samples were prepared by releasing the spider from the START strand through the addition of a 27-base TRIGGER strand, immediately followed by the addition of 10mM $ZnCl_2$ to a final concentration of 1 mM. Spiders were allowed to traverse the product or substrate TRACK array in solution for 15, 30, or 60 min (depending on the experiment) at room temperature before the origami were deposited on mica. Samples (2 µL) were deposited onto a freshly cleaved mica surface (Ted Pella, Inc.) and left to adsorb for 3 min. Buffer (1× TA-Mg, 400 µL) was added to the liquid cell and the sample was scanned in tapping mode on a Pico-Plus AFM (Molecular Imaging, Agilent Technologies) with NP-S tips (Veeco, Inc.). Each sample was scanned for 2-3 hrs before being discarded (therefore "30 minutes after" means that the sample spent 30 minutes in solution followed by up to 3 hours on mica). Note that the reduced but non-negligible deoxyribozyme cleavage rate in TA-Mg raises the possibility that spiders could move during the this imaging period; however, given the apparent difficulty of spider movement on mica-bound origami even in the presence of Zn^{2+} (see AFM Imaging for Movie) and the consistent trends in the time-lapse experiments (Figure 2.2), we conclude that very little movement takes place during the imaging period. All imaging by AFM was carried out at room temperature.

AFM Imaging for Movie. The sample (2 µL) was deposited onto a freshly cleaved mica surface and left to bind for 2 min. Then 1 µL of TRIGGER strand was added to the sample on the surface and after 2 min 270 µL of buffer and 30uL of 10mM $ZnCl_2$ was added to the sample cell. The four images were taken over a 26-minute time frame with about 10 min between the saving of each scan. (It should be noted that many prior and subsequent attempts were made to capture another AFM movie using various optimizations of our buffer, and protocol, without success.) Although we were only able to capture one movie, reported in Figure 2.4, we are convinced that it is not an artifact. The origami with the moving spider is substrate face-up while the three origami in the

same image are substrate side down (see below for a discussion of how the face of the origami affects spider analysis). As a result spiders on the three adjacent origami are stationary over the time course of the movie. In addition the spider's motion follows the TRACK in each frame (therefore it is not randomly diffusing, because it neither moves backwards nor off the TRACK). If the AFM tip were merely pushing the spider forward we would not expect the spider to turn in the transition from frame 3 to frame 4.

AFM Time Lapse Experiments. There is one seeming contradiction in our report that we would like to address here. If we were to suggest (as we do in Figure 2.4) that the spider can walk on origami deposited on mica, then how could we expect to obtain viable statistics from time lapse experiments imaged for up to 3 hours? We assume that under these conditions, most spiders get stuck on the origami, while some small percentage of spiders are able to continue moving. We find that we can differentiate between samples deposited at 15 minutes from those deposited at 30 and 60 minutes (see Table A1.1). These results help to explain why obtaining the AFM movie was so difficult.

Statistical Analysis of AFM images. We divided our flattened AFM images into 1 x 1 um images and numbered them. Within each of these images, we assigned a roman character to each origami (thus each origami we analyzed could be uniquely identified by a number and letter (i.e. "EABD Before 1e", or "EABD 30 min 3a"; Figures A1.12-A1.19). The origami arrays were classified by the following criteria: orientation (is the origami "face-up" or "face-down"?), number of spiders (0,1, multiple), location of spiders (START, TRACK, STOP, CONTROL), image quality (do imaging errors or sample impurities make the classification difficult?). This process was conducted independently by three people, for each data set excluding the EABD 15 minute and EABD 60 minute data sets, which were conducted by two people. The classifications were then compared: if two or more people agreed on the origami classification it was held, otherwise the origami was discarded from further analysis. By this method, we sought to ensure that our results are neither subjective nor irreproducible. While it is possible that some putative spiders were actually image artifacts or molecular contaminants, it is

unlikely that this inaccuracy in our measurements could affect the main trends in our data or the qualitative conclusions we drew from them.

An origami that is “face up” is one that displays its substrates and spiders on the face opposite the mica; an origami that is “face down” displays its substrates on the face that rests on the mica (Figures 2.12-2.13). Orientation was determined by landscape asymmetries in the positions of the TRACK and marker. By analyzing the statistics of origami classification, we concluded that the probability of an origami landing on one face or the other was approximately equal. However, we discovered that “face down” origami appeared to have a larger number of spiders at the STOP. We conducted a double-blind study in which 6 researchers were given an AFM image of origami and asked to classify these according to our criteria. We discovered that in the absence of spiders, all “face-up” origami were classified as vacant while a significant portion of “face-down” origami were classified as displaying a spider at the STOP site, when in fact there was none. Due to this “false positive” effect, we did not count “face-down” origami in our statistics. Approximately 50% of “face-up” origami were unoccupied by any spiders, and between 0 and 7% displayed more than one spider on the TRACK. Because the quantity of multiply occupied origami was small compared to the quantity of unoccupied and singly-occupied origami, we only considered singly-occupied origami to simplify our analysis (Figure 2.2).

Experimental results for all four landscapes with substrate TRACKS showed that the fraction of spiders at the START diminishes with a concomitant increase in spiders observed on the STOP positions (Figures 2.2c,g and 2.3). Our shortest track (ABD, spanning 48 nm) efficiently delivers spiders to the STOP, with less than 20% of spiders on the TRACK after 30 min (Figure 2.2c). If the TRACK was omitted on the ABD landscape, spiders were equally distributed between the STOP and CONTROL sites after 30 min, implying that the track is needed for efficient delivery to the STOP site (Figure A1.20). On longer TRACKS (such as EABD, spanning ~ 90 nm) ~15% of spiders are delivered to the STOP within 15 min after release. Longer incubation times (30 and 60 min) increase the efficacy of delivering spiders to the STOP to up to 70%,

(Figure 2.2c,g). Even at 60 min, however, we observed between 10-15% of spiders still on the TRACK. This outcome could be attributed to the distribution of spider velocities resulting from the stochastic nature of individual walks and possibly from backward steps onto product, initiating an unbiased random walk on product. We observed no significant difference in the efficacy of "turn right" and "turn left" actions (paths EABD and EABC, respectively) 30 min after release (Figure 2.2c,g).

PAGE ACTIVITY ASSAYS OF $NICK_{3.4A+1}$• $(Cy3)_4$

The cleavage activity of spiders under various conditions in bulk solution was tested as follows. Reactions were initiated by combining **$NICK_{3.4A+1}$• $(Cy3)_4$** (34 nM) with 4A substrate (5'-/5bio//iSp18//iSp18//iSp18//TTT TTT TTT TTC ACT AT(rA) GGA AGA G-Cy5, 34 nM) in the presence of either 1× SSC (15 mM sodium citrate (Mallinckrodt Inc.), pH 7.0, 150 mM NaCl) or 1× TA-Mg, and 0, 1, 2 or 10 mM $ZnSO_4$ (all reported concentrations are final). All reactions were supplemented with an oxygen scavenger system (1× OSS) consisting of 25 nM protocatechuate dioxygenase, 2.5 mM protocatechuate, and 1 mM Trolox as described[100]. Reactions (10 µL) were quenched after 0, 5, or 30 minutes with 2.5 µL of 0.25 M EDTA, and characterized by denaturing PAGE (Figure A1.11a). Fluorescence from Cy5 and Cy3 was detected on a Typhoon 9410 Variable Mode Imager (Amersham Biosciences) and the fraction of cleaved substrate quantified in ImageQuant 5.2 (Molecular Dynamics). Substrate was cleaved at least five-fold more slowly in 1× SSC + 1 mM Zn^{2+} than in 1× TA-Mg + 1 mM Zn^{2+}, while cleavage in 1× SSC + 10 mM Zn^{2+} was only about two-fold slower than in 1× TA-Mg + 1mM Zn^{2+} (Figure A1.11b). The maximal extent of cleavage is also about 9-fold lower in SSC + 1 mM Zn^{2+} than in TA-Mg + 1 mM Zn^{2+}, consistent with a significant fraction of inactive deoxyribozyme-substrate complexes. This discrepancy among buffers is likely due to partial complexation of Zn^{2+} ions by citrate: from a direct Zn^{2+} concentration measurement in buffer using the low-affinity (30 µM) indicator dye Newport Green PDX (Molecular Probes) we estimate the free Zn^{2+} concentration in SSC buffer to be approximately 3-fold lower than in TA-Mg buffer at 1 mM total Zn^{2+}. Nevertheless, these assays demonstrate that spiders are active under the buffer conditions used in Single

Molecule Fluorescence Microscopy imaging (see below). A limiting factor for increasing the Zn^{2+} concentration above 1 mM is slow spontaneous $Zn(OH)_2$ precipitation at the near-neutral pH used in our studies. We therefore varied the buffer conditions in our Single Molecule Fluorescence Microscopy imaging experiments between 1× SSC with 0-10 mM $ZnSO_4$, 1× HBS buffer (10 mM HEPES, 150 mM NaCl, pH 7.4) with 0-5 mM $ZnSO_4$, and 1× TA-Mg, carefully monitoring (and avoiding) any $Zn(OH)_2$ precipitation.

Further studies have examined the buffer-dependence of 8-17 spider leg cleavage activity in **NICK3.4A+1**; cleavage rates varied from 0.25-1.5 min^{-1} in TA-Mg, TA-Mg with 1 mM Zn^{2+}, and HEPES with 1 mM Zn^{2+} (Taylor, Pei, Stojanovic, unpublished results). In particular, the non-negligible cleavage rate in TA-Mg with no Zn^{2+} has implications for the AFM experiments prior to adding TRIGGER, as discussed above. Finally, these solution-based cleavage assays and the SPR assays (discussed above and in Figure A1.10), while useful for detection of cleavage activity under various conditions, may not be in quantitative agreement with the cleavage rate at the surface of an origami tile, where the locally high density of substrates and other surface effects may have a large impact on the rate-limiting step of this reaction.

SUPER-RESOLUTION PARTICLE TRACKING WITH FLUORESCENCE MICROSCOPY

Overview. For more facile real-time observation of the movement of individual spiders along tracks, we applied super-resolution imaging by total internal reflection fluorescence (TIRF) video microscopy[78]. Four biotin molecules were attached to the underside of the origami to facilitate its immobilization on the avidin-conjugated quartz slide. Experiments were performed using EAC, EABC, and EABD tracks. Spiders were covalently labeled with on average 2.3 Cy3 fluorophores (λ_{ex} 568 nm), and the STOP position was labeled with 6 Cy5 fluorophores (λ_{ex} 672 nm). This labeling scheme allowed us to colocalize spider position relative to its STOP using two-color single-molecule high-resolution colocalization (SHREC)[101] and monitor their relative movement by single-particle tracking[81]. In a typical experiment, spider-loaded tracks were incubated with TRIGGER in the absence of Zn^{2+} ions and then immobilized on the slide

(Figure 2.6). Within 20 min of commencing fluorescence imaging, we added Zn^{2+} to promote spider movement via substrate cleavage. As the 8-17 deoxyribozyme's activity depends sensitively on buffer conditions[102], we optimized our conditions for a combination of best catalytic activity and SMFM imaging quality, obtaining best results from SSC or HEPES with increased Zn^{2+} concentrations and no Mg^{2+} (see below and Figures A1.10-A1.11). The position of a spider on its origami path relative to the START was extracted over time by fitting the diffraction-limited point-spread functions (PSFs) to two-dimensional Gaussians in an up to 80-min sequence of wide-field images (time resolution 15-30 s) with a precision (standard deviation) of 10-30 nm. We controlled for focal drift and developed a consistent set of criteria to distinguish moving spiders from stationary ones as detailed below and in Figures 2.5 and A1.22.

Preparation of Avidin-Coated Microscope Slides. Two 1-mm holes were drilled in each microscope slide (fused silica) to allow for buffer exchange. The slides were immersed in boiling "piranha" solution (5% (v/v) ammonium hydroxide, 14% (v/v) hydrogen peroxide) for at least 20 min, then sonicated for 30 min in 1 M KOH, and flamed for several seconds with a propane torch. The slides were then aminosilanized by immersing them in a 5% (v/v) solution of 3-aminopropyltriethoxysilane (Sigma-Aldrich) in acetone for 1 h, rinsed with acetone, and dried for 1 h at 80 °C. A layer of the bifunctional crosslinking agent *para*-phenylene diisothiocyanate (PDITC) was covalently coupled to the aminosilanized surface by incubating the slides for 2 h in a solution of 0.2 % (w/v) PDITC, 10% (v/v) pyridine in *N,N*-dimethylformamide (spectroscopic grade). The slides were rinsed thoroughly with methanol and acetone, and 70 µL of 0.5 mg/mL avidin (Sigma-Aldrich) was added to each slide, covered with a glass coverslip (VWR), and allowed to incubate for 2 h at room temperature in a closed container above a water bath to avoid drying out. The coverslips were removed, and the slides were washed thoroughly with deionized water, followed by 1 M NaCl plus 40 mM NaOH, and again washed with deionized water, then dried under nitrogen. A flow channel about 2-3 mm wide was made between the drilled holes with two strips of double-sided tape, a coverslip was placed on the tape, and the edges were sealed with Epoxy glue

(Hardman Adhesives). A schematic of the resulting surface structure is shown in Figure 2.6. Slides were stored in an evacuated desiccator at 4 °C for up to four weeks.

Fluorescence Microscopy. For the EAC, EABC, and EABD tracks**,** spider-origami complexes at 10 nM in the annealing buffer were combined with an equal volume of 1 μM to 10 μM TRIGGER strand in water and incubated for 30-60 min on ice in the absence of Zn^{2+} ions. The mixture was then diluted to 10 pM in the imaging buffer: for the EAC track, 1× SSC or HBS; for EABC and EABD tracks,1× TA-Mg. (Note that for the EABC and EABD tracks, the reduced but non-negligible deoxyribozyme cleavage rate in TA-Mg raises the concern that pre-incubation with TRIGGER may allow some spider movement prior to imaging; however, the reduced temperature would be expected to inhibit such movement.) All buffers used for fluorescence imaging were supplemented with 1×-5× OSS to reduce the rate of signal loss through fluorophore photobleaching. The spider-origami complexes were immobilized on avidin-coated microscope slides for imaging.

Samples were imaged at room temperature by a prism-based total internal reflection fluorescence microscope with a 1.2 NA 60× water objective (IX71, Olympus). Cy3 excitation was provided by a 532-nm green laser (ultra-compact diode-pumped Nd:YAG laser GCL-025-S, CrystaLaser) and Cy5 excitation by a 638-nm red diode laser (Coherent CUBE 635-25C, Coherent Inc.). The Cy3 and Cy5 emission signals were separated by a dichroic mirror with a cutoff of 610 nm (Chroma) and projected side by side onto an ICCD camera chip (iPentamax HQ Gen III, Roper Scientific, Inc.). Relay lenses matched the microscope image with the camera focal plane and the IX71 internal 1.6x magnifier (final effective pixel length 133 nm) was used during collection of all traces except EAC 1 and 2 (Figure A1.23) in which no magnifier was used (effective pixel length 196 nm). The donor channel image was passed through a band pass filter (HQ580/60m, Chroma) and the acceptor channel was passed through a long pass filter (HQ655LP, Chroma). A cleanup filter (z640/20, Chroma) was placed at the output of the red laser to reject any extraneous or infrared light. A Newport ST-UT2 vibration isolation table was used in all experiments. After introducing imaging buffer without

oxygen scavenger to the slide flow channel, a small fluorescent background was observed; this was bleached briefly by exposing the slide to excitation light from both lasers until the background stabilized. The origami sample with oxygen scavenger was then introduced into the sample channel in the dark, allowed to incubate for 2-10 minutes, and the excess flushed out with fresh imaging buffer. The sample was imaged at room temperature with excitation from both lasers using a 1- to 2.5-s signal integration time and a 12.5- to 27.5-s delay (i.e., 2-4 frames per min). After 0-20 min of imaging, depending on the experiment, the same imaging buffer containing or lacking $ZnSO_4$ was introduced into the flow channel, and the sample was imaged for an additional 60-70 min. For substrate-covered EAC tracks in SSC, the concentration of $ZnSO_4$ introduced was either 0 or between 1 and 10 mM. For substrate-covered EABD and EABC tracks, the $ZnSO_4$ concentration was 0 or 1 mM $ZnSO_4$. For substrate-covered EAC tracks in 1× HBS, the $ZnSO_4$ concentration was 0 or 5 mM.

EAC track origami with product-covered tracks were prepared and imaged as described above for the substrate-covered EAC track origami in SSC buffer. Due to concerns about releasing spiders from the START prematurely on product tracks where the walk is independent of cleavage activity, experiments were also conducted in which the spider-origami assemblies were not incubated with TRIGGER 30-60 min prior to imaging, as described above, but instead SSC imaging buffer containing 1 mM $ZnSO_4$ and 10 μM TRIGGER was added to the sample channel 10-15 min before imaging. In both types of experiments, $ZnSO_4$ was not introduced until immediately prior to imaging by fluorescence microscopy.

Fitting and Filtering of Particle Tracking Data. Point spread functions (PSFs) of fluorescence emission from individual spiders and origami were imaged by fluorescence microscopy, and their relative positions tracked through time by fitting 2-D Gaussian functions to the PSFs. First, PSFs from Cy3 (spider) and Cy5 (origami) were imaged on spectrally separated halves of the ICCD camera using WinView32 software (Roper Scientific, Inc.). PSFs were identified in the ICCD output and paired with their corresponding partner using methods described previously[103], resulting in intensity traces such as those shown in Figures A1.23-A1.24 that reflect the total photon count

per movie frame for each PSF over time. The Cy3 and Cy5 channels were registered with a locally weighted mean mapping[101] using fluorescent beads that appear in both channels (Fluospheres, red fluorescent (580/605), 0.2 μm, Molecular Probes FluoSpheres F8801), to establish with ~50-nm accuracy that the Cy3 PSF in each pair was located within 200 nm of its Cy5 partner. To ensure adequate signal intensity and duration for tracking, candidate PSF pairs were only included in the analysis if they met both of the following criteria:

1. Cy3 and Cy5 signal of more than 1,000 photon counts per frame for at least 25 min (1-33% of all pairs fulfilled this criterion per experiment)
2. No erratic signal intensities such as from excessive blinking or nearby unresolved PSFs (23-95% of all remaining pairs fulfilled this criterion per experiment)

Traces that were discarded based on low or absent signal intensity from either Cy3 or Cy5 likely resulted from incompletely labeled spider-origami complexes, fragmented or disassembled complexes, or other fluorescent contaminants. Each PSF in the remaining pairs (0.4-22% of all candidate pairs) was fit, frame-by-frame, with a two-dimensional Gaussian function (Figure 2.5d) of the form:

$$f(x,y) = z_0 + A\exp\left\{\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} \qquad (1)$$

The position values σ_x and σ_y from Gaussian fitting of each Cy5 PSF (Figure 2.5e,g) were subtracted, frame-by-frame, from those of its corresponding Cy3 partner. The resulting difference trajectory was then plotted against time for each Cy3-Cy5 pair to show the motion of each spider relative to its Cy5-labeled STOP position (Figure 2.5f,h). This subtraction served as a necessary internal drift control since, as shown in Figure 2.5, there was often significant drift through the x-y plane in the course of a typical 30-80 min movie. Brief aberrant position measurements, such as those caused by transient binding of nearby fluorescent contaminants, were identified by a large distance from the median position (> 3 standard deviations in the x or y direction) or sudden

displacements of >100 nm within a single frame, and removed. Focal drift throughout an experiment, if severe, sometimes resulted in an apparent motion of Cy3 relative to Cy5 (data not shown). This focal drift was evident visually from the original video image as well as from very asymmetric PSF shapes during Gaussian fitting. Such traces were also discarded.

Probable moving spiders were selected using the following criteria:

1. Relative motion of Cy3 and Cy5 > 45 nm, corresponding to 2-3 times the standard deviation in individual position measurements (33-44% of all fitted pairs fulfilled this criterion per experiment)
2. No discontinuities in position, i.e., sudden jumps in position of 45 nm or greater (89-100% of all fitted pairs fulfilled this criterion per experiment)
3. Apparent movement < 45 nm prior to zinc addition (88-100% of all fitted pairs fulfilled this criterion per experiment)

This process is illustrated in Figure A1.22 for representative experiments from the EAC, EABC, and EABD constructs. The resulting spiders (22-39% of all fitted traces) are included in Figure A1.23. Examples of spider trajectories that did not satisfy all three of these criteria are also shown in Figures A1.23 (EAC Tier 2) and A1.24 (all stationary spiders observed on the EAC track in HBS). A statistical summary of this filtering process for the EAC, EABD, and EABC tracks is presented in Table A1.3.

Representation of Spider Trajectories. To smooth the trajectories for presentation, a 16-frame rolling average was applied separately to the trajectories of Cy3 and Cy5 before subtracting them for drift correction (black line in Figures A1.23b and A1.24b). The error bars shown in Figures A1.23b and A1.24b are the standard deviations of the raw trajectory from the temporally corresponding points in the smoothed trajectory. For ease of viewing in Figure 2.7a and Figure 2.8, the trajectories were instead smoothed with an 8-frame rolling average followed by a 4-frame sequential average before drift correction.

Measurement of Displacement. Net displacement was determined as follows for motion of each spider on the EAC track. An initial position (x_0, y_0) was defined as the arithmetic mean of the first 16 position measurements after the time t_{zinc} at which $ZnSO_4$ or control buffer lacking zinc ions was added (t = 0 min in Figures A1.23-A1.24). For traces containing data prior to t_{zinc}, the initial position was instead calculated as the mean of the 16 position measurements centered on t_{zinc} (i.e., the interval from frame -7 to frame 8, where t_{zinc} occurs at frame 0). The center time coordinate of this averaged initial position (x_{start}, y_{start}) was designated t_{start} (i.e., the interval from frame -7 to frame 8, where t_{zinc} occurs at frame 0). The distance of (x_{start}, y_{start}) from each subsequent position measurement (x_i, y_i) was then calculated to obtain the spider's net displacement over time (green line, Figures A1.23c and A1.24c). As has been noted in similar distance determinations[101], these displacement measurements are artificially increased when equal to or less than the noise level (hence why displacement typically does not equal 0 nm near t = 0 min). Therefore, an analogous displacement vs. time curve was calculated from the smoothed trajectory (black line in Figures A1.23b and A1.24b) and was plotted as a black line in Figures A1.23c and A1.24c. This smoothed displacement has a value of zero at t_{start}, resulting in a systematic deviation from the noise-inflated raw curve at low displacements. The time of stopping t_{stop} was defined as the time coordinate of the first local maximum in the smoothed displacement curve that approaches within 20 nm of the global maximum in the smoothed displacement curve (considering only the interval from t_{zinc} to the end of the trace). The value of 20 nm is a typical standard deviation in our position measurements. The total net displacement d (inset box, Figures A1.23b and A1.24b) was then defined as the smoothed displacement value at t_{stop}. The time of travel Δt was defined as the difference $t_{stop} - t_{start}$, and the mean magnitude of velocity was calculated for each EAC spider as $v = d/\Delta t$ (box, Figures A1.23b and A1.24b). The resulting displacement vs. time plots are shown in Figures A1.23c and A1.24c.

Interpretation. In some traces we observed movement before addition of Zn^{2+}; we could not determine whether these represented spiders walking in the absence of Zn^{2+} or were due to other causes. We also observed several moving traces that exhibit net

displacements significantly smaller than others, which similarly is consistent with spiders having finished (part of) their tracks early, taking the wrong direction after walking in the absence of Zn^{2+}, prematurely stopping or stalling on the track, and/or taking backward steps onto product. These issues are discussed in more detail below. In the following, we enumerate all independent lines of evidence that these time traces represent genuine walking spiders:

1. The highest density of PSFs we observed in each channel with $\geq$ 1,000 photon counts over at least 25 minutes was 0.03 μm^{-2}. Given this density, the probability that a Cy3 and Cy5 PSF will colocalize to within 200 nm of one another by coincidence is 0.9%[104]. However, in each experiment we observe that, on average, 31% of PSFs in one channel are colocalized with a PSF in the other channel. This strongly suggests that the majority of signals originate from spiders bound to origami.
2. We find most of the trajectories longer than 45 nm to be consistent in length and shape with a progressive walk on the respective track design (Figures A1.22 and A1.23b). In particular the trajectories observed on the EAC track in SSC buffer, which are nearly linear and often stop nearly 100 nm from the starting position (EAC 1, 2, 4, 5, 9, 12, 13, and 15), are in good agreement with expectations based on the track design.
3. Comparison of experiments to negative controls (such as in Figures A1.22d and A1.24), rules out instrument drift as the sole source of the observed spider motion.
4. Moving and non-moving spiders are seen alongside each other in experiments conducted in the presence of both Zn^{2+} ion and release strand (Figure A1.22a-c), providing further fiduciary markers and a strong argument against instrument drift as the cause for movement.
5. Ensemble MSD (Figure 2.9a) and RMSD plots (Figure 2.7c) of the 15 Tier 1 EAC spiders (Figure A1.23) are consistent with an approximately 100-nm walk across the prescribed linear substrate track.

Especially when considered in combination with the results from our AFM studies, the fluorescence microscopy data are most consistent with processive walking of individual spiders on DNA origami. While the stopping distances are not strong evidence (filtering precluded walks shorter than 45nm, and photobleaching may have precluded having many walks longer than 100nm), this interpretation is confirmed by control experiments lacking zinc in the buffer and on product tracks, as discussed below.

The large percent of spiders moving less than 45 nm (22-67 % of all PSF pairs fit to Gaussians in a given experiment) likely results from some combination of the following: 1) immobile contaminants that fluoresce in both channels, thus having the appearance of a colocalized Cy3-Cy5 pair; 2) a substantial fraction of inactive or slowly cleaving spider legs, especially in SSC + 1 mM Zn^{2+}, 3) failed or delayed release of a spider from the START position, 4) spiders binding initially at the STOP instead of the START position (though precautions against this were taken in the assembly of origami-spider complexes), or 5) undirected, random diffusion of a spider on previously cleaved or damaged substrate. As we cannot distinguish between these possibilities, the estimated percent of non-moving spiders must be taken as an upper bound.

Most of the trajectories from the EAC track show clearly biased, generally linear motion with few or no discontinuities in displacement (Figure A1.22). However, some spiders exhibit non-monotonically increasing displacements with time (e.g. EAC 5) that could have resulted, for example, from spiders taking steps backwards onto cleavage product. Furthermore, some trajectories exhibit unexplained irregularities in the 2-D motion trajectory, displacement or velocity measurements (Figure A1.23, EAC 16-19). These issues are described in the captions above the respective traces in Figure A1.23. For instance, the net displacement values of EAC 16 and 18 at $t > 0$ are less than 45 nm and therefore less reliable. Some putative spiders (EAC 17, 18, and 19) also show significant displacement before addition of Zn^{2+} at $t = 0$. Slight focal drift or an instrumental perturbation might have resulted in an apparent displacement between the PSFs in these traces, particularly during the addition of Zn^{2+}-containing buffer. Although this addition was performed slowly and carefully (generally at a rate of 1 mL/min or less), it occasionally brought the image out of focus. Such slight focal drift could also

affect the measurement of net displacement values. For example, similar influences could have given rise to the few trajectories that showed an apparent net displacement larger than the track length of 110 nm (e.g., EAC 10). Alternatively, although the 8-17 deoxyribozyme legs are inactive in 1× SSC lacking $ZnSO_4$ (see Figure A1.11), the spider might still exhibit slow diffusion on a surface of substrate. It is also possible that some origami assemblies exhibit rotational dynamics relative to the slide that contribute to the observed motion of PSFs. Finally, the calculation of net displacement for some spiders is likely biased by early photobleaching which may prevent observation of the entire trajectory of the spider (see, for example, EAC 3).

Calculation of Ensemble Mean Square Displacement and Root Mean Square Displacement. To characterize the ensemble behavior of spiders, ensemble mean square displacement (MSD, Figure 2.9) and root mean square displacement (RMSD, Figure 2.7c-d) versus time plots were generated. To calculate the individual displacements plotted in Figure 2.7c-d, and used to calculate the MSD and RMSD, an initial position (x_{start},y_{start}) was first calculated as the arithmetic mean of the 16 points of the raw trajectory closest to the time at which $ZnSO_4$ or control buffer lacking zinc ions was added, t_{zinc} (i.e., the interval from frame -7 to frame 8 if data were taken before t_{zinc}, or the interval from frame 1 to frame 16, if data acquisition began at t_{zinc}). Trajectory data were averaged separately for Cy3 and Cy5 in (sequential) one-minute intervals, and the averaged trajectory of Cy5 was subtracted from that of Cy3 to correct for microscope stage drift. Each displacement value was then calculated as the distance of each averaged position (x_i, y_i) from the initial position (x_{start},y_{start}). This same procedure was applied to data acquired at the rate of two, three, and four frames per minute. The displacement of each spider for each 1-minute time interval was squared and then averaged across all spiders within a given dataset to yield the ensemble MSD as a function of time. The square root of the ensemble MSD was calculated for each time interval to yield the ensemble RMSD as a function of time. Note that outliers were removed from the raw data as described above (Filtering and Fitting of Particle Tracking Data).

For comparison with Tier 1 EAC spiders, MSD and RMSD versus time plots were also generated from the 7 EAC spiders in a no-zinc control experiment on the EAC substrate track in 1× SSC. These spiders were subjected to the same selection criteria as the Tier 1 EAC spiders except that they were not required to move $\geq$45 nm for inclusion in the MSD plot (by this criterion, no moving spiders were observed in this control). Both of these MSD plots are shown in Figure 2.9a, and the RMSD plots shown in Figure 2.7c.

In an attempt to determine the relative impact of substrate cleavage on the motion of EAC spiders (compared to that of the presence of START and STOP sites), control experiments were conducted on EAC tracks covered with cleavage product instead of substrate in 1× SSC and 1 mM $ZnSO_4$. For consistency, identical experimental procedures were applied, including addition of Zn^{2+} immediately prior to imaging (although product walks are not expected to be zinc-dependent). To reduce the risk of bias in comparing these two types of experiments, we employed a less stringent set of selection criteria than those described above. Specifically, all spider trajectories with Cy3 and Cy5 signal intensity above an arbitrary cutoff were retained. Individual data points in a trajectory were discarded if the ellipticity E exceeded 0.3 ($E = 1 - w_{minor}/w_{major}$, where w_{minor} and w_{major} are the full widths at half maximum along the major and minor axes of the fitted 2-D Gaussian function, respectively). Position measurements greater than three standard deviations from the median of all position measurements within a trace in either the x- or y-direction (or 500 nm from the position of the spider when zinc was added, whichever is smaller) were regarded as outliers and discarded. An application of these more inclusive criteria first to our substrate-covered track data resulted in 85 traces that were converted to the ensemble MSD and RMSD versus time plots described above; the results are shown in Figures 2.7d and 2.9b (see Table A1.4 for full statistics of selection based on these criteria). The roughly twofold difference in steepness from and less pronounced curvature than the substrate MSD plot in Figure 2.9a are likely due to the inclusion of a larger number of slow-moving or stationary spiders in Figure 2.9b.

We found that the ensemble MSD versus time plot for the product-covered linear EAC track in 1× SSC and 1 mM $ZnSO_4$, generated using the more inclusive selection criteria above, dramatically depends on whether the TRIGGER was added 10-15 min (short incubation) or 30-60 min before the experiment (long incubation, similar to our protocol for the linear substrate track). In the former case, the MSD plot of 18 spiders increases non-linearly with a concave up slope curvature greater than that seen for the substrate track, while in the latter case, linear behavior with a much shallower slope is observed in an MSD plot of 29 spiders (Figure 2.9b). Since Zn^{2+} is not predicted to be required for diffusive walking on a product surface, a long pre-incubation with TRIGGER is expected to allow many spiders to prematurely walk and possibly be captured by the STOP site prior to the onset of imaging, resulting in a much lower net displacement over the time window of observation. However, when the TRIGGER is added 10-15 min before the experiment, the spider release from the START position may become rate-limiting to effect an initial delay followed by Brownian diffusion of the released spiders along the track. This possibility prevents a direct comparison of the MSD plot of the latter experiment with that of the linear substrate track in Figure 2.9b (see also discussion of Monte Carlo simulations below). We therefore conclude that we cannot distinguish the behavior of spiders on substrate- and product-covered tracks with confidence from these experiments except insofar as they respond differently to pre-incubation with the TRIGGER.

As an additional control, MSD versus time plots (Figure 2.9c) were created for the EAC spiders in HBS buffer shown in Figure A1.24. The MSD plot begins with the addition of HBS buffer containing 0 mM (EAC 1-21H) or 5 mM (EAC 1-16HZ) $ZnSO_4$. As for the MSD of the Tier 1 EAC substrate track spiders observed in 1× SSC, the presence of zinc increases the slope of the MSD versus time plot for spiders in 1× HBS, suggesting that the movement of spiders on the EAC substrate track is zinc-dependent in these buffers.

Monte Carlo simulations of spiders on EAC track. To aid in the interpretation of our experimental results, Monte Carlo simulations of simplified models of spiders walking on EAC tracks were conducted as follows. The spider consists of three legs, each of which

can exist in an unbound state or bind a specific substrate or product within a 2-dimensional array based on the EAC track dimensions. The three legs are constrained to bind substrates within 10 nm (an estimated effective leg span) of all other bound legs, and can bind any such substrate with equal probability as long as that substrate is not already bound by another leg. The spider's body position is taken as the arithmetic mean of the positions of all legs bound to the substrate array.

At each time step of the simulation, each leg acts independently to perform one or more of the following actions:

- If bound to a substrate, it can cleave it or not.
- If it is bound to a substrate or product, it can dissociate or remain bound.
- If it is unbound, it binds a substrate or product within 10 nm of other bound legs within the same timestep of the simulation.

Each of the first two of these actions has an associated probability P_i that can be related to an effective first-order rate constant k_i according to $P_i = (1-\exp(-k_i*t))$, where t is the length of a timestep, chosen here as 1 second. There are thus three adjustable probability parameters: the probability of cleaving a bound substrate (P_{cleave}), of dissociating from a bound substrate ($P_{off,substrate}$), and of dissociating from a bound product ($P_{off,product}$). The legspan is a fourth adjustable parameter. Note that, for simplicity, it is assumed that hybridization to a new site is instantaneous compared to a timestep, and independent of whether that substrate has been cleaved.

At the beginning of each simulation, the spider is positioned with all three legs bound to substrates (or products) within 10 nm of one end of the track corresponding to the START position. At the opposite end of the track are six non-cleavable substrates which constitute the STOP site. The spider's legs are then allowed to freely cleave, dissociate, and bind substrates and products. The legs must remain within 10 nm of the START end until the spider is released by a TRIGGER event which occurs with a probability $P_{release}$. Each simulation ran for 35 min (2100 time steps).

Probabilities for cleavage of and dissociation from substrates were determined from effective first-order rate constants: $k_{cleave} = 1\ min^{-1}$, $k_{off,substrate} = 0.002\ s^{-1}$, $k_{off,product} = 0.2$

s^{-1}. These rate constants are within one order of magnitude of those determined by bulk fluorescence experiments in 10 mM HEPES, 150 mM NaCl and 1 mM ZnCl2, pH 7.4 (data not shown), preserve the experimentally observed ratio $k_{off,substrate}/k_{off,product}$, and are compatible with sound principles of nucleic acid thermodynamics and kinetics[105].

Ensemble MSD versus time curves (Figure 2.9d) were calculated based on the current spider's body position relative to its first observed position, and are an average over either 20, 80 (thin lines), or 1,000 (thick lines) simulated spider trajectories. When allowed to walk on a linear EAC track containing cleavable substrates, simulated spiders yielded an MSD plot with positive (concave up) curvature, similar to the slightly positive curvature seen in the experimental plots (Figure 2.9a-c). In contrast, when walking on an EAC track with cleavage product, the spiders yield an MSD curve that first increases linearly, then asymptotically approaches a maximum value corresponding to the STOP position (spiders undergoing an undirected walk are still able to bind and become trapped at the non-cleavable STOP). However, positive MSD curvature is also obtained from a product track simulation if the spider is released from the START with a half-life of 10 min (rather than immediately), so we cannot rule out a lag phase as contributing to the positive curvature of the experimentally observed ensemble MSD versus time plots. Furthermore, simulations of only 20-80 spiders yielded a fairly broad range of MSD behaviors, often obscuring the idealized curvature and slope; this suggests that our experimental MSD plots, constructed from 85 or fewer spiders, may not represent the fully converged behavior of the system.

The qualitative features of MSD versus time plots generated from these simulations, such as curvature (linear or concave-up) and maximal extent of increase, are robust to variations in k_{off}, and k_{cleave} at least one order of magnitude about their experimentally observed values. However, the precise values of the parameters can affect the slope of the MSD versus time plots. The relative slopes of product and substrate walks are quite sensitive to the effective legspan parameter. Furthermore, the introduction of unequal association probabilities for substrate and product can affect the slope and curvature of these plots. Still, these simulations show that the observed ensemble MSD versus time

behavior for the linear EAC substrate track is consistent with the proposed mechanism of spider locomotion based on reasonable kinetic parameters, even if other mechanisms cannot be conclusively ruled out.

2.3 Results

The DNA walkers chosen for this work, called “molecular spiders”, comprise an inert body and multiple catalytic “legs”. Specifically, here we use three-legged spiders with a streptavidin body. Spider legs are adapted from DNA enzyme 8-17 that binds and cleaves single-stranded oligodeoxynucleotide substrates with a single ribose moiety into two shorter products that have a lower affinity for the enzyme[106]. In the context of substrates that are immobilized at sites on a surface, spider behavior can be modeled using local rules[107]: a leg bound to substrate will cleave it at a low rate; a leg bound to product will detach at an intermediate rate; and a free leg will quickly bind (with little or no bias) a nearby substrate or product. For a multipedal spider positioned at the interface between regions of product and substrate, these rules predict that after a given leg cleaves and then lifts, it will by trial-and-error search out a nearby substrate to bind, thus moving the spider’s body toward the substrate region while enlarging the product region behind it. On 2D surfaces or in a 3D matrix, such spider movement results in a random walk with memory of visited sites, while on a 1D linear track it results in directed motion as the substrate is consumed. Crucially, unlike related “burnt bridge” Brownian ratchet mechanisms used in DNA walkers[66,94,95,97,108] and observed in nature[109], these local rules predict that multipedal spiders will not readily dissociate even from tracks consisting exclusively of product strands, and indeed will perform a rapid unbiased random walk there until they again encounter substrate.

Considering spider legs to be simultaneously sensors that detect nearby oligonucleotides and actuators that modify their environment to inhibit reverse motion, we exploited this sensor-actuator feedback to design prescriptive landscapes that direct

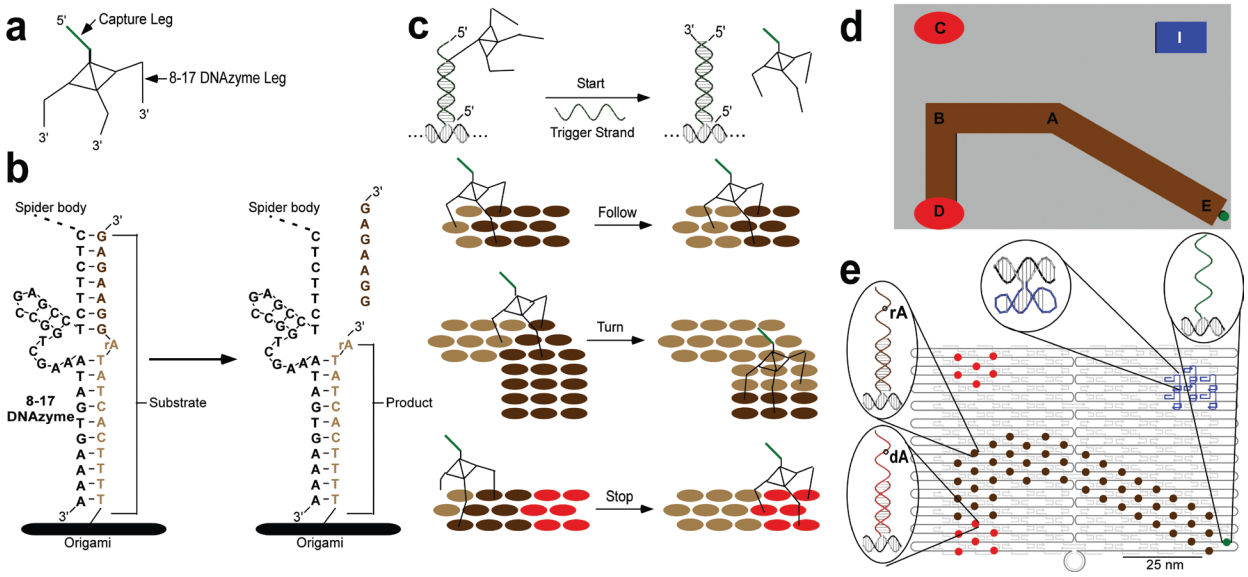


Figure 2.1. Schematics of deoxyribozyme-based molecular walker and prescriptive origami landscapes. (**a**) The NICK3.4A$_{3+1}$ spider consists of a streptavidin core that displays a 20 base ssDNA that positions the spider at the start (green), and three deoxyribozyme legs. (**b**) The 8-17 deoxyribozyme cleaves its substrate at an RNA base creating two shorter products (seven and eleven bases). Dissociation from these products allows legs to associate with the next substrate. (**c**) Spider actions: after release by a 27-base ssDNA trigger, the spider follows the substrate track, turns, and continues to a stop site (red). (**d**) Schematic of the DNA origami landscape with positions A-E labeled; track EABD is shown. (**e**) A representative origami landscape shows the start position (green), the substrate track (brown), stop and control sites (red), and a topographical marker (blue).

the spiders' motion along a predefined path (Figures 2.1c and d). A spider traversing this landscape of oligonucleotide substrates can sense the set of available cues within its reach and take action accordingly. Prescriptive landscapes were constructed using the DNA origami scaffolding technique[53]. The scaffold consists of a 7249-nucleotide single-stranded DNA folded by 202 distinct staple strands into a rectangular shape roughly 65x90x2 nm in size and with 6-nm feature resolution (Figure 2.1e, Figure A1.1). Each staple can be extended on its 5′ end with probes that recruit substrates, products, goal and control strands[69].

We designed pseudo-one dimensional tracks on origami of about spider width (three adjacent rows of substrates, Figure 2.1d). Tracks are coded by a sequence of points (A, B, C, D, E; i.e., on an ABD landscape the spider starts at A, and passes through B before ending at D). Staples were modified to position: (1) A START oligonucleotide, used to position a spider at the start of the experiment, that is complementary to a TRIGGER oligonucleotide used to release the spider[110] (the "start" action); (2) Substrate TRACK probes to capture the 5′ extension on substrates forming the TRACK (directing the "follow" and "turn" actions); (3) STOP probes complementary to the 5′ extension on STOP strands (non-chimeric and uncleavable analogs of the substrate) that do not influence directional movement but trap spiders to prevent them from walking backwards after completing the track (the "stop" action); (4) CONTROL probes (identical to the STOP, but disconnected from the track), used to assess the extent to which free-floating spiders are captured directly from solution; and (5) MARKER oligonucleotides based on inert dumbbell hairpins, aiding in origami classification within atomic force microscopy (AFM) images (Figure 2.1e). To position spiders at START sites, we replaced one of the four catalytic legs of the NICK-4.4A[99] spider with a tethering oligonucleotide (Figures A1.2-A1.5) partially complementary to the START oligonucleotide.

To estimate the efficiency of spider motion directed by the TRACK, we defined and tested four paths with no (EAC), one (ABD), or two (EABD, EABC) turns (Figures A1.6-A1.9, 2.2). Our basic procedure consisted of: (1) Assembling the origami; (2) attaching the spider to the START site; (3) adding TRACK, STOP, and CONTROL strands to

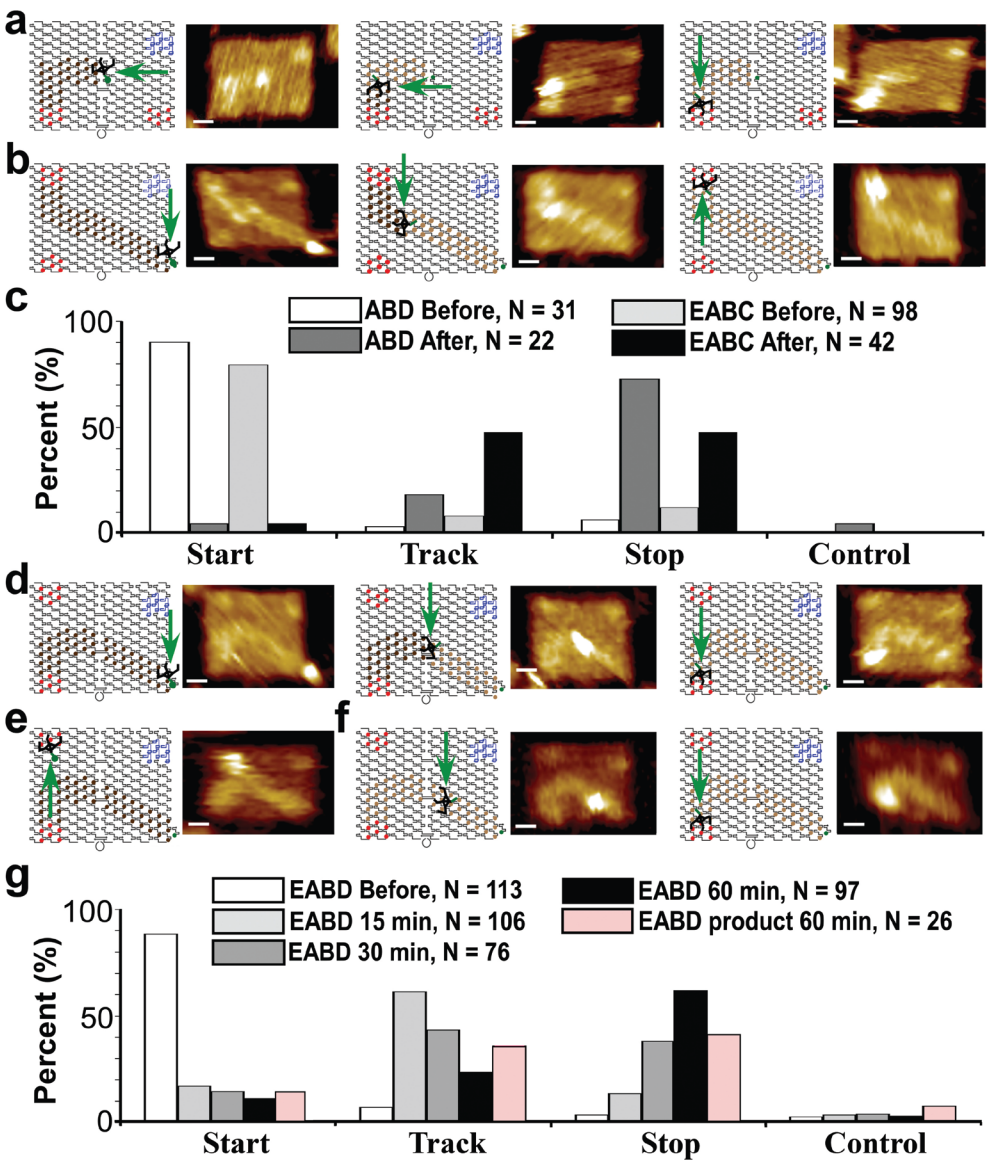


Figure 2.2. Results of spider movement along three tracks with schematics and AFM images of the spider at the start, on the track, and at the stop site. (**a**) ABD track. (**b**) EABC track. (**c**) Graph of ABD and EABC spider statistics before and 30 minutes after release. (**d**) EABD track. (**e**) EABD track with spider on control. (**f**) EABD product-only track. (**g**) Graph of the EABD spider statistics before, and 15, 30 and 60 min after release, and 60 min after release on the EABD product-only track. All AFM images are 144 x 99.7 nm, the scale bar is 20 nm. Legend text indicates the number of origami with a single spider that were counted for the given sample.

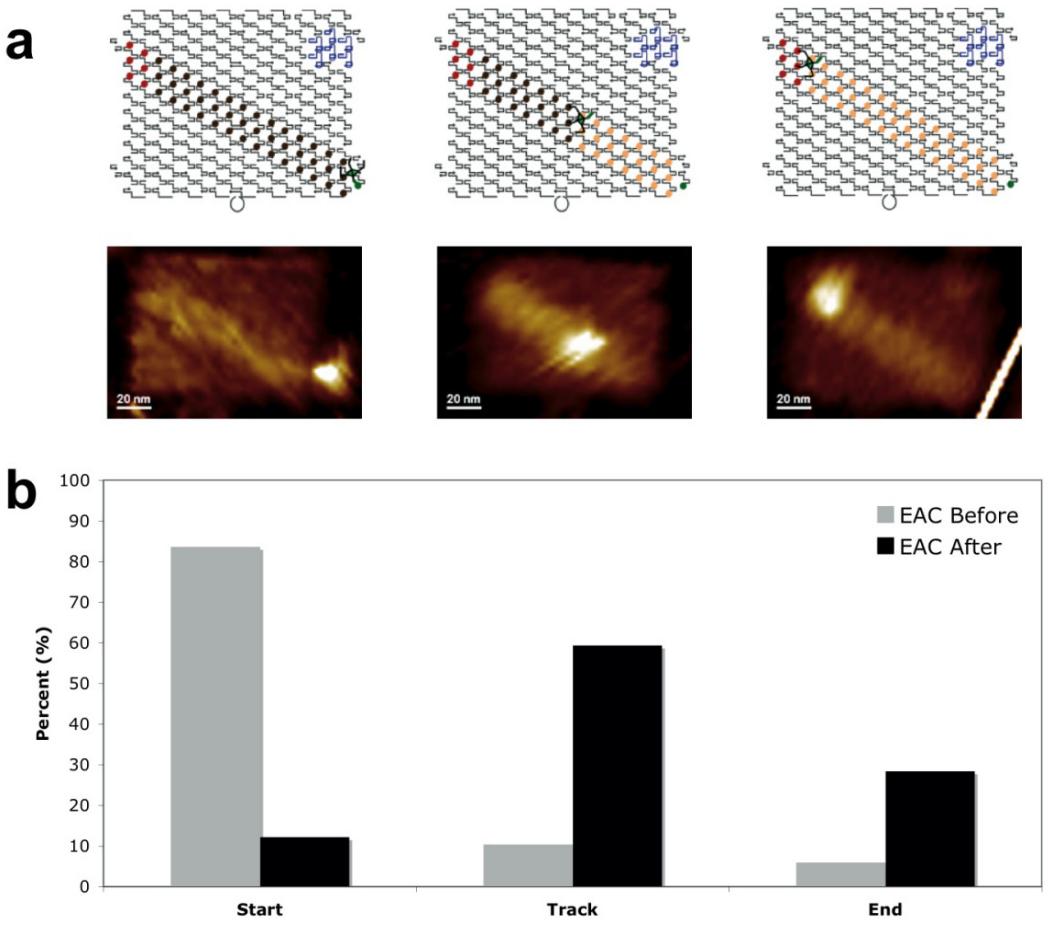


Figure 2.3. Schematics, AFM images and Graph of EAC before vs. after. (**a**) Schematics and AFM images of the EAC walk before addition of TRIGGER and 30 min after addition of TRIGGER. (**b**) Statistical graph of EAC before vs. after.

complete the landscape; and (4) initiating an experiment by releasing the spider through addition of TRIGGER and 1 mM Zn^{2+} cofactor[102] (Figures A1.10-A1.11). We sampled the origami solution before and after spider release, and imaged individual samples by AFM to determine the locations of spiders. We scored only "face-up" origami (substrates projected away from mica) to avoid artifacts and developed procedures to minimize readout bias (see Materials and Methods for details).

In all samples imaged before spider release, 30-40% of the assembled origami carry at least one spider, 80-95% of which are singly occupied, and of these 80-90% bound their spider at the START position (Table A1.1 and Figures A1.12-A1.19). Upon adding trigger, all four landscapes with substrate tracks showed that the fraction of spiders at the START diminishes with a concomitant increase in spiders observed on the STOP sites (Figures 2.2c,g and 2.3). A spider's ability to reach the STOP sites decreased with increased TRACK length and with decreased time of incubation in solution. In time-lapse experiments on a long path (EABD, spanning ~ 90 nm) we observed a gradual increase of up to 70% of spiders on STOP sites within 60 min (Figure 2.2c,g). A short path (ABD, ~ 48 nm) was completed to the same extent within 30 min.

We captured one series of AFM images of a spider moving along an origami track (Figure 2.4). The rate of spider movement (~90 nm over 30 min, with approximately 6 nm per three parallel cleavage events) was consistent with the processive cleavage rates (~1 min^{-1}) of spiders on a 2D surface as obtained by SPR (Figure A1.10). More systematic sequential imaging proved difficult due to mica's inhibitory effects on the spider.

We can eliminate deviations from the proposed mechanism of spider motion as major contributors to these results. First, to test that spiders can indeed traverse product tracks by means of unbiased random walks, we challenged spiders with EABD origami in which the substrate was replaced by product on the TRACK. Spiders still reached the STOP sites albeit more slowly (Figure 2.2f,g), as expected from purely Brownian spider movement even if individual steps are somewhat faster[107]. Second, we wished to confirm that spiders don't often 'jump'; if all three legs simultaneously dissociate

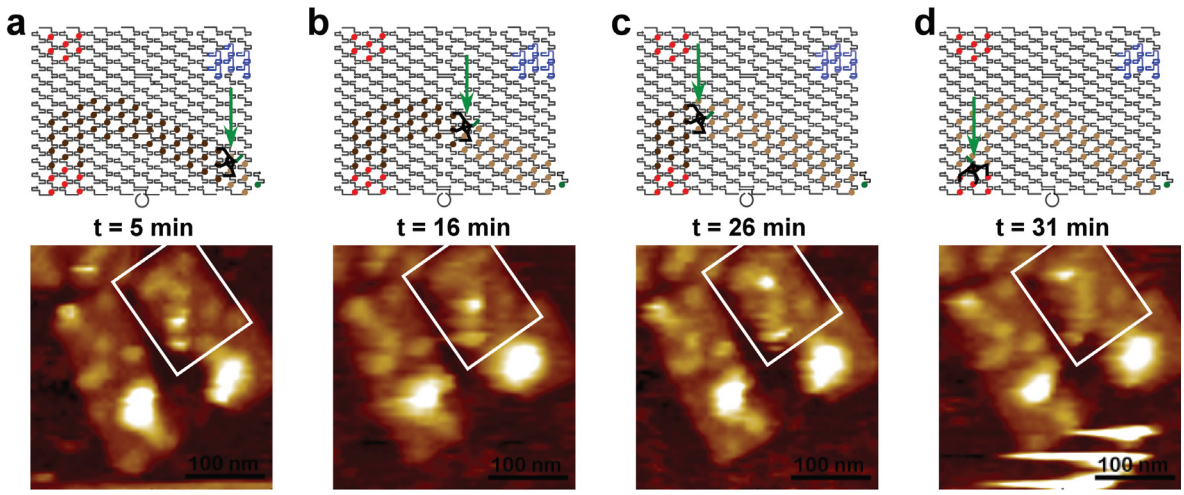


Figure 2.4. AFM movie of spider movement. Schematics and AFM images of the spider moving along the EABD track at 5 min (**a**), 16 min (**b**), 26 min, (**c**) and 31 min (**d**) after trigger was added. AFM images are 300 x 300 nm and the scale bar is 100 nm.

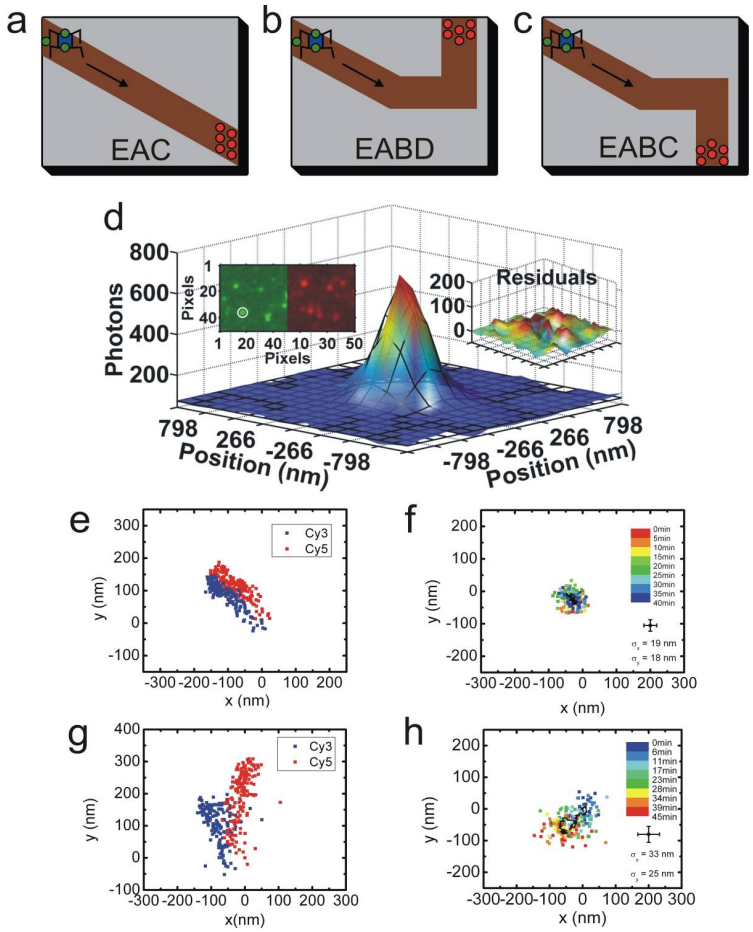


Figure 2.5. Schematic representation of the EAC (**a**), EABD (**b**), and EABC (**c**) constructs for fluorescence microscopy. The spider is labeled with 2-3 Cy3 molecules (green) and the STOP with up to 6 Cy5 molecules (red). PSFs from spider-origami pairs are imaged over time and fit, frame-by-frame, to a 2-D Gaussian function; the fit has low residuals (**d**). The coordinates of each PSF exhibit significant drift through time (**e, g**) which is corrected by subtracting the coordinates of Cy5 from its proximal Cy3. The resulting coordinate plots (**f, h**) track the motion of each spider relative to its STOP position. In absence of Zn^{2+} but in presence of TRIGGER in SSC buffer, primarily stationary spiders are observed (**e, f**); the standard deviations σ_x and σ_y give an estimate of precision in position measurements. In contrast, a spider incubated with TRIGGER and zinc in SSC (**g, h**) shows a distinctly biased pattern of motion when subtracted (**h**). The trajectory in (**e**, **f**) corresponds to trajectory 4 in Figure A1.22d, and the trajectory in (**g**, **h**) corresponds to EAC 2 (Figure A1.23).

before any leg reattaches, a spider could completely dissociate from the origami and subsequently reattach elsewhere at random. Evidence against frequent jumping (or an excess of spiders in solution during the initial assembly stage) comes from the low level of spider occupancy at CONTROL sites in both substrate and product track experiments (Figure 2.2c,e,g) and the stable proportions of unoccupied and multiply-occupied origami (Table A1.1; both before and after the addition of trigger, 5-10% of origami displayed more than one spider on its track). In contrast, when spiders were released on ABD landscapes with no TRACK strands, after 30 min we observed an equal distribution between STOP and CONTROL sites (Figure A1.20 and Table A1.2), as expected for a process that involves spider dissociation from and random rebinding to the origami. In independent ensemble experiments using surface plasmon resonance, we observed that up to 15% of spiders may dissociate from a non-origami 2D product-covered surface within 60 min under flow conditions (Figure A1.21). On similar substrate-covered surfaces, spiders show an average processivity of ~200 substrates before being removed by flow (Figures A1.10 and A1.21). Together, these results rule out that spiders move predominantly by jumping; there is insufficient jumping even on product tracks to explain the 50-70% occupation of the STOP sites after walks on ABD, EABC, and EABD substrate tracks.

For a more facile real-time observation of the movement of individual spiders, we applied particle tracking by super-resolution total internal reflection fluorescence (TIRF) video microscopy[78] (Figure 2.5). Four biotin molecules were attached to the underside of the origami for immobilization on the avidin-coated quartz slide (Figure 2.6). Spiders were covalently labeled with on average 2.3 Cy3 fluorophores, and STOP sites were labeled with 6 Cy5 fluorophores. The labeling allowed us to monitor changes in spider position relative to the STOP site by two-color fluorescent particle tracking[81,101]. In a typical experiment, spider-loaded tracks were incubated with TRIGGER and immobilized on the slide (Figure 2.6), then Zn^{2+} was added to promote spider movement via substrate cleavage. Recognizing that the 8-17 activity depends on buffer conditions[102], we obtained the best results from SSC or HEPES with increased Zn^{2+} concentrations but without Mg^{2+} (Figures A1.10-A1.11).

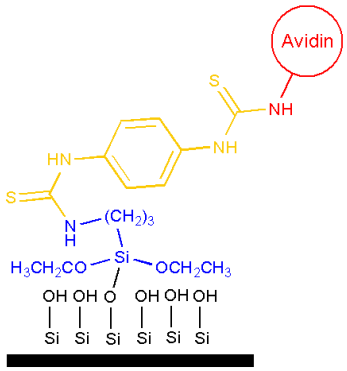


Figure 2.6. Preparation of microscope slides. Surface coating of the microscope slide, showing the aminosilane (blue), PDITC (yellow), and covalently bound avidin layers (red).

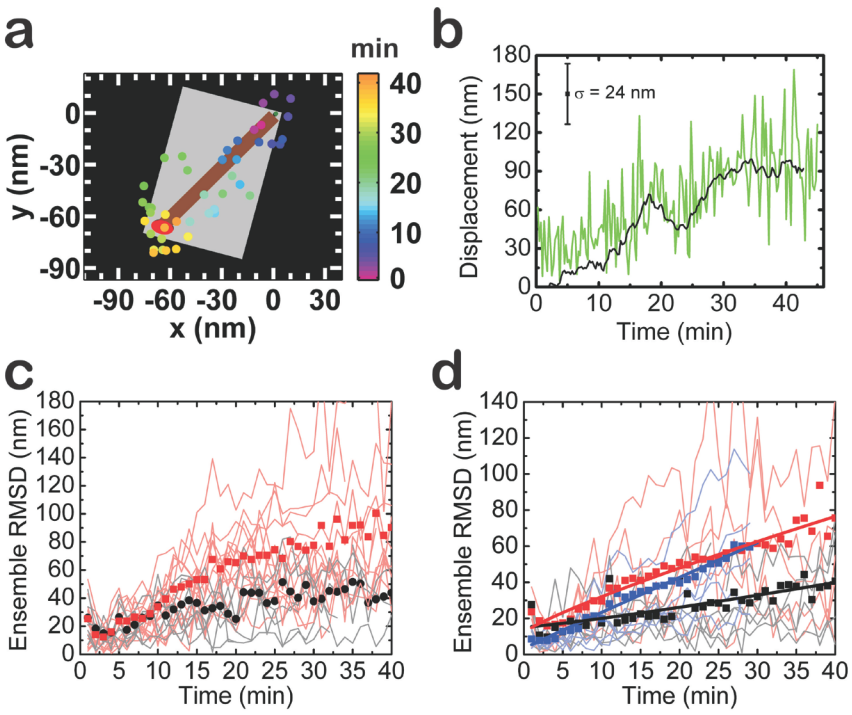


Figure 2.7. Spiders imaged on origami tracks in real-time using super-resolution TIRF microscopy. (**a**) Position-time trajectory of a selected spider (EAC 2, Cy3-labeled) on the EAC substrate track. The position as a function of time is represented by color-coded dots (see Materials and Methods for details). A small green dot represents the START and a large red oval represents the Cy5-labeled STOP site. $ZnSO_4$ was added at time zero. (**b**) Displacement of the spider trajectory in panel **a** from its initial position as a function of time. The green line represents displacement calculated using averaged position measurements of 1 min intervals, and the black line represents the displacement from a rolling 4-min average (see Materials and Methods). (**c**) Ensemble root mean square displacement (RMSD) of exemplary spiders on the EAC substrate track in the presence (red, corresponding to the 15 Tier 1 Spiders in Figure A1.23) and absence (black, 7 spiders) of Zn^{2+}, with the corresponding displacements used to calculate each ensemble RMSD for each buffer condition (similarly colored line graphs). (**d**) Ensemble RMSD for spiders on EAC tracks satisfying simple filtering criteria. Curves are shown for spiders on EAC substrate track (red, 85 spiders), EAC product track with TRIGGER introduced to the sample 10-15 min before imaging (blue, 18 spiders), and EAC product track with TRIGGER introduced 30-60 min before imaging (black, 29 spiders). EAC substrate and 10-15 min trigger product RMSD plots are fit to a power law function, and the EAC 30-60 min trigger product RMSD is fit to a straight line. Individual displacements are shown with colors corresponding to the respective ensemble RMSD plots. All Figure 2.7 data were obtained in SSC buffer.

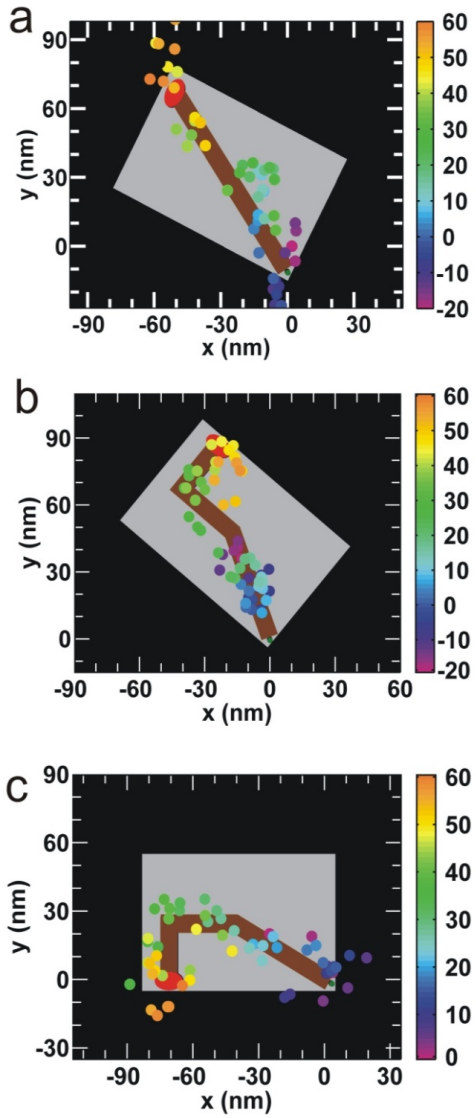


Figure 2.8. Particularly clean trajectory plots for individual spiders on the EAC (**a**), EABC (**b**), and EABD (**c**) tracks. The EAC trace was collected in 1× HBS + 5 mM $ZnSO_4$ and corresponds to spider EAC 5H (Figure A1.24), while the EABC and EABD traces were collected in 1× TA-Mg + 1 mM $ZnSO_4$ and correspond to spiders EABC 1 and EABD 1 (Figure A1.23). The color bars on the left indicate the time in minutes. Zinc was added at time 0. Among the x-y plots for EABC and EABD traces, some were consistent with the prescribed turn (as shown here); however, our resolution was not sufficient to extract features of these landscapes such as turn angles with satisfactory confidence.

Our resolution was not sufficient to reliably detect turns, so we focused on EAC landscapes. Individual particle traces showed a distribution of behaviors that may result from variations across molecules, idiosyncrasies of the sample preparation, the stochastic nature of the observed process, photobleaching, and/or instrument measurement error (Figures 2.7a,b, 2.8, and A1.22-A1.24; Table A1.3). Despite this variability, moving traces commonly had net displacements between 60 and 140 nm and their mean velocity varied between 1 and 6 nm/min, within error consistent with track length (~90 nm) and deoxyribozyme cleavage rate (~1 min^{-1}/leg), respectively.

To confirm that our particle traces reflect genuine spider movement, we performed tests with and without Zn^{2+} and/or TRIGGER, both on substrate and product tracks. In each case, RMSD plots varied in a way consistent with the expected corresponding behavior of spiders on origami tracks, despite the inherent noise associated with single particle tracking over tens-of-nanometer length scales and tens-of-minute time scales (Figure 2.7c,d). For instance, RMSD plots indicated substantially more movement on substrate tracks in the presence of Zn^{2+} and trigger than in their individual absence (Figures 2.7c, 2.8, 2.9, A1.24; and Table A1.4). On product tracks, results were consistent with an unbiased random walk with no dependence on Zn^{2+}. When product tracks were pre-incubated with TRIGGER 30-60 min prior to addition of Zn^{2+} and onset of imaging (as were substrate tracks), little or no movement was observed (Figure 2.7d), consistent with spiders having been released and having diffused toward or to the STOP sites prior to imaging. In contrast, when TRIGGER and Zn^{2+} were both added shortly prior to imaging, substantial movement was observed (Figure 2.7d), consistent with our AFM results for spiders on product tracks (Figure 2.2f,g) and with Monte Carlo simulations of spider movement (Figure 2.9).

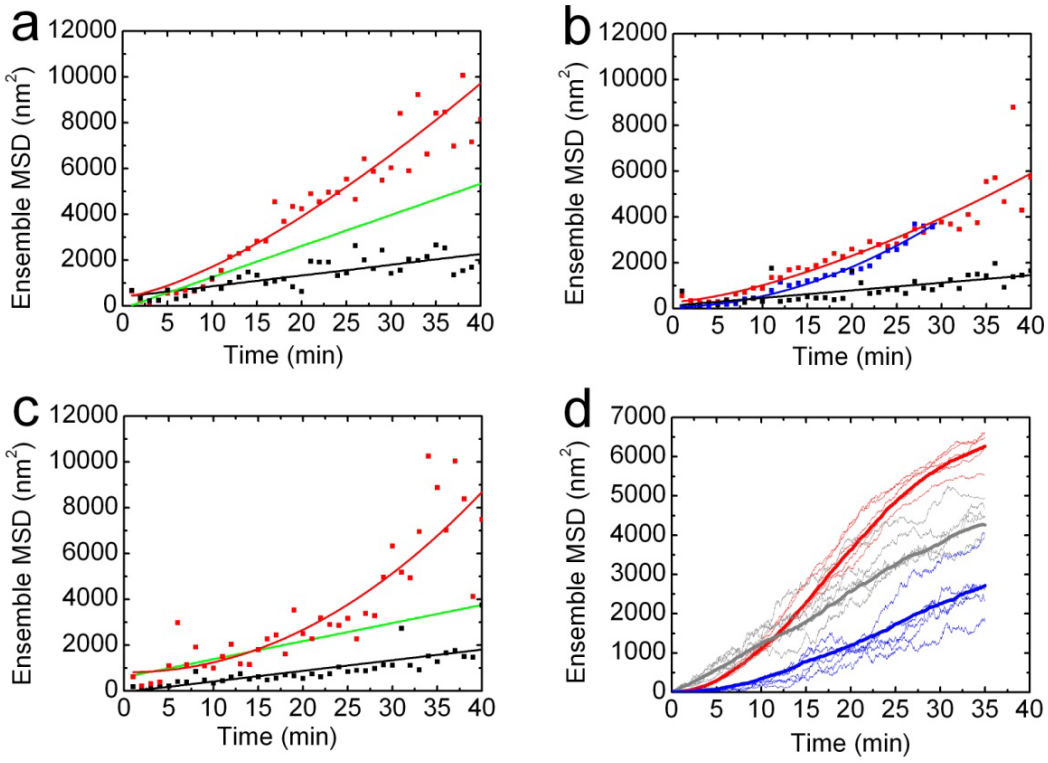


Figure 2.9. (a) Ensemble mean square displacement (MSD) versus time calculated from 15 individual Tier 1 spiders on the EAC substrate track (red squares; EAC 1-15 in Figure A1.23) in 1× SSC. A power law function (red curve) is fit to the MSD from 1-30 min, and a linear function (green curve) is fit to the first 12 min. For further comparison, an MSD plot is shown that is derived from 7 spiders (black circles; traces found in Figure A1.22d) from a no-Zn^{2+} control experiment in which spiders are not expected to walk. In both types of experiments, the origami-spider complexes were incubated with TRIGGER for 30-60 min prior to imaging. (b) Ensemble MSD versus time plots comparing behavior on the substrate-covered (red) and product-covered EAC tracks with TRIGGER added either 30-60 min (long incubation, black) or 10-15 min before (short incubation, blue) imaging by fluorescence microscopy in 1× SSC in the presence of zinc. The MSD values were calculated from 85 individual spiders on the substrate-covered EAC track, 29 spiders on the product-covered EAC track incubated for 30-60 min with TRIGGER, and 18 spiders on the product-covered EAC track incubated for 10-15 min with TRIGGER selected according to intensity, ellipticity, and outlier cutoff criteria stated in Materials and Methods. All three are fit to power law functions (solid curves). The MSD plot for the substrate track is assembled from the same datasets as the red plot in panel (a), but the more relaxed selection criteria result in the inclusion of more stationary or slowly moving spiders, resulting in a shallower curve. (c) Ensemble MSD versus time plots of spider movement on the EAC substrate track observed in 1×

HBS with 1× HBS buffer containing either 0 mM (black curve) or 5 mM (red curve) $ZnSO_4$ added at time t = 0 min. In both types of experiments, the sample was incubated with TRIGGER for 30-60 min prior to the beginning of the experiment. A power law function (red curve) is fit to the MSD with 5 mM zinc from 1-30 min, and a straight line (green curve) is fit to the first 15 min of the MSD with 5 mM zinc. A straight line (black) is also fit to the MSD in 0 mM zinc. (d) Simulated MSD versus time plots calculated as described in the Materials and Methods from 1,000 spiders (thick lines) or separate trials of fewer spiders (thin lines) for the substrate-covered EAC track (red), and for the product-covered EAC track without (gray) or with a delayed release (blue) from the START region ($t_{1/2release}$ = 0 or 10 min). For the smaller trials, 80 spiders per trial were used for the substrate-covered track, while 20 spiders per trial were used for the product track with and without delayed release (to approximate the numbers of experimental spiders observed in each case).

2.4 Discussion

In this work, we have shown that the interactions between a processive molecular spider and a precisely defined track on two-dimensional (2D) DNA origami leads to directed processive motion. These walkers exploit a thermodynamic gradient to autonomously execute a program of motion using the base pairing properties of DNA. They thus recapitulate some of the features of natural protein motors, but *via* a completely synthetic mechanism that utilizes *in vitro* selected DNAzyme legs and a non-natural DNA track.

Our AFM measurements provide results consistent with random DNA-based walkers guided by their landscapes for as far as 100 nm, for up to 50 cleavage steps, at speeds of roughly 3 nm/min. The ability to obtain programmed behavior from the interaction of simple molecular robots with a complex modifiable environment suggests that exploiting stochastic local rules and programming the environment are effective ways to minimize the limitations that molecular construction places on the complexity of robotic behavior at the nanoscale.

Interestingly, a TRACK constructed of cleavable substrates results in equally fast (Figure 2.9b) or faster (Figure 2.2g) progress towards of the GOAL than a TRACK composed of cleavage product, despite the fact that individual steps are expected to be much faster on product than on substrate (since $k_{off,prod} > k_{cleave} > k_{off,sub}$). This is consistent with analytical predictions for multipedal spiders on an idealized one-dimensional track[107]. Furthermore, the mean-squared displacement of both experimental and simulated spiders on substrate tracks follows a concave-up trajectory (Figure 2.9), suggestive of superdiffusive behavior; that is, in the power law relationship

$$M.S.D. \sim t^{\alpha} \qquad (1)$$

the exponent α is greater than unity[111]. In contrast, the predicted behavior on cleavage product has a nearly linear dependence, with $\alpha \sim 1$ (Figure 2.9d, gray curve). While experimental observations of spiders on cleavage product with simultaneous addition of TRIGGER and Zn^{2+} suggest $\alpha > 1$ as well (Figure 2.9b), simulations predict that this

could result from delayed release from the START position (Figure 2.9d). Thus, it appears that cleavage of substrates leads to superdiffusive behavior of spiders that accelerates progress towards the GOAL, albeit only slightly.

The single-particle fluorescence tracking results are in good general agreement with the AFM assays, but augment them by revealing a wide diversity of walking trajectories, as expected from a complex, stochastic process and predicted by our Monte Carlo modeling. The mean velocity, track length, and calculated cleavage rate are comparable to the results from SPR and AFM. We observe a variety of non-designed behaviors among some spiders as well, including (1) significant movement towards the GOAL on product tracks, (2) occasional movement prior to addition of the catalytically necessary Zn^{2+} ion (3) immobility. Behavior (1) is predicted by Monte Carlo modeling, and its influence may be reduced by exploring the parameter space, i.e., the relative rates of binding, cleavage, and dissociation, as well as the effective legspan. Behavior (2) constitutes a form of leakage that is unexpected based on the design principles and simulations, and may either result from particle tracking artifacts such as focal drift or real deviations from the proposed walking mechanism. In Chapter 4, we provide evidence that legs may be passed between substrates without cleavage having occurred, which we predict would lead to this kind of leakage. Finally, behavior (3) could result from kinetic traps, or spiders having diffused to the GOAL by behavior (2) prior to the addition of Zn^{2+}. All of these observed deviations provide important information for the guidance of more well-behaved molecular walkers in the future.

The performance of molecular spiders is decidedly weak compared to natural motors. At mean velocities of 3 nm/min, spiders fall about three orders of magnitude short of natural protein motors such as myosin V[92], and theoretical work suggests similar disparities in terms of stall force[112]. This may ultimately be due to the fact that spiders, unlike many natural motors, take steps *via* passive diffusion without a genuine power stroke. Addressing this limitation may require a fundamental reworking of the design of DNA-based walkers, such as introducing synchronized changes in the conformation of rigid legs.

Still, spiders incorporate some interesting behaviors not typically observed in natural protein motors. In addition to following prescribed paths, molecular spiders bias their own behavior by modifying the landscape on which they walk. When further improved, processes like this could be used to couple the behavior of multiple walkers through their interactions with a common landscape. For instance, one spider could modify features of the landscape so as to repel or attract another spider, leading to collective behavior[113–115]. Recent theoretical work suggests that simple types of collective behavior may be feasible with the current design of spiders[116].

Like protein motors, molecular spiders act in response to cues from their environment rather than any internalized instructions. Integration of logic and memory into the robot's body, or coupling to delocalized molecular computing circuits, would enhance the robot's ability to respond appropriately to changes in its environment. Since sophisticated synthetic computing circuits have already been demonstrated using DNA[62,63,117,118], such synergy between walkers as local agents and molecular computers as distributed decision-makers may materialize in the near future.

CHAPTER 3:

CHEMICALLY SENSITIVE SUPER-RESOLUTION FINGERPRINTING OF NANOSCALE OLIGONUCLEOTIDE ARRAYS ON DNA ORIGAMI[6]

3.1 Introduction

The DNA origami method[53,55] has laid the foundation for a multitude of nanoscale devices that permit control over dynamic chemical or optoelectronic processes[58–61,119,120], including one example described in detail in Chapter 2. Since they are constructed from soft biopolymers, self-assembling from often hundreds of unique components, many of these devices are chemically heterogeneous and susceptible to damage or distortion by mechanical imaging techniques like atomic force microscopy (AFM). Furthermore, while DNA origami holds promise for the nanoscale positioning of multiple interacting components such as enzymes, the distance scales involved – typically only a few nanometers) – render it difficult to reliably resolve the components of individual assemblies in order to assess yield. Finally, there exists a gap between static single-particle characterization of nanomaterials by AFM or TEM and the functional characterization of such materials, which is typically carried out in bulk with little or no information on the variation between copies of an assembly. Such information would be valuable in comparisons between competing designs.

Fluorescence nanoscopy combines high spatial resolution and tunable chemical specificity with relatively low invasiveness[78,86,88,121–123], and therefore holds promise for the spatiotemporal imaging and quality control of functional nanomaterials[88,120–122,124,125]. Stochastic reconstruction microscopies[88,121,123,126,127] show particular promise

[6] Alexander Johnson-Buck designed, performed, and analyzed the DNA-PAINT and single-molecule kinetics experiments, as well as all alignment and heterogeneity analysis and simulation of PAINT reconstructions. Jeanette Nangreave and Alexander Johnson-Buck designed the DNA origami tiles for analysis. Jeanette Nangreave synthesized the DNA origami tiles and performed characterization by AFM with assistance from Shuoxing Jiang. Do-Nyun Kim and Mark Bathe performed CanDo structure prediction and local concentration modeling.

for chemically specific two- and three-dimensional imaging in the near term, since they can be readily carried out using widely available fluorescence microscopes and have been applied to imaging both fixed and living cells. Recently, stochastic super-resolution fluorescence techniques have been applied to the visualization of isolated DNA features on DNA origami[88] and distributions of reactivity patterns along 150-700-nm gold nanorods[121]. However, this family of techniques has not yet seen widespread application in nanotechnology, nor has its capacity for the quantitative imaging of chemically and functionally heterogeneous nanodevices been fully realized.

Here, we employ the single-particle fluorescence nanoscopy technique PAINT (points accumulation for imaging in nanoscale topography)[86,88] to acquire quantitative two-dimensional maps of heterogeneous oligonucleotide features on DNA origami pegboards. We show that PAINT has sufficient resolution (~10 nm) to reliably distinguish dense (>10^4 μm^{-2}) sub-100-nm patterns of features. We employ two-color PAINT to quantitatively image enzyme-catalyzed modifications of surface features of single origami over time, and to show that single origami pegboards exhibit stable spatial patterns of binding to specific probes, or interaction "fingerprints." Finally, we present experimental and modeling evidence suggesting that these fingerprints may arise from variations in feature spacing that locally modulate the probe binding kinetics. This work highlights the power of fluorescence nanoscopy in the quality control on individual soft nanodevices that interact with and position reagents in solution.

3.2 Materials and Methods

Unless otherwise noted, all chemicals were purchased from Fisher and all oligonucleotides were purchased from Integrated DNA Technologies (IDT). Analysis of PAINT experiments, including plotting and reconstruction, was performed using home-written MATLAB code unless stated otherwise.

Preparation of DNA origami scaffolds. Rectangular DNA origami arrays consist of an M13mp18 viral DNA scaffold (New England Biolabs) and 202 ssDNA staples as previously described[120]. For all structures assembled here, staples 1-12 and 205-216 were omitted to prevent inter-array base stacking interactions that result in undesirable

aggregation (Figure 3.1). Of the remaining staples, several were modified at their 5′-end with an additional sequence, 5′-ACC TCT CAC CCA CCA TTC ATC, to which the substrate **S** (5′-GAT GAA TGG TGG GTG AGA GGT TTT TCA CTA TrAG GAA GAG) can bind (Table 3.1). The arrays were annealed in either 2x HBS buffer (300 mM NaCl, 20 mM HEPES, pH 7.4) or 5x SSC (750 mM NaCl, 75 mM Trisodium Citrate, pH 7.4) buffer, with a 1:3 ratio of M13 to staple strands and a final concentration of 10 nM (M13). There is no apparent difference in the assembly of arrays using these two buffer conditions. The arrays were annealed over 12 hours from 94^{o}C-25^{o}C using a PCR thermocycler (Eppendorf). The template origami **R** or **L** was incubated with a 3:1 ratio of substrate **S** to available binding sites on the origami prior to imaging. Integrity of the ribose moiety of **S** was verified by subjecting this oligonucleotide to denaturing PAGE in 8 M urea alongside an equivalent sample that had been incubated for 15 minutes in sodium phosphate buffer, pH 12, at 75°C and subsequently staining with SYBR Gold (Invitrogen) per the manufacturer's instructions (Figure 3.16).

Atomic force microscopy characterization of DNA origami scaffolds and assembled pegboards. 2 μL of annealed sample was deposited on a freshly cleaved mica surface (Ted Pella, Inc.) and left to adsorb for two minutes. After adsorption, 400 μL of buffer (1x TAE-Mg^{2+}: 40mM Tris, 20 mM acetic acid, 12.5mM Mg2+, pH 7.6) was added to the liquid cell and the sample was scanned in peak-force mode, using ScanAsyst in liquid probes, on a Veeco Multimode 8 AFM. All imaging by AFM was carried out at room temperature. The resulting AFM images were processed/flattened and analyzed with NanoScope Analysis software (Veeco, version 1.40). To determine the yield of DNA origami scaffold formation, ~1μM x ~1μM AFM images were evaluated. Each DNA origami structure in the AFM images was assigned to one of the following three categories, based on the height features present in the images: 1) well-formed tile with clear evidence of fairly complete track, 2) well-formed tile with defective or missing track, or 3) broken or deformed tile. Only those DNA origami tiles with clearly discernible boundaries were considered (i.e. not cut-off like those at the edges of the AFM images, not obscured by impurities in the sample, and not stacked/clustered together). Estimated yields for **R** and **L** origami are shown in Table 3.3.

Preparation of PAINT probes and other fluorescently labeled DNA. Oligonucleotides were ordered with terminal amine modifications for fluorescent labeling: probe **α-NH_2**, 5′-/5aminoC6/ATA GTG AAA; probe **β-NH_2**, 5′-/5aminoC6/CTC TTC CTA; **S-NH_2**, 5′-GAT GAA TGG TGG GTG AGA GGT TTT TCA CTA TrAG GAA GAG /3AmMO/. **S-NH_2** was ordered HPLC purified, and all three oligonucleotides were used as provided without further purification. The oligonucleotides were labeled with *N*-hydroxysuccinimidyl ester derivatives of Cy3 or Cy5 (GE Healthcare) by overnight incubation in $NaHCO_3$, pH 8.3, followed by ethanol precipitation and thorough washing with 80% ethanol until the supernatant was colorless. Denaturing polyacrylamide gel electrophoresis revealed no detectable free dye. Labeling efficiency was quantified by absorbance at 280 nm and either 550 nm (Cy3) or 650 nm (Cy5) using a Beckman DU 640B Spectrophotometer, and exceeded 85% for all strands except for **β-Cy3**, for which it was 30%. Labeling efficiency less than unity does not hamper PAINT experiments due to the continuous exchange of unlabeled probes for labeled ones. Cy3-labeled 8-17 DNAzyme (**8-17-Cy3**), 5′-/5Cy3/TCT CTT CTC CGA GCC GGT CGA AAT AGT GAA AA, was ordered with HPLC purification and used as-is for binding kinetics assays.

Preparation of microscope slide surface for fluorescence microscopy and PAINT. Quartz microscope slides (3” x 1” x 1 mm, G. Finkenbeiner) were prepared as described[128]. Briefly, ~1 mm holes were drilled approximately 3 cm apart to create inlet and outlet ports for a flow channel. The slides were cleaned by sonicating in 1 M KOH, followed by heating in a solution of 5% hydrogen peroxide (Acros, 202460010) and 5% ammonium hydroxide (Acros, 205840025). The slides were rinsed thoroughly with deionized water and flamed for approximately 1 min using a propane torch. To prepare the surface for conjugation to NeutrAvidin, the slides were silanized by incubating for 1 hour in a 5% (v/v) solution of (3-aminopropyl)triethoxysilane (Sigma-Aldrich, A3648) in acetone, rinsed thoroughly with acetone, and cured at 80°C for 1 hour. The bisfunctional cross-linking agent para-diisothiocyanate (PDITC, Acros, 417510050) was then conjugated to the free amines of the aminosilane by immersing the slides in a 0.2 % (w/v) solution of PDITC in a 1:10 mixture of pyridine:*N,N*-dimethylformamide for 2 hours.

The slides were washed thoroughly with methanol (Acros, 610090040) followed by acetone. Finally, to conjugate NeutrAvidin by its surface amines to the PDITC, a 0.5 mg/mL solution of NeutrAvidin (Invitrogen, A-2666) in 50 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0 was applied to each slide and incubated in a humid environment for 2 hours. The slides were washed with a solution of 1 M NaCl and 40 mM NaOH for 1 minute to quench free isothiocyanate, rinsed thoroughly with deionized water, and dried under nitrogen. A fluidic channel between the two drilled holes was formed over the NeutrAvidin-coated portion of each slide using double-sided tape (Scotch, permanent 1/2") and coverslips (VWR, 24 x 30 mm, No. 1.5), then sealed with 5-minute Epoxy (Hardman Adhesives, 4001). The slides were stored in a desiccated chamber at 4 °C for up to 4 weeks. Prior to an experiment, inlet and outlet ports were constructed on a slide using sterile 200 µL pipet tips inserted into the drilled holes (Eppendorf) and ~ 5-cm lengths of microbore tubing (Cole-Parmer, EW06418-05), and sealed with Epoxy.

Total internal reflection fluorescence microscope. Assembly kinetics and all PAINT experiments were carried out on an inverted total internal reflection (TIRF) fluorescence microscope with a 1.2 NA 60x water-immersion objective (IX71, Olympus) in an environmentally controlled room at 20 ± 3 °C. Cy3 excitation was provided by a 532-nm green laser (ultra-compact diode-pumped Nd:YAG laser GCL-025-S, CrystaLaser, 5 W/cm^2 for kinetic measurements and 60 W/cm^2 for PAINT measurements) and Cy5 excitation by a 638-nm red diode laser (Coherent CUBE 635-25C, 4 W/cm^2 for kinetic measurements, and Olympus LAS/640/100-D, 100 W/cm^2 for PAINT measurements). Excitation was continuous in all experiments. The Cy3 and Cy5 emission signals were separated by a dichroic mirror with a cutoff wavelength of 610 nm (Chroma) and projected side-by-side onto an ICCD camera chip (iPentamax HQ Gen III, Roper Scientific, Inc.). Relay lenses matched the microscope image with the camera focal plane and the IX71 internal 1.6x magnifier (final effective pixel length 133 nm). The Cy3 channel image was passed through a band pass filter (HQ580/60m, Chroma) and the Cy5 channel was passed through a long pass filter (HQ655LP, Chroma). A Newport ST-UT2 vibration isolation table was used in all experiments.

Characterization of kinetics of origami pegboard assembly and S cleavage by 8-17 DNAzyme. A 10 nM solution of **R** template origami in 5x HEPES-buffered saline (HBS; 1x HBS ≡ 150 mM NaCl, 10 mM HEPES-KOH, pH 7.0-7.4) was diluted to 100 pM in 1x HBS, flowed into the channel of a NeutrAvidin-coated slide, and allowed to bind via the biotin-NeutrAvidin interaction for 10 min. The excess origami was washed out twice with 1x HBS. While monitoring the fluorescence of Cy3 at the slide surface using the TIRF microscope, a solution of 100 nM **S-Cy3** in 1x HBS containing oxygen scavenger system[100] (OSS ≡ 2.5 mM 3,4-dihydroxybenzoic acid, Sigma P5630; 1 mM Trolox, Acros 218940050; and 25 nM protocatechuate dioxygenase, Sigma-Aldrich P8279) was injected into the slide channel with a dead time of less than 10 s. To limit photobleaching, the excitation light was passed through a neutral density filter (OD 2.0, Newport Model 5215) and a shuttered illumination scheme was used with 0.5-s exposures separated by 14.5-s dark periods. The **S-Cy3** was injected during a dark period. The mean fluorescence signal from 382 origami was plotted as a function of time (Figure 3.4a) and fit to the single-exponential model $y = C(1 - e^{k'_{obs}t})$.

To measure the cleavage of **S-Cy3** by 8-17 DNAzyme (**8-17**) at the ensemble level, a 1x HBS solution containing 1 μM **8-17** (5′- CTC TTC TCC GAG CCG GTC GAA ATA GTG AAA A, used as-is from IDT), 1 mM $ZnSO_4$, and OSS was added to the slide already containing **R** origami saturated with **S-Cy3** while observing *via* the same shuttered illumination scheme described above. Upon the addition, the Cy3 fluorescence signal from each origami began to attenuate. The signal was averaged across all origami and plotted as a function of time. The decay was not well modeled by a single-exponential decay function, but was well fit to the double-exponential model $y = C_1\left(e^{k'_{obs,1}t}\right) + C_2\left(e^{k'_{obs,2}t}\right)$ (Figure 3.4b). The signal decrease due to photobleaching is minimal under these illumination conditions, as is evidenced by the nearly horizontal signal intensity prior to **8-17** addition at time *t*=0. The decline in signal is not significantly different from the time course measured by PAINT under the same cleavage conditions (Figure 3.11c).

In order to perform the time-course measurements of **S** cleavage using PAINT, it was necessary to remove the **8-17** DNAzyme after each interval of cleavage to make **S**

available for binding by α and β. To determine the kinetics of **8-17** dissociation from **S**-loaded **R** origami, a mixture of 5 nM **R** template origami and 3.8 µM **S** was incubated in 250 mM NaCl, 25 mM HEPES-KOH, pH 7.4 at room temperature for 10 minutes. The origami were then diluted to 100 pM in 1x HBS, flowed into the channel of a NeutrAvidin-coated slide, and allowed to bind for 10 minutes. Excess sample was flushed away by two washes with 1x HBS. Then, a solution containing 1x HBS , 100 nM **8-17-Cy3**, and OSS was added to the slide channel until apparent saturation was achieved, as judging by the increase in Cy3 fluorescence intensity of each origami (20 minutes). Finally, the dissociation kinetics of **8-17** were measured by monitoring the decrease in Cy3 fluorescence upon addition of 1x HBS containing OSS and 1 µM unlabeled **S** to compete with origami-bound **S** for **8-17-Cy3** under shuttered, attenuated illumination as described above. The intensity of Cy3 from many origami was averaged and modeled well by a single-exponential model (Figure 3.4c). According to the resulting rate constant, approximately 80% of bound **8-17** is expected to dissociate from full-length **S** over the course of 1 hour.

*Characterization of PAINT probe binding kinetics on DNA origami pegboards***.** A mixture of 5 nM **R** template origami and 3.8 µM **S** (a 6:1 ratio between **S** and binding sites for **S** on the origami) was incubated in 250 mM NaCl, 25 mM HEPES-KOH, pH 7.4 at room temperature for 10 minutes. The origami was then diluted to 100 pM in 1x HBS, flowed into the channel of a NeutrAvidin-coated slide, and allowed to bind via the biotin-NeutrAvidin interaction for 10 minutes. Excess sample was flushed away by two washes with 1x HBS.

The slide was mounted on the TIRF microscope, and a solution containing OSS, 1x HBS, and 1 or 2.5 nM each of **α-Cy3** and **β-Cy5** was added to the slide channel. After a 2-minute incubation to permit equilibration of [O_2], the binding of the PAINT probes was visualized under excitation at 532 nm and 640 nm with a camera exposure time of 1 s. To limit photobleaching, excitation power was reduced to ~10% of the power used in PAINT experiments. Doubling the excitation power did not yield significantly different first-order rate constants.

Intensity time traces for Cy3 and Cy5 were analyzed using the hidden Markov modeling software package vbFRET[129] to extract idealized trajectories. A single exponential decay function, $y = Ce^{kt}$, was fit to the histograms of dwell times in the bound and unbound states to yield the dissociation rate constant k_{off} and pseudo-first-order association rate constant k_{on}', respectively. The values of k_{on}' were plotted as a function of concentration, fit to linear increase functions, yielding the second-order association constant k_{on} as the slope. The results from duplicate trials are shown in Table 3.2.

PAINT nanoscopy of DNA origami pegboards. **R** or **L** origami template was loaded with **S** and immobilized on a NeutrAvidin-coated slide as in the characterization of PAINT probe kinetics, above. The slide was mounted on the TIRF microscope, and a solution containing OSS, 1x HBS, and 10-20 nM of **α-Cy3** and/or **β-Cy5** (or **α-Cy5** and **β-Cy3**) was added to the slide channel. After a 2-minute incubation to allow equilibration of $[O_2]$, the binding of the PAINT probes was visualized under excitation at 532 nm and 640 nm with a camera exposure time of 1 s. Imaging proceeded for 1000-4000 s.

Generation of PAINT reconstructions. Individual origami tiles were located in the field of view by the repeated appearance of Cy3 and/or Cy5 signal in the same location. Specifically, a fluctuation map (Figure 3.5) was generated by subtracting each movie frame from the preceeding frame, taking the absolute value of the intensity differences, and averaging across all movie frames to obtain the average frame-to-frame fluctuation in intensity at each pixel. This allowed us to distinguish origami, as sites of repeated PAINT probe binding, from other bright fluorescent contaminants. The origami appeared as bright diffraction-limited spots in this fluctuation map, and were localized by Gaussian fitting to obtain coordinates of all origami in a field of view.

For each origami, traces of intensity as a function of time (Figure 3.2e) were generated as follows. In each movie frame, a 5x5-pixel square region ***A*** centered on each origami was defined. The background fluorescence, determined from the median of the 24 pixels immediately surrounding ***A***, was subtracted from each pixel within ***A***. The background-corrected intensity values within ***A*** were summed to obtain the total

fluorescence intensity of probe(s) bound to an origami tile in a given movie frame. During the generation of intensity traces, the region ***A*** was re-defined if the microscope stage drifted by more than one pixel, or 133 nm, in the x or y direction (see below).

The microscope stage drifted by 50-250 nm in the x-y plane throughout a typical experiment (Figure 3.6). To correct for this, each movie was divided into 100- to 200-s bins and the intensity of all the frames within each bin was averaged. The time-averaged image from each bin was cross-correlated with time-averaged image from the first bin in the movie with 100-fold up-sampling using the MATLAB script dftregistration[130]. Linear interpolation yielded an estimate of the microscope stage drift in each movie frame (Figure 3.6b).

During movie frames in which a DNA origami's fluorescence intensity exceeded a threshold (500-2000 photons/s, depending on the experiment and excitation intensity), the intensity profile was fit to a 2-dimensional Gaussian function of the form

$$I = Ce^{-\frac{(x-\mu_x)^2}{2\sigma_x{}^2}-\frac{(y-\mu_y)^2}{2\sigma_y{}^2}} + b \qquad (1)$$

to extract the centroid (μ_x,μ_y), as well as parameters for localization error estimation, including point-spread function widths σ_x and σ_y and a more precise photon count ($2\pi C\sigma_x\sigma_y$). A 7x7-pixel fitting box centered on the origami was used for fitting. Gaussian fits were filtered for quality and were rejected if any of the following criteria were met:

1. Either σ_x and σ_y exceeded a cutoff of 2 pixels (266 nm).
2. The residual between the fit and the actual intensity profile within the fitting box exceeded 25% of the total volume of the Gaussian fit.

This helped to reduce the influence of aberrant fits resulting from nonspecific binding of probes to the slide surface in the vicinity of the origami. Fitting error was estimated as described[80], using the parameters derived from each fit as well as the standard deviation of the background signal and the effective pixel size of 133 nm. Multiple fits from the same binding event were combined by taking the median of all centroid measurements ($\mu_{x,i}$,$\mu_{y,i}$) for that event to avoid multiple counting of a single binding event.

To generate PAINT reconstructions, the set of all position measurements (μ_x,μ_y) for an origami was enumerated. The microscope stage drift was subtracted from each position measurement. Then, a reconstruction was generated as a sum of Gaussian functions on a grid of 4-nm square pixels, with the centroid of each Gaussian representing a drift-corrected position measurement. The width parameters σ_x and σ_y for the reconstruction Gaussians were defined as the median error for all position measurements for a given origami.

When characterizing dense fields of targets, the quality of the reconstructions is often limited by sampling density rather than localization error. Therefore, on the basis of Poisson statistics, we found it convenient to define a sampling radius σ_{sample} and used it as a lower bound for the reconstruction Gaussian widths σ_x and σ_y:

$$\sigma_{sample} = \sqrt{\frac{\lambda}{\pi\rho}} = \sqrt{\frac{\lambda A}{N\pi}} \qquad (2)$$

Where λ is the desired number of localizations per sampling area of $\pi\sigma^2$, and ρ=*N*/*A* is the actual sampling density consisting of *N* localizations over the object area *A*. With λ = 2, 85% of the available binding sites will lie within σ_{sample} of a localization, assuming equal sampling probability of all binding sites. Using λ = 2, we calculated σ_{sample} for each origami and used it as a lower bound for the error values used in the reconstruction. For instance, if 100 localizations are counted for an object approximately 60 x 100 nm in size, $\sigma_{sample} = \sqrt{\frac{12000}{100\pi}}$ = 6 nm, which is comparable to the theoretical localization error in our experiments. In other words, resolution was limited by sampling unless *N* exceeded ~100. Consistent with this observation, simulated PAINT images of origami show rapid improvement in reconstruction quality, as quantified by deviation from an idealized reconstruction, as *N* increases from 10 to 200, with marginal improvements thereafter (Figure 3.7).

For two-color PAINT reconstructions, the binding distributions of α and β binding had to be registered in the same coordinate space due to the fact that they were detected *via* different sets of optics projecting the image onto separate regions of the CCD camera. First, a coarse third-order polynomial mapping was found between the two

channels using Gaussian fitting of fiduciary markers with fluorescence visible in both channels (FluoSpheres 580/605, Invitrogen, F-8810). The registration error[101], calculated as the average distance between the calculated and actual positions of the Cy5 centroid based on the polynomial mapping from Cy3 coordinates, was 10-20 nm. An initial two-color overlay of the PAINT reconstructions was generated using this mapping. To further fine-tune the registration, the PAINT reconstructions from Cy3 and Cy5 were registered directly by cross-correlation[130] (Figure 3.12a-d). To reduce the influence of uneven binding (heterogeneous binding or sampling noise) on registration, the reconstruction in each channel was saturated at 10% of its maximal intensity value for purposes of registration. For all experiments except for those involving cleavage of substrate by a DNA enzyme, the Cy3 and Cy5 reconstructions were normalized such that their total integrated intensity was equal to unity for final visualization.

Identification of origami pegboard patterns from PAINT reconstructions. Samples consisting of **R** or **L** origami were synthesized and their identities concealed. Each sample was imaged in the presence of 10 nM **β-Cy5** for 1000 s, and reconstructions generated from the resulting 20-60 localizations per origami as described above. Origami reconstructions were visually inspected and classified as linear, rectangular, or other as follows:

1. Linear: between 50 and 150 nm in length, less than 50 nm wide, only one main segment of intensity visible.
2. Rectangular: between 50 and 150 nm in size in both dimensions, at least three sides of a rectangle visible.
3. Other: neither of the above two criteria satisfied.

The results of this classification are shown in Figure 3.8e.

PAINT time course of S cleavage by 8-17 DNAzyme on individual R origami. **R** origami were assembled with **S** and deposited on NeutrAvidin slides as described above. The origami were first imaged in the presence of 1x HBS containing OSS and 10 nM each of **α-Cy3** and **β-Cy5** for 33 minutes. Then, a solution of 1x HBS containing 1 μM **8-17** and 1 mM $ZnSO_4$ was added to the slide channel and incubated for 2 minutes. The slide

channel was flushed with 1x HBS containing 1 μM unlabeled **S** to stop the reaction and sequester any remaining **8-17** in solution. After 60 minutes, the sample was imaged again in the presence of **α-Cy3** and **β-Cy5** for 33 minutes. A solution of 1 μM **8-17** and 1 mM $ZnSO_4$ was once again added to the slide channel and incubated for 8 minutes. The slide channel was flushed with 1 μM unlabeled **S** again and incubated for 60 minutes. Finally, the sample was imaged for the third time in the presence of **α-Cy3** and **β-Cy5**. Two-color PAINT images of origami were reconstructed and analyzed as described above (Figure 3.11b). The total number of binding events for **α-Cy3** and **β-Cy5** (N_α and N_β, respectively) were compared at different time points to quantify the fraction of **S** that had been cleaved after each incubation with **8-17** and Zn^{2+} (Figure 3.11c).

Model-free 2D alignment of R origami reconstructions. To characterize the population-level heterogeneity of PAINT reconstructions of **R** origami, the refine2d functionality of EMAN v1.9 was used. Reconstructions from **β-Cy5** binding to 198 **R** origami pooled from three independent experiments were cropped to equal-size square images and subjected to 10 iterations of model-free alignment assuming between 2 and 10 classes. All runs converged before the 10^{th} iteration. Inspection of the output revealed that fewer than 9 classes resulted in some smearing or loss of features (e.g., disappearance of the empty region in the center of the **R** rectangle), while more than 8 classes produced more nearly degenerate classes. Although the results from using 9 classes (Figure 3.9) show some possible degenerate class averages (e.g. panels g and h), this was the lowest number of classes that recapitulated the diversity of reconstructions observed. Approximately 45% of the origami (panels a, c, d, and f) fall into classes that approximate an open rectangular shape, which is similar to the percentage of apparently rectangular shapes observed in the **R** vs. **L** comparison (Figure 3.8) and the yield estimates by AFM (Table 3.3). Although limited contrast hinders characterization of **S** pattern completeness by AFM, several AFM images of **R** origami show patterns that resemble the class averages (lower right corner of each panel in Figure 3.9), suggesting that at least some of the heterogeneity across origami is due to incomplete assembly of the origami scaffold or pegboard. This is consistent with the fact that the distribution of

total binding events per origami within a single experiment is broader than would be expected for identical, perfectly assembled origami (Figure 3.10).

Characterization of spatial homogeneity of PAINT probe binding to individual R origami. When a rectangular **S** pattern is divided into four quadrants containing equal areas of **S**, if binding is homogeneous, the number of probe binding events observed in each quadrant follows a Poisson distribution with a single expectation value across all quadrants. The homogeneity of binding across the four quadrants can thus be characterized using

$$\chi^2 = \sum_{i=1}^{4} \frac{(I_i - \bar{N})^2}{\bar{N}} \quad (2)$$

Where I_i is the observed number of binding events in quadrant *i* and $\bar{N}$ is the expected number of binding events per quadrant, estimated as the average across all four quadrants. Assuming a Poisson-distributed probability of binding to each quadrant, χ^2 can be approximated by a chi-squared distribution with 3 degrees of freedom[131].

To automatically divide each origami into equal quadrants, the two registered reconstructions from Cy3 and Cy5 probe binding were summed to yield a combined reconstruction R_{comb}. Again, to reduce the influence of uneven binding, R_{comb} was saturated at 10%. The reconstruction R_{comb} was aligned to the cardinal axes by finding its edges *via* the Sobel method in the MATLAB Image Processing Toolbox and rotating the edge map in 1-degree increments to find the angle of maximal cross-correlation to a 60-by-100-nm rectangular mask (Figure 3.12e-f). The angles of rotation for all origami in a given movie are uniformly distributed, as expected for origami deposited in random orientations (Figure 3.3). The aligned R_{comb} was then divided into four quadrants so as to minimize χ^2. Finally, the same rotation and division were applied to the raw Cy3 and Cy5 reconstructions, the number of localizations falling into each quadrant counted, and χ^2 calculated for each channel (Figure 3.12g-h).

Verifying sequence dependence of binding heterogeneity. **R** origami were imaged with the probe set (**α-Cy3** + **β-Cy5**) for about 66 min. The same origami were then imaged in the presence of the inversely labeled probe set (**β-Cy3** + **α-Cy5**) for about 66 minutes.

Hence, there is an approximately 1 h delay between the reconstructions generated from the two probe sets. Reconstructions from Cy3- and Cy5-labeled probes were generated and registered as described above, and the reconstruction from each channel was normalized to a total intensity of unity. No angular rotation was performed. The reconstruction from Cy5 was then subtracted from the Cy3 reconstruction to yield a difference map (Figure 3.14). The two difference maps were aligned by cross-correlating the combined reconstructions R_{comb} for the two probe sets, and the 2-D correlation coefficient $R(\Delta I_{\alpha\beta}, \Delta I_{\beta\alpha})$ between the difference maps was calculated.

Simulation of PAINT reconstructions. Numerical simulations to predict the properties of PAINT reconstructions as a function of imaging parameters, and to interpret experimental results, were conducted as follows. Virtual PAINT probes were allowed to bind at random with uniform probability to virtual binding sites with spatial patterns defined by the designs shown in Figure 3.2. Each PAINT probe was "localized" by perturbing the coordinates of the binding site with random Gaussian-distributed variables with standard deviations defined by the localization error (6-10 nm, reflective of typical experimental values). A reconstruction was generated according to the procedure used for experimental PAINT measurements as described above (Figure 3.7a). In cases of two-color simulations, two independent reconstructions were generated for each origami. For Figure 3.13d-e, the simulated origami arrays were randomly oriented and the reconstructions subjected to the same automated alignment, registration, and analysis procedures as the experimental reconstructions. In simulations of 1000 **R** origami with 100 binding events per tile, the alignment error was ± 3.3° (s. d.).

3.3 Results

As targets for fluorescence nanoscopy, 60 x 90 nm rectangular DNA origami tiles[53] were synthesized, each bearing 42 or 126 identical single-stranded overhangs for the attachment of substrate (**S**) oligonucleotides *via* a 20-base-pair DNA duplex (Figure 3.1, Figure 3.2a-d, Table 3.1). In addition, each tile had 4-5 overhangs bearing biotins on the

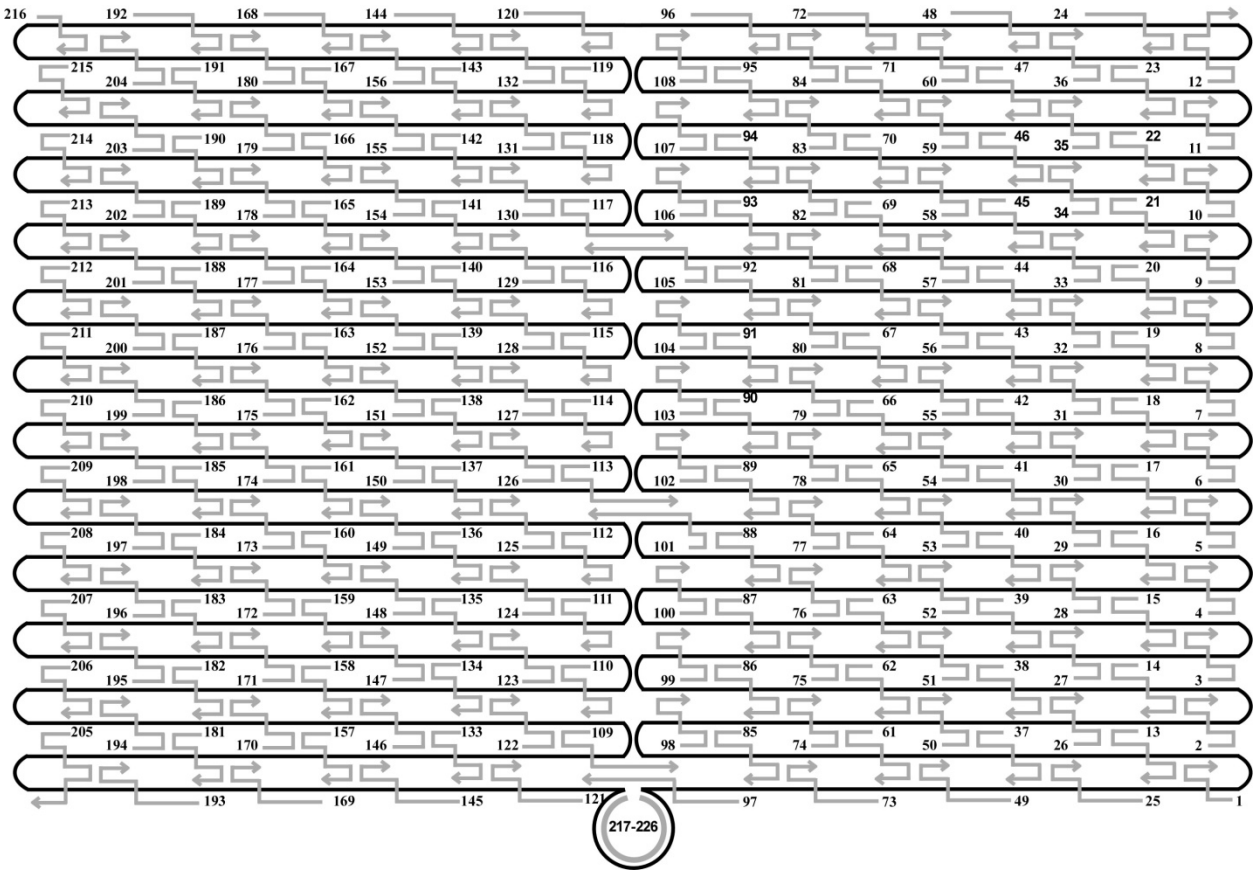


Figure 3.1. Schematic representation of a rectangular origami tile. The continuous black line represents the circular M13 viral genome and the gray lines correspond to unmodified staples with arrows pointing toward the 3′ ends. Each staple is labeled with a number at the 5′ end that corresponds to the sequences listed in Table 3.1.

Table 3.1: Staple Sequences

1	TTTTCGATGGCCCACTACGTAAACCGTC
2	TATCAGGGTTTTCGGTTTGCGTATTGGGAACGCGCG
3	GGGAGAGGTTTTTGTAAAACGACGGCCATTCCCAGT
4	CACGACGTTTTTGTAATGGGATAGGTCAAAACGGCG
5	GATTGACCTTTTGATGAACGGTAATCGTAGCAAACA
6	AGAGAATCTTTTGGTTGTACCAAAAACAAGCATAAA
7	GCTAAATCTTTTCTGTAGCTCAACATGTATTGCTGA
8	ATATAATGTTTTCATTGAATCCCCCTCAAATCGTCA
9	TAAATATTTTTTGGAAGAAAAATCTACGACCAGTCA
10	GGACGTTGTTTTTCATAAGGGAACCGAAAGGCGCAG
11	ACGGTCAATTTTGACAGCATCGGAACGAACCCTCAG
12	CAGCGAAAATTTTACTTTCAACAGTTTCTGGGATTTTGCTAAACTTTT
13	TGGTTTTTAACGTCAAAGGGCGAAGAACCATC
14	CTTGCATGCATTAATGAATCGGCCCGCCAGGG
15	TAGATGGGGGGTAACGCCAGGGTTGTGCCAAG
16	CATGTCAAGATTCTCCGTGGGAACCGTTGGTG
17	CTGTAATATTGCCTGAGAGTCTGGAAAACTAG
18	TGCAACTAAGCAATAAAGCCTCAGTTATGACC
19	AAACAGTTGATGGCTTAGAGCTTATTTAAATA
20	ACGAACTAGCGTCCAATACTGCGGAATGCTTT
21	CTTTGAAAAGAACTGGTCCTCTTTTGAGGAACAAGTTTTCTTGT CTCATTATTTAATAAA
22	ACGGCTACTTACTTAGTCCTCTTTTGAGGAACAAGTTTTCTTGT CCGGAACGCTGACCAA
23	GAGAATAGCTTTTGCGGGATCGTCGGGTAGCA
24	ACGTTAGTAAATGAATTTTCTGTAAGCGGAGT
25	ACCCAAATCAAGTTTTTTGGGGTCAAAGAACG
26	TGGACTCCCTTTTCACCAGTGAGACCTGTCGT
27	GCCAGCTGCCTGCAGGTCGACTCTGCAAGGCG
28	ATTAAGTTCGCATCGTAACCGTGCGAGTAACA
29	ACCCGTCGTCATATGTACCCCGGTAAAGGCTA
30	TCAGGTCACTTTTGCGGGAGAAGCAGAATTAG
31	CAAAATTAAAGTACGGTGTCTGGAAGAGGTCA
32	TTTTTGCGCAGAAAACGAGAATGAATGTTTAG
33	ACTGGATAACGGAACAACATTATTACCTTATG
34	CGATTTTAGAGGACAGTCCTCTTTTGAGGAACAAGTTTTCTTGT ATGAACGGCGCGACCT
35	GCTCCATGAGAGGCTTTCCTCTTTTGAGGAACAAGTTTTCTTGT TGAGGACTAGGGAGTT
36	AAAGGCCGAAAGGAACAACTAAAGCTTTCCAG
37	AGCTGATTACAAGAGTCCACTATTGAGGTGCC
38	CCCGGGTACTTTCCAGTCGGGAAACGGGCAAC
39	GTTTGAGGGAAAGGGGGATGTGCTAGAGGATC
40	AGAAAAGCAACATTAAATGTGAGCATCTGCCA
41	CAACGCAATTTTTGAGAGATCTACTGATAATC

42	TCCATATACATACAGGCAAGGCAACTTTATTT
43	CAAAAATCATTGCTCCTTTTGATAAGTTTCAT
44	AAAGATTCAGGGGGTAATAGTAAACCATAAAT
45	CCAGGCGCTTAATCATTCCTCTTTTGAGGAACAAGTTTTCTTGT TGTGAATTACAGGTAG
46	TTTCATGAAAATTGTGTCCTCTTTTGAGGAACAAGTTTTCTTGT TCGAAATCTGTACAGA
47	AATAATAAGGTCGCTGAGGCTTGCAAAGACTT
48	CGTAACGATCTAAAGTTTTGTCGTGAATTGCG
49	GTAAAGCACTAAATCGGAACCCTAGTTGTTCC
50	AGTTTGGAGCCCTTCACCGCCTGGTTGCGCTC
51	ACTGCCCGCCGAGCTCGAATTCGTTATTACGC
52	CAGCTGGCGGACGACGACAGTATCGTAGCCAG
53	CTTTCATCCCCAAAAACAGGAAGACCGGAGAG
54	GGTAGCTAGGATAAAAATTTTTAGTTAACATC
55	CAATAAATACAGTTGATTCCCAATTTAGAGAG
56	TACCTTTAAGGTCTTTACCCTGACAAAGAAGT
57	TTTGCCAGATCAGTTGAGATTTAGTGGTTTAA
58	TTTCAACTATAGGCTGGCTGACCTTGTATCAT
59	CGCCTGATGGAAGTTTCCATTAAACATAACCG
60	ATATATTCTTTTTTCACGTTGAAAATAGTTAG
61	GAGTTGCACGAGATAGGGTTGAGTAAGGGAGC
62	TCATAGCTACTCACATTAATTGCGCCCTGAGA
63	GAAGATCGGTGCGGGCCTCTTCGCAATCATGG
64	GCAAATATCGCGTCTGGCCTTCCTGGCCTCAG
65	TATATTTTAGCTGATAAATTAATGTTGTATAA
66	CGAGTAGAACTAATAGTAGTAGCAAACCCTCA
67	TCAGAAGCCTCCAACAGGTCAGGATCTGCGAA
68	CATTCAACGCGAGAGGCTTTTGCATATTATAG
69	AGTAATCTTAAATTGGGCTTGAGAGAATACCA
70	ATACGTAAAAGTACAACGGAGATTTCATCAAG
71	AAAAAAGGACAACCATCGCCCACGCGGGTAAA
72	TGTAGCATTCCACAGACAGCCCTCATCTCCAA
73	CCCCGATTTAGAGCTTGACGGGGAAATCAAAA
74	GAATAGCCGCAAGCGGTCCACGCTCCTAATGA
75	GTGAGCTAGTTTCCTGTGTGAAATTTGGGAAG
76	GGCGATCGCACTCCAGCCAGCTTTGCCATCAA
77	AAATAATTTTAAATTGTAAACGTTGATATTCA
78	ACCGTTCTAAATGCAATGCCTGAGAGGTGGCA
79	TCAATTCTTTTAGTTTGACCATTACCAGACCG
80	GAAGCAAAAAAGCGGATTGCATCAGATAAAAA
81	CCAAAATATAATGCAGATACATAAACACCAGA
82	ACGAGTAGTGACAAGAACCGGATATACCAAGC
83	GCGAAACATGCCACTACGAAGGCATGCGCCGA

84	CAATGACACTCCAAAAGGAGCCTTACAACGCC
85	CCAGCAGGGGCAAAATCCCTTATAAAGCCGGC
86	GCTCACAATGTAAAGCCTGGGGTGGGTTTGCC
87	GCTTCTGGTCAGGCTGCGCAACTGTGTTATCC
88	GTTAAAATTTTAACCAATAGGAACCCGGCACC
89	AGGTAAAGAAATCACCATCAATATAATATTTT
90	TCGCAAATGGGGCGCGAGCTGAAATAATGTGT
91	AAGAGGAACGAGCTTCAAAGCGAAGATACATT
92	GGAATTACTCGTTTACCAGACGACAAAAGATT
93	CCAAATCACTTGCCCTGACGAGAACGCCAAAA
94	AAACGAAATGACCCCCAGCGATTATTCATTAC
95	TCGGTTTAGCTTGATACCGATAGTCCAACCTA
96	TGAGTTTCGTCACCAGTACAAACTTAATTGTA
97	GAACGTGGCGAGAAAGGAAGGGAACAAACTAT
98	CCGAAATCCGAAAATCCTGTTTGAAGCCGGAA
99	GCATAAAGTTCCACACAACATACGAAGCGCCA
100	TTCGCCATTGCCGGAAACCAGGCATTAAATCA
101	GCTCATTTTCGCATTAAATTTTTGAGCTTAGA
102	AGACAGTCATTCAAAAGGGTGAGAAGCTATAT
103	TTTCATTTGGTCAATAACCTGTTTATATCGCG
104	TTTTAATTGCCCGAAAGACTTCAAAACACTAT
105	CATAACCCGAGGCATAGTAAGAGCTTTTTAAG
106	GAATAAGGACGTAACAAAGCTGCTCTAAAACA
107	CTCATCTTGAGGCAAAAGAATACAGTGAATTT
108	CTTAAACATCAGCTTGCTTTCGAGCGTAACAC
109	ACGAACCAAAACATCGCCATTAAATGGTGGTT
110	CGACAACTAAGTATTAGACTTTACAATACCGA
111	CTTTTACACAGATGAATATACAGTAAACAATT
112	TTAAGACGTTGAAAACATAGCGATAACAGTAC
113	GCGTTATAGAAAAAGCCTGTTTAGAAGGCCGG
114	ATCGGCTGCGAGCATGTAGAAACCTATCATAT
115	CCTAATTTACGCTAACGAGCGTCTAATCAATA
116	AAAAGTAATATCTTACCGAAGCCCTTCCAGAG
117	TTATTCATAGGGAAGGTAAATATTCATTCAGT
118	GAGCCGCCCCACCACCGGAACCGCGACGGAAA
119	AATGCCCCGTAACAGTGCCCGTATCTCCCTCA
120	CAAGCCCAATAGGAACCCATGTACAAACAGTT
121	CGGCCTTGCTGGTAATATCCAGAACGAACTGA
122	TAGCCCTACCAGCAGAAGATAAAAACATTTGA
123	GGATTTAGCGTATTAAATCCTTTGTTTTCAGG
124	TTTAACGTTCGGGAGAAACAATAATTTTCCCT
125	TAGAATCCCTGAGAAGAGTCAATAGGAATCAT

126	AATTACTACAAATTCTTACCAGTAATCCCATC
127	CTAATTTATCTTTCCTTATCATTCATCCTGAA
128	TCTTACCAGCCAGTTACAAAATAAATGAAATA
129	GCAATAGCGCAGATAGCCGAACAATTCAACCG
130	ATTGAGGGTAAAGGTGAATTATCAATCACCGG
131	AACCAGAGACCCTCAGAACCGCCAGGGGTCAG
132	TGCCTTGACTGCCTATTTCGGAACAGGGATAG
133	AGGCGGTCATTAGTCTTTAATGCGCAATATTA
134	TTATTAATGCCGTCAATAGATAATCAGAGGTG
135	CCTGATTGAAAGAAATTGCGTAGACCCGAACG
136	ATCAAAATCGTCGCTATTAATTAACGGATTCG
137	ACGCTCAAAATAAGAATAAACACCGTGAATTT
138	GGTATTAAGAACAAGAAAAATAATTAAAGCCA
139	ATTATTTAACCCAGCTACAATTTTCAAGAACG
140	GAAGGAAAATAAGAGCAAGAAACAACAGCCAT
141	GACTTGAGAGACAAAAGGGCGACAAGTTACCA
142	GCCACCACTCTTTTCATAATCAAACCGTCACC
143	CTGAAACAGGTAATAAGTTTTAACCCCTCAGA
144	CTCAGAGCCACCACCCTCATTTTCCTATTATT
145	CCGCCAGCCATTGCAACAGGAAAAATATTTTT
146	GAATGGCTAGTATTAACACCGCCTCAACTAAT
147	AGATTAGATTTAAAAGTTTGAGTACACGTAAA
148	ACAGAAATCTTTGAATACCAAGTTCCTTGCTT
149	CTGTAAATCATAGGTCTGAGAGACGATAAATA
150	AGGCGTTACAGTAGGGCTTAATTGACAATAGA
151	TAAGTCCTACCAAGTACCGCACTCTTAGTTGC
152	TATTTTGCTCCCAATCCAAATAAGTGAGTTAA
153	GCCCAATACCGAGGAAACGCAATAGGTTTACC
154	AGCGCCAACCATTTGGGAATTAGATTATTAGC
155	GTTTGCCACCTCAGAGCCGCCACCGATACAGG
156	AGTGTACTTGAAAGTATTAAGAGGCCGCCACC
157	GCCACGCTATACGTGGCACAGACAACGCTCAT
158	ATTTTGCGTCTTTAGGAGCACTAAGCAACAGT
159	GCGCAGAGATATCAAAATTATTTGACATTATC
160	TAACCTCCATATGTGAGTGAATAAACAAAATC
161	CATATTTAGAAATACCGACCGTGTTACCTTTT
162	CAAGCAAGACGCGCCTGTTTATCAAGAATCGC
163	TTTTGTTTAAGCCTTAAATCAAGAATCGAGAA
164	ATACCCAAGATAACCCACAAGAATAAACGATT
165	AATCACCAAATAGAAAATTCATATATAACGGA
166	CACCAGAGTTCGGTCATAGCCCCCGCCAGCAA
167	CCTCAAGAATACATGGCTTTTGATAGAACCAC

168	CCCTCAGAACCGCCACCCTCAGAACTGAGACT
169	GGAAATACCTACATTTTGACGCTCACCTGAAA
170	GCGTAAGAGAGAGCCAGCAGCAAAAAGGTTAT
171	CTAAAATAGAACAAAGAAACCACCAGGGTTAG
172	AACCTACCGCGAATTATTCATTTCCAGTACAT
173	AAATCAATGGCTTAGGTTGGGTTACTAAATTT
174	AATGGTTTACAACGCCAACATGTAGTTCAGCT
175	AATGCAGACCGTTTTTATTTTCATCTTGCGGG
176	AGGTTTTGAACGTCAAAAATGAAAGCGCTAAT
177	ATCAGAGAAAGAACTGGCATGATTTTATTTTG
178	TCACAATCGTAGCACCATTACCATCGTTTTCA
179	TCGGCATTCCGCCGCCAGCATTGACGTTCCAG
180	TAAGCGTCGAAGGATTAGGATTAGTACCGCCA
181	CTAAAGCAAGATAGAACCCTTCTGAATCGTCT
182	CGGAATTATTGAAAGGAATTGAGGTGAAAAAT
183	GAGCAAAAACTTCTGAATAATGGAAGAAGGAG
184	TATGTAAACCTTTTTTAATGGAAAAATTACCT
185	AGAGGCATAATTTCATCTTCTGACTATAACTA
186	TCATTACCCGACAATAAACAACATATTTAGGC
187	CTTTACAGTTAGCGAACCTCCCGACGTAGGAA
188	TTATTACGGTCAGAGGGTAATTGAATAGCAGC
189	CCGGAAACACACCACGGAATAAGTAAGACTCC
190	TGAGGCAGGCGTCAGACTGTAGCGTAGCAAGG
191	TGCTCAGTCAGTCTCTGAATTTACCAGGAGGT
192	TATCACCGTACTCAGGAGGTTTAGCGGGGTTT
193	GAAATGGATTATTTACATTGGCAGACATTCTG
194	GCCAACAGTCACCTTGCTGAACCTGTTGGCAA
195	ATCAACAGTCATCATATTCCTGATTGATTGTT
196	TGGATTATGAAGATGATGAAACAAAATTTCAT
197	TTGAATTATGCTGATGCAAATCCACAAATATA
198	TTTTAGTTTTTCGAGCCAGTAATAAATTCTGT
199	CCAGACGAGCGCCCAATAGCAAGCAAGAACGC
200	GAGGCGTTAGAGAATAACATAAAAGAACACCC
201	TGAACAAACAGTATGTTAGCAAACTAAAAGAA
202	ACGCAAAGGTCACCAATGAAACCAATCAAGTT
203	TGCCTTTAGTCAGACGATTGGCCTGCCAGAAT
204	GGAAAGCGACCAGGCGGATAAGTGAATAGGTG
205	AAACCCTCTTTTACCAGTAATAAAAGGGATTCACCAGTCACACGTTTT
206	GATGGCAATTTTAATCAATATCTGGTCACAAATATC
207	AAAACAAATTTTTTCATCAATATAATCCTATCAGAT
208	ACAAAGAATTTTATTAATTACATTTAACACATCAAG
209	TAAAGTACTTTTCGCGAGAAAACTTTTTATCGCAAG

210	TATAGAAGTTTTCGACAAAAGGTAAAGTAGAGAATA
211	GCGCATTATTTTGCTTATCCGGTATTCTAAATCAGA
212	TACATACATTTTGACGGGAGAATTAACTACAGGGAA
213	AGCACCGTTTTTTAAAGGTGGCAACATAGTAGAAAA
214	ACAAACAATTTTAATCAGTAGCGACAGATCGATAGC
215	AGGGTTGATTTTATAAATCCTCATTAAATGATATTC
216	TTTTTATAAGTATAGCCCGGCCGTCGAG
217	AACATCACTTGCCTGAGTAGAAGAACT
218	TGTAGCAATACTTCTTTGATTAGTAAT
219	AGTCTGTCCATCACGCAAATTAACCGT
220	ATAATCAGTGAGGCCACCGAGTAAAAG
221	ACGCCAGAATCCTGAGAAGTGTTTTT
222	TTAAAGGGATTTTAGACAGGAACGGT
223	AGAGCGGGAGCTAAACAGGAGGCCGA
224	TATAACGTGCTTTCCTCGTTAGAATC
225	GTACTATGGTTGCTTTGACGAGCACG
226	GCGCTTAATGCGCCGCTACAGGGCGC

All sequences in Table 3.1 are listed from 5′ to 3′ and correspond to unmodified staples. Biotin modifications were performed as follows: for **R**, **D5**, **D10**, and **D20** tiles, staples 53, 57, 103, 160, and 164 were divided into two 18 nucleotide long staples and one of the resulting fragments was modified with a biotin molecule at the 5′ end. For **L** tiles, staples 3, 11, 206, and 214 were similarly divided and modified.

For **R** tiles, the following staples were modified with the substrate binding probe sequence at the 5′ end: 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 50, 51, 52, 58, 59, 60, 61, 62, 63, 69, 70, 71, 74, 75, 76, 82, 83, 84, 85, 86, 87, 93, 94, 95, 98, 99, 100, 106, 107, 108, 109, 110, 111, 117, 118, 119, 122, 123, 124, 130, 131, 132, 133, 134, 135, 141, 142, 143, 146, 147, 148, 154, 155, 156, 157, 158, 159, 165, 166, 167, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204

For **L** tiles, the following staples were modified with the substrate binding probe sequence at the 5′ end: 13, 14, 25, 26, 27, 28, 37, 38, 39, 51, 52, 53, 62, 63, 64, 76, 77, 78, 87, 88, 89, 101, 102, 103, 112, 113, 114, 126, 127, 128, 137, 138, 139, 151, 152, 153, 162, 163, 164, 176, 177, 178

For **D5** tiles, the following staples were modified with the substrate binding probe sequence at the 5′ end: 13-52, 54-57,59-102,104-148,150-211

For **D10** tiles, the following staples were modified with the substrate binding probe sequence at the 5′ end: 13, 15, 17, 19, 21, 23, 37, 39, 41, 43, 45, 61, 63, 65, 67, 69, 71, 85, 87, 89, 91, 93, 95, 109, 111, 113, 115, 117, 119, 133, 135, 137, 139, 141, 143, 157, 159, 161, 163, 165, 167, 181, 183, 185, 187, 189, 191

For **D20** tiles, the following staples were modified with the substrate binding probe sequence at the 5′ end: 27, 31, 35, 75, 79, 83, 123, 127, 131, 171, 175, 179

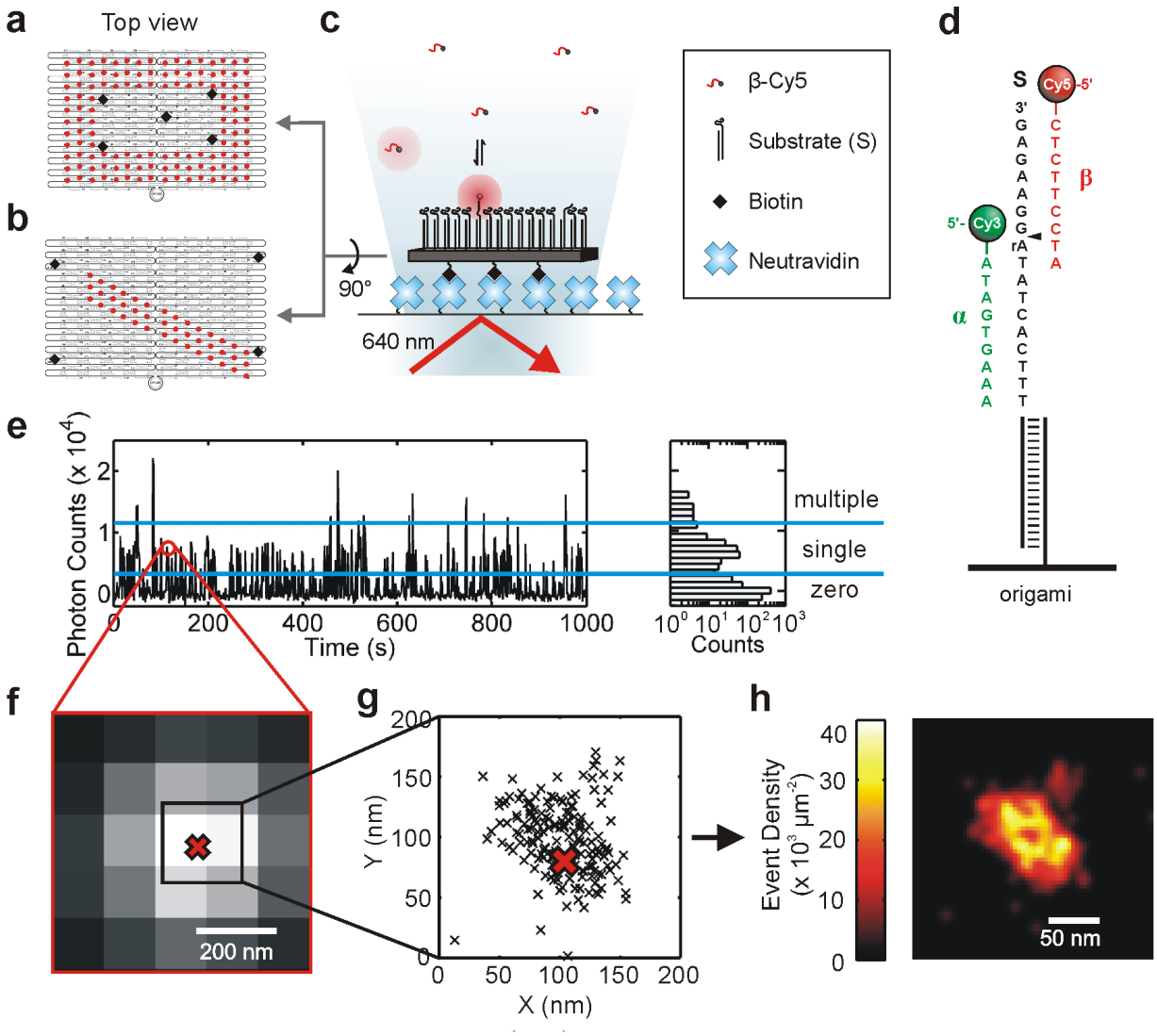


Figure 3.2. Origami tile designs used in this study: (a) rectangular origami **R** bearing 126 substrates (red circles) and 5 biotin molecules (black diamonds) for immobilization on a NeutrAvidin-coated microscope slide; and (b) linear origami **L** bearing 42 substrates and 4 biotin molecules. Substrates and biotins are displayed on opposite faces of the tile. (c) Scheme for PAINT experiments. DNA origami were immobilized on a NeutrAvidin-coated fused silica slide on a TIRF microscope *via* multiple biotin-NeutrAvidin interactions. Imaging occurred in the presence of single-stranded DNA probes that were fluorescently labeled at their 5′-end. As probes bind reversibly to the substrates on the origami tile, they enter the evanescent field of excitation light and are localized. (d) Sequences of the substrate (S) and fluorescently labeled PAINT probes **α-Cy3** and **β-Cy5**. In some experiments, only **β-Cy5** was used; in others, **β** was labeled with Cy3 and **α** was labeled with Cy5. **S** contains an RNA base (rA) to allow for enzymatic cleavage at the site indicated by the black triangle. (e) Fluorescence intensity time trace and histogram showing repeated binding of **β-Cy5** to a single **R** origami tile.

Only binding events with intensity between the two horizontal blue lines were used in the reconstruction, as these have a high probability of originating from individual **β-Cy5** molecules (rather than $\geq$ 2 bound simultaneously). For ease of viewing, only 1,000 s are shown from a 3,000-s experiment. (f) Wide-field diffraction-limited fluorescence image of the **β-Cy5** binding event circled in (e). The intensity profile is fit with a 2-D Gaussian function to localize the binding event (red X). (g) Coordinates of 174 localizations of **β-Cy5** binding PAINT reconstruction of an **R** origami. The red X corresponds to the localization of the binding event shown in (f). (h) PAINT reconstruction of the origami shown in (g). Each experiment yielded reconstructions for ~20-100 origami.

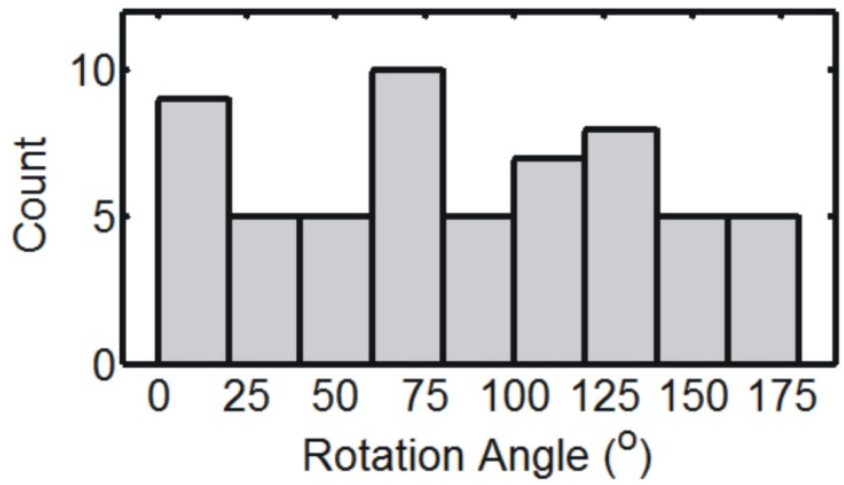


Figure 3.3. Angles of Rotation for 117 R Origami in One Movie. Unidirectional rotation angle providing optimal alignment with a 60 x 100 nm rectangular mask for 117 **R** origami. The distribution is isotropic ($\chi^2(8, N = 117) = 4.9, P = 0.76$), consistent with random deposition of origami on the slide surface.

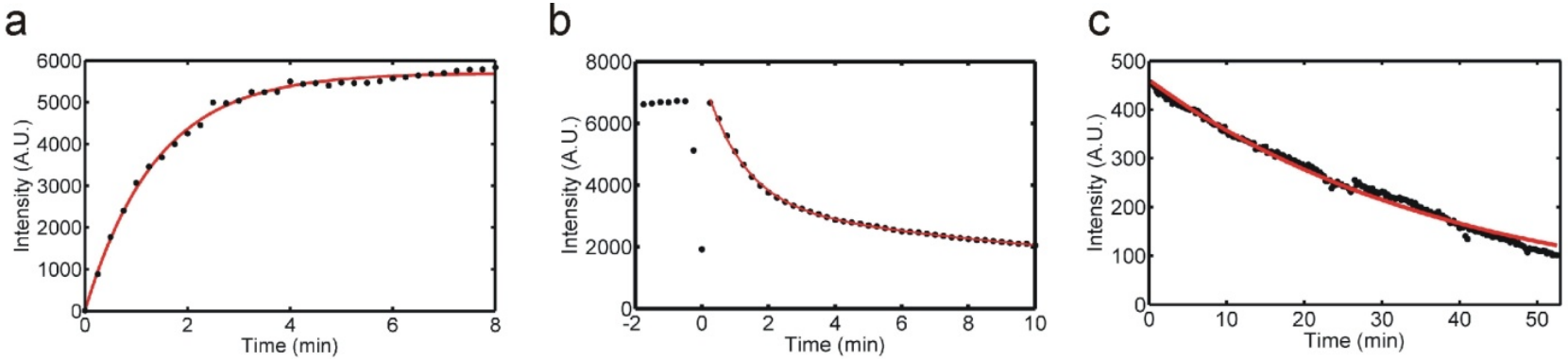


Figure 3.4. Kinetics of Substrate Assembly, Substrate Cleavage, and Dissociation of 8-17 DNAzyme from Substrate on R Origami Pegboards. **a,** Fluorescence time course of 100 nM Cy3-labeled substrate (**S-Cy3**) binding to surface-immobilized **R** origami (black dots). This is a ~40-fold lower concentration of substrate and ~100-fold lower concentration of origami than that used in preparing samples for PAINT imaging. Fitting to the single exponential model $y = C(1-e^{-kt})$ yields an apparent pseudo-first-order rate constant $k'_{obs} = 0.72\ min^{-1}$ (red curve, $R^2 = 0.998$). The y-coordinate is the mean Cy3 fluorescence intensity of 382 origami. **S-Cy3** was added at time $t = 0$ with a dead time of < 10 s. **b,** Fluorescence time course of cleavage of **S-Cy3** on **R** origami in the presence of 1 μM 8-17 DNAzyme and 1 mM $ZnSO_4$. Fitting to a double exponential model $y = C_1(e^{-k1*t}) + C_2(e^{-k2*t})$ to the interval from 0.25-10 min (red curve, $R^2 > 0.999$) yields apparent rate constants of 0.83 min^{-1} (relative amplitude 0.56) and 0.05 min^{-1} (relative amplitude 0.44). The y-axis is the mean Cy3 fluorescence intensity from 259 origami. **S-Cy3** was added at time $t = 0$ with a dead time of < 10 s. The temporary drop in intensity around time $t = 0$ was caused by the sample going out of focus due to mechanical perturbations immediately prior to the addition. **c**, Fluorescence time course of dissociation of Cy3-labeled **8-17** DNAzyme from unlabeled **S** immobilized on **R** origami. Fitting to the single-exponential model $y = C(1-e^{-kt})$ yields an apparent first-order rate constant of 0.025 min^{-1} (red line, $R^2 = 0.994$). The y-axis is the mean Cy3 fluorescence intensity from 181 origami.

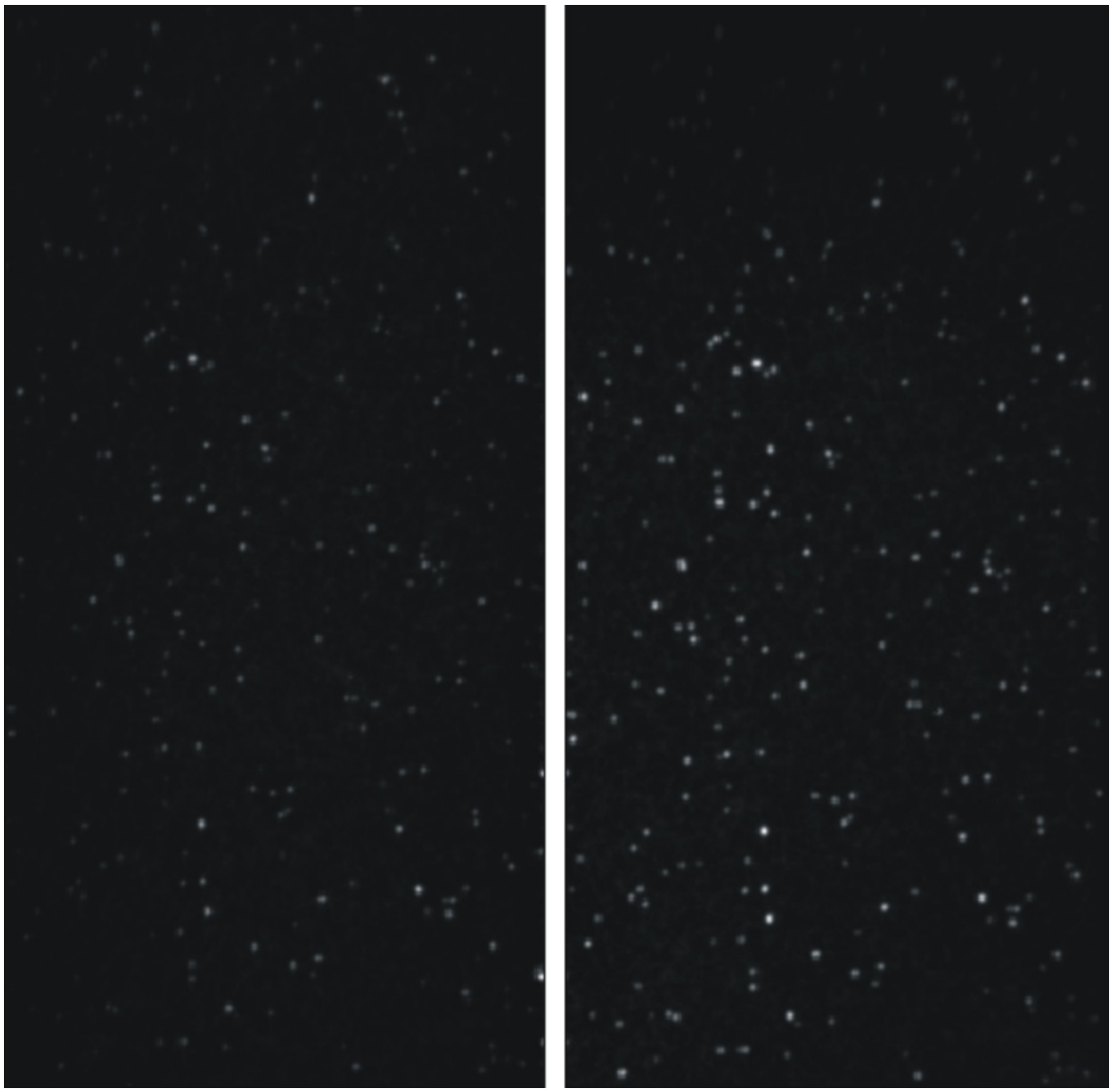


Figure 3.5. Fluctuation Map of a Representative Field of View Containing Origami. Fluctuation map for of a representative 34 x 68 μM field of view for **α-Cy3** (left half) and **β-Cy5** (right half). The intensity of each pixel in the fluctuation map is proportional to the average frame-to-frame intensity fluctuation of that pixel in the raw movie. Origami are thus localized as the sites of repeated appearance and disappearance of Cy3 and Cy5 signal, which appear as bright spots in the fluctuation map.

Probe Strand	$k_{on,R}$ (x 10^7 M^{-1} s^{-1})	k_{off} (s^{-1})
α-Cy3	1.98 ± 0.04	0.34 ± 0.01
β-Cy5	2.03 ± 0.03	0.22 ± 0.03

Table 3.2. Kinetics of Probe Strand Binding to R Origami. Error bars are 1 s.e.m from duplicate measurements. $k_{on,R}$ represents the apparent second-order rate constant of probe binding to an **R** origami tile with up to 126 copies of **S**. Kinetic characterization was conducted under ~10% maximal illumination to limit photobleaching, and results were insensitive to a two-fold change in excitation intensity (data not shown).

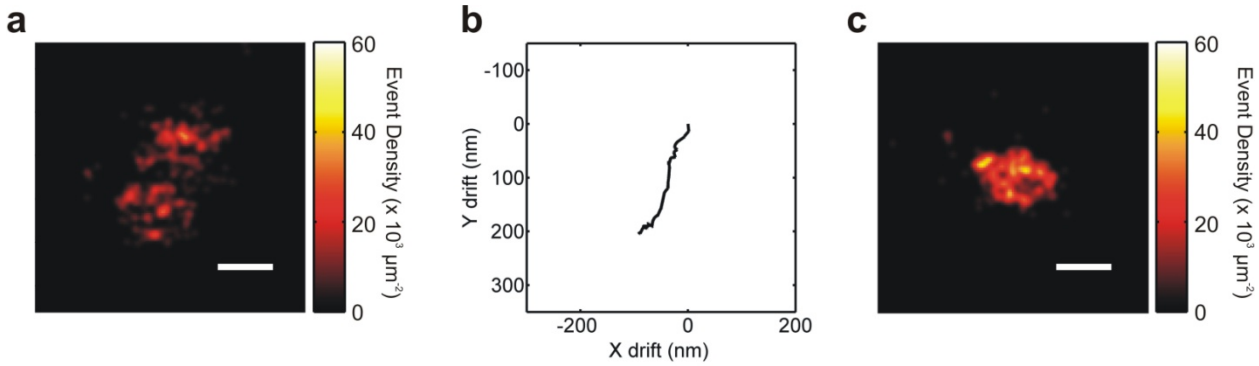


Figure 3.6. Impact of Stage Drift Correction on Reconstruction Quality**. a,** Monochromatic PAINT reconstruction of one **R** origami without accounting for X-Y drift of the microscope stage. **b**, Drift of the microscope stage as determined by cross-correlation between consecutive ~200-s bins of the original movie. **c,** Final reconstruction obtained by subtracting the drift (**b**) from the raw coordinates in (**a**). Scale bars in **a** and **c** are 100 nm.

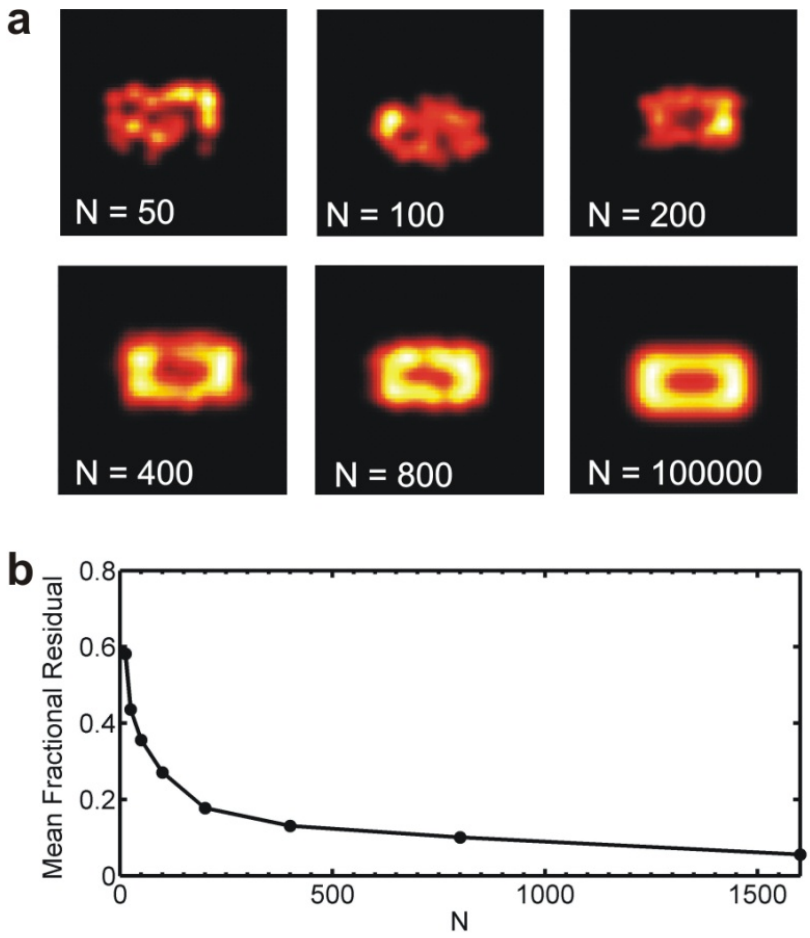


Figure 3.7. Impact of Sampling on Reconstruction Quality. **a**, Representative simulated reconstructions of **R** origami with varying numbers of PAINT binding events. **b**, Mean fractional residual of reconstructions as a function of the number of localizations, N. Residuals were calculated by subtracting the intensity profile of each reconstruction from that of an ideal reconstruction (N = 100000, panel **a**), with all reconstructions normalized to a maximal value of 1. The residuals are expressed as a fraction of the total intensity of the ideal reconstruction.

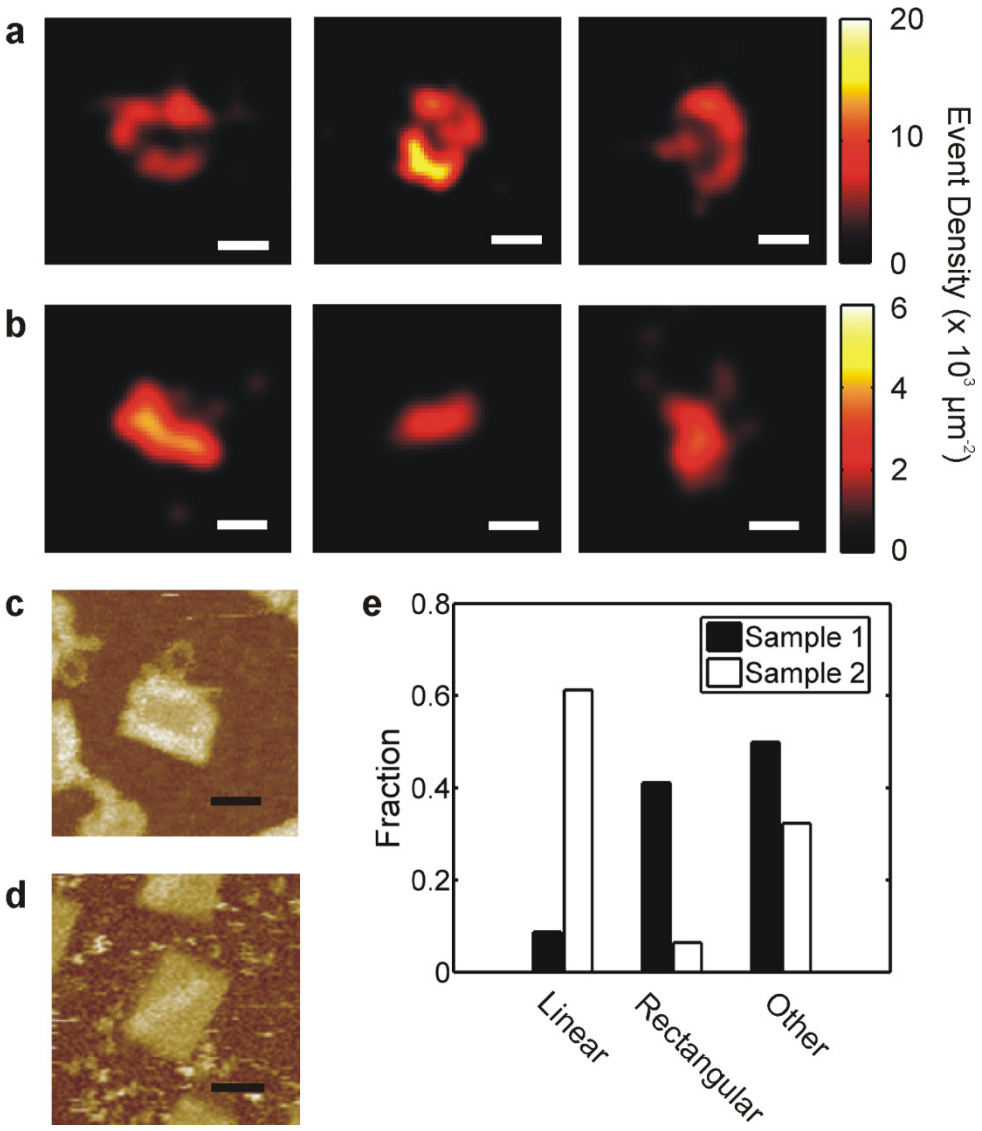


Figure 3.8. One-color PAINT reconstructions of **R** (a) and **L** (b) origami tiles imaged in the presence of 10 nM **β-Cy5** with 20-60 binding events per tile. Scale bars: 50 nm. (c), (d) Atomic force micrographs of **R** (c) and **L** (d) origami. Scale bars: 50 nm. (e) Results from a blind experiment in which two origami samples of unknown identity (either **L** or **R**) were imaged in the presence of 10 nM **β-Cy5** and classified according to their morphology: linear (e.g. Figure 3.8b), rectangular (e.g., Figure 3.8d), or other (42 and 27 origami were examined from samples 1 and 2, respectively). The "other" category likely included malformed origami tiles, aggregates of multiple origami, or origami with spatially heterogeneous binding of **β-Cy5** (see Figure 3.13). Samples 1 and 2 were correctly identified as **R** and **L**, respectively.

few dozen localizations per origami (Figure 3.8e). For both patterns, a significant fraction (~33% for **L**, ~50% for **R**) of reconstructions could not be classified as linear or rectangular, in agreement with independent estimates of assembly yield from AFM images (Table 3.3). Furthermore, a model-free alignment of 198 reconstructions of **R** origami, each comprising 100-300 localizations (Figure 3.9), using standard single particle analysis software EMAN revealed several class averages resembling the desired rectangular structure (45-55% of origami), with most of the remaining class averages resembling aggregated or incompletely assembled origami. Many defects revealed in the PAINT images have counterparts in AFM images (Figure 3.9), suggesting that they are due to imperfect tile or pegboard assembly. This is consistent with the fact that the number of binding events per origami is distributed more broadly than would be predicted for binding to a set of identical, fully assembled pegboards (Figure 3.10). We note, however, that PAINT monitors the single-stranded DNA regions involved in interactions with external reagents that are too soft to be visible by AFM.

To demonstrate sequence-specific imaging, the **R** pattern was evaluated simultaneously in the presence of **α-Cy3** and **β-Cy5**. The resulting binding patterns were reconstructed and registered in the same coordinate space, resulting in a two-color overlay (Figure 3.11b, Figure 3.12). Unlike other fluorescence nanoscopy techniques, PAINT is insensitive to photobleaching and labeling efficiency due to the vast reserve of probes in solution that are readily exchanged for origami-bound probes, enabling quantitative imaging over hours. Since **S** contains a single ribonucleotide (Figure 3.2d), it can be site-specifically cleaved by an **8-17** DNAzyme in the presence of Zn^{2+} such that, after cleavage, the **β-Cy5** binding frequency is expected to diminish relative to that of **α-Cy3**. Incubation with the deoxyribozyme results in a time-dependent decrease in **β-Cy5** relative to **α-Cy3** binding, consistent with but going beyond ensemble-averaged measurements (Figures 3.11b-c, 3.4) by demonstrating PAINT's ability to spatiotemporally monitor enzymatic remodeling reactions on individual origami nanodevices.

The use of two probe strands also provides a means of assessing the homogeneity of binding to nanostructures. Surprisingly, we found several cases where one probe bound uniformly across the pattern of **S** and the other did not, even for well-formed **R** patterns

Status	R (N = 592)	L (N = 564)
Well-formed tile with clear evidence of fairly complete S **pattern**	30.1% (178/592)	65.2% (368/564)
Well-formed tile with defective or missing S **pattern**	21.8% (129/592)	7.3% (41/564)
Broken or deformed tile	48.1% (285/592)	27.5% (155/564)

Table 3.3. Yield of R and L Tiles and Patterns Determined by AFM.

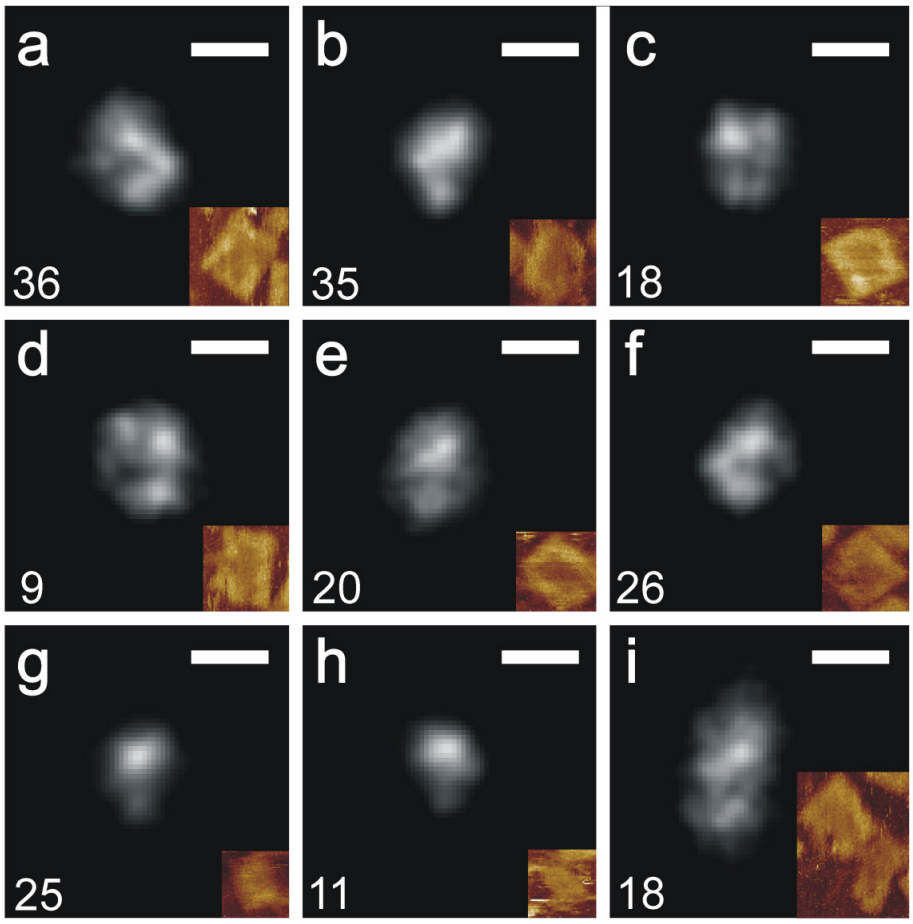


Figure 3.9. **a-i,** Nine class averages demonstrating the observed morphological variety in reconstructions of 198 **R** origami from **β-Cy5** binding. The number of reconstructions in a given class is indicated at the bottom center of the panel. Classes **a, c, d, f, and** possibly **e** represent well-formed origami immobilized with a relatively flat geometry parallel to the plane of the microscope slide. Scale bars (upper right of each frame): 100 nm. AFM images of **R** individual origami bearing **S** patterns comparable to each PAINT class average are shown in the lower right corner of each panel, at the same scale as the corresponding PAINT image. The class averages in panels **g** and **h** may represent origami fragments and/or incomplete immobilization (e.g., freely rotating due to being attached by only one biotin). Class average **i** may correspond to various side-by-side aggregates of origami.

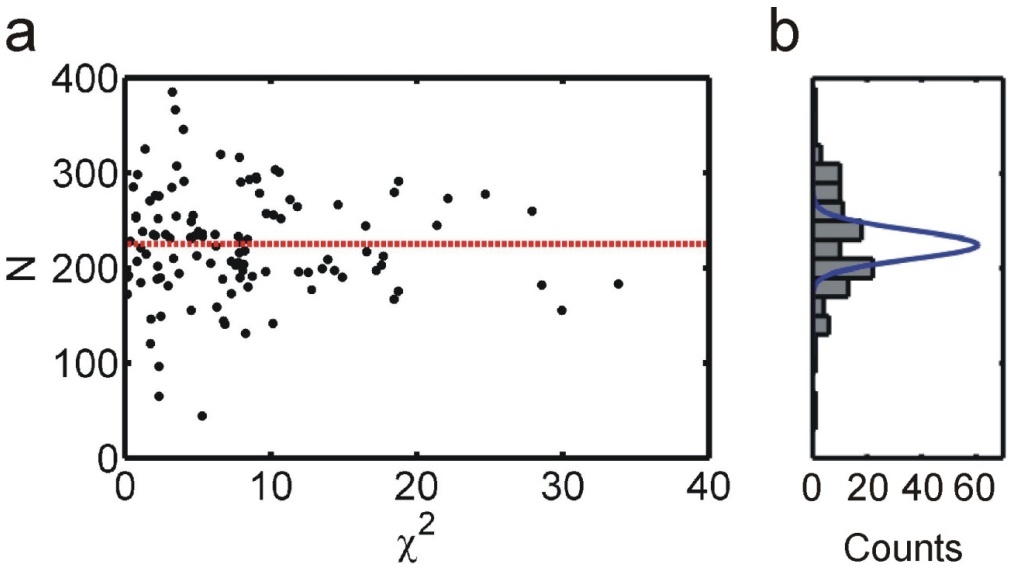


Figure 3.10. Number of β-Cy5 Binding Events Per Origami Versus Spatial Heterogeneity of Binding. **a,** Chi-squared value vs. number of **β-Cy5** binding events (N) per origami for 114 **R** origami in one movie. The Pearson correlation coefficient between χ^2 and N is -0.03, demonstrating that there is no correlation between overall probe binding efficiency and measured heterogeneity. The red dashed line is the mean binding events per origami, 225 +/- 57 (s.d.). **b,** Histogram of the values of N shown in panel **a** (gray bars) as compared to the predicted Poisson distribution for identical origami. The observed distribution is significantly broader than the theoretical distribution ($\chi^2(113) = 1625$, $p < 10^{-10}$), suggesting considerable differences in the number or availability of assembled **S** strands across different origami tiles.

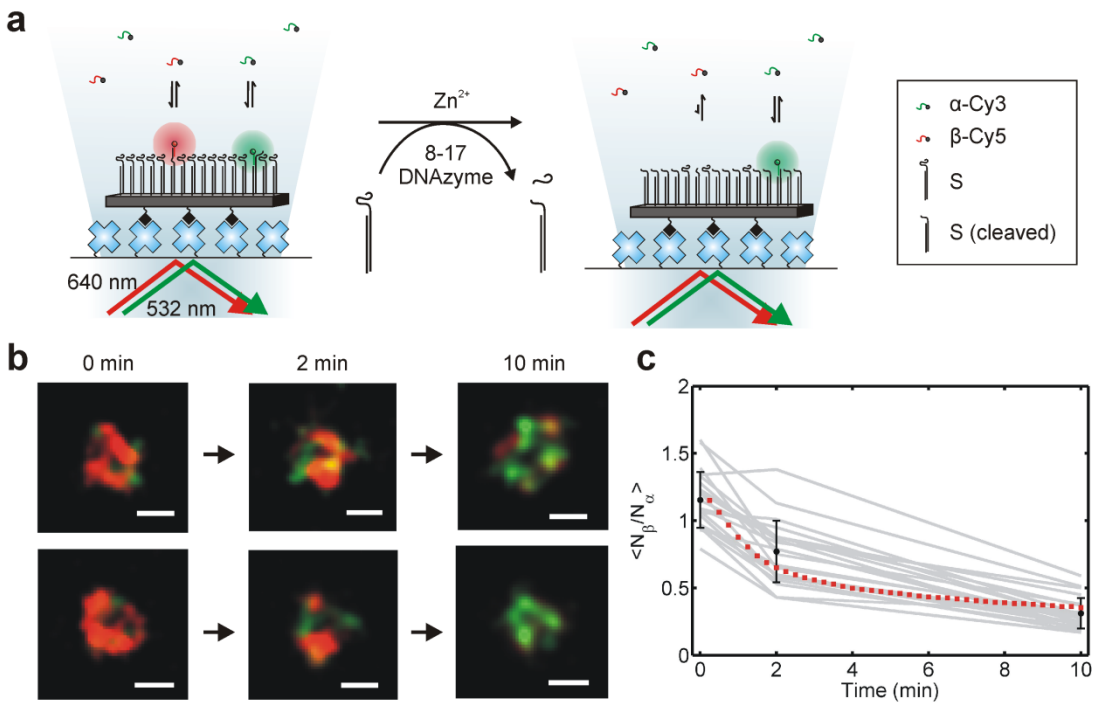


Figure 3.11. (a) Experiment for monitoring chemical changes by two-color PAINT. **R** origami tiles were imaged in the presence of 10 nM each of **α-Cy3** and **β-Cy5**. Incubation with 1 μM 8-17 deoxyribozyme (DNAzyme) and 1 mM Zn^{2+} results in the cleavage of **S**. The cleavage product can bind probe α, but not probe β, resulting in a change in the PAINT readout. (b) Two-color reconstructions of three individual **R** origami tiles after 0, 2, or 10 min total incubation with 8-17 DNAzyme and Zn^{2+} (**α-Cy3**, green; **β-Cy5**, red, scale bars 50 nm). (c) Mean ratio of β binding events to α binding events for 21 **R** origami after 0, 2, or 10 min total incubation with 1 μM 8-17 DNAzyme and 1 mM $ZnSO_4$ (black circles, error bars 1 s.d.). The time courses for 21 individual origami tiles are also shown (gray lines). An ensemble time course for the cleavage of substrate on R origami under identical conditions, normalized to the initial value of $<N_\beta/N_\alpha>$, is shown for comparison (red squares, Figure 3.4).

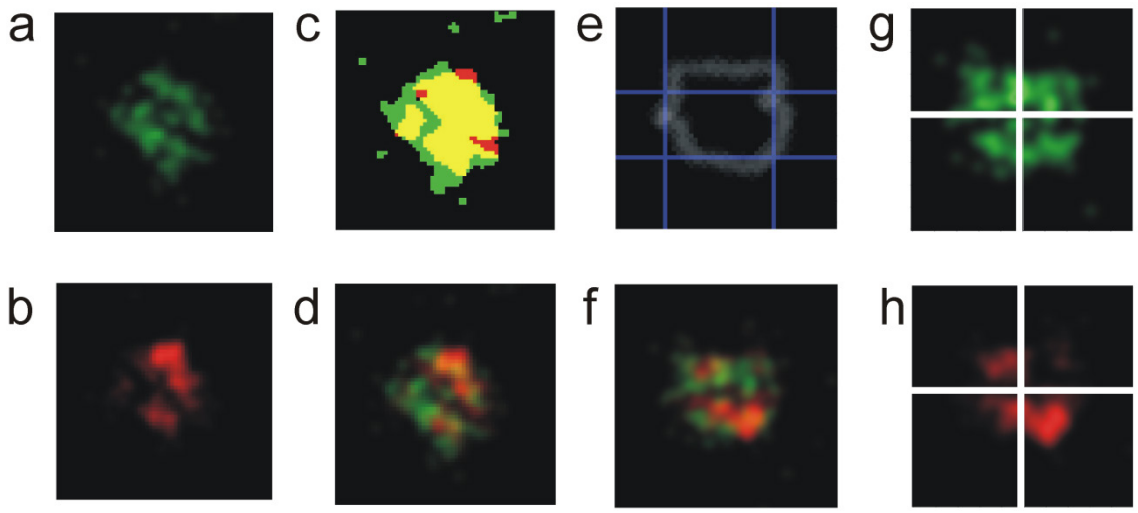


Figure 3.12. Registration and Alignment of Two-Color PAINT Reconstructions of R Origami**. a**, **b**, **c**, **d**, Drift-corrected reconstructions from Cy3 (**a**) and Cy5 (**b**) binding to a single **R** origami tile are registered by passing each through a binary filter with a threshold of 20% maximal intensity, then finding the maximal cross-correlation between the filtered images (**c**), resulting in a two-color overlay (**d**). **e**, **f**, An edge map of the origami (**e**, gray) is aligned with a 60 x 100 nm rectangular mask (**e**, blue) to align the reconstruction with the cardinal axes (**f**). The reconstruction is then divided into four quadrants, with the boundaries chosen so as to divide the area of the origami as equally as possible between the four quadrants. **g**, **h**, Finally, the divisions are applied to each channel separately to calculate χ^2 for the binding distribution of each probe.

(Figure 3.13a). To rule out fluorophore-specific imaging errors as the source of non-uniform binding, a set of **R** origami was imaged first with the probe combination **α-Cy3** + **β-Cy5** and subsequently with the labels inverted, i.e., **β-Cy3** + **α-Cy5**. A chi-squared test of homogeneity across different quadrants of the rectangular pattern (Figure 3.12) revealed that the binding patterns of **α-Cy3** and **α-Cy5** to the origami in Figure 3.11 are indistinguishable from homogeneous binding, while the binding of **β-Cy5** and **β-Cy3** cannot be explained by a homogeneous model ($\chi^2(3, N > 99) > 15$, $P < 0.002$). Furthermore, the intensity difference profile, calculated by subtracting the Cy5 reconstruction from the Cy3 reconstruction (Figure 3.14), appears to invert upon switching the probe labels (Figure 3.11b,c), with a 2-dimensional correlation coefficient of -0.67. Taken together, these observations suggest a "fingerprint" of sequence-specific binding patterns for this tile, with more heterogeneous binding to **β** than to **α**, and that this fingerprint persists throughout the ~1 h time lag between imaging with **α-Cy3** + **β-Cy5** and **β-Cy3** + **α-Cy5**. Importantly, these patterns cannot be trivially attributed to a fraction of pre-cleaved **S**, which would lack the **β**-binding sequence (Figure 3.16). Furthermore, binding heterogeneity is not significantly correlated with the total number of binding events (Figure 3.10), which implies that some well-assembled origami with intact **S** nevertheless bind probes unevenly. Together, these data suggest that the accessibility of **β**-binding sequence varies across the surface of the origami somewhat independently of the accessibility of **α**-binding sequence.

We therefore hypothesized that local interactions between adjacent **S** strands exert a differential influence on **α** and **β** binding. To test this possibility, we measured the kinetics of **α-Cy3** and **β-Cy5** binding to origami with spacings of ~5, 10, or 20 nm between nearest-neighbor **S** strands (Figure 3.17). We found that **β-Cy5** binding is slowed by ~25% relative to that of **α-Cy3** at a spacing of 5 nm between **S** strands, but not 10 or 20 nm (Figure 3.18a). This is consistent with a model in which interactions between nearby **S** strands compete with **β** binding, inhibiting rather than enhancing its binding relative to **α** (Figure 3.19). Thus, if the spacing of **S** varies across the surface of a tile, there may be regions in which **β-Cy5** binding is inhibited relative to that of **α-Cy3**, resulting in heterogeneous binding fingerprints such as those in Figure 3.13.

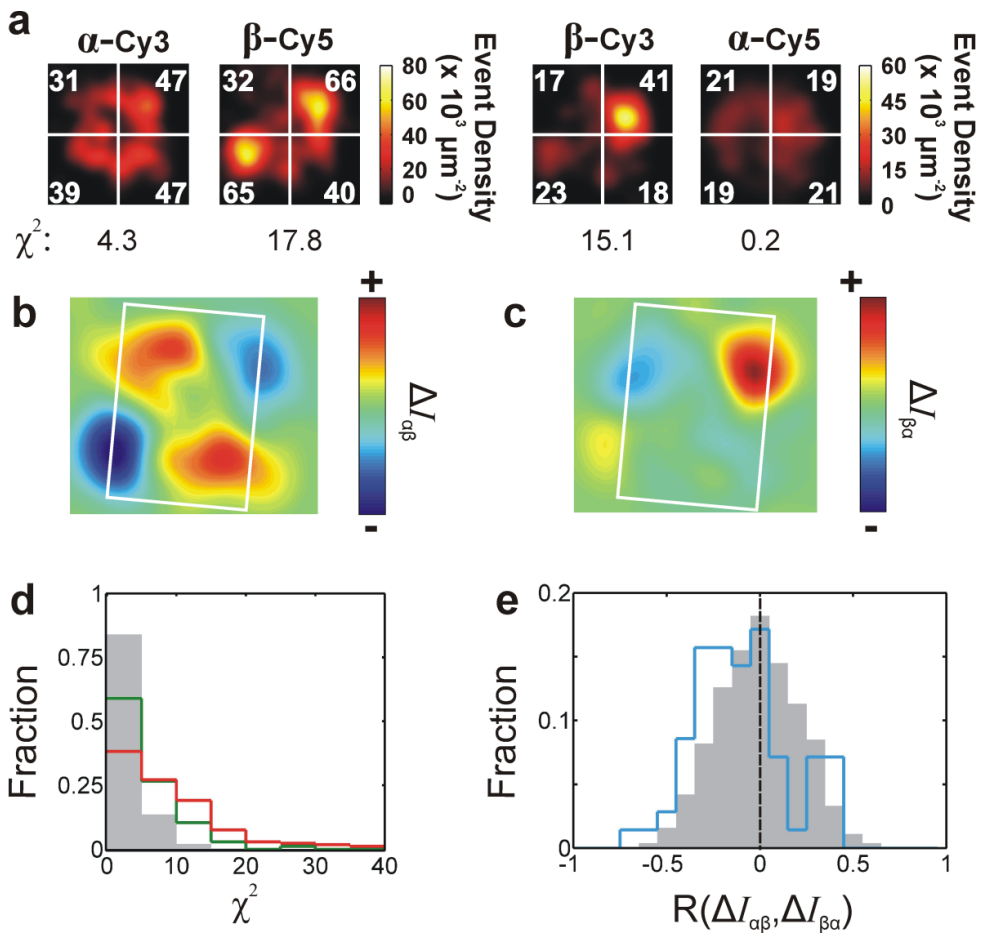


Figure 3.13. (a) DNA-PAINT reconstructions of the same **R** origami tile using two different sets of probes: **α-Cy3** + **β-Cy5**, and **β-Cy3** + **α-Cy5**, and quantification of binding uniformity by chi-squared analysis of the distribution of binding events between origami quadrants. The number of binding events observed in a 60-min period is indicated in each quadrant. The distributions of **α-Cy3** and **α-Cy5** binding can be explained by a homogeneous model, while that of **β-Cy5** and **β-Cy3** cannot (df = 3, $P < 0.001$). Reconstructions are 125 × 125 nm^2. (b),(c) Intensity difference maps, calculated by subtracting the Cy5 reconstruction from the Cy3 reconstruction, for the origami tile shown in (a) as imaged by the two probe sets. White rectangular outlines depict typical origami dimensions as measured by AFM (60 x 90 nm). In (b), $\Delta I_{\alpha\beta} = I_{\alpha\text{-}Cy3} - I_{\beta\text{-}Cy5}$, while in (c), $\Delta I_{\beta\alpha} = I_{\beta\text{-}Cy3} - I_{\alpha\text{-}Cy5}$. The difference maps in (b) and (c) have a correlation coefficient R = -0.67. (d) Histograms of χ^2 for the binding distributions of probes **α** (green line) and **β** (red line) to 173 **R** origami, as compared to the distribution predicted from 1,000 simulated **R** origami (gray shaded region). (e) Histogram of correlation coefficients between difference maps $\Delta I_{\alpha\beta}$ and $\Delta I_{\beta\alpha}$ for 70 **R** origami tiles (blue line; μ = -0.10, s.e.m. = 0.03) as compared to the results from 1,000 simulated tiles (gray shaded region; μ = -0.008, s.e.m. = 0.008). The black dashed line indicates the mean value of the simulated distribution. The experimental distribution is significantly skewed toward negative values compared to the simulated distribution ($t(69) = 3.1$, two-tailed P = 0.003).

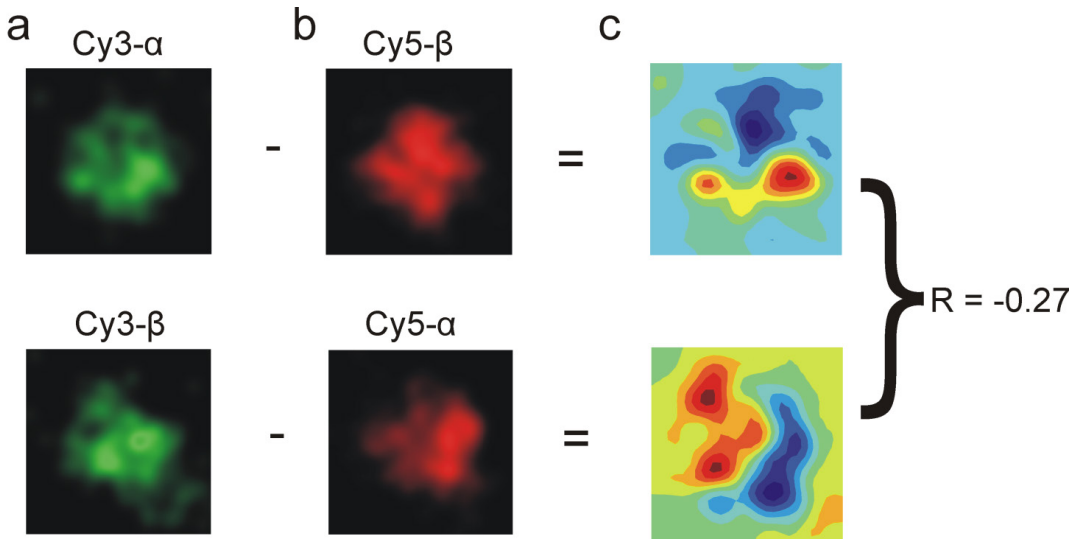


Figure 3.14. Calculation of Difference Maps and Correlation Coefficients from the Binding of Inversely Labeled Probes. PAINT reconstructions from the same origami using two sets of probes, (**α-Cy3** + **β-Cy5**) and (**β-Cy3** + **α-Cy5**), are used to calculate intensity difference maps to investigate the dependence of binding distributions on probe sequence. For each probe set, the normalized Cy5 reconstruction of a single origami (**b**) is subtracted from the normalized Cy3 reconstruction (**a**), yielding an intensity difference map (**c**). The correlation coefficient between the difference maps from the two probe sets is then calculated. A negative correlation coefficient is expected if there is sequence-dependent heterogeneity of binding that persists over the approximately 1 h between imaging experiments.

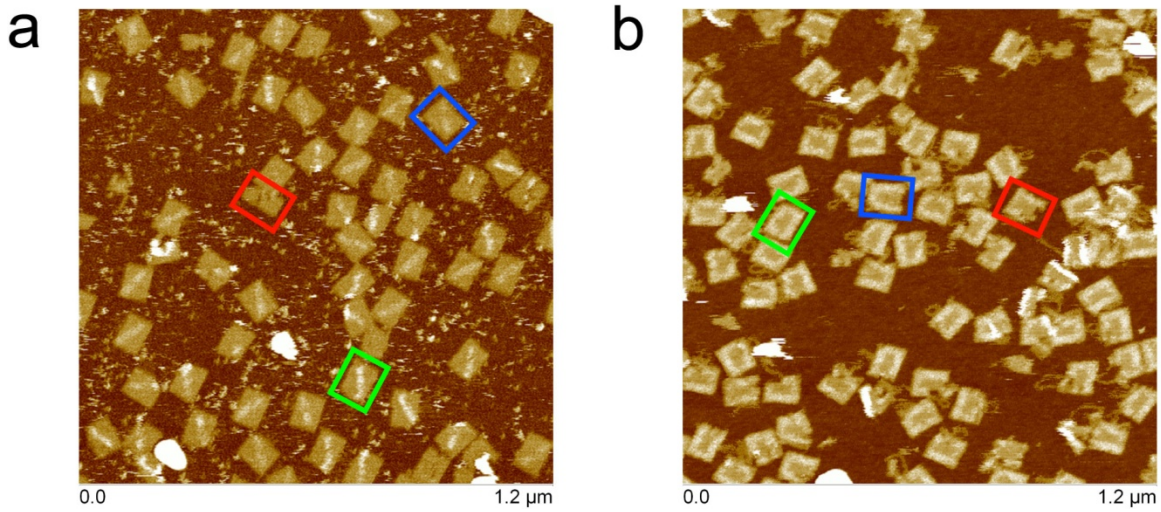


Figure 3.15. Determination of Assembly Yield by AFM. **a, b,** Representative AFM images of R **(a)** and L **(b)** origami, respectively, used for classification of assembly yield. In the images, the green, blue, and red rectangles depict origami tiles with different statuses. The green rectangles denote origami tiles that are clearly well formed with evidence of a fairly complete **S** pattern. The blue rectangles indicate origami tiles that are well formed with defective or missing **S** pattern. The red rectangles represent origami tiles that are broken or deformed. Each DNA origami structure in the AFM images shown here (and additional images not shown) was assigned to one of the three previously described categories. See Table 3.3 for the results of this analysis.

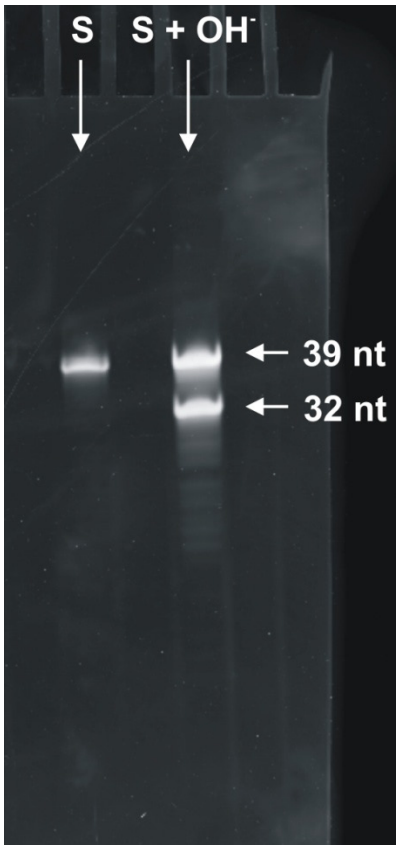


Figure 3.16. Denaturing Polyacrylamide Gel Characterization of Substrate. Comparison of substrate (**S**, lane 1) to an alkaline hydrolysis ladder of the substrate (**S** + OH^-, lane 2) in a 20% polyacrylamide gel containing 8 M urea. The bands are visualized using SYBR Gold. The upper band is full-length **S** (39 nt), while the lower band is the longer of two cleavage products (32 nt). No cleavage product band is detected in lane 1. The shorter product (7 nt) is not visibly stained.

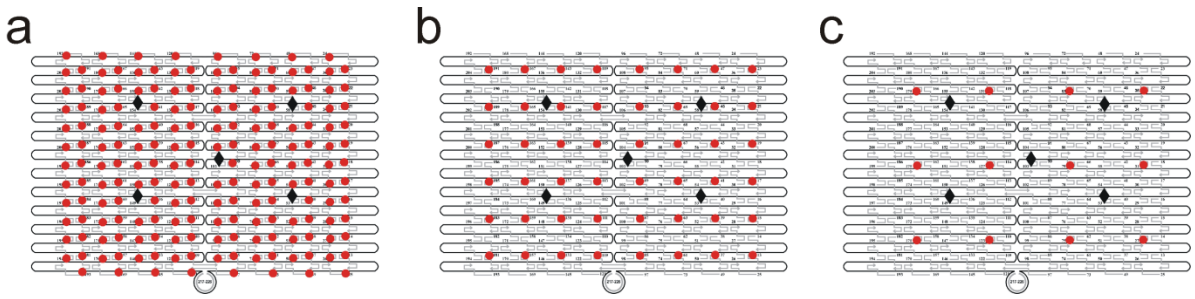


Figure 3.17. Origami for measuring dependence of PAINT probe binding kinetics on substrate density. Origami with different distances between nearest-neighbor **S** strands: (a) **D5** (5 nm), (b) **D10** (10 nm), and (c) **D20** (20 nm). Origami tiles bear 187, 48, or 12 copies of **S** (red dots) and five biotins (black diamonds) on opposite faces of the tile.

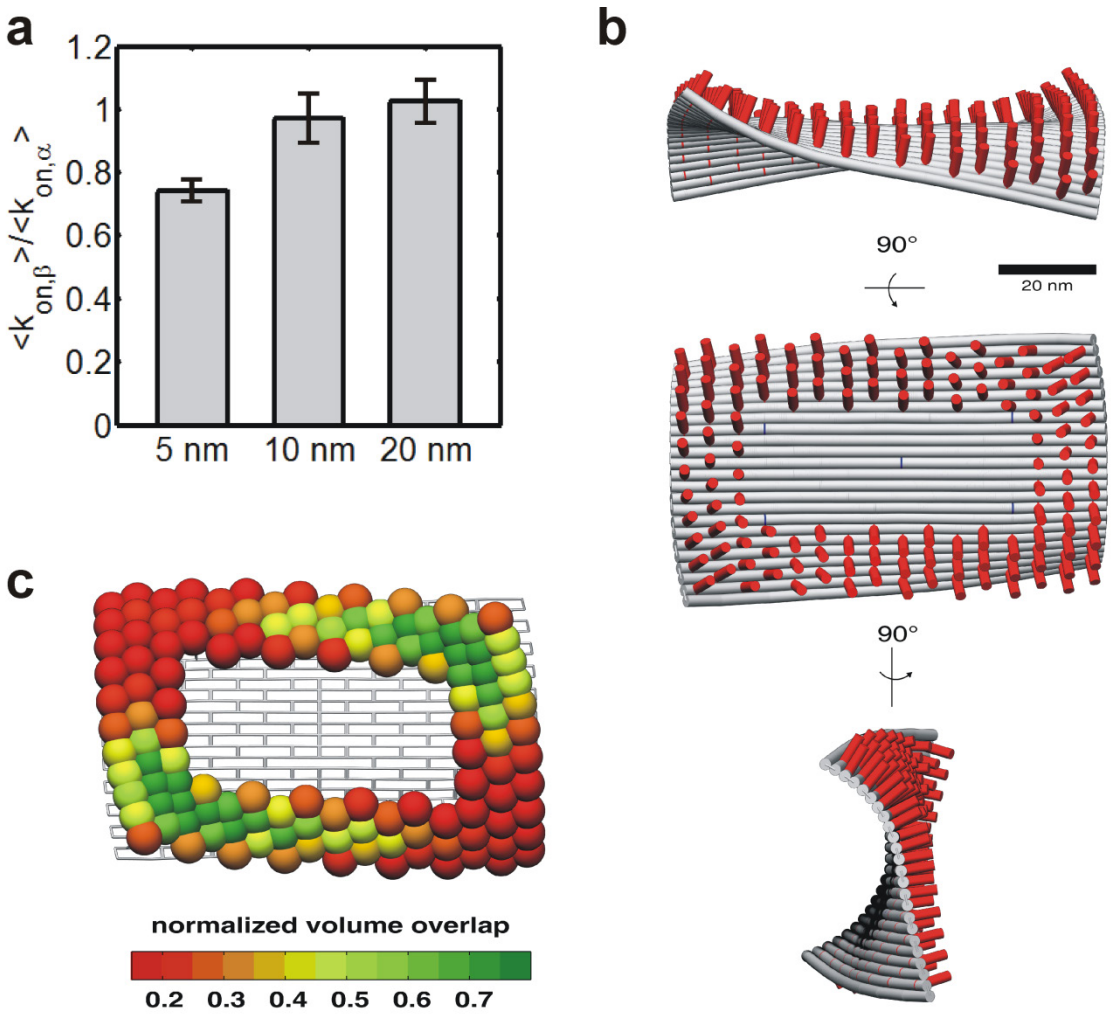


Figure 3.18 (a) Relative association rate constants of **α-Cy3** and **β-Cy5** to origami with approximate distances of 5, 10, or 20 nm between adjacent **S** strands. Error bars: 1 s.e.m. (b) Three-dimensional solution structure of **R** origami tile predicted by CanDo with constrained biotin positions (see Materials and Methods). Red cylinders represent **S** positions. (c) Normalized effective volume overlap of neighboring free **S** strands on **R** origami based on the CanDo structural model in panel (b).

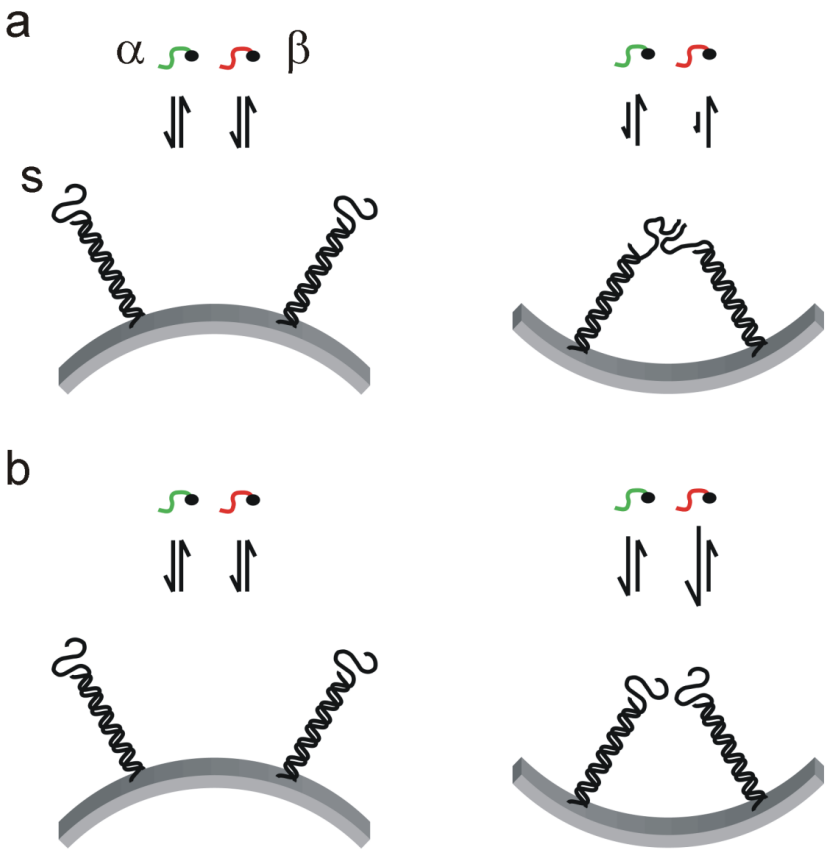


Figure 3.19: Two competing models for the effects of local substrate concentration on probe-substrate binding. (a) Inhibitory model and (b) cooperative model. Relative lengths of on-off rate arrows illustrate reduced versus enhanced on-rates. In the inhibitory model, we hypothesize that weak, non-Watson-Crick interactions between nearest-neighbor **S** strands reduce the effective on-rate of PAINT probe binding to **S**, with this effect enhanced when the tile is bent such that strands splay inwards (right) versus outwards (left). In contrast, in the cooperative model we hypothesize that proximity of nearest-neighbor substrates increases the effective on-rate by, for instance, increasing the probability of a productive encounter, with this effect reduced when the tile is bent such that target sites splay outwards (left) versus inwards (right). In both models, the proximity of substrates is expected to influence the binding of **β** to a greater extent than **α** because the ssDNA sequence to which **β** binds is located distal to alpha on **S**.

A variety of factors could produce variation in **S** spacing across the surface of an origami array, including global bend/twist or distortion of the origami tile[68,70,132] and incomplete tile or staple assembly. To investigate the possible impact of tile distortion on **S** spacing, we used the finite-element model CanDo to predict the three-dimensional solution conformation of the **R** origami tile, accounting for constraints imposed by surface immobilization *via* biotin. The model predicts a saddle-like conformation with significant curvature (Figures A1.12b, 3.20), consistent with previous reports[68,70,132]. Using a simple model of free **S** as a flexible, freely jointed chain with root-mean-square end-to-end distance of 3.6 nm connected to the origami surface by a rigid double-stranded DNA rod of ~7-nm length (Figure 3.21), the effective local concentration of **S** is predicted to vary 2- to 4-fold between different corners of the tile (Figures 3.18c, 3.21). The predicted variation in local concentration bears close resemblance to some of the most heterogeneous patterns we observe (Figure 3.22).

3.4 Discussion

In this work, we employed multicolor PAINT to acquire quantitative, 2D maps of chemical properties of individual DNA origami tiles, revealing their previously unobservable, stable, idiosyncratic fingerprints of interaction with reagents in solution. In addition, we have shown that the low invasiveness and insensitivity to photobleaching make PAINT suitable for spatiotemporal monitoring of subtle chemical modifications to individual nanostructures. Since it reveals previously hidden properties of non-rigid features of DNA origami that can be functionalized but yield little contrast for AFM and electron microscopy, PAINT complements these more established analytical tools. PAINT should thus find broad application in the characterization of the growing toolkit of soft, internally complex, nanoscale devices with applications in fields as diverse as organic synthesis, optoelectronics and molecular robotics[45,120,133].

The predictions from CanDo and local concentration modeling are consistent with binding patterns observed in PAINT reconstructions (Figure 3.22) and a competitive inhibition model in which nearest-neighbor strands interact via non-canonical or nonspecific binding interactions that are enhanced or diminished by structurally-induced changes in inter-strand proximity. Relaxation of the position restraints at the biotin

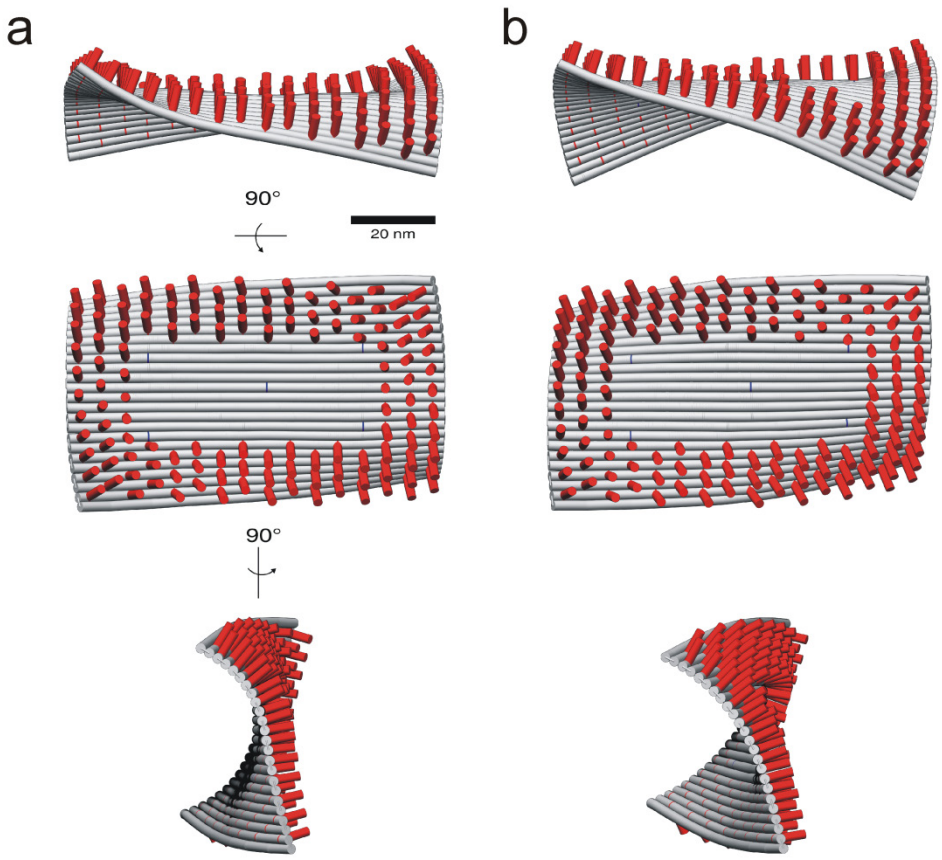


Figure 3.20: Predicted three-dimensional solution shape of R origami in three orthogonal views (a) when the pegboard is fixed at biotin binding sites (blue bands) and (b) when the pegboard is constrained only in the middle. Red cylinders represent 20-base-pair double-stranded DNA segments at substrate locations that are assumed to be normal to the surface.

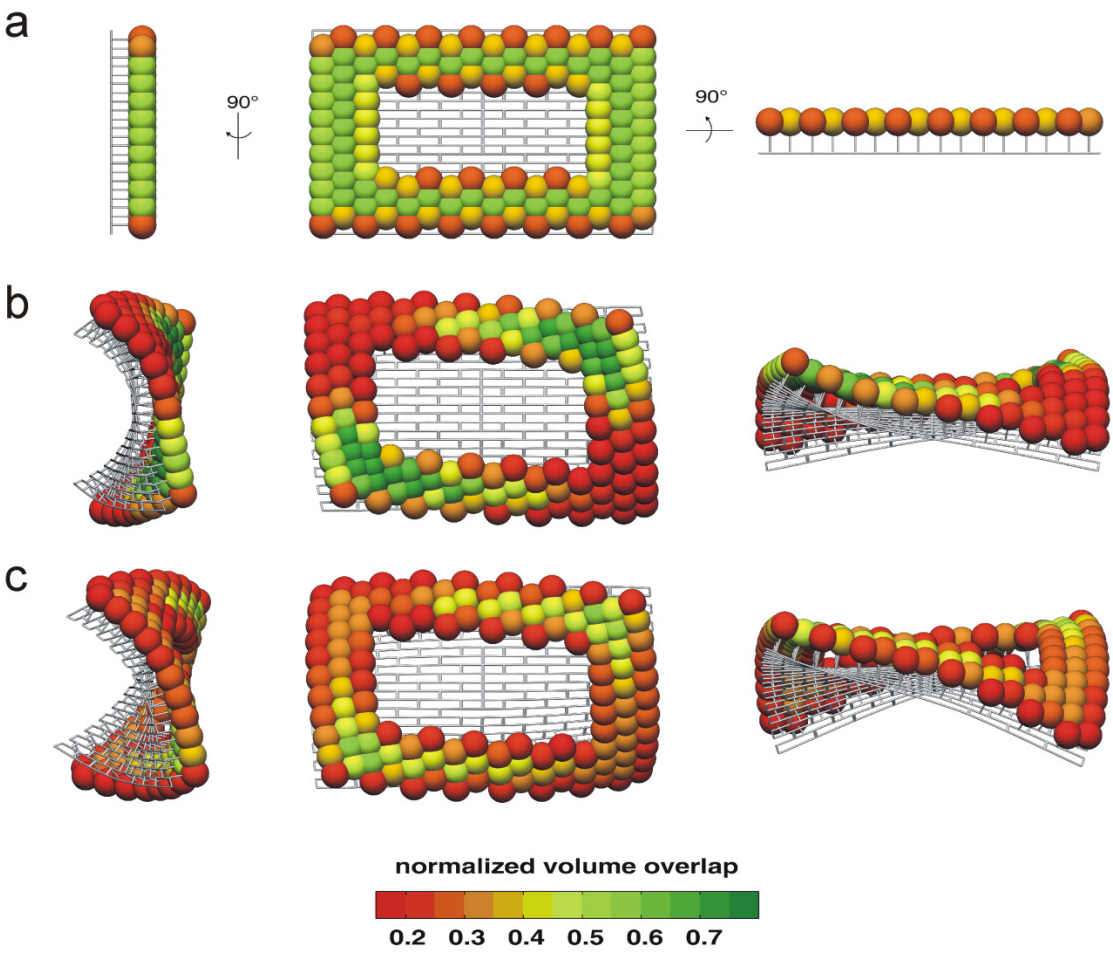


Figure 3.21. Effective relative substrate concentration on R origami. The effect of three-dimensional solution shape on effective local substrate concentration is characterized by calculating the effective volume overlap between adjacent spheres with a radius of 3.6 nm centered at the tip of 20-base-pair DNA duplexes. (a) Homogeneous pattern of the volume overlap when a flat conformation of the pegboard is assumed and (b-c) heterogeneous volume overlap patterns of curved pegboards whose three-dimensional solution shapes were computed using CanDo[132,134] (b) with and (c) without constraints at the biotin binding sites.

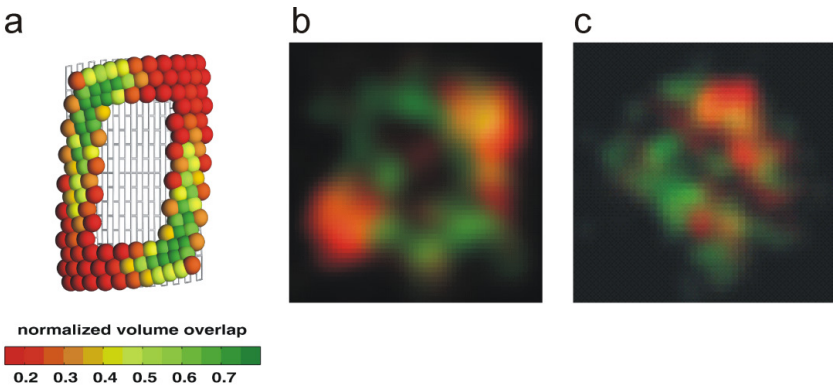


Figure 3.22. Side-by-side comparison of predicted effective substrate concentration and two-color PAINT images. (a) Predicted pattern of effective local concentration for **R** pattern of substrates with constrained biotin positions (from Figure S16, panel (b)). (b), (c) Two **R** origami with especially heterogeneous patterns of **β-Cy5** binding (red; **α-Cy5** binding distribution in green).

positions predicts less pronounced curvature at the corners (Figure 3.20b) and consequently less variation in **S** concentration across the tile (Figure 3.21c), suggesting that the number and orientation of surface-bound biotins can influence local variations in **S** spacing. Furthermore, since both PAINT (Figure 3.10) and AFM (Figure 3.15, Table 3.3) images show evidence of incomplete tile assembly, it is also possible that variations in assembly play a role in generating the fingerprints observed (Figure 3.14), as structural defects could also generate local variations in spacing between **S** strands.

Previous studies have presented mixed evidence for spatially-dependent oligonucleotide binding to origami, with a slight (10-40%) preference for binding towards the outer edges[69,88]. Our findings further show that individual origami tiles bearing dense arrays of targets can have stable fingerprints of sequence-specific interactions, with binding kinetics varying as much as twofold between different corners or edges of the tile (Figure 3.13b,c,e). **S** has little self-complementarity (Figure 3.2d), but the locally high concentration (~2-8 mM by our model) of **S** may lead to non-Watson-Crick interactions such as G-tetrads[135] between neighboring **S** strands that, even if transient, may compete with probe binding in a sequence-specific fashion. Such strand-strand interactions could have implications for other devices as well, including the spiders presented in Chapter 2. For instance, local variations in strand-strand interactions could compete (weakly) with the binding of spider leg to substrate, potentially exerting an influence on the walker's movement. In the next chapter, we examine more systematically the impact of substrate density on the kinetics of hybridization with oligonucleotide probes from solution.

CHAPTER 4:

MODULATION OF HYBRIDIZATION KINETICS ON ORIGAMI-TEMPLATED OLIGONUCLEOTIDE ARRAYS[7]

4.1 Introduction

In recent years there has been a growing interest in DNA nanodevices that exploit precise control over positioning of components by DNA hybridization[56,59,61,120,136]. Many of these devices are dynamic, with hybridization reactions or conformational changes playing a crucial role in their function[59,61,120]. In the near future, similar devices may be combined with DNA computing[62] for enhanced control over their timing and operation. However, for such approaches to be successful, a quantitative understanding of the factors influencing the kinetics and thermodynamics of DNA hybridization to targets on DNA origami is needed.

One previous study showed little deviation between the kinetics of hybridization in solution and at the surface of a DNA origami tile[88]. However, this work did not take into account the possible effects of sterical crowding, hopping between nearby target molecules, or nonspecific electrostatic interactions in a dense array of nearby target molecules. As illustrated in Chapter 3, immobilization of oligonucleotides on a DNA tile can have unanticipated effects on binding kinetics. Furthermore, owing to their complex composition and often small copy numbers of components, the functional behavior of a DNA nanodevice can vary significantly between copies of the device[137]. It is therefore important to understand the reproducibility of behavior between assemblies as well as their bulk behavior.

With these considerations in mind, we here use single-particle fluorescence resonance energy transfer (spFRET) microscopy to investigate the kinetics of hybridization reactions on individual surface-immobilized DNA origami arrays with

[7] Alexander Johnson-Buck designed, performed, and analyzed all kinetic measurements, as well as Monte Carlo simulations of kinetics. Jeanette Nangreave and Shuoxing Jiang synthesized all DNA origami.

different distances between adjacent targets, as well as in bulk solution. We use single DNA origami as sub-zeptomole nanoreactors to show that the kinetics of DNA hybridization to dense arrays of oligonucleotide targets deviates significantly from the kinetics in bulk solution. By systematically varying the spacing of targets and the properties of probes, we show that the rate of probe association is only slightly slowed in dense target arrays, but the rate of dissociation can be reduced by up to an order of magnitude. We present evidence for at least two distinct mechanisms for the slowing of dissociation: direct passing of probes between adjacent targets, and nonspecific interactions with the DNA origami tile.

4.2 Materials and Methods

Unless otherwise noted, all oligonucleotides were ordered from Integrated DNA Technologies (IDT).

Preparation of DNA origami scaffolds. Rectangular DNA origami arrays consist of an M13mp18 viral DNA scaffold (New England Biolabs) and 202 ssDNA staples as previously described[120]. For all structures assembled here, staples 1-12 and 205-216 were omitted to prevent inter-array base stacking interactions that result in undesirable aggregation (Figure 3.1). Of the remaining staples, several were modified at their 5′-end with an additional sequence, 5′-ACC TCT CAC CCA CCA TTC ATC, to which the target strand **T** (5′-GAT GAA TGG TGG GTG AGA GGT TTT TCA CTA TrAG GAA GAG) can bind (Figure 4.3). The arrays were annealed in 1× TA-Mg Buffer (40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM Mg^{2+}, pH 7.6) with a 1:3 ratio of M13 to staple strands and a final concentration of 10 nM (M13). The arrays were annealed over 12 hours from 94^{o}C-25^{o}C using a PCR thermocycler (Eppendorf).

Preparation of target and probe oligonucleotides. The target oligonucleotide **T-NH_2** (5'-GAT GAA TGG TGG GTG AGA GGT TTT TCA CTA TrAG GAA GAG /3AmMO/) was ordered with a 3′-terminal amine modification and HPLC purified by the manufacturer, then labeled with an *N*-hydroxysuccinimidyl ester derivative of Alexa Fluor 647 (Invitrogen) by overnight incubation in 0.1 M $NaHCO_3$, pH 8.3, followed by ethanol precipitation and thorough washing with 80% ethanol until the supernatant was

colorless, yielding **T-AF647**. Denaturing polyacrylamide gel electrophoresis revealed no detectable free dye. The probe oligonucleotides **DRz** (5′-/5Cy3/TCT CTT CTC CGA GCC GGT CGA AAT AGT GAA AA), **D11** (5′-/5Cy3/CTC TTC CTA TA), and **D11+F** (5′-/5Cy3/TCT CTT CCT ATA CGC TGA AAG GTG ACG GCA AA) were ordered HPLC-purified by the manufacturer and used as-is. Labeling efficiency was quantified by absorbance at 280 nm and either 550 nm (Cy3) or 650 nm (Alexa Fluor 647) using a Beckman DU 640B Spectrophotometer, and was >95% for all ssDNA strands. The strands **T** (5'-GAT GAA TGG TGG GTG AGA GGT TTT TCA CTA TrAG GAA GAG), **T*** (5’-GAT GAA TGG TGG GTG AGA GGT AAA TCA TCG AAG ACT CTA), and **Tcomp** (5'-CCT CTC ACC CAC CAT TCA TC) were ordered gel purified and used as supplied.

SINGLE-ORIGAMI KINETIC ASSAYS

Single-origami kinetic experiments were carried out on an inverted total internal reflection fluorescence (TIRF) microscope with a 1.2 NA 60x water-immersion objective (IX71, Olympus) in an environmentally controlled room at 20 ± 3 °C. Fluorescence excitation was provided by a 532-nm green laser (ultra-compact diode-pumped Nd:YAG laser GCL-025-S, CrystaLaser, 1 W/cm^2). The Cy3 and Alexa Fluor 647 emission signals were separated by a dichroic mirror with a cutoff wavelength of 610 nm (Chroma) and projected side-by-side onto an ICCD camera chip (iPentamax HQ Gen III, Roper Scientific, Inc.). The Cy3 channel image was passed through a band pass filter (HQ580/60m, Chroma) and the Alexa Fluor 647 channel was passed through a long pass filter (HQ655LP, Chroma). A Newport ST-UT2 vibration isolation table was used in all experiments. In all measurements, an oxygen scavenger containing oxygen scavenger system[100] (OSS ≡ 2.5 mM 3,4-dihydroxybenzoic acid, Sigma P5630; 1 mM Trolox, Acros 218940050; and 25 nM protocatechuate dioxygenase, Sigma-Aldrich P8279) was included in the imaging buffer to reduce photobleaching.

Microscope slides with a flow channel were prepared using double-sided tape (Scotch) and treated with biotinylated BSA and streptavidin as described[138] to prepare the surface for immobilization of biotinylated DNA origami. A solution containing 20-100 pM origami was incubated in the presence of 1x HBS (150 mM NaCl, 25 mM HEPES-KOH, pH 7.4) at room temperature for 10 minutes, and excess sample was flushed

away by two washes with 1x HBS. A solution of 200 nM **T-AF647** was added to the slide channel and incubated for 15 minutes before flushing the excess away by two washes with 1x HBS. Fluorescence from the AF647 label of the target **T** was visible even under 532-nm excitation, enabling us to focus on and locate origami prior to beginning FRET measurements.

Association Kinetics. To limit photobleaching, a shuttered illumination scheme was used: the sample was illuminated for 0.5-s intervals separated by 29.5-s dark periods. After an initial waiting period, a solution of 25, 50, 75, or 100 nM **DRz**, **D11**, or **D11+F** in 1x HBS was added to the slide during the beginning of a dark period with a dead time of 5 s. FRET from Cy3 to AF647 resulted in an approximately 5-fold increase in AF647 fluorescence upon binding of the probe to the target.

Dissociation Kinetics. For dissociation kinetics experiments, the length of dark periods was increased to 119.5 s. The same exposure time of 0.5 s was used. During the dark period after the first measurement, a solution of 500 nM unlabeled **T** was added as a chase.

The fluorescence intensity of each origami was normalized to its maximal value in a given experiment. The mean intensity across all origami was plotted as a function of time and fit to the single exponential decay models $y = C(1 - e^{k'_{obs}t})$ and $y = C_1 e^{k'_{obs,1}t}$ for association and dissociation measurements, respectively.

SOLUTION KINETIC ASSAYS

All measurements were performed at 22°C on an Aminco-Bowman Series 2 Luminescence Spectrometer at a time resolution of 1 or 6 s, exciting at 520 nm (4 nm bandwidth) and detecting at 690 nm (16 nm bandwidth). As in the single-origami kinetic assays, all measurements were taken in the presence of oxygen scavenger and 1x HBS. Under these conditions, no photobleaching was observed over the course of 1 h.

Association kinetics. To a 99.5-µL solution of 25, 50, 75, or 100 nM **DRz**, **D11**, or **D11+F** (final concentration) was added 0.5 µL of a pre-equilibrated solution of 1 µM **T-AF647** (final concentration 5 nM) and 4 µM **TComp**, and the solution was mixed well by pipetting. T**Comp** was used to block the portion of **S** that normally hybridizes to

overhangs on the DNA origami. The increase in A647 fluorescence due to FRET was monitored until the signal was stable, and subsequently fit with a single exponential function.

Dissociation kinetics. To a 97.5-μL solution of 25 nM **DRz**, **D11**, or **D11+F** (final concentration) was added 2.5 μL of a pre-equilibrated solution of 1 μM **T-AF647** (final concentration 25 nM) and 4 μM **TComp**. After equilibrium was reached, a 25-μL chase solution of 2.5 μM unlabeled **T** (final concentration 500 nM) was added to the reaction and mixed well. The decay was fit with a single exponential function.

Monte Carlo simulations of probe binding to and dissociation from targets on DNA origami. Origami were modeled as a collection of 187, 48, 12, or 4 targets, each capable of binding one probe. Probe association was modeled as a single-exponential increase with a pseudo-first-order rate constant k'_{obs} taken from experiment (Figure 4.4). Probe dissociation was modeled as a single-exponential decrease, again with a rate constant taken from experiment for a given origami construct. In association experiments, bound probes were allowed to dissociate, whereas in dissociation experiments, dissociation was considered irreversible. Each run was divided into 1,000 timesteps spanning five half-lives of the reaction. During each timestep, each target has an opportunity to bind and/or release a probe according to the probability of reaction $P(r) = (1-\exp(-k*\Delta t))$, where k is the rate constant and Δt is the timestep. For each condition, 1,000 runs, each representing a single origami trajectory, were performed.

4.3 Results

Upon binding of a probe labeled with a FRET donor (Cy3) to a target oligonucleotide (**T**) labeled with a FRET acceptor (AF647), the donor is brought into close proximity with the acceptor, resulting in energy transfer from the excited donor to the acceptor (Figure 4.1a). The increase in acceptor signal was detected using total internal reflection fluorescence (TIRF) microscopy (Figure 4.1b-c), enabling kinetic characterization of individual DNA nanoarrays. Furthermore, by adding a chase consisting of unlabeled **T** in solution, we monitored the kinetics of probe dissociation from an individual target array

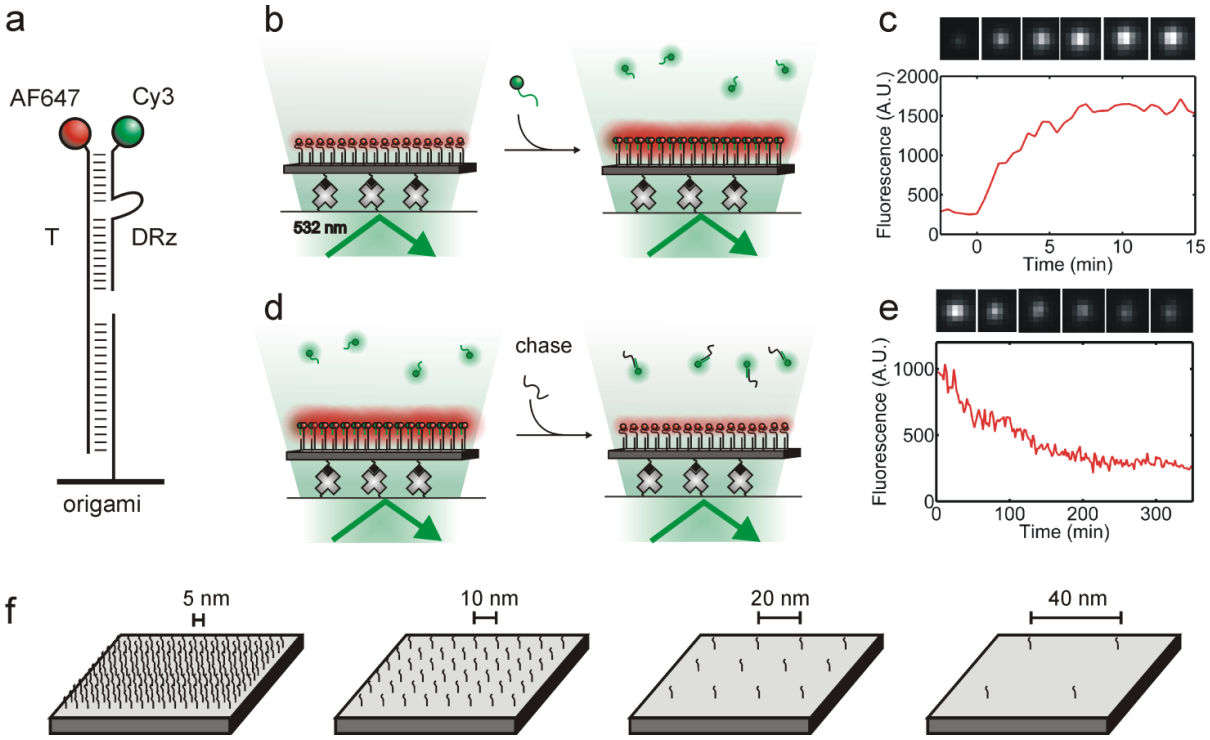


Figure 4.1. (a) A Cy3-labeled DNA probe (**DRz**; alternatively **D11** or **D11+F**) binds to multiple copies of the AF647-labeled target (**T-AF647**) oligonucleotide within an origami-templated array. (b) Upon addition of 25-100 nM probe strand, the binding of the probe to the target is visualized by FRET on a fluorescence microscope. (c) The binding of multiple (4-187) copies of the probe leads to a gradual increase in the acceptor (AF647) signal on a single DNA origami array. (d, e) Upon removal of excess probe and addition of unlabeled **T** as a chase, bound probe dissociates from the origami array, resulting in a loss of FRET signal from a single DNA origami array. (f) Schematics of the four DNA origami arrays used in this study, each bearing multiple copies of **T** spaced by 5, 10, 20, or 40 nm.

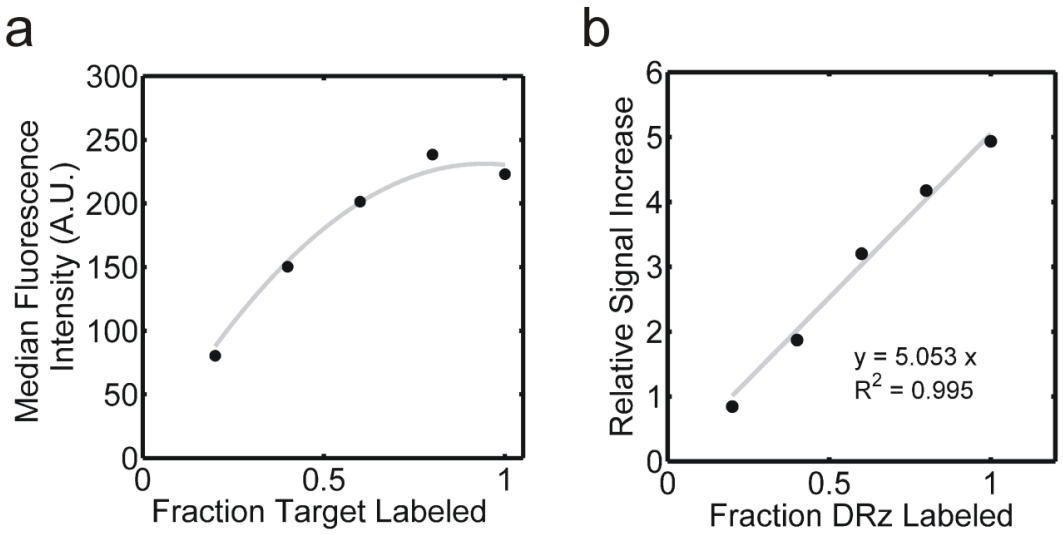


Figure 4.2. (a) Median fluorescence intensity per origami tile as a function of the fraction of **T** labeled with AF647. The gray curve is a best-fit quadratic polynomial. (b) Fold increase in AF647 fluorescence from FRET upon **DRz** binding as a function of the fraction of **DRz** labeled with Cy3. The gray line is a linear regression fit with the y-intercept constrained to the origin.

through the loss of FRET (Figure 4.1d-e). The chase was added at a concentration of 500 nM, which is at least a 5-fold excess over the amount of probe present. Increasing the chase concentration tenfold to 5 μM does not alter the apparent rate constant of probe dissociation (rate constants within 1 s.e.m. for all probes), suggesting that the chase does not actively displace the probe from **T**. Importantly, although AF647 self-quenches at high target densities, reducing the fluorescent signal per target, the increase in FRET signal depends linearly on the amount of fluorescently labeled probe within the range of distances we consider here (Figure 4.2). Kinetics were measured for nucleotide arrays with four different target spacings ranging from ~5 to ~40 nm (Figure 4.1f, Figure 4.3). We used a probe with the 8-17 deoxyribozyme sequence (**DRz**, Figure 4.1a, Materials and Methods) because of its frequent use in DNA nanotechnology[61,120] as well as its two independent binding arms, which may give rise to non-standard kinetic behavior in dense target arrays. Measuring the kinetics of **DRz** binding to **T** under pseudo-first-order conditions, we find only a slight, approximately twofold slowing of association at distances < 20 nm (Figures 4.4a, 4.5). Furthermore, the relative increase in AF647 fluorescence (~5-fold) upon probe binding is consistent for all target spacings (Figure 4.6). This suggests that, even at target spacings of 5 nm, steric hindrance and/or nonspecific target-target interactions only slightly hinder probe binding. The binding approaches deterministic behavior at high probe densities due to the large number of probes binding to each array, with little variation between individual origami tiles (Figure 4.4b); if each origami trajectory is fit individually, the observed pseudo-first-order rate constant for association of **DRz** to the target with 5 nm spacing is 0.46 ± 0.11 (s.d.). As a control, a second probe **D11** having similar binding kinetics as **DRz** but forming only a single 11-base pair helical domain with **T** was also characterized, and exhibited similarly modest differences from solution kinetics and reproducible behavior between origami (Figure 4.4c,d).

In contrast, the rate constant of **DRz** dissociation from origami arrays deviates dramatically from the solution case, spanning more than an order of magnitude (Figure 4.7a-b). Even at the 40-nm spacing, there is a 3- to 4-fold reduced rate constant of dissociation compared to in solution. Furthermore, the apparent rate constant at a target

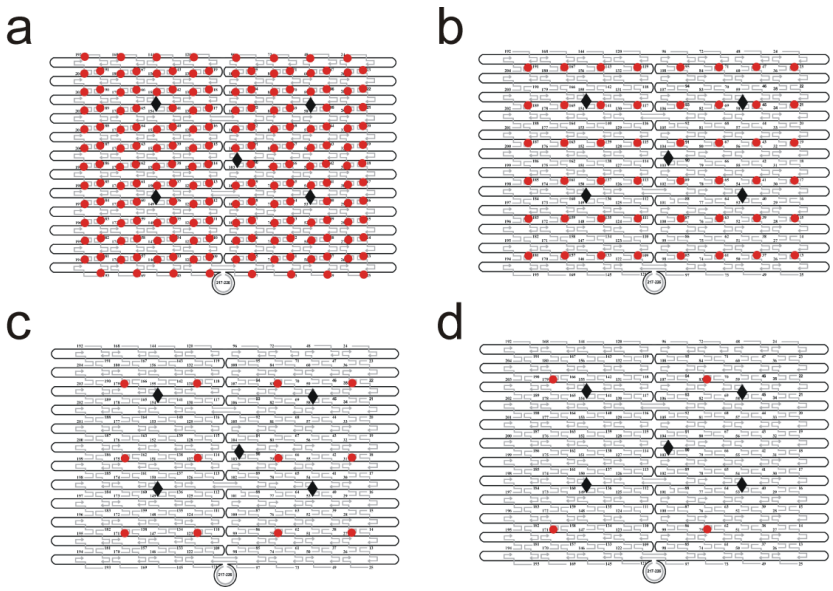


Figure 4.3. (a) (b), (c), (d), Schematics of the origami-templated target arrays used in this study, with spacings of approximately 5 nm (a), 10 nm (b), 20 nm (c), and 40 nm (d) between adjacent targets. Red circles correspond to target positions, and black diamonds indicate positions of biotinylated staples for immobilization prior to fluorescence microscopy. Biotin moieties and target molecules project from opposite faces of the rectangular tile. The continuous black line represents the circular M13 viral genome and the gray lines correspond to unmodified staples.

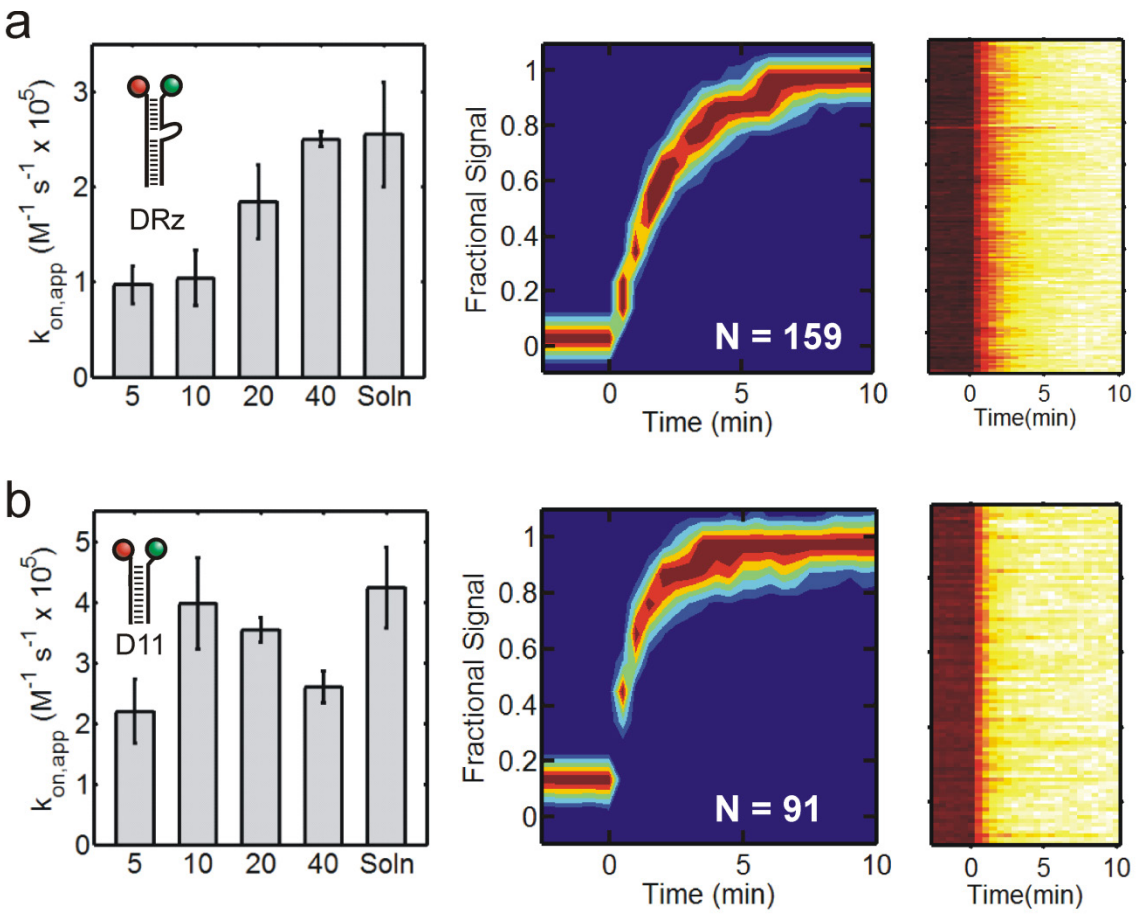


Figure 4.4. (a) Kinetics of probe **DRz** binding to target on origami with different spacings (5-40 nm) and in solution ("Soln"). (b) Probability density map (left) and single-origami trajectories (right) of AF647 signal increase upon binding of 75 nM **DRz** to target on origami spaced by 5 nm. (c) Kinetics of probe **D11** binding to target on origami and in solution. (d) Probability density map (left) and single-origami trajectories (right) of AF647 signal increase upon binding of 75 nM **D11** to target on origami spaced by 5 nm. Error bars are 1 s.e.m. In (b) and (d), the probe was added at time t = 0 min.

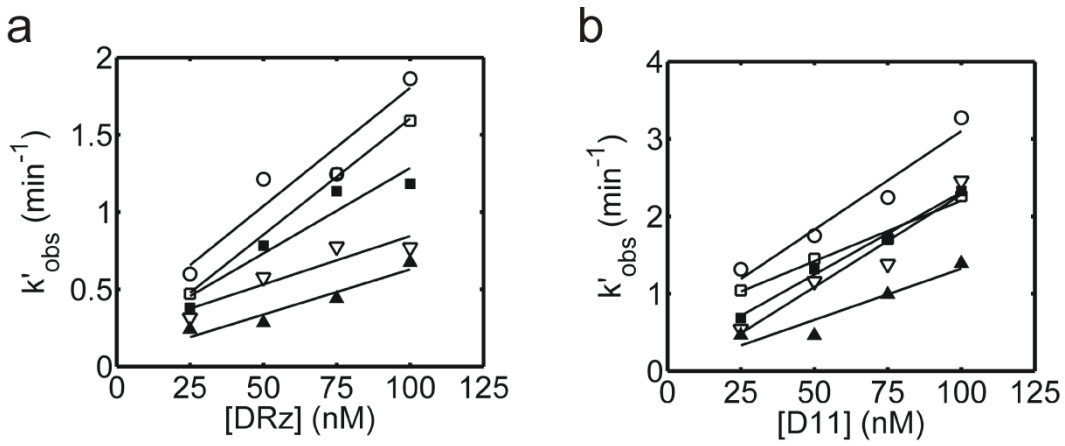


Figure 4.5. Pseudo-first-order kinetics for the binding of **DRz** (a) and **D11** (b) to origami with targets spaced by 5 nm (filled triangles), 10 nm (open triangles), 20 nm (filled squares), 40 nm (open squares), or to target molecules in solution (open circles). Best-fit linear regression lines are shown.

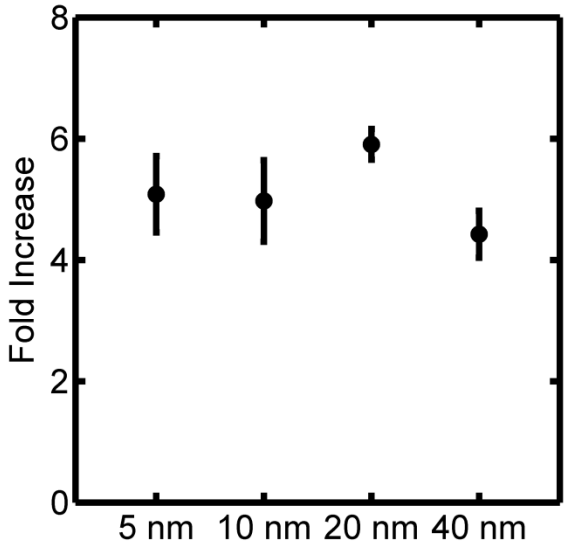


Figure 4.6. Fold increase in AF647 signal upon target binding for origami with different distances between neighboring target molecules. Error bars represent 1 s.e.m. from at least 3 trials.

spacing of 5 nm is about four times smaller than at 40 nm, showing a clear distance dependence. In contrast, the rate constant of **D11** dissociation from origami was only about 35% less than in solution, and showed no distance dependence (Figure 4.7c-d). To investigate whether the differences between **DRz** and **D11** result from their different lengths (32 and 11 nt, respectively), a third probe sequence **D11+F** was designed with (1) the same binding sequence as **D11** and (2) the same overall length and base composition as **DRz**. Like **DRz** on the 40-nm-spaced origami array, **D11+F** exhibits a rate constant of dissociation from origami about four-fold slower than from isolated targets in solution; however, like **D11**, this rate constant does not decrease further with decreased distances between targets (Figure 4.7b). Hence, there is evidence of at least two mechanisms for the slowing of oligonucleotide dissociation from origami-templated target arrays: a distance-dependent walking or hopping mechanism observed for **DRz** but not **D11** or **D11+F**; and a nonspecific mechanism (e.g. electrostatic) by which longer DNA strands (**DRz** and **D11+F**) are retained more strongly than shorter ones (**D11**). Each mechanism slows dissociation by a factor of up to 3-4.

To further investigate the nature of the distance-dependent effect seen for **DRz**, we measured the rate constant of **DRz** dissociation from origami with 5-nm target site spacing, but prepared with a 1:15 mixture of target oligonucleotide and an inert control oligonucleotide **T*** that competes for sites on the origami but possesses a scrambled sequence that is not complementary to **DRz**. The **DRz** strand was observed to dissociate about three times faster (0.03 min^{-1} compared to 0.009 min^{-1}) from the **T**+**T*** array than the original **T**-only array (Figure 4.8a,b), showing that the distance-dependent slowing of dissociation depends on the proximity of complementary target molecules rather than non-specific interactions with surface-bound DNA.

4.4 Discussion

The results presented here show clear deviations from solution behavior in hybridization reactions at the surface of DNA origami. The moderate slowing of association at high oligonucleotide densities (Fig. 4a) is not unprecedented; surface plasmon resonance (SPR) measurements of hybridization showed a similar effect, albeit at somewhat

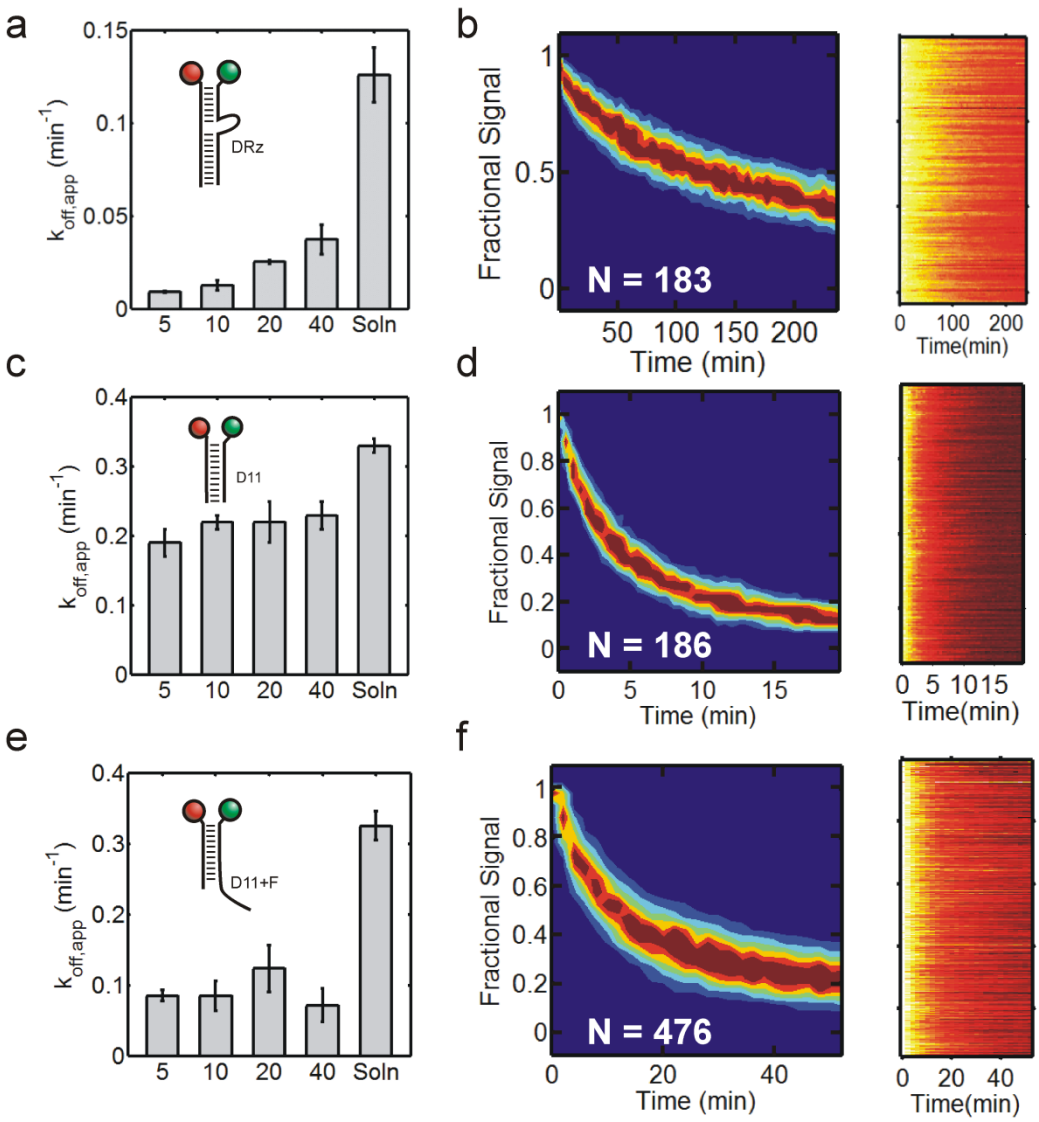


Figure 4.7. (a) Kinetics of probe **DRz** dissociation from target on origami with different spacings (5-40 nm) and in solution ("Soln"). (b) Probability density map (left) and single-origami trajectories (right) of AF647 signal decrease upon dissociation of **DRz** from target on origami spaced by 5 nm. (c) Kinetics of probe **D11** dissociation from target on origami and in solution. (d) Probability density map (left) and single-origami trajectories (right) of AF647 signal decrease upon dissociation of **D11** from target on origami spaced by 5 nm. (e) Kinetics of probe **D11+F** dissociation from target on origami and in solution. (f) Probability density map (left) and single-origami trajectories (right) of AF647 signal decrease upon dissociation of **D11+F** from target on origami spaced by 5 nm. Error bars are 1 s.e.m.

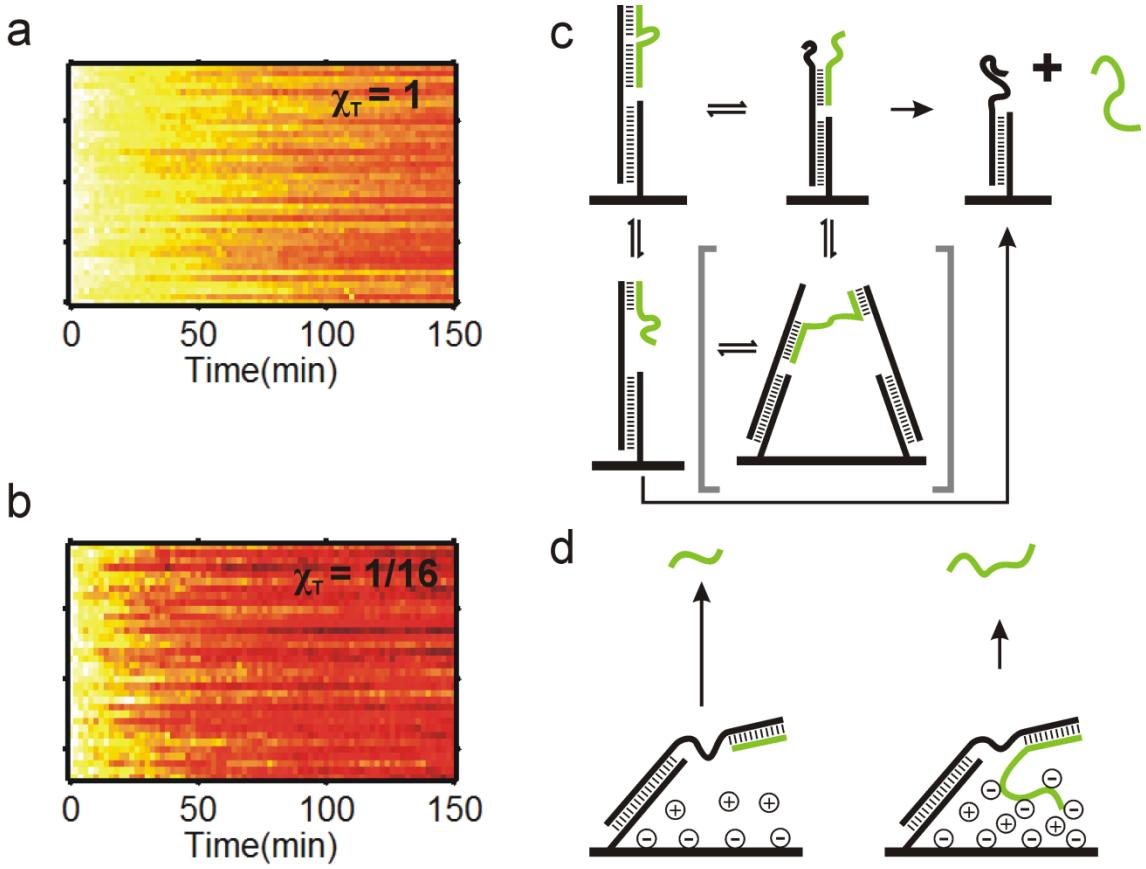


Figure 4.8. (a) Single-origami trajectories of AF647 signal decrease upon dissociation of **DRz** from target spaced by 5 nm. (N=40, k_{obs} = 0.009 min^{-1}) (b) Single-origami trajectories of AF647 signal decrease upon dissociation of **DRz** from target arrays identical to in (a), but prepared with a 1:15 mixture of target **T** and a binding-inert control strand **T***. (N=34, k_{obs} = 0.03 min^{-1}). (c) Proposed kinetic scheme of **DRz** (green) dissociation from origami-bound target. When other substrate molecules are within reach, a **DRz** probe (green) may walk between them (in brackets), slowing its overall rate of dissociation from the origami tile. (d) Schematic representation of nonspecific slowing of dissociation from origami-bound targets based on probe length. Short probe **D11** (left), lacking extensive interactions with the origami surface, dissociates nearly as rapidly as from the target in solution. Longer probes such as **DRz** and **D11+F** (right) may interact with the origami surface *via* loosely bound counterions, slowing its dissociation.

higher target densities (2-12 x 10^{12} cm^{-2} compared to 0.07-3 x 10^{12} cm^{-2} in the present study), as did fluorescence-based assays of hybridization to targets immobilized on microparticles[139]. Another study showed an ionic strength-dependent modulation of thermodynamic selectivity for perfectly matched over mismatched probes as the spacing between targets was decreased from ~40 nm to ~5 nm.[140] However, to our knowledge the dramatic slowing of oligonucleotide dissociation from surface-bound target arrays we report here (Figure 4.7a) is novel. The results in Figures 4.7a,b and 4.8 suggest a mechanism of walking or hopping for the **DRz** molecule, in which dissociation from the origami tile is slowed by the ability of **DRz** molecules to bridge or hop between adjacent target molecules (Figure 4.8c). In this model, the degree to which dissociation is retarded should depend on the local concentration of targets on the surface of the origami, and this is indeed observed (Figure 4.7). In contrast, the data do not support a model in which the distance- dependent slowing of dissociation from the origami surface depends on recapture of loosely associated probes, because **D11** and **D11+F** do not undergo any distance-dependent slowing of dissociation (Figure 4.7). As **D11** and **D11+F** can each form only a single helical stem with the target, this duplex must melt before either of these probes can form new base pairing interactions with another target, and hence they are not expected to be capable of the walking behavior illustrated in Figure 4.8c.

Nevertheless, even when target strands are separated by 40 nm, at which distance no direct passing of probes between targets is likely, both **DRz** and **D11+F** exhibit 3- to 4-fold reduced dissociation rate constants relative to the solution case (Figure 4.7). In contrast, the shorter probe **D11** dissociates almost as rapidly from the origami surface as from targets in solution. This effect is apparently independent of target spacing, and we therefore propose that it results from nonspecific interactions between the bound probe and the origami tile or its ionic environment (Figure 4.8d). Monovalent cations are known to bind diffusely to nucleic acids, screening the negative charge of the phosphate backbone and stabilizing more compact conformations[141–143]. Furthermore, origami structures possess a high density of negatively charged phosphates, and their folding is strongly influenced by metal cations[54]. They can also bind stably to the negatively

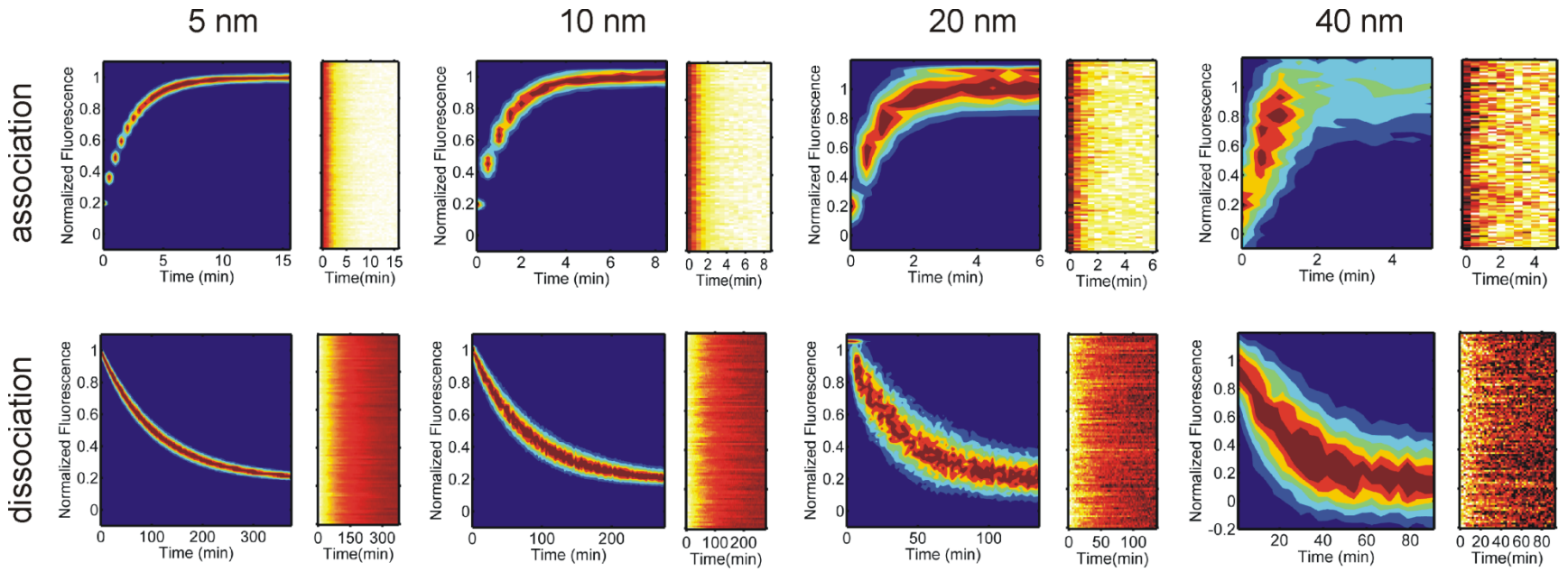


Figure 4.9. Monte Carlo kinetic simulations showing the expected variation in individual origami behavior based on the number of targets per origami. Probability density maps of intensity versus time (left panels) and single origami trajectories (right panels) from Monte Carlo simulations of **DRz** binding to, and dissociation from, origami with 5, 10, 20, or 40 nm between neighboring target molecules (i.e., with 187, 48, 12, or 4 targets per origami) are shown. Rate constants were set to the values from runs shown in Figure 4.10. Association simulations were carried out according to the predicted pseudo-first-order rate constant at 75 nM **DRz**. N = 1000 for all simulations.

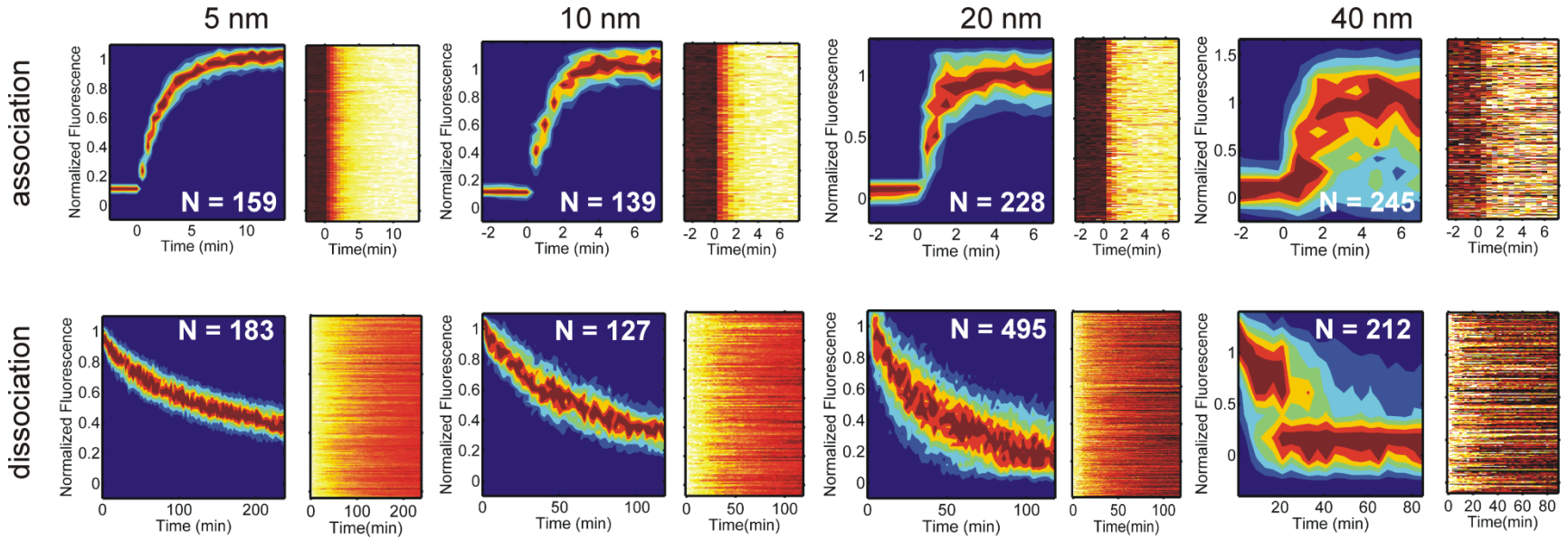


Figure 4.10. Probability density maps of intensity versus time (left panels) and single origami trajectories (right panels) of **DRz** binding to, and dissociation from, origami with 5, 10, 20, or 40 nm between neighboring target molecules. The association reactions shown here were carried out in the presence of 75 nM **DRz**. The number of origami observed in each reaction (N) is shown in the respective panel.

charged surfaces of materials such as mica and SiO_2 when an appropriate counterion is provided in solution, as is common protocol in their characterization by AFM[144,145]. It is possible that similar sandwich-like interactions between longer probes and the origami lead to delayed dissociation from origami-bound targets. Such an interaction would be expected to depend on the amount of charged surface a probe possesses, which for an oligonucleotide probe depends on its length. Hence, this model is consistent with the observation of a more pronounced slowing for 32-mers **DRz** and **D11+F** than for the 11-mer **D11**.

There is little variation between the kinetic properties of individual origami, particularly when the total number of targets per origami is high. This suggests that the assembly yield is high and reproducible enough to guarantee a performance that is predictable and almost deterministic, which may be a useful property for interfacing with molecular computing systems[62,63,65,117], especially if objects composed of DNA origami are to be used as molecular automata that must perform consistently as individual devices. Monte Carlo simulations (Figure 4.9) suggest that most or all of the variation between kinetic trajectories of individual origami (Figure 4.10) can be attributed to statistical noise (i.e., variation in the exponentially distributed wait times for a single reaction to occur) and noise in the measurement. As expected, as the number of targets per array decreases, the reaction trajectories vary to a greater extent between arrays (Figure 4.10).

In summary, we have shown that the patterning of binding targets on individual DNA origami can have a significant effect on the apparent kinetics and, by extension, thermodynamics of probe binding to the targets. The most dramatic effect is the slowing of dissociation from the array, particularly for longer probes and those capable of multivalent binding. In this sense, DNA origami arrays act as a “sponge,” in some cases retaining probe molecules >10 times longer than do the corresponding targets in solution. As quantitatively predictable performance becomes more important for DNA nanodevices, such as those coordinated or synchronized by DNA computing circuits, it will be necessary to understand to what extent reactions on an origami array deviate from those in solution.

CHAPTER 5:

SUMMARY AND OUTLOOK

1.1 Summary of results

The field of DNA nanotechnology continues to proliferate rapidly, as evidenced by the ongoing publication of novel design paradigms and devices using either DNA origami scaffolding[56,60,67,119,137,146,147] or discrete oligonucleotide building blocks[51,52]. These devices now routinely incorporate not only DNA, but other functional materials such as motor proteins[146], enzymes[56], inorganic nanoparticles[60,147], and lipids[137] to broaden the repertoire of spatially controlled processes achievable with DNA-based materials. At the same time, the functional and structural capabilities of DNA itself are being stretched beyond previous expectations.

As the field matures, it will be increasingly important to functionally couple isolated devices to other systems such as living cells, synthetic chemical computing circuits, and optically or electrically active materials. Efforts to design such interfaces are in fact already underway. For instance, a compound nanotube constructed from DNA origami was recently shown to emulate properties of natural ion channels when embedded in a synthetic lipid bilayer, and could even gate the current across the bilayer when its central channel was obstructed by ssDNA strands[137]. In another application, DNA origami was used to construct a logic-gated drug delivery platform that could selectively deliver antibody fragments to cell types displaying certain combinations of surface antigens, triggering cellular signaling pathways *via* a clever union of structural DNA nanotechnology and DNA computing[119]. The devices are not only becoming more structurally and compositionally complex, but also more useful, and may see widespread practical application in medicine and materials engineering very soon[45].

Yet, as we have seen, complexity often begets heterogeneity. In the case of the DNA-based ion channel described above, single-molecule electrochemical measurements allowed the authors to detect persistent differences between individual copies of the channel[137]. Likewise, single-molecule TEM measurements of the aforementioned drug-delivery platform permitted improvements of the design that led to a higher yield of initially closed structures and hence more selective delivery[119]. It is clear that a single-molecule understanding of DNA nanotechnology is needed to provide complete feedback during the design process and maximize the efficacy of the final devices.

Through the work described in this dissertation, we have expanded the single-molecule toolkit available to researchers in DNA nanotechnology, as well as revealed properties of particular DNA nanodevices that are interesting in their own right. In Chapter 2, we described single-molecule characterization of a synthetic DNA-based walker, the molecular spider, as it interfaced with detailed instructions of movement programmed using DNA origami. While AFM provided high-throughput population-level characterization of spiders, single-particle tracking using TIRF microscopy was necessary to achieve the high temporal resolution necessary to make detailed comparisons with the Monte Carlo model of spider walking. Both single-particle tracking and Monte Carlo modeling showed significant variation between individual spiders as well as non-designed behaviors such as significant net movement toward the GOAL on cleavage product. In addition, the single-particle tracking allowed us to characterize non-designed behaviors not predicted from the simple Monte Carlo model, such as immobility and movement prior to the addition of the zinc ion cofactor. In addition to these particular observations of molecular spiders, the synergy between the high-throughput structural information of AFM, the time-resolved longitudinal observation of single-particle tracking, and the mechanistic predictions of numerical modeling is a good model for the characterization of dynamic DNA nanodevices in the future.

In chapter 3, we presented a dramatic expansion of the capabilities of super-resolution fluorescence microscopy as well as their application to single-molecule imaging in DNA nanotechnology. We provided what I believe to be the first spatiotemporal nanoscale characterization of soft, dense chemical features on individual

DNA nanostructures. This is significant, because many reported DNA nanodevices make use of dense arrays of features that are too closely spaced or too delicate to reliably detect and distinguish in a non-perturbative fashion using AFM or TEM[56,59,67,69], including the spider tracks in Chapter 2. In addition, we demonstrated the first use of PAINT to follow chemical reactions over time on DNA origami, and indeed the first time-lapse measurements of modifications to soft features on single DNA nanostructures. We were intrigued to discover the presence of "fingerprints" of interactions between origami and solution that vary greatly between individual tiles. The evidence we presented suggests that these fingerprints are the product of weak interactions between densely spaced features on the origami surface, a property which we expect to be present in many such devices. The comparison of these experimental results to CanDo models of origami conformation was crucial, as it enabled us to identify global curvature and variations in local concentration of surface features as the probable source of the striking global patterns of PAINT probe binding observed for some tiles. The results here have implications for spider walking: specifically, they suggest that spacing footholds too closely may cause them to interact with each other, competing with spider legs and possibly even influencing the motion of walkers. Furthermore, I anticipate that the ability to spatiotemporally monitor dynamic interactions with dense surface features will be valuable in future interfaces between DNA nanostructures and molecular computing circuits[62].

Finally, in Chapter 4, we presented a thorough characterization of hybridization kinetics in dense fields of targets immobilized on single DNA origami. We showed that DNA hybridization in the milieu of a DNA tile deviates systematically from the corresponding solution reactions in at least two primary ways. First, DNA strands with two independent binding domains may be directly passed between nearby binding targets on the tile, resulting in an approximately three- to fourfold slowing of dissociation from the tile. Second, DNA probes appear to be retained through nonspecific interactions with the origami tile surface in a manner dependent on length, resulting in an additional three- to fourfold slowing of dissociation for the largest probes we examined (32 base pairs). The combined effects of these two behaviors can reduce the apparent rate constant of dissociation from an ssDNA target by an order of magnitude.

The rate constant of association is also decreased, albeit less dramatically, at the origami surface. These results show that immobilization of binding targets on a DNA origami can influence hybridization kinetics considerably, and may have an impact on the behavior of walkers like the molecular spiders presented in Chapter 2. For example, the monopedal walking of individual spider legs between adjacent substrates could give rise to the 'leaky' behavior of walking in the absence of zinc, since it would reduce the energetic barrier that must be overcome to dissociate from uncleaved substrate. Future designs of tracks for DNA walkers should thus take such possibilities into consideration by, for instance, increasing the spacing between footholds. The use of single-origami assays also permits us to detect the very reproducible, nearly deterministic kinetic behavior in a sub-zeptomole reactor of defined nanoscale dimensions. This behavior could be exploited in the development of, for instance, quantitative localized biosensors ('DNA nanoarrays') that depend on precise positioning of two or more components.

Taken together, the above work represents a large-scale advance in the characterization of nanoscale DNA structures and devices. We have affirmed the importance of the stochasticity and heterogeneity endemic to all molecular machines, and provided additional approaches to quantify them. These tools and lessons should be integrated and expanded upon as we look forward to the promising future of bionanotechnology.

1.2 Outlook

The molecular walker system presented in Chapter 2 showcases the promise of rational control of complex behaviors in bionanomachines. However, it and other walkers like it fall far short of the sophisticated motor proteins natural selection has crafted over billions of years of evolution. The results in Chapters 3 and 4 suggest sub-optimal features in the design of the spider systems characterized in Chapter 2, namely substrates that are spaced too closely and likely give rise to 'leakage', i.e., walking without catalysis. This is expected to impose a limit on processivity, since it decreases the definition of the substrate-product interface upon which biased motion depends. This, and other features particular to oligonucleotide interactions at the surface of a

DNA nanostructure, should be integrated into more specific models of future DNA-based walkers.

Within the current design paradigm of spiders, the processivity may also ultimately be limited by the small number of sensor-actuator components (legs). The results from Chapter 4 show that as the number of reacting components on a single nanodevice increases, the behavior becomes more deterministic and reproducible between devices. Consistent with this observation, additional Monte Carlo simulations using the model from Chapter 2 (not shown) predict that using a DNA origami tile itself as a spider with dozens of legs can lead to more structured and processive walking on a two-dimensional surface. Enhancing the reproducibility of behaviors between individual walkers may improve the prospects of using spiders as autonomous agents capable of interacting with each other and navigating a variety of environments. Such efforts should begin with detailed modeling and careful design based on the results in this dissertation, as well as extensive characterization of any new designs by single-particle tracking. An attractive property of origami for this purpose is that they can be readily observed for hours at a time by labeling them site-specifically with hundreds of fluorophores and monitoring them at low excitation intensity (Chapter 4), which should permit long-term ultra-high-resolution tracking with minimal risk of photobleaching.

The two-color DNA-PAINT approach presented in Chapter 3 is a first step towards enabling real-time chemically specific imaging of structurally and compositionally complex DNA nanodevices. In the future, a primary goal should be improvement of temporal and spatial resolution. At present, these are primarily limited by background fluorescence from unbound probes. This background signal restricts the concentration of PAINT probes that can be practically used in super-resolution imaging of nanostructures, as any increase in concentration must come at the cost of localization accuracy. Yet, probe concentration (or the association rate constant) must be increased in order to achieve higher temporal resolution, which requires more binding events per unit time. Background fluorescence could be suppressed by a variety of approaches, from probes incorporating quenching-dequenching platforms reminiscent of molecular beacons[148] to analytical devices such as zero-mode waveguides[149]. Combining these approaches with higher concentrations of probes with shorter residence times could

push DNA-PAINT imaging into time resolutions of minutes or seconds, depending on the particular application, or spatial resolutions of one nanometer or less.

Since DNA nanodevices are increasingly incorporating materials other than DNA, it will be increasingly important to adapt PAINT to the detection of a wider variety of targets. In addition to DNA oligonucleotides, antibodies, aptamers, and lipophilic and charged probes should be explored as means of detecting proteins, lipid modifications, and local electrostatic or fields in the vicinity of DNA nanostructures.

The single-origami kinetics results in Chapter 4 show how severely solution-based models of hybridization kinetics can break down at the surface of a DNA nanostructure. In the future, we should investigate whether other processes deviate similarly from canonical solution behavior in this environment. For instance, it will be interesting to know whether nuclease-catalyzed hydrolysis or ssDNA-catalyzed strand displacement reactions can be modulated or tuned by controlling the spacing between substrates on an origami surface. Such effects have been reported for the restriction endonuclease *Dpn*II in nanografted monolayers of dsDNA with controlled density (~10-30 nm spacing)[150]. As DNA origami enables still more precise control over substrate spacing and orientation, it will be interesting to see whether it yields similar results. If such properties are discovered and quantified, they may even prove useful in designing new classes of devices, such as timed-release mechanisms for drug delivery or delay mechanisms in DNA computing circuits. It will also be important to investigate whether ionic strength or the identity of metal cations present in solution have an impact on deviations from solution kinetics, as the model in Figure 4.8d predicts.

Finally, the near-determinism and non-canonical kinetics observed in DNA-templated arrays bearing hundreds of reacting components could itself be exploited in the development of novel devices. For instance, one can conceive of origami-templated oligonucleotide arrays functioning as highly sensitive point-like biosensors for monitoring intracellular concentrations of metal ions, metabolites, short interfering RNAs, or proteins in live cells, particularly in conjunction with *in vitro* selected aptamers. Alternatively, the ability of immobilized targets to exchange captured probes may be useful as a sort of molecular ‘conveyor belt’ with controlled kinetic properties that, for

instance, captures dilute components from solution and transfers them to a nearby enzyme for processing.

The potential of bionanotechnology is limited only by our understanding of biophysics and our imagination. It is my hope that the single-molecule approaches and results presented in this dissertation have honed the former and provided fuel for the latter. I look forward to the many surprising and delightful discoveries that surely lie in the future of this rapidly advancing field.

APPENDIX 1

SUPPLEMENTARY INFORMATION TO CHAPTER 2

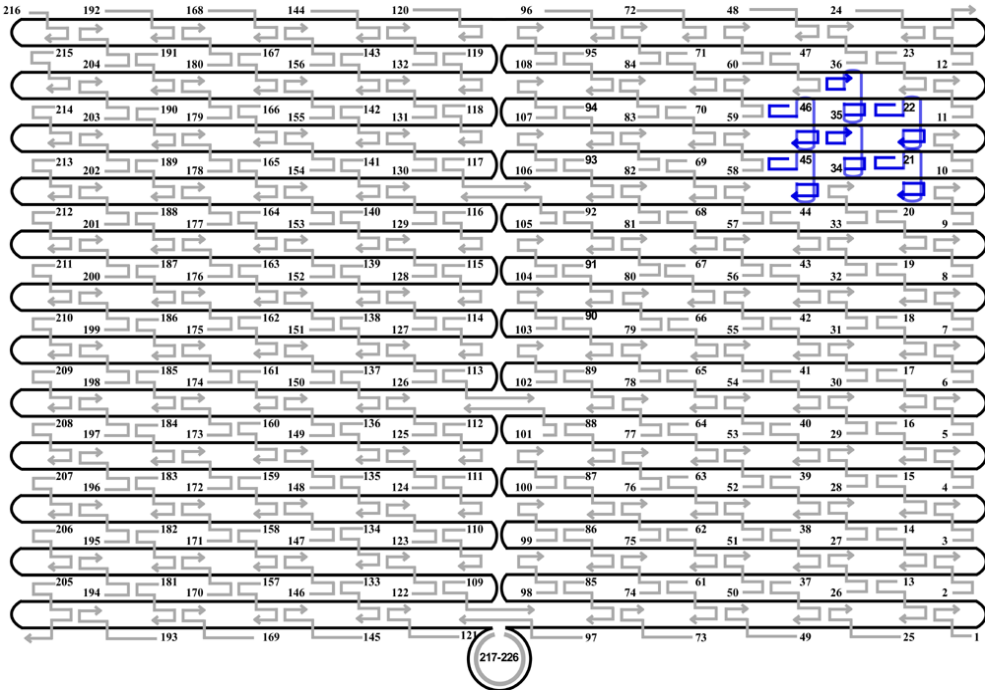


Figure A1.1. Schematic of the rectangular shaped DNA origami structure with the staple strand location and numbering marked. This is a representation of a plain origami structure with the marker included. In this drawing, the continuous black colored strand represents circular M13 viral genome and all the staple strands are shown in grey with arrows pointing the 3′- ends of the sequences. Numbers denote the sequence of the strands below. The blue strands denote the dumbbell hairpins used as a marker to aid in identification of origami by AFM.

DNA Sequences

The M13mp18 sequence can be found at the following web-address http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/sequences/m13mp18.txt.

Name	Sequence
1	TTTTCGATGGCCCACTACGTAAACCGTC
2	TATCAGGGTTTTCGGTTTGCGTATTGGGAACGCGCG
3	GGGAGAGGTTTTTGTAAAACGACGGCCATTCCCAGT
3A	GGGAGAGGTTTTTGTAAAAC
3B	Biotin GACGGCCATTCCCAGT
4	CACGACGTTTTTGTAATGGGATAGGTCAAAACGGCG
5	GATTGACCTTTTGATGAACGGTAATCGTAGCAAACA
6	AGAGAATCTTTTGGTTGTACCAAAAACAAGCATAAA
7	GCTAAATCTTTTCTGTAGCTCAACATGTATTGCTGA
8	ATATAATGTTTTCATTGAATCCCCCTCAAATCGTCA
9	TAAATATTTTTTGGAAGAAAAATCTACGACCAGTCA
10	GGACGTTGTTTTTCATAAGGGAACCGAAAGGCGCAG
11	ACGGTCAATTTTGACAGCATCGGAACGAACCCTCAG
11A	ACGGTCAATTTTGACAGCAT
11B	Biotin CGGAACGAACCCTCAG
12	CAGCGAAAATTTTACTTTCAACAGTTTCTGGGATTTTGCTAAACTTTT
13	TGGTTTTTAACGTCAAAGGGCGAAGAACCATC
14	CTTGCATGCATTAATGAATCGGCCCGCCAGGG
15	TAGATGGGGGGTAACGCCAGGGTTGTGCCAAG
16	CATGTCAAGATTCTCCGTGGGAACCGTTGGTG
17	CTGTAATATTGCCTGAGAGTCTGGAAAACTAG
18	TGCAACTAAGCAATAAAGCCTCAGTTATGACC
19	AAACAGTTGATGGCTTAGAGCTTATTTAAATA
20	ACGAACTAGCGTCCAATACTGCGGAATGCTTT
21	CTTTGAAAAGAACTGGTCCTCTTTTGAGGAACAAGTTTTCTTGT CTCATTATTTAATAAA
22	ACGGCTACTTACTTAGTCCTCTTTTGAGGAACAAGTTTTCTTGT CCGGAACGCTGACCAA
23	GAGAATAGCTTTTGCGGGATCGTCGGGTAGCA
24	ACGTTAGTAAATGAATTTTCTGTAAGCGGAGT
25	ACCCAAATCAAGTTTTTTGGGGTCAAAGAACG
26	TGGACTCCCTTTTCACCAGTGAGACCTGTCGT
27	GCCAGCTGCCTGCAGGTCGACTCTGCAAGGCG
28	ATTAAGTTCGCATCGTAACCGTGCGAGTAACA
29	ACCCGTCGTCATATGTACCCCGGTAAAGGCTA
30	TCAGGTCACTTTTGCGGGAGAAGCAGAATTAG
31	CAAAATTAAAGTACGGTGTCTGGAAGAGGTCA
32	TTTTTGCGCAGAAAACGAGAATGAATGTTTAG
33	ACTGGATAACGGAACAACATTATTACCTTATG

34 CGATTTTAGAGGACAGTCCTCTTTTGAGGAACAAGTTTTCTTGT ATGAACGGCGCGACCT
35 GCTCCATGAGAGGCTTTCCTCTTTTGAGGAACAAGTTTTCTTGT TGAGGACTAGGGAGTT
36 AAAGGCCGAAAGGAACAACTAAAGCTTTCCAG
37 AGCTGATTACAAGAGTCCACTATTGAGGTGCC
38 CCCGGGTACTTTCCAGTCGGGAAACGGGCAAC
39 GTTTGAGGGAAAGGGGGATGTGCTAGAGGATC
40 AGAAAAGCAACATTAAATGTGAGCATCTGCCA
41 CAACGCAATTTTTGAGAGATCTACTGATAATC
42 TCCATATACATACAGGCAAGGCAACTTTATTT
43 CAAAAATCATTGCTCCTTTTGATAAGTTTCAT
44 AAAGATTCAGGGGGTAATAGTAAACCATAAAT
45 CCAGGCGCTTAATCATTCCTCTTTTGAGGAACAAGTTTTCTTGT TGTGAATTACAGGTAG
46 TTTCATGAAAATTGTGTCCTCTTTTGAGGAACAAGTTTTCTTGT TCGAAATCTGTACAGA
47 AATAATAAGGTCGCTGAGGCTTGCAAAGACTT
48 CGTAACGATCTAAAGTTTTGTCGTGAATTGCG
49 GTAAAGCACTAAATCGGAACCCTAGTTGTTCC
50 AGTTTGGAGCCCTTCACCGCCTGGTTGCGCTC
51 ACTGCCCGCCGAGCTCGAATTCGTTATTACGC
52 CAGCTGGCGGACGACGACAGTATCGTAGCCAG
53 CTTTCATCCCCAAAAACAGGAAGACCGGAGAG
53A CTTTCATCCCCAAAAA
53B Biotin CAGGAAGACCGGAGAG
54 GGTAGCTAGGATAAAAATTTTTAGTTAACATC
55 CAATAAATACAGTTGATTCCCAATTTAGAGAG
56 TACCTTTAAGGTCTTTACCCTGACAAAGAAGT
57 TTTGCCAGATCAGTTGAGATTTAGTGGTTTAA
57A TTTGCCAGATCAGTTG
57B Biotin AGATTTAGTGGTTTAA
58 TTTCAACTATAGGCTGGCTGACCTTGTATCAT
59 CGCCTGATGGAAGTTTCCATTAAACATAACCG
60 ATATATTCTTTTTTCACGTTGAAAATAGTTAG
61 GAGTTGCACGAGATAGGGTTGAGTAAGGGAGC
62 TCATAGCTACTCACATTAATTGCGCCCTGAGA
63 GAAGATCGGTGCGGGCCTCTTCGCAATCATGG
64 GCAAATATCGCGTCTGGCCTTCCTGGCCTCAG
65 TATATTTTAGCTGATAAATTAATGTTGTATAA
66 CGAGTAGAACTAATAGTAGTAGCAAACCCTCA
67 TCAGAAGCCTCCAACAGGTCAGGATCTGCGAA
68 CATTCAACGCGAGAGGCTTTTGCATATTATAG
69 AGTAATCTTAAATTGGGCTTGAGAGAATACCA
70 ATACGTAAAAGTACAACGGAGATTTCATCAAG
71 AAAAAAGGACAACCATCGCCCACGCGGGTAAA

72	TGTAGCATTCCACAGACAGCCCTCATCTCCAA
73	CCCCGATTTAGAGCTTGACGGGGAAATCAAAA
74	GAATAGCCGCAAGCGGTCCACGCTCCTAATGA
75	GTGAGCTAGTTTCCTGTGTGAAATTTGGGAAG
76	GGCGATCGCACTCCAGCCAGCTTTGCCATCAA
77	AAATAATTTTAAATTGTAAACGTTGATATTCA
78	ACCGTTCTAAATGCAATGCCTGAGAGGTGGCA
79	TCAATTCTTTTAGTTTGACCATTACCAGACCG
80	GAAGCAAAAAAGCGGATTGCATCAGATAAAAA
81	CCAAAATATAATGCAGATACATAAACACCAGA
82	ACGAGTAGTGACAAGAACCGGATATACCAAGC
83	GCGAAACATGCCACTACGAAGGCATGCGCCGA
84	CAATGACACTCCAAAAGGAGCCTTACAACGCC
85	CCAGCAGGGGCAAAATCCCTTATAAAGCCGGC
86	GCTCACAATGTAAAGCCTGGGGTGGGTTTGCC
87	GCTTCTGGTCAGGCTGCGCAACTGTGTTATCC
88	GTTAAAATTTTAACCAATAGGAACCCGGCACC
89	AGGTAAAGAAATCACCATCAATATAATATTTT
90	TCGCAAATGGGGCGCGAGCTGAAATAATGTGT
91	AAGAGGAACGAGCTTCAAAGCGAAGATACATT
92	GGAATTACTCGTTTACCAGACGACAAAAGATT
93	CCAAATCACTTGCCCTGACGAGAACGCCAAAA
94	AAACGAAATGACCCCCAGCGATTATTCATTAC
95	TCGGTTTAGCTTGATACCGATAGTCCAACCTA
96	TGAGTTTCGTCACCAGTACAAACTTAATTGTA
97	GAACGTGGCGAGAAAGGAAGGGAACAAACTAT
98	CCGAAATCCGAAAATCCTGTTTGAAGCCGGAA
99	GCATAAAGTTCCACACAACATACGAAGCGCCA
100	TTCGCCATTGCCGGAAACCAGGCATTAAATCA
101	GCTCATTTTCGCATTAAATTTTTGAGCTTAGA
102	AGACAGTCATTCAAAAGGGTGAGAAGCTATAT
103	TTTCATTTGGTCAATAACCTGTTTATATCGCG
103A	TTTCATTTGGTCAATA
103B	Biotin ACCTGTTTATATCGCG
104	TTTTAATTGCCCGAAAGACTTCAAAACACTAT
105	CATAACCCGAGGCATAGTAAGAGCTTTTTAAG
106	GAATAAGGACGTAACAAAGCTGCTCTAAAACA
107	CTCATCTTGAGGCAAAAGAATACAGTGAATTT
108	CTTAAACATCAGCTTGCTTTCGAGCGTAACAC
109	ACGAACCAAAACATCGCCATTAAATGGTGGTT
110	CGACAACTAAGTATTAGACTTTACAATACCGA
111	CTTTTACACAGATGAATATACAGTAAACAATT
112	TTAAGACGTTGAAAACATAGCGATAACAGTAC
113	GCGTTATAGAAAAAGCCTGTTTAGAAGGCCGG
114	ATCGGCTGCGAGCATGTAGAAACCTATCATAT
115	CCTAATTTACGCTAACGAGCGTCTAATCAATA

116 AAAAGTAATATCTTACCGAAGCCCTTCCAGAG
117 TTATTCATAGGGAAGGTAAATATTCATTCAGT
118 GAGCCGCCCCACCACCGGAACCGCGACGGAAA
119 AATGCCCCGTAACAGTGCCCGTATCTCCCTCA
120 CAAGCCCAATAGGAACCCATGTACAAACAGTT
121 CGGCCTTGCTGGTAATATCCAGAACGAACTGA
122 TAGCCCTACCAGCAGAAGATAAAAACATTTGA
123 GGATTTAGCGTATTAAATCCTTTGTTTTCAGG
124 TTTAACGTTCGGGAGAAACAATAATTTTCCCT
125 TAGAATCCCTGAGAAGAGTCAATAGGAATCAT
126 AATTACTACAAATTCTTACCAGTAATCCCATC
127 CTAATTTATCTTTCCTTATCATTCATCCTGAA
128 TCTTACCAGCCAGTTACAAAATAAATGAAATA
129 GCAATAGCGCAGATAGCCGAACAATTCAACCG
130 ATTGAGGGTAAAGGTGAATTATCAATCACCGG
128 AACCAGAGACCCTCAGAACCGCCAGGGGTCAG
132 TGCCTTGACTGCCTATTTCGGAACAGGGATAG
133 AGGCGGTCATTAGTCTTTAATGCGCAATATTA
134 TTATTAATGCCGTCAATAGATAATCAGAGGTG
135 CCTGATTGAAAGAAATTGCGTAGACCCGAACG
136 ATCAAAATCGTCGCTATTAATTAACGGATTCG
137 ACGCTCAAAATAAGAATAAACACCGTGAATTT
138 GGTATTAAGAACAAGAAAAATAATTAAAGCCA
139 ATTATTTAACCCAGCTACAATTTTCAAGAACG
140 GAAGGAAAATAAGAGCAAGAAACAACAGCCAT
141 GACTTGAGAGACAAAAGGGCGACAAGTTACCA
142 GCCACCACTCTTTTCATAATCAAACCGTCACC
143 CTGAAACAGGTAATAAGTTTTAACCCCTCAGA
144 CTCAGAGCCACCACCCTCATTTTCCTATTATT
145 CCGCCAGCCATTGCAACAGGAAAAATATTTTT
146 GAATGGCTAGTATTAACACCGCCTCAACTAAT
147 AGATTAGATTTAAAAGTTTGAGTACACGTAAA
148 ACAGAAATCTTTGAATACCAAGTTCCTTGCTT
149 CTGTAAATCATAGGTCTGAGAGACGATAAATA
150 AGGCGTTACAGTAGGGCTTAATTGACAATAGA
151 TAAGTCCTACCAAGTACCGCACTCTTAGTTGC
152 TATTTTGCTCCCAATCCAAATAAGTGAGTTAA
153 GCCCAATACCGAGGAAACGCAATAGGTTTACC
154 AGCGCCAACCATTTGGGAATTAGATTATTAGC
155 GTTTGCCACCTCAGAGCCGCCACCGATACAGG
156 AGTGTACTTGAAAGTATTAAGAGGCCGCCACC
157 GCCACGCTATACGTGGCACAGACAACGCTCAT
158 ATTTTGCGTCTTTAGGAGCACTAAGCAACAGT
159 GCGCAGAGATATCAAAATTATTTGACATTATC
160 TAACCTCCATATGTGAGTGAATAAACAAAATC
160A TAACCTCCATATGTGA

160B Biotin GTGAATAAACAAAATC
161 CATATTTAGAAATACCGACCGTGTTACCTTTT
162 CAAGCAAGACGCGCCTGTTTATCAAGAATCGC
163 TTTTGTTTAAGCCTTAAATCAAGAATCGAGAA
164 ATACCCAAGATAACCCACAAGAATAAACGATT
164A ATACCCAAGATAACCC
164B Biotin ACAAGAATAAACGATT
165 AATCACCAAATAGAAAATTCATATATAACGGA
166 CACCAGAGTTCGGTCATAGCCCCCGCCAGCAA
167 CCTCAAGAATACATGGCTTTTGATAGAACCAC
168 CCCTCAGAACCGCCACCCTCAGAACTGAGACT
169 GGAAATACCTACATTTTGACGCTCACCTGAAA
170 GCGTAAGAGAGAGCCAGCAGCAAAAAGGTTAT
171 CTAAAATAGAACAAAGAAACCACCAGGGTTAG
172 AACCTACCGCGAATTATTCATTTCCAGTACAT
173 AAATCAATGGCTTAGGTTGGGTTACTAAATTT
174 AATGGTTTACAACGCCAACATGTAGTTCAGCT
175 AATGCAGACCGTTTTTATTTTCATCTTGCGGG
176 AGGTTTTGAACGTCAAAAATGAAAGCGCTAAT
177 ATCAGAGAAAGAACTGGCATGATTTTATTTTG
178 TCACAATCGTAGCACCATTACCATCGTTTTCA
179 TCGGCATTCCGCCGCCAGCATTGACGTTCCAG
180 TAAGCGTCGAAGGATTAGGATTAGTACCGCCA
181 CTAAAGCAAGATAGAACCCTTCTGAATCGTCT
182 CGGAATTATTGAAAGGAATTGAGGTGAAAAAT
183 GAGCAAAAACTTCTGAATAATGGAAGAAGGAG
184 TATGTAAACCTTTTTTAATGGAAAAATTACCT
185 AGAGGCATAATTTCATCTTCTGACTATAACTA
186 TCATTACCCGACAATAAACAACATATTTAGGC
187 CTTTACAGTTAGCGAACCTCCCGACGTAGGAA
188 TTATTACGGTCAGAGGGTAATTGAATAGCAGC
189 CCGGAAACACACCACGGAATAAGTAAGACTCC
190 TGAGGCAGGCGTCAGACTGTAGCGTAGCAAGG
191 TGCTCAGTCAGTCTCTGAATTTACCAGGAGGT
192 TATCACCGTACTCAGGAGGTTTAGCGGGGTTT
193 GAAATGGATTATTTACATTGGCAGACATTCTG
194 GCCAACAGTCACCTTGCTGAACCTGTTGGCAA
195 ATCAACAGTCATCATATTCCTGATTGATTGTT
196 TGGATTATGAAGATGATGAAACAAAATTTCAT
197 TTGAATTATGCTGATGCAAATCCACAAATATA
198 TTTTAGTTTTTCGAGCCAGTAATAAATTCTGT
199 CCAGACGAGCGCCCAATAGCAAGCAAGAACGC
200 GAGGCGTTAGAGAATAACATAAAAGAACACCC
201 TGAACAAACAGTATGTTAGCAAACTAAAAGAA
202 ACGCAAAGGTCACCAATGAAACCAATCAAGTT
203 TGCCTTTAGTCAGACGATTGGCCTGCCAGAAT

204 GGAAAGCGACCAGGCGGATAAGTGAATAGGTG
205 AAACCCTCTTTTACCAGTAATAAAAGGGATTCACCAGTCACACGTTTT
206 GATGGCAATTTTAATCAATATCTGGTCACAAATATC
206A GATGGCAATTTTAATCAATA
206B Biotin TCTGGTCACAAATATC
207 AAAACAAATTTTTTCATCAATATAATCCTATCAGAT
208 ACAAAGAATTTTATTAATTACATTTAACACATCAAG
209 TAAAGTACTTTTCGCGAGAAAACTTTTTATCGCAAG
210 TATAGAAGTTTTCGACAAAAGGTAAAGTAGAGAATA
211 GCGCATTATTTTGCTTATCCGGTATTCTAAATCAGA
212 TACATACATTTTGACGGGAGAATTAACTACAGGGAA
213 AGCACCGTTTTTTAAAGGTGGCAACATAGTAGAAAA
214 ACAAACAATTTTAATCAGTAGCGACAGATCGATAGC
214A ACAAACAATTTTAATCAGTA
214B Biotin GCGACAGATCGATAGC
215 AGGGTTGATTTTATAAATCCTCATTAAATGATATTC
216 TTTTTATAAGTATAGCCCGGCCGTCGAG
217 AACATCACTTGCCTGAGTAGAAGAACT
218 TGTAGCAATACTTCTTTGATTAGTAAT
219 AGTCTGTCCATCACGCAAATTAACCGT
220 ATAATCAGTGAGGCCACCGAGTAAAAG
221 ACGCCAGAATCCTGAGAAGTGTTTTT
222 TTAAAGGGATTTTAGACAGGAACGGT
223 AGAGCGGGAGCTAAACAGGAGGCCGA
224 TATAACGTGCTTTCCTCGTTAGAATC
225 GTACTATGGTTGCTTTGACGAGCACG
226 GCGCTTAATGCGCCGCTACAGGGCGC

The following three sequences are attached to the 5′ end of the staple sequences, as a probe, for the START position, binding of the cleavable substrate, and binding of the non-cleavable substrate. For fluorescence microscopy, strands 3A, 3B, 11A, 11B, 206A, 206B, 214A, 214B were incorporated into the origami and CONTROL staples were replaced with staples lacking the non-cleave-able substrate probes.

Spider START (green)
5′- GATGTCTACTTGCGTCAGGTTCTCGGC[staple]

Spider Cleavable Substrate Probes (brown)
5′- CCTCTCACCCACCATTCATC[staple]

Spider Non-Cleavable Substrate Probes (for STOP and CONTROL; red)
5′- GGTTCAGTTCGTTGAGCCAG[staple]

Spider Cleavable Substrate
5′- GATGAATGGTGGGTGAGAGGTTTTTCACTATrAGGAAGAG

Spider Non-Cleavable Substrate (STOP and CONTROL)
5′- CTGGCTCAACGAACTGAACC TTTTTCACTATAGGAAGAG

Spider Non-Cleavable Substrate (STOP) for fluorescence microscopy
5′- CTGGCTCAACGAACTGAACC TTTTTCACTATAGGAAGAG-Cy5

Spider TRIGGER Strand
5′- GCCGAGAACCTGACGCAAGTAGACATC

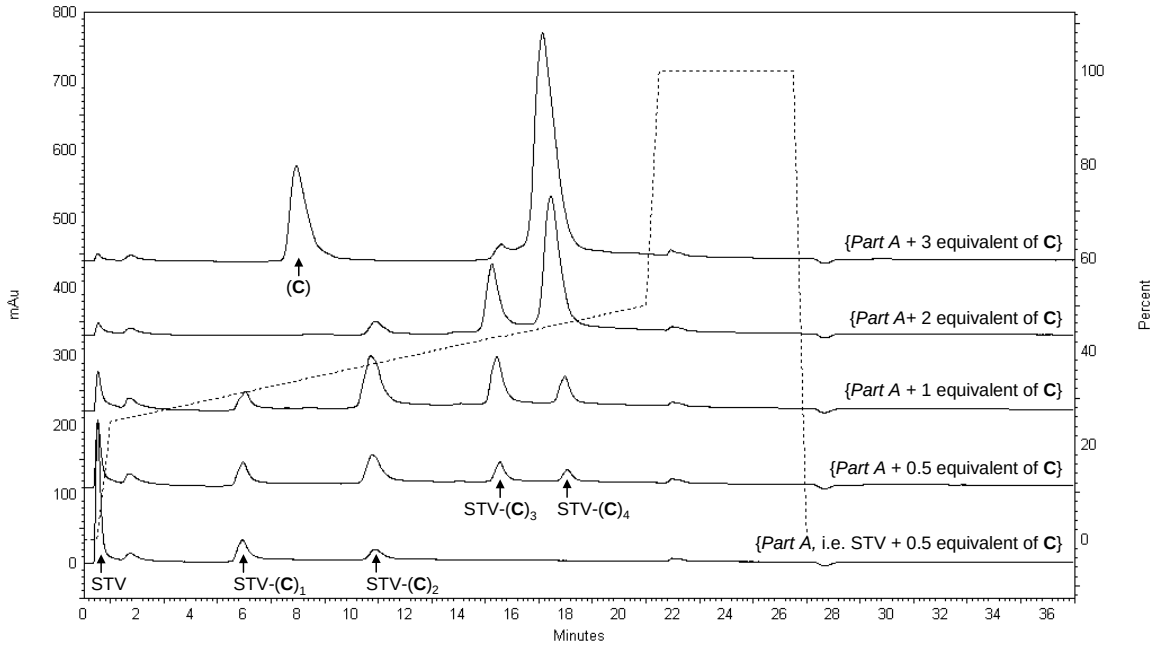


Figure A1.2. IE HPLC trace showing: 'Part A' mixture for **NICK$_{3.4A+1}$** from which "STV-(**C**)$_1$" was isolated (lowest trace); Other traces show the 'Part A' mixture with increasing amounts of **C** added (Note: "equivalent" amounts were based on the reported lyophilized amount of product supplied and not determined by absorption at 260 nm, which likely accounts for the introduction of a systematic error in the actual number of equivalents as observed by excess oligonucleotide present in the top trace). The 260nm/280nm ratios for the peaks of the middle trace are (left-to-right) 1.06, 1.28, 1.39, and 1.44, consistent with each consecutive peak containing a higher ratio of DNA-to-streptavidin than the peak preceding it. The 260nm/280nm ratio for peak "(**C**)" is 1.94, consistent with the absorption characteristics of pure DNA. See right y-axis for buffer B gradient (dotted line) as a percentage of buffers A plus B. Buffer A was composed of 20 mM TRIS, and buffer B, 20 mM TRIS/1 M NaCl, both adjusted to pH 7.4. The total flow rate of buffer A and B was 1 min^{-1}.

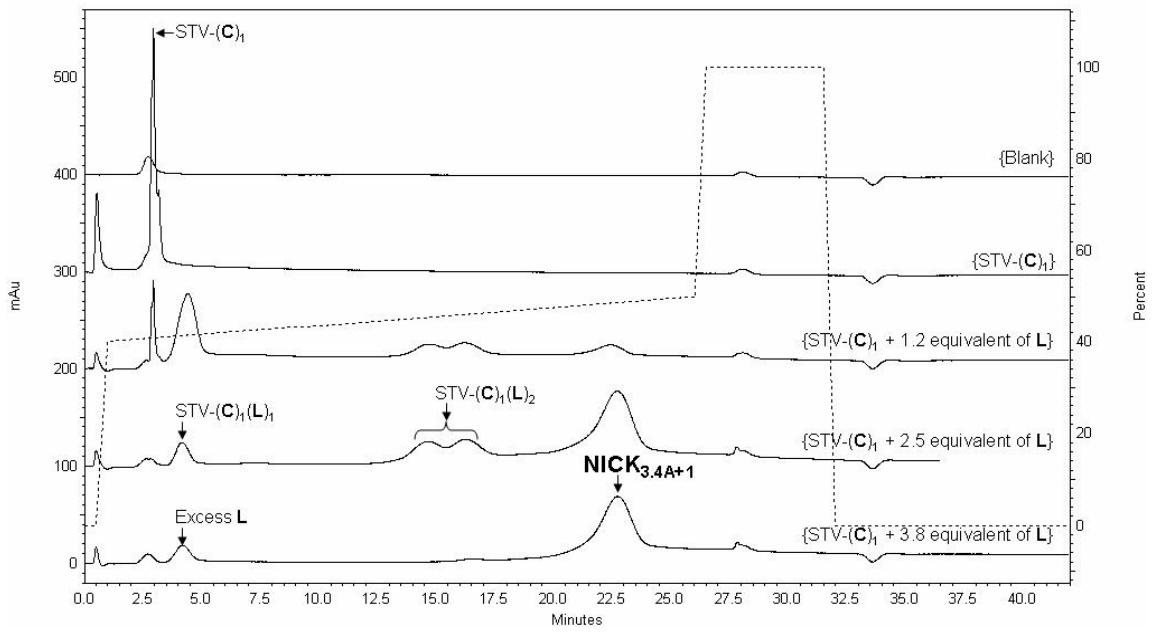


Figure A1.3. IE HPLC trace showing titration of STV-$(C)_1$ with increasing equivalents of L. 260nm/280nm ratios are STV-$(\mathbf{C})_1$ 1.09; STV-$(\mathbf{C})_1(\mathbf{L})_1$ 1.38; STV-$(\mathbf{C})_1(\mathbf{L})_2$ 1.53 and 1.49 (taken at the two apparent maxima respectively for STV-$(\mathbf{C})_1(\mathbf{L})_2$); **NICK$_{3.4A+1}$** 1.59; and **L** 2.05 (see caption for Figure A1.2 for explanation of absorption wavelength ratio 260/280). See right y-axis for buffer B gradient (dotted line) as a percentage of buffers A plus B. Buffer A was composed of 20 mM TRIS, and buffer B, 20 mM TRIS/1 M NaCl, both adjusted to pH 7.4. The total flow rate of buffer A and B was 1 min^{-1}.

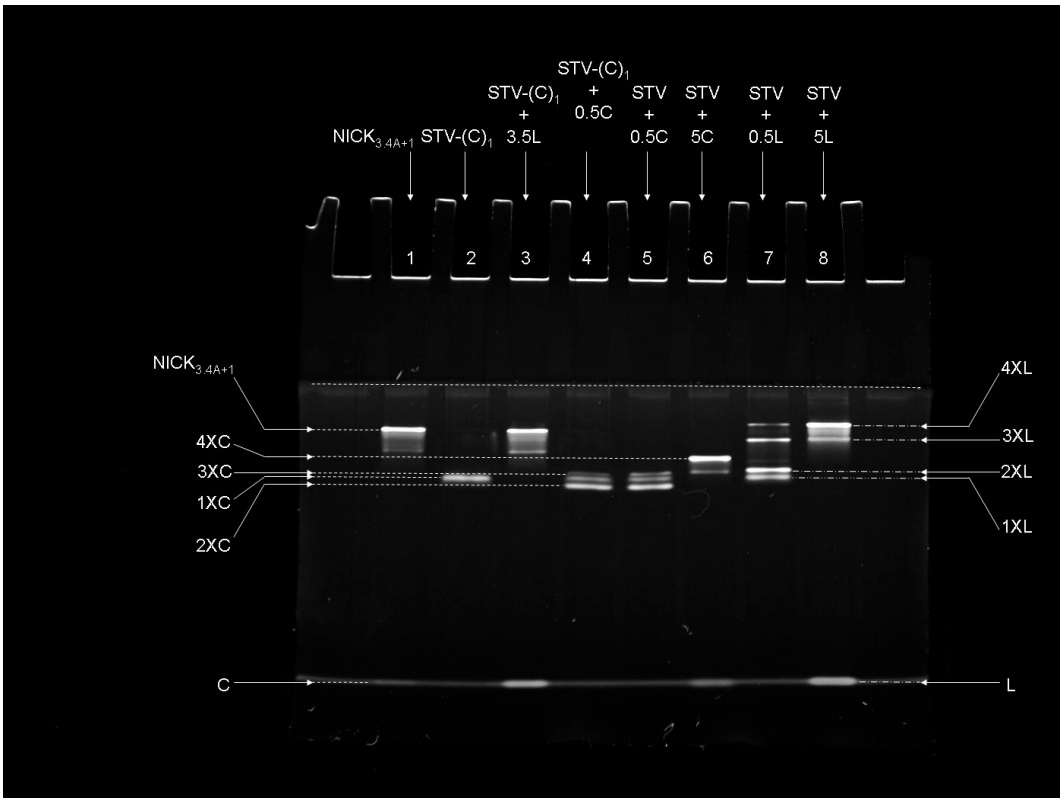


Figure A1.4. Gel characterization of spider assembly. PAGE characterization of **NICK$_{3.4A+1}$** showing that isolated **NICK$_{3.4A+1}$ (**lane 1**)** contains the strand C, i.e. the capture strand 5′ - GCC GAG AAC CTG ACG CAA GT/iSp18//iSp18//3Bio/ - 3′, and strand L, i.e. the deoxyribozyme or "leg" strand 5′ - /5BioTEG//iSp18//iSp18/TCT CTT CTC CGA GCC GGT CGA AAT AGT GAA AA - 3′ in a ratio of 1:3. Native stacking gel with a 12% acrylamide separation layer and a 4% acrylamide stacking layer; running buffer is TRIS-glycine. Bands were stained with SYBR Gold (Invitrogen). STV is streptavidin. 1XC is an assembly consisting of one streptavidin conjugated to one capture strand C, 2XC is an assembly consisting of one streptavidin conjugated to two capture strands, etc (assignments of bands 1XC, 2XC, 3XC, and 4XC are made based on results shown in Figure A1.2). Lane 1 is the isolated **NICK$_{3.4A+1}$** assembly; Lane 2 is the isolated streptavidin-(mono)capture strand conjugate (STV-(C)$_1$) used to form **NICK$_{3.4A+1}$** by adding the "leg" strand L to the three remaining biotin binding sites; Lane 3 is the unpurified result on adding 3.5 equivalents of "leg" strand, L, to STV-(C)$_1$; Lane 4 is the titration of a half equivalent of C with STV-(C)$_1$ showing migration distances of STV-(C)$_n$ (where n = 1-3); Lane 5 is the titration of a half equivalent of C with STV; Lane 6 the titration of a excess C with STV; Lane 7 is the titration of a half equivalent of L with STV, where 1XL is an assembly consisting of one streptavidin conjugated to one "leg" strand L, 2XL is an assembly consisting of one streptavidin conjugated to two "leg" strands L etc. (assignments of bands 1XL, 2XL, 3XL, and 4XL are made based on results shown in Figure A1.5); Lane 8 is the titration of excess L with STV.

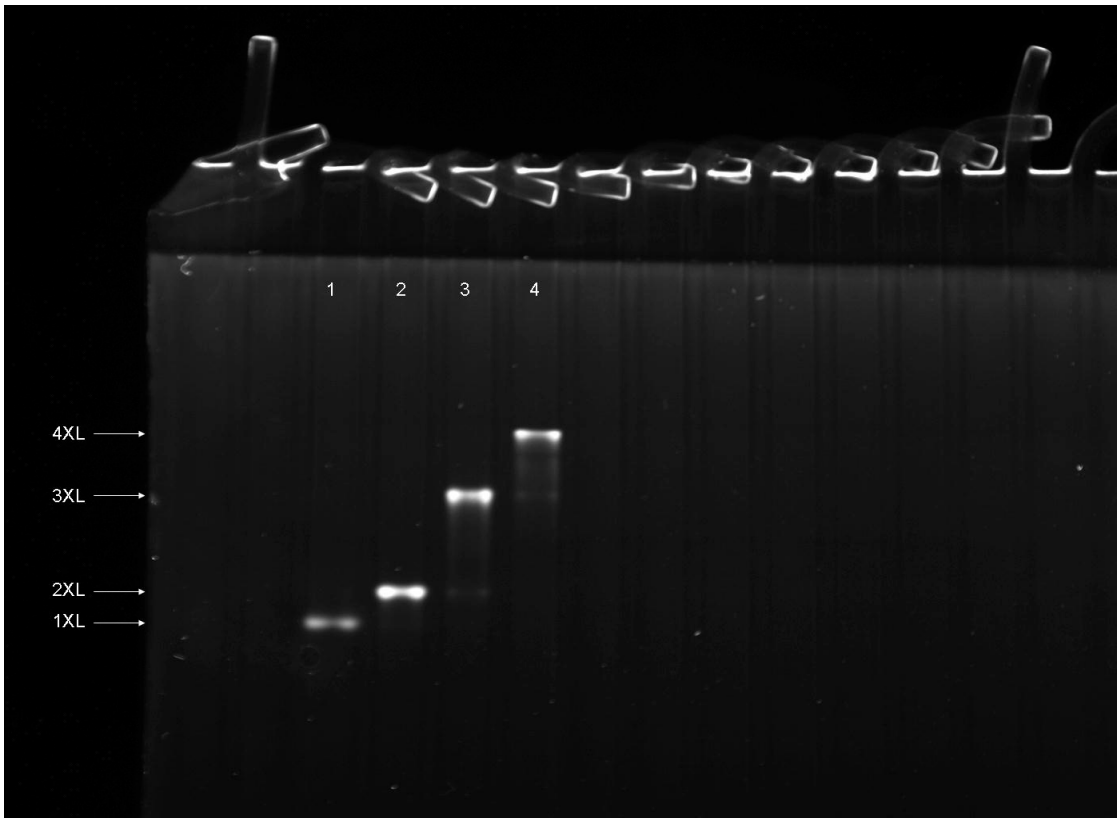


Figure A1.5. Gel characterization of assembled spiders. PAGE characterization of **NICK$_{n.4A}$** (where n = 1-to-4) supporting assignments of lane 8 in Figure A1.4. Native stacking gel with a 10% acrylamide separation layer and a 4% acrylamide stacking layer; running buffer is TRIS-glycine. Bands were stained with SYBR Gold (Invitrogen). STV-(**L**)$_n$ conjugates used in this gel were isolated and characterized as previously described above (Figure A1.2). 1XL is an assembly consisting of one streptavidin conjugated to one "leg" (i.e. deoxyribozyme strand L), 2XL is an assembly consisting of one streptavidin conjugated to two "legs," etc.

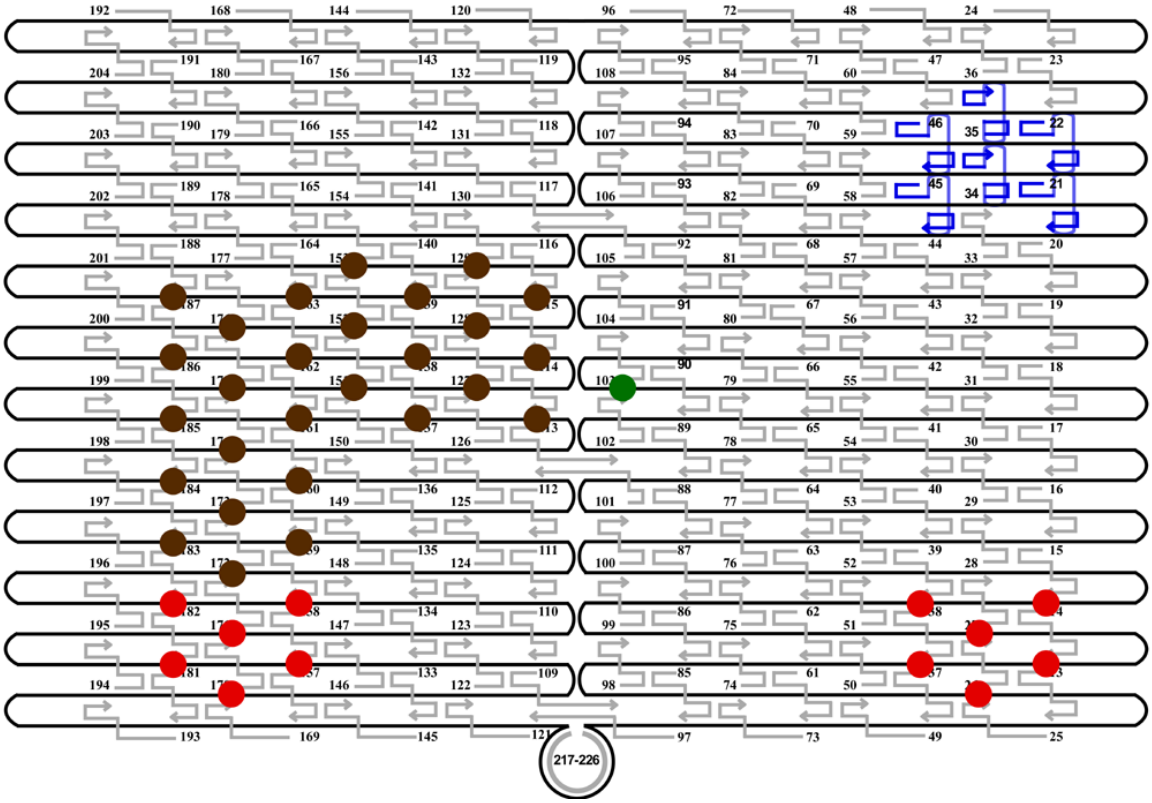
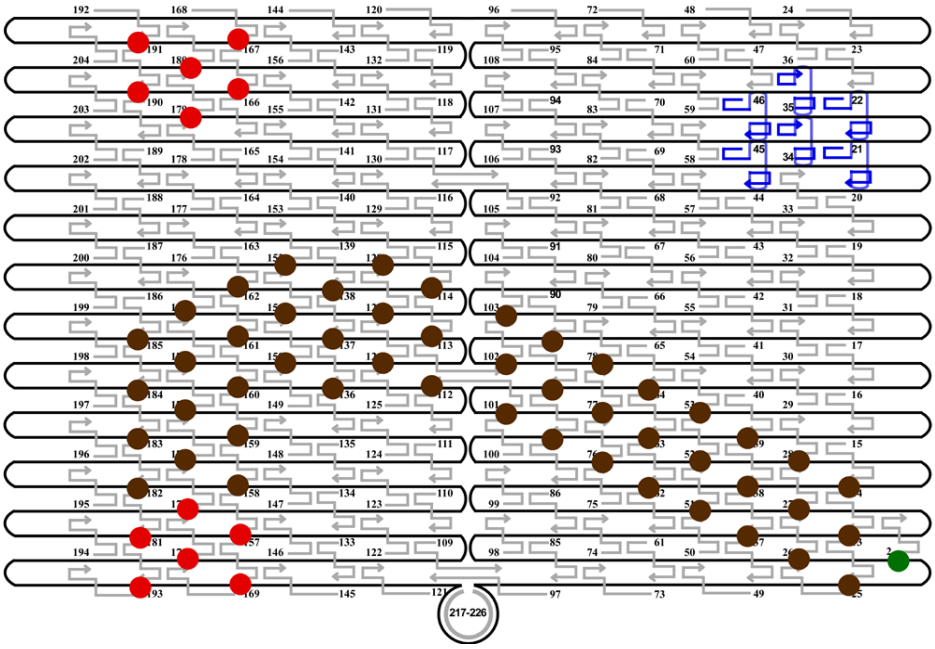


Figure A1.6. Schematic of the ABD origami design. Green represents the START position, brown the probes for the substrate, and red the probes for the STOP and CONTROL.

a



b

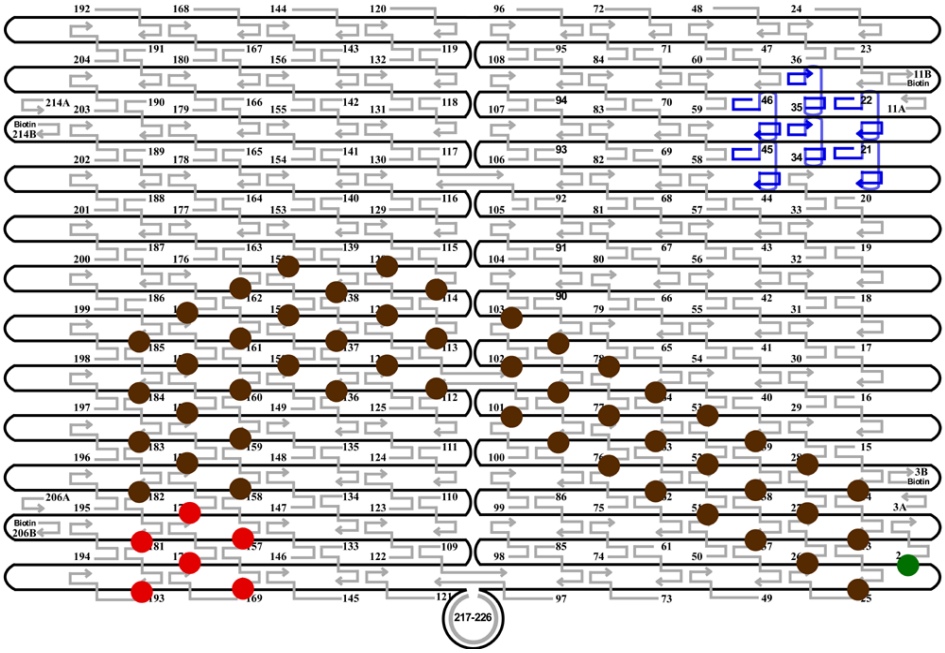
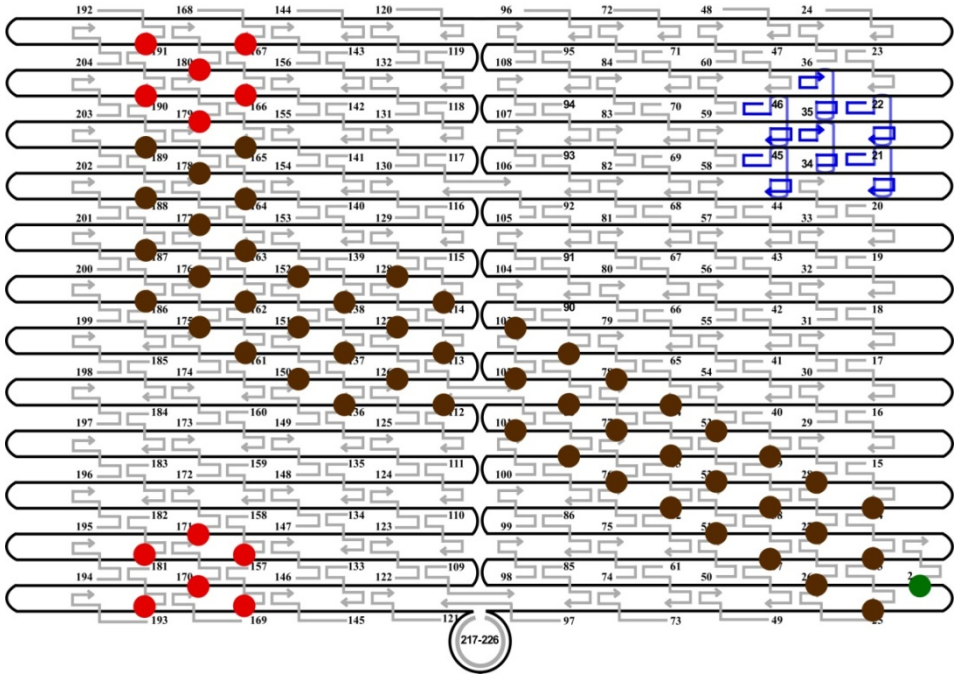


Figure A1.7. Schematic of the EABD origami design. Green represents the START position, brown the probes for the substrate, and red the probes for the STOP and CONTROL. **a,** AFM design and **b,** fluorescence microscopy design.

a



b

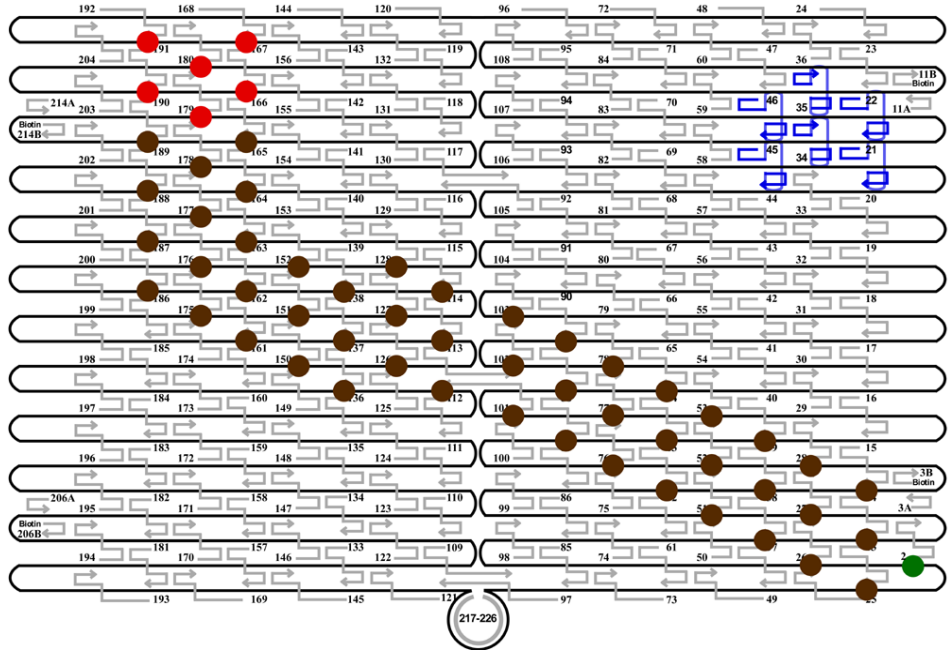
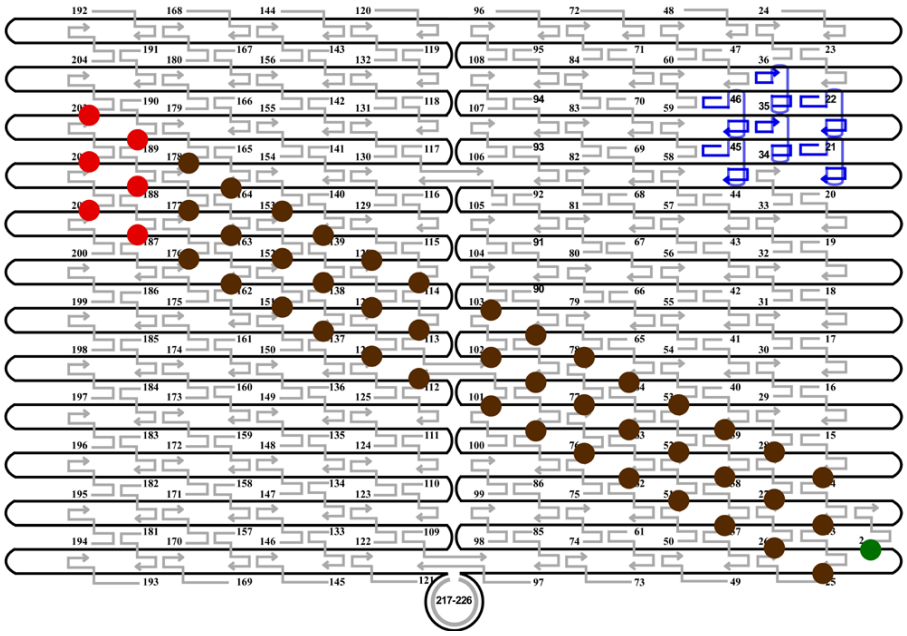


Figure A1.8. Schematic of the EABC origami design. Green represents the START position, brown the probes for the substrate, and red the probes for the STOP and CONTROL. **a,** AFM design and **b,** fluorescence microscopy design.

a



b

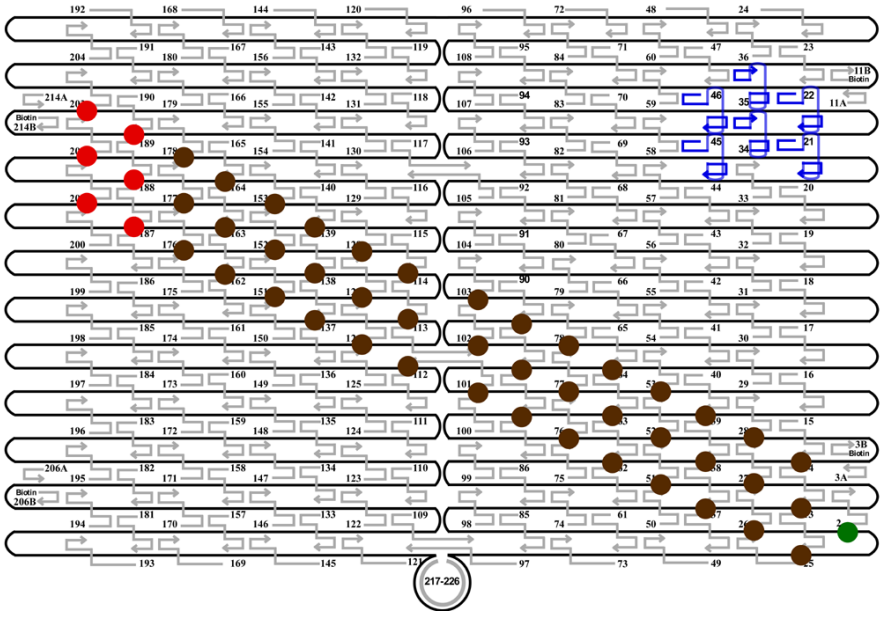


Figure A1.9. Schematic of the EAC origami design. Green represents the START position, brown the probes for the substrate, and red the probes for the STOP and CONTROL. (**a**) AFM design and (**b**) fluorescence microscopy design.

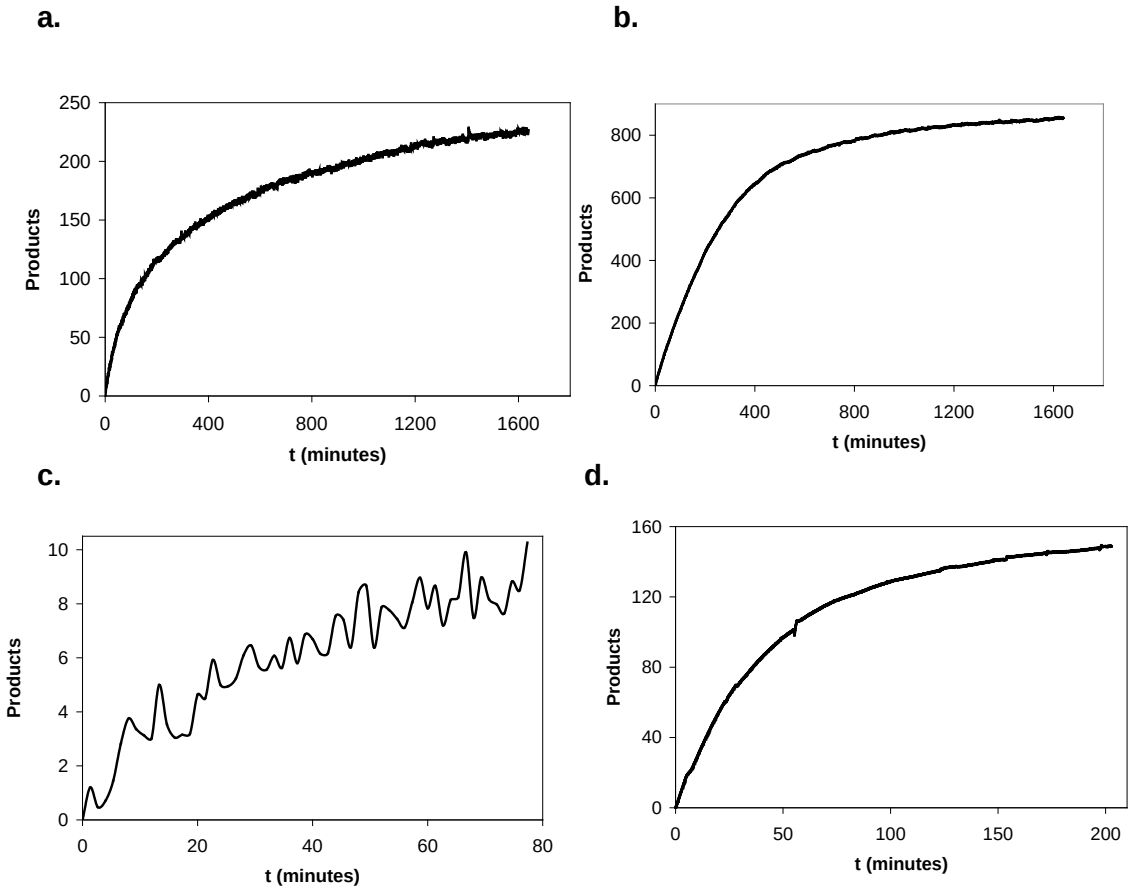


Figure A1.10. Spider cleavage sensorgrams. (**a**) Sensorgram (y-axis is products released per spider, the number of products released was obtained by conversion of SPR response unit (RU) to mass using the standard formula 1,000 RU= 1ng·mm^{-2}) of **NICK$_{3.4A+1}$** spider on the 2D monolayer surface showing the real-time substrate cleavage at a 1:291 ratio of spider to substrate with a cleavage rate of 1.42 min^{-1} per spider in 1× TA-Mg buffer with 1mM $ZnCl_2$. (**b**) Sensorgram of **NICK$_{3.4A+1}$** spider on the pseudo-2D matrix surface showing the real-time substrate cleavage at a 1:990 ratio of spider to substrate with a cleavage rate of 2.81 min^{-1} per spider in 1× TA-Mg buffer with 1mM $ZnCl_2$. (**c**) Sensorgram of **NICK$_{3.4A+1}$·(Cy3)$_4$** spider on the pseudo-2D matrix surface showing the real-time substrate cleavage at a 1:50 ratio of spider to substrate with a cleavage rate of 0.18 min^{-1} per spider in 1× SSC with 2 mM $ZnSO_4$. (**d**) Sensorgram of **NICK$_{3.4A+1}$·(Cy3)$_4$** spider on the pseudo-2D matrix surface showing the real-time substrate cleavage at a 1:180 ratio of spider to substrate with a cleavage rate of 2.72 min^{-1} per spider in HBS buffer with 1mM $ZnSO_4$. All cleavage reactions were monitored with a flow rate of 20 µL/min.

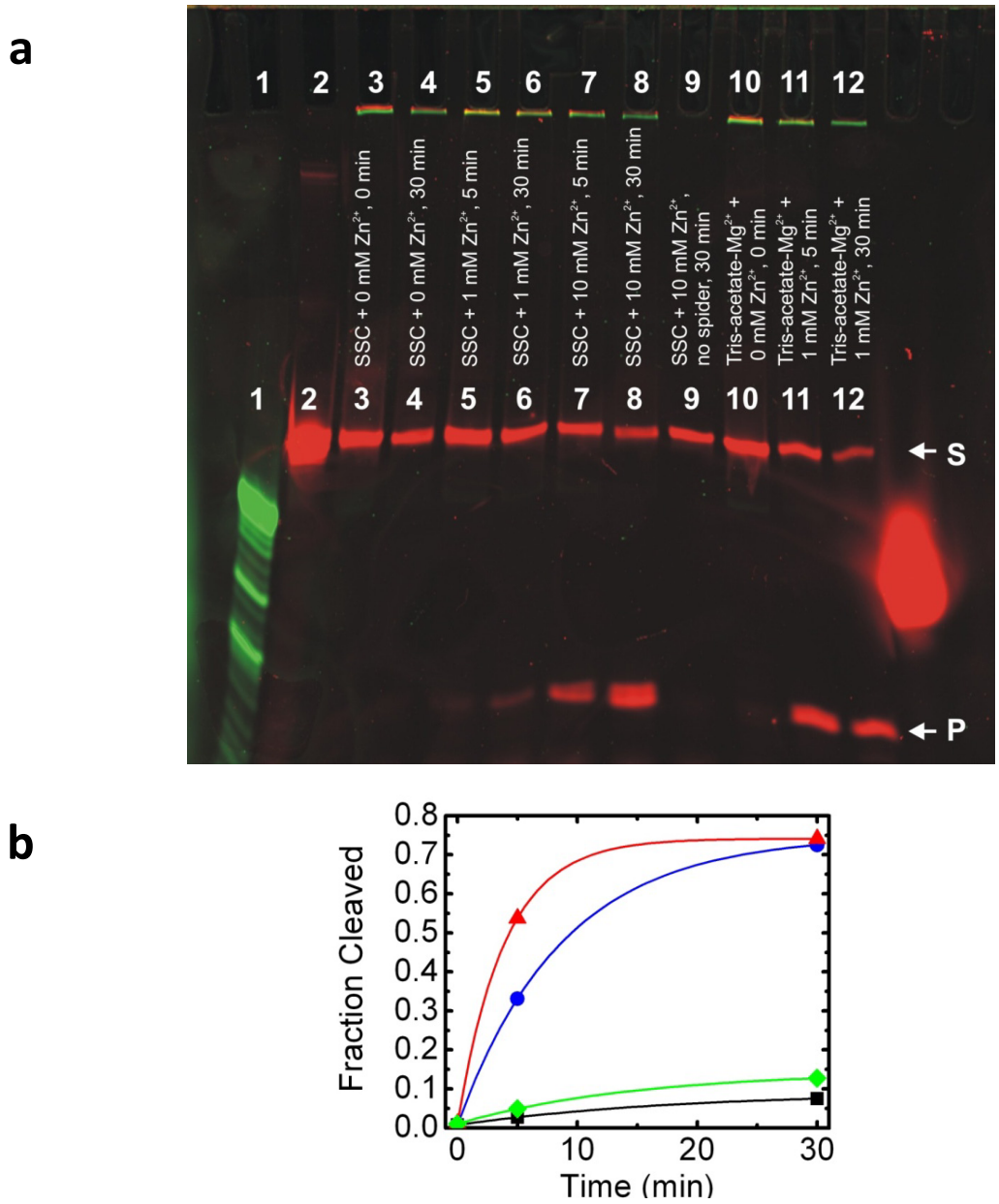
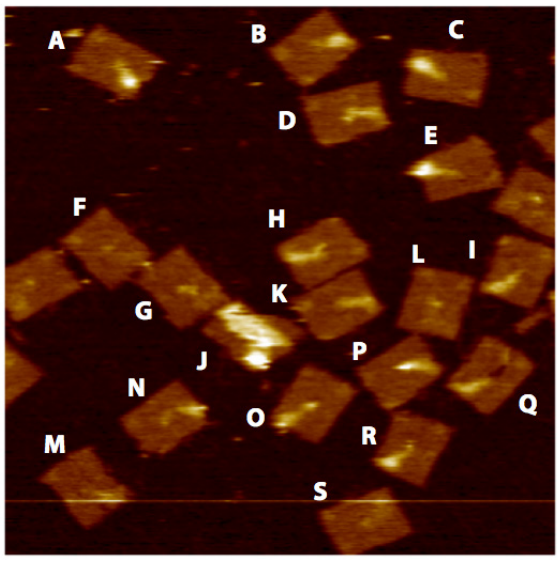


Figure A1.11. PAGE Characterization of Spider Activity in Solution. (**a**) Fluorescence scan of Cy3 and Cy5 in polyacrylamide gel (24% acrylamide). Lane 1 contains an alkali hydrolysis RNA ladder (sequence: 5'-pUGCGUUAGUAGGUUGUAUAGUU-Cy3). Lane 2 contains Cy5-substrate incubated at pH 12 for 5 min at 70°C. Lanes 3-12 contain the products of reactions between spider and Cy5-substrate (S) to form product (P) under the conditions shown in the respective lanes. No cleavage was detected after 30 minutes in absence of either $ZnSO_4$ (lane 4) or spider (lane 9). (**b**) Fraction of substrate cleaved versus incubation time in TA-Mg + 1 mM Zn^{2+} (red triangles). SSC + 10 mM Zn^{2+} (blue circles), SSC + 2 mM Zn^{2+} (green diamonds) or SSC + 1 mM Zn^{2+} (black squares). The cleavage assay in SSC + 2 mM Zn^{2+} is not shown in (**a**) but was performed in an identical manner to the other assays in a separate experiment. Each curve is fit to a single exponential decay function.

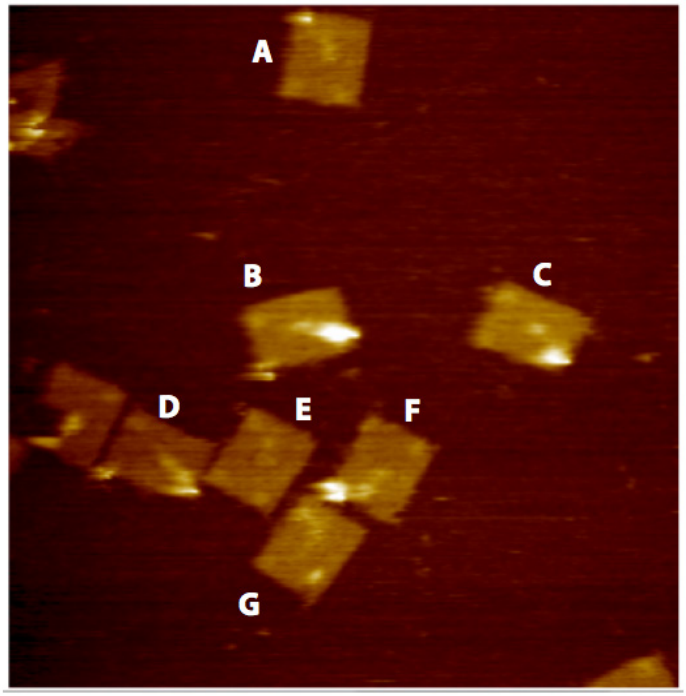
Track	Before							After						
	Total spiders	START (% of total)	TRACK (% of total)	STOP (% of total)	CONTROL (% of total)	Multiples	None	Total spiders	START (% of total)	TRACK (% of total)	STOP (% of total)	CONTROL (% of total)	Multiples	None
ABD 30 min	31	28 (90.3%)	1 (3.2%)	2 (6.5%)	0 (0%)	0	67	22	1 (4.5%)	4 (18.2%)	16 (72.7%)	1 (4.5%)	0	42
EABD 15 min								106	19 (17.9%)	68 (64.2%)	15 (14.1%)	4 (3.8%)	15	165
EABD 30 min	113	98 (86.7%)	8 (7.1%)	4 (3.5%)	3 (2.7%)	18	159	76	11 (14.5%)	33 (43.4%)	29 (38.2%)	3 (3.9%)	13	135
EABD 60 min								97	11 (11.3%)	23 (23.7%)	60 (61.9%)	3 (3.1%)	10	179
EABD product 60 min								26	4 (15.4%)	9 (34.6%)	12 (46.2%)	1 (3.9%)	2	37
EABC 30 min	98	78 (79.6%)	8 (8.2%)	12 (12.4%)	0 (0%)	14	149	42	2 (4.8%)	20 (47.6%)	20 (47.6%)	0 (0%)	8	70
EAC 30 min	67	56 (83.6%)	7 (10.4%)	4 (6%)	N/A	7	233	74	9 (12.2%)	44 (59.4%)	21 (28.4%)	N/A	8	197

Table A1.1. Data and statistics of “face-up” origami arrays. The number of spiders is the total number of spiders found at START, TRACK, STOP and CONTROL sites on singly-occupied origami.



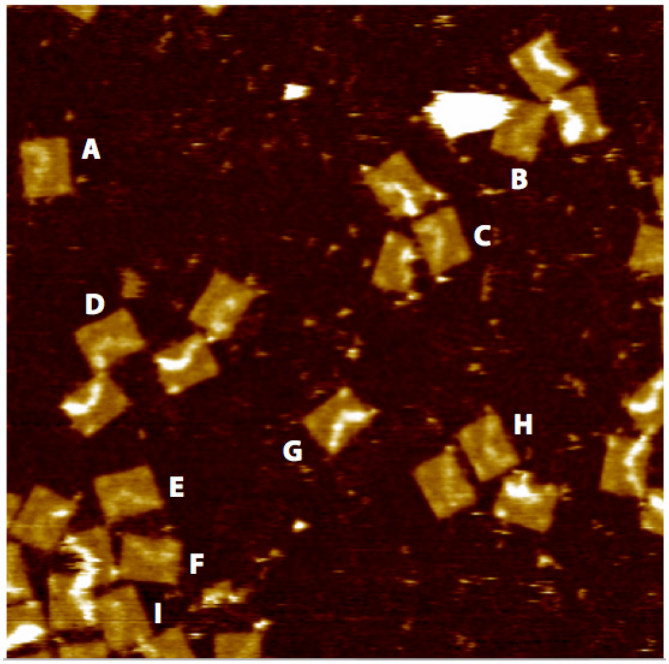
	up	Down	#Spiders	START	TRACK	STOP	CONTROL	inscrutable
a	0	1	1	0	0	1	0	0
b								1
c								1
d	0	1	0	0	0	0	0	0
e	0	1	1	0	0	1	0	0
f								1
g	1	0	1	1	0	0	0	0
h								1
i	0	1	0	0	0	0	0	0
j								1
k	0	1	0	0	0	0	0	0
l	1	0	1	1	0	0	0	0
m								1
n	0	1	1	0	0	1	0	0
o	0	1	1	0	0	1	0	0
p	0	1	1	1	0	0	0	0
q								1
r	0	1	1	0	0	1	0	0
s	1	0	0	0	0	0	0	0

Figure A1.12. Wide Field AFM images and classifications used for statistical analysis of ABD design. AFM images of the spider before release.



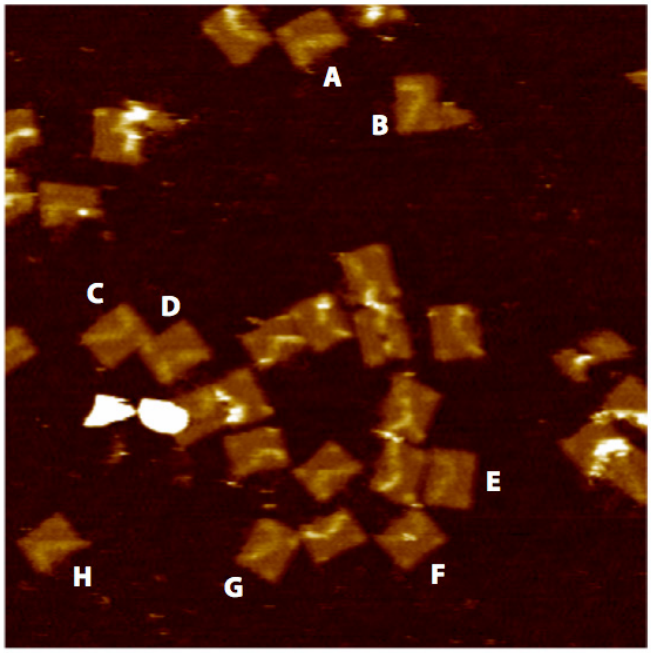
	up	down	#Spiders	START	TRACK	STOP	CONTROL	inscrutable
a	1	0	1	0	0	1	0	0
b	0	1	1	0	0	1	0	0
c	0	1	1	0	0	1	0	0
d	0	1	0	0	0	0	0	0
e								1
f	0	1	1	0	0	1	0	0
g	1	0	0	0	0	0	0	0

Figure A1.13. Wide Field AFM images and classifications used for statistical analysis of ABD design. AFM images of the spider after release.



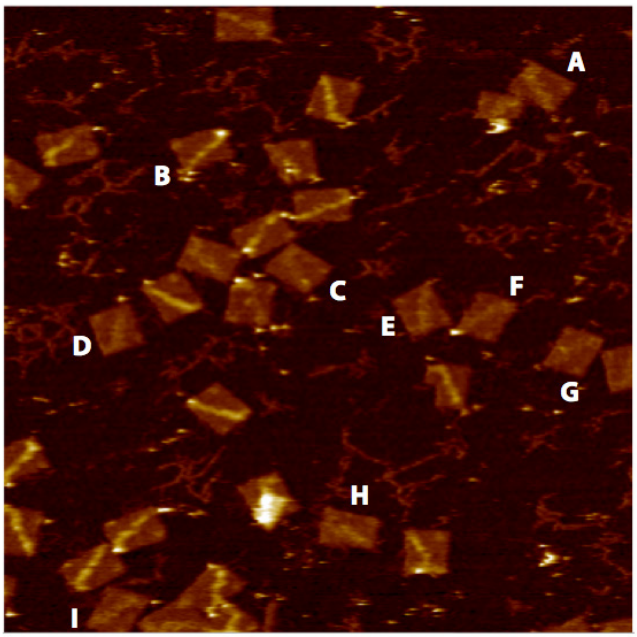
	up	Down	#Spiders	START	TRACK	STOP	CONTROL	inscrutable
a	1	0	0	0	0	0	0	0
b								1
c	1	0	1	1	0	0	0	0
d	1	0	1	0	0	1	0	0
e	1	0	0	0	0	0	0	0
f	1	0	0	0	0	0	0	0
g	1	0	1	1	0	0	0	0
h	1	0	1	1	0	0	0	0
i								1

Figure A1.14. Wide Field AFM images and classifications used for statistical analysis of EABD design. AFM images of the spider before release.



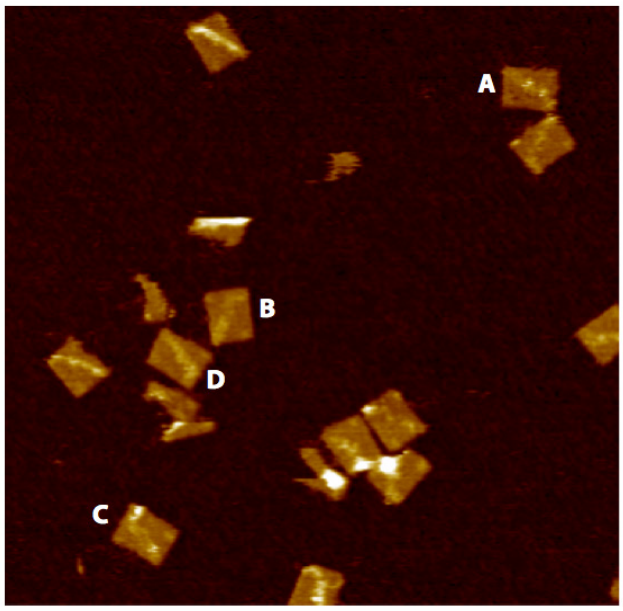
	up	down	#Spiders	START	TRACK	STOP	CONTROL	inscrutable
A	1	0	0	0	0	0	0	0
B								1
C	1	0	0	0	0	0	0	0
D	1	0	0	0	0	0	0	0
E	1	0	0	0	0	0	0	0
F	1	0	1	0	1	0	0	0
G	1	0	0	0	0	0	0	0
H	1	0	0	0	0	0	0	0

Figure A1.15. Wide Field AFM images and classifications used for statistical analysis of EABD design. AFM images of the spider after release.



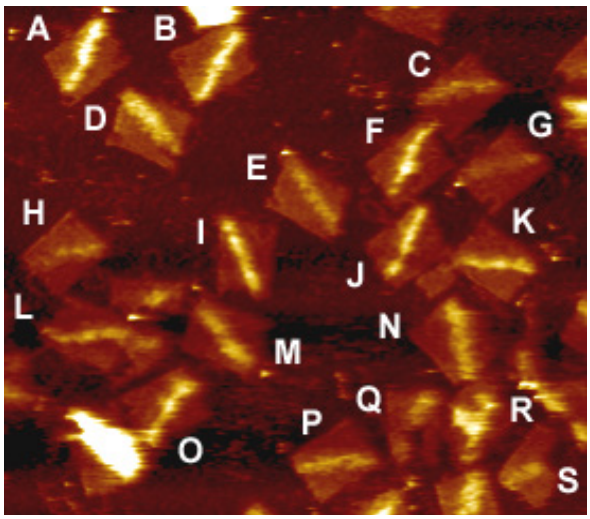
	up	down	#Spiders	START	TRACK	STOP	CONTROL	inscrutable
a	1	0	0	0	0	0	0	0
b	1	0	0	0	0	0	0	0
c	1	0	0	0	0	0	0	0
d	1	0	1	1	0	0	0	0
e	1	0	0	0	0	0	0	0
f	1	0	1	1	0	0	0	0
g								1
h	1	0	0	0	0	0	0	0
i	1	0	0	0	0	0	0	0

Figure A1.16. Wide Field AFM images and classifications used for statistical analysis of EABC design. AFM images of the spider before release.



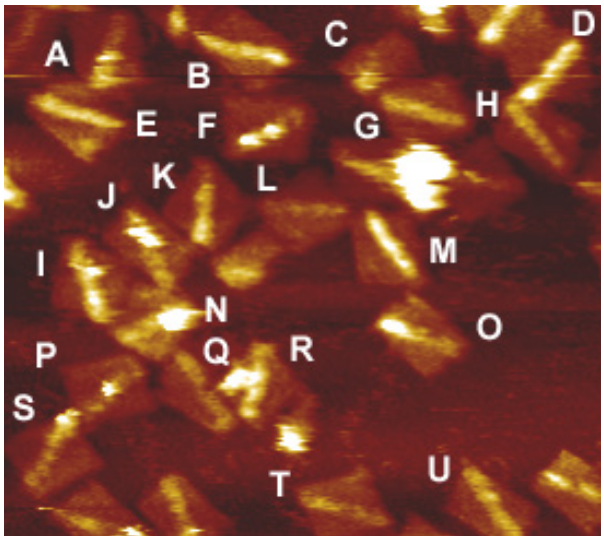
	up	down	#Spiders	START	TRACK	STOP	CONTROL	inscrutable
a	1	0	2	0	2	0	0	0
b	1	0	0	0	0	0	0	0
c	1	0	2	0	1	1	0	0
d	1	0	0	0	0	0	0	0

Figure A1.17. Wide Field AFM images and classifications used for statistical analysis of EABC design. AFM images of the spider after release.



	up	down	# spiders	START	TRACK	STOP	CONTROL	inscrutable
a		1						0
b		1						0
c	1	0	0	0	0	0	0	0
d	1	0	0	0	0	0	0	0
e	1	0	0	0	0	0	0	0
f		1				0	0	0
g	1	0	1	1	0	0	0	0
h	1	0	0	0	0	0	0	0
i		1						0
j		1						0
k		1						0
l		1						0
m								1
n	1	0	0	0	0	0	0	0
o		1						0
p	1	0	0	0	0	0	0	0
q								1
r								1
s								1

Figure A1.18. Wide Field AFM images and classifications used for statistical analysis of EAC design. AFM images of the spider before spider is released.



	up	down	# spiders	START	TRACK	STOP	CONTROL	inscrutable
a								1
b								1
c								1
d		1						0
e		1						0
f		1						0
g	1		0	0	0	0	0	0
h	1		0	0	0	0	0	0
i								1
j	1		1		1			0
k		1						0
l	1	0	0	0	0	0	0	0
m		1						0
n								1
o		1						0
p		1						0
q	1	0	0	0	0	0	0	0
r								1
s		1						0
t	1	0	0	0	0	0	0	0
u	1	0	0	0	0	0	0	0

Figure A1.19. Wide Field AFM images and classifications used for statistical analysis of EAC design. AFM images of the spider after release.

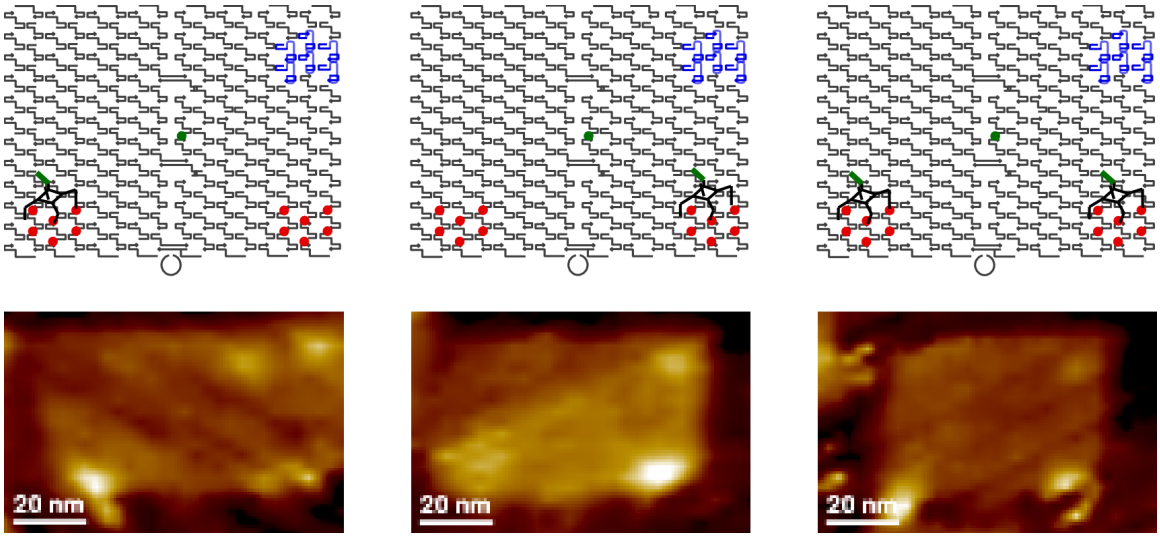


Figure A1.20. Schematics and AFM images of spider release control. The spider was released without the TRACK present and allowed to traverse the array for 30 minutes in solution. The images below show the spider at the STOP and CONTROL of this array and an instance where two spiders were seen occupying both positions on one array.

Table A1.2. Data of Spider Release without the TRACK

Track	START	STOP	CONTROL
ABD	7	48	45

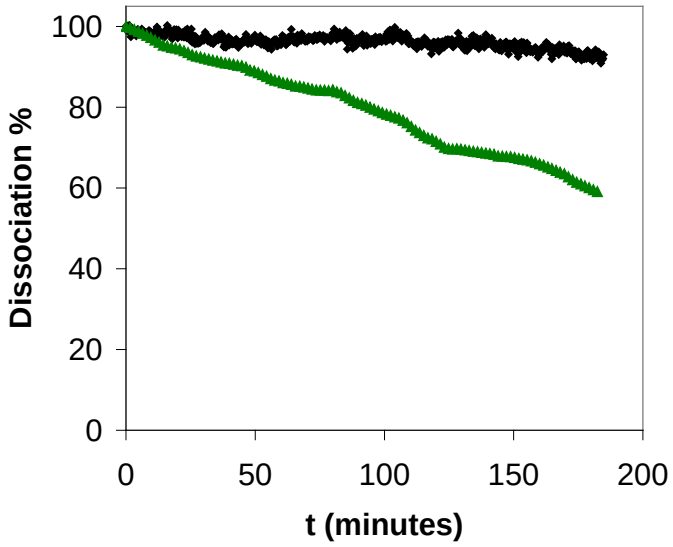
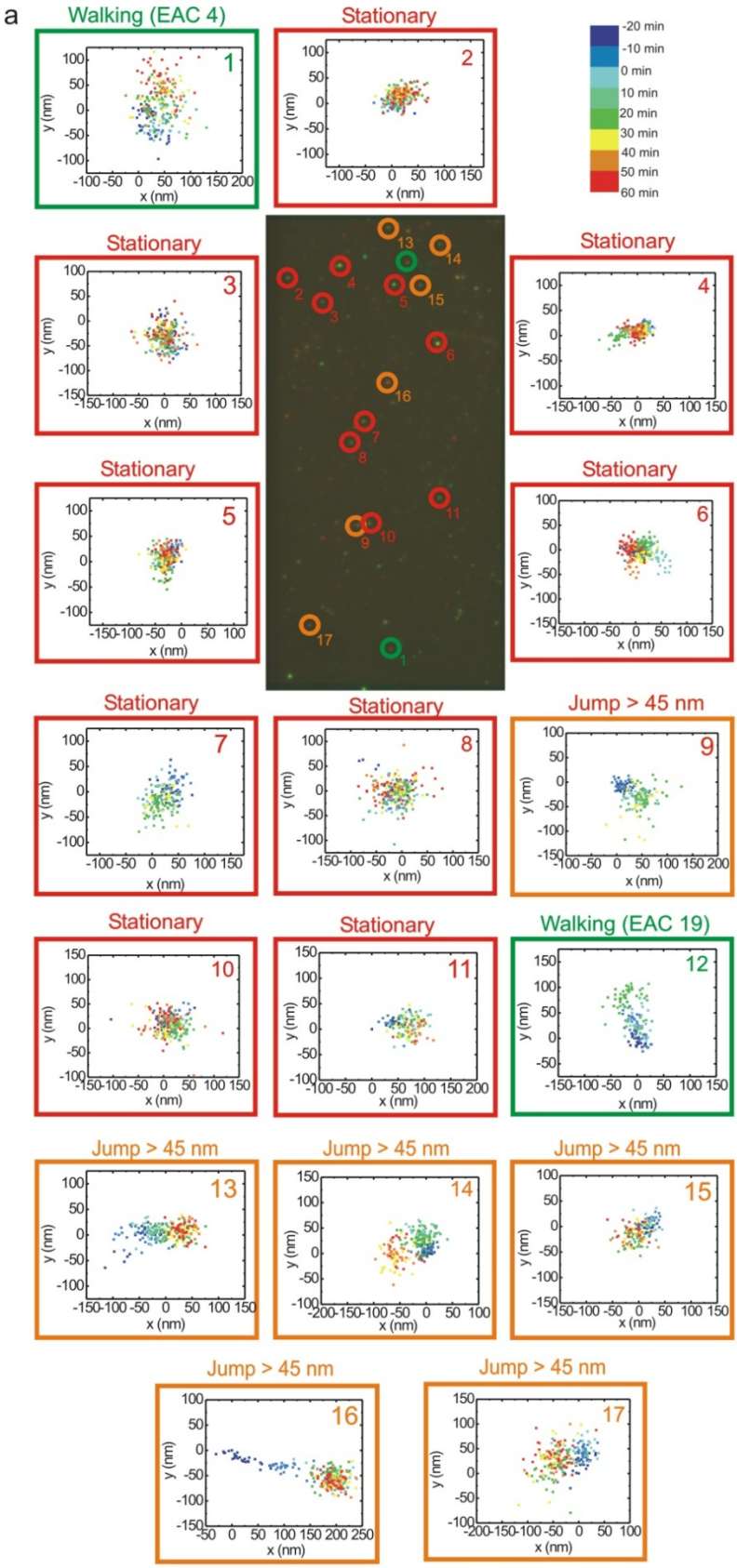
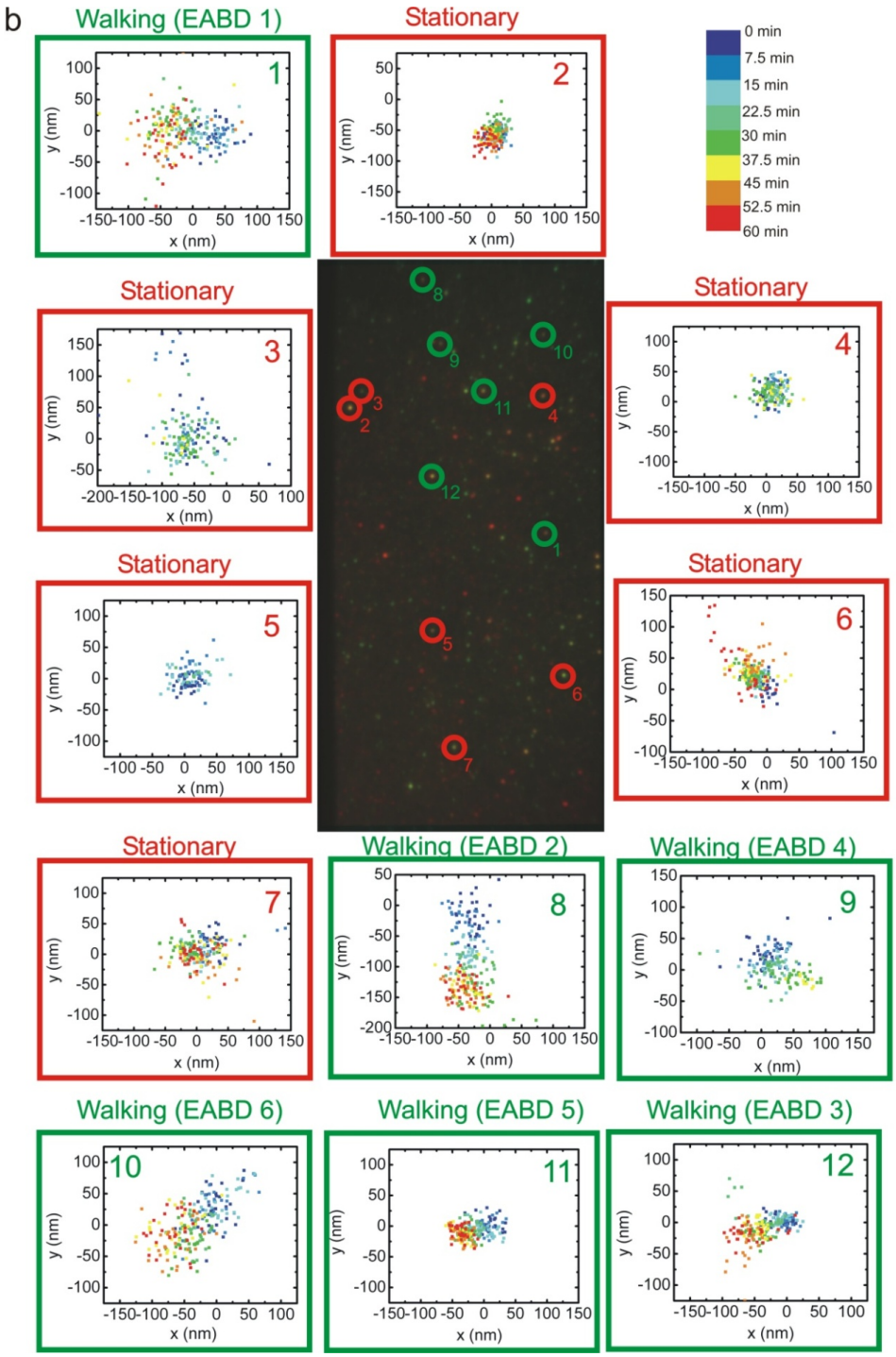
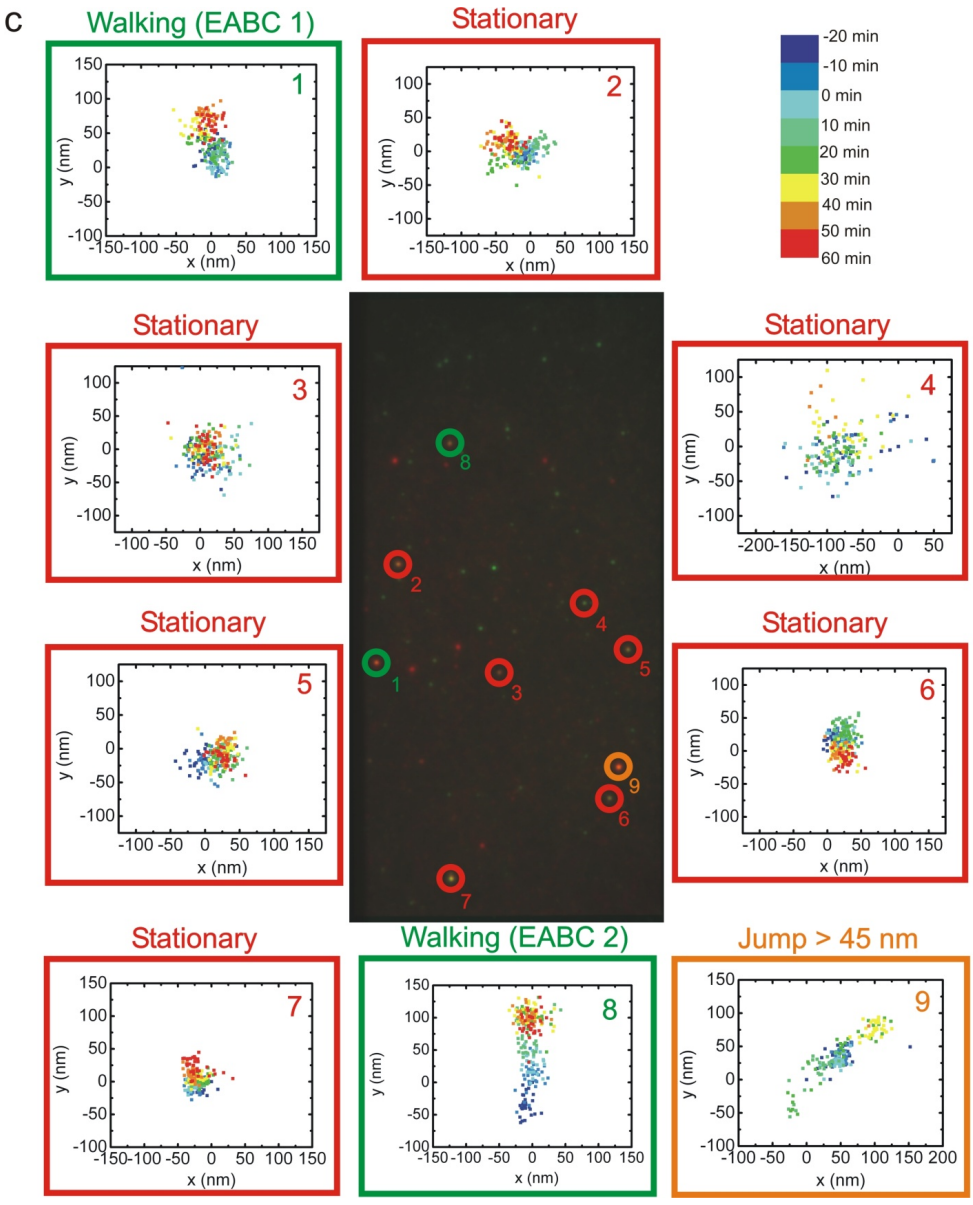


Figure A1.21. Dissociation curves for **NICK$_{3.4A+1}$** spider from non-cleavable substrate (black trace, 1:89 ratio of spider to substrate) and product (green trace, 1:97 ratio of spider to product) on the 2D monolayer surfaces.

Figure A1.22. Example CCD camera images from one of each type of experiment: EAC (**a**), EABD (**b**), EABC (**c**), and EAC in the absence of zinc (**d**). Both EAC experiments are performed in SSC buffer, and the EABC and EABD experiments in TA-Mg buffer as described in the text. The numbered circles mark the coordinates of the PSFs whose positions over time were determined using the Gaussian fitting technique described in Figure 2.5 as displayed in their corresponding numbered trajectory graphs. The trajectory graphs include spiders that walked continuously with a net displacement > 45 nm (green) as determined from the criteria in the text; PSFs that exhibited discontinuous displacement(s) > 45 nm and were thus determined to not be analyzable spiders (orange); and spiders or PSFs that remained stationary or displayed movement $\leq$ 45 nm (or 2-3 standard deviations, red). The lack of movement in the (-) zinc control (**d**) is consistent with the fact that cleavage activity is dependent on zinc, and supports the notion that movements seen in the experiments with zinc addition are not optical artifacts. Additionally, the presence of many apparently stationary spiders in the (+) zinc experiments (**a-c**) strongly suggests that the motion of adjacent spiders does not result from systematic instrument drift.


a
Walking (EAC 4)
Stationary
-20 min
-10 min
0 min
10 min
20 min
30 min
40 min
50 min
60 min
Stationary
Stationary
Stationary
Stationary
Stationary
Stationary
Jump > 45 nm
Stationary
Stationary
Walking (EAC 19)
Jump > 45 nm
Jump > 45 nm
Jump > 45 nm
Jump > 45 nm
Jump > 45 nm
y (nm)
x (nm)


b
Walking (EABD 1)
Stationary
0 min
7.5 min
15 min
22.5 min
30 min
37.5 min
45 min
52.5 min
60 min
Stationary
Stationary
Stationary
Stationary
Stationary
Walking (EABD 2)
Walking (EABD 4)
Walking (EABD 6)
Walking (EABD 5)
Walking (EABD 3)
y (nm)
x (nm)


C
Walking (EABC 1)
Stationary
-20 min
-10 min
0 min
10 min
20 min
30 min
40 min
50 min
60 min
Stationary
Stationary
Stationary
Stationary
Stationary
Walking (EABC 2)
Jump > 45 nm
y (nm)
x (nm)

d

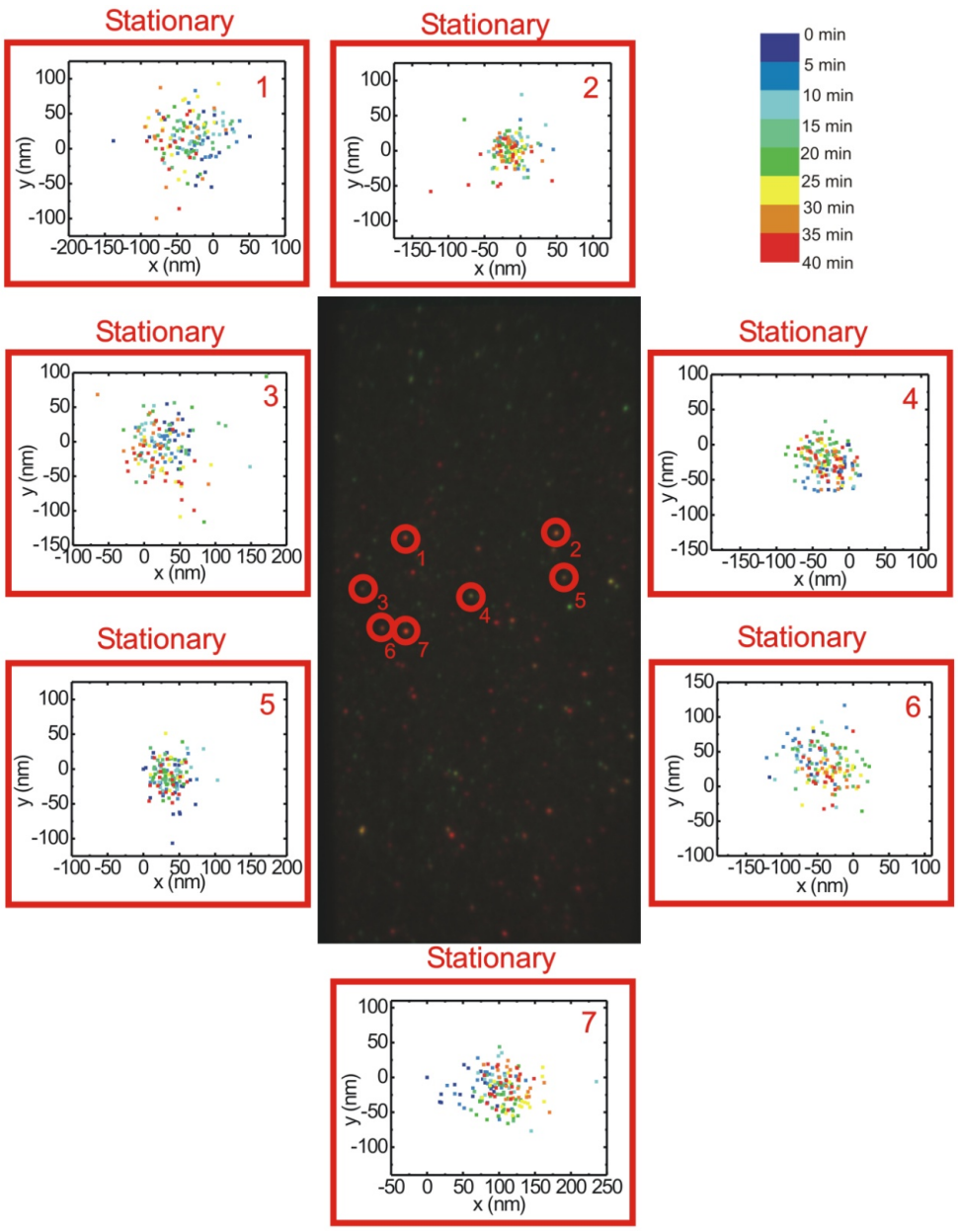
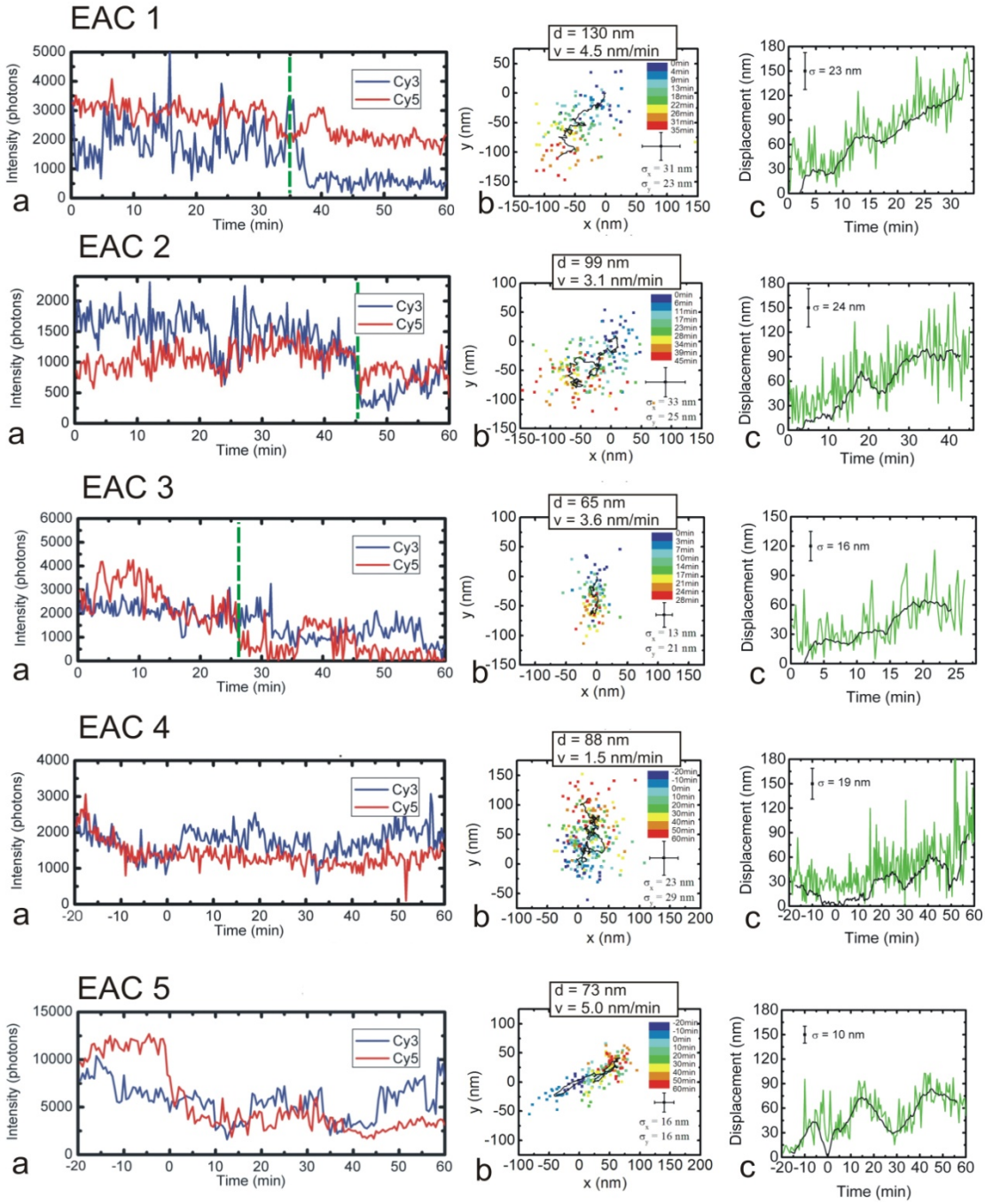
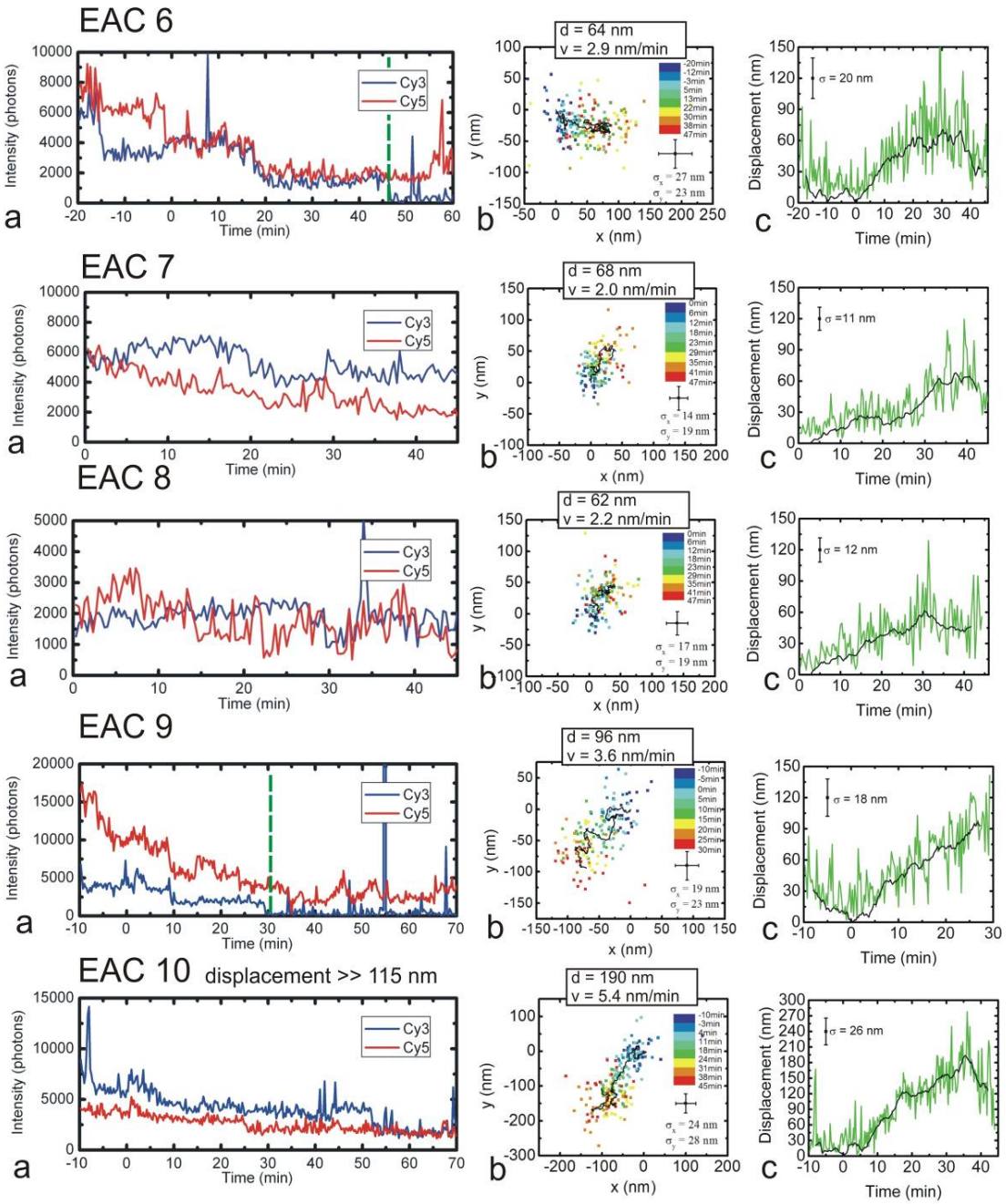
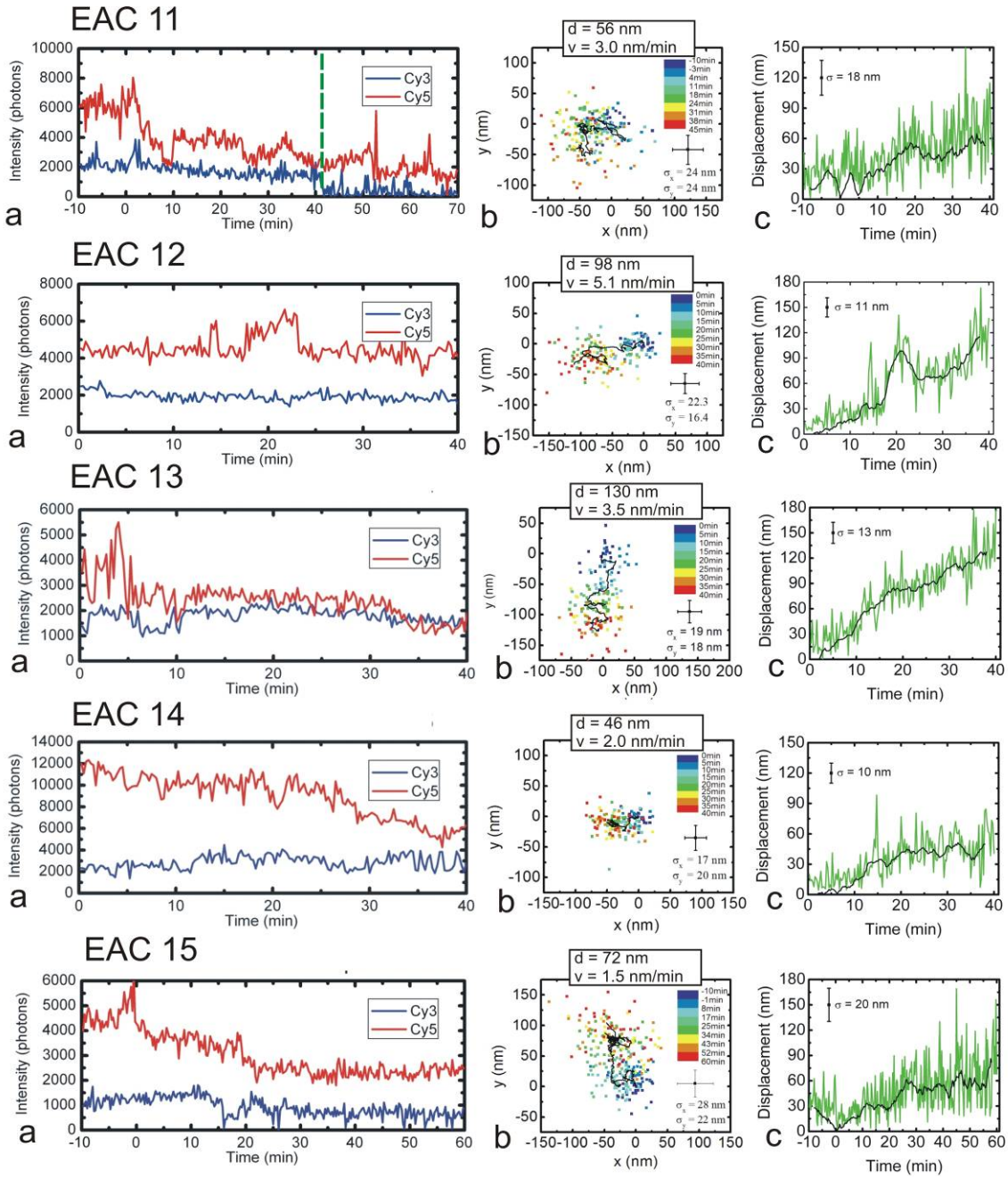

Stationary
1
Stationary
2
0 min
5 min
10 min
15 min
20 min
25 min
30 min
35 min
40 min
Stationary
3
Stationary
4
Stationary
5
Stationary
6
Stationary
7
y (nm)
x (nm)

Figure A1.23. Fluorophore emission intensity over time (**a**) and (non-averaged) 2-dimensional trajectories of the motion of individual spiders relative to the Cy5 PSF (**b**) for the EAC (EAC 1-19), EABD (EABD 1-6), and EABC (EABC 1-2) substrate tracks as viewed by fluorescence microscopy. The EAC traces in this figure were imaged in SSC buffer, and the EABD and EABC traces in TA-Mg buffer as described in the text. EAC traces are divided into "Tier 1" (EAC 1-15) and "Tier 2" (EAC 16-19), traces in the latter group having a lower probability of representing single walking spiders than the former due to reasons state above each trace and discussed in the text. In (**a**), the dashed green line represents the point after which the trace is no longer analyzed due to photobleaching. The black line in (**b**) represents the smoothed trajectory obtained by applying a 16-frame rolling average as described in the text. The position of the origin is arbitrarily chosen as the start of the trajectory. Plots of displacement versus time for the raw trajectory (**c,** green line) and smoothed trajectory (**c,** black line) are also shown for each trace. The addition of 1 mM $ZnSO_4$ (1-10 mM $ZnSO_4$ for EAC traces) occurred at t = 0 min. The intensity traces show the number of photons collected from each Cy3 or Cy5 point spread function over time, with sharp drops in photon count upon photobleaching of individual fluorophores. Only those time intervals with adequate tracking precision for both fluorophores – generally with more than 1,000 photon counts per frame – are shown in the 2-D trajectories (**b**), and it is these intervals which were analyzed to produce Fig. 4. In the 2-D trajectories, the axes represent spatial dimensions in the fluorescence microscopy image after drift correction. Also shown in panel b are values of net displacement (d) and mean velocity (v) for the EAC track, and mean velocity for the long leg (prior to the 90-degree turn) of the EABC and EABD tracks.

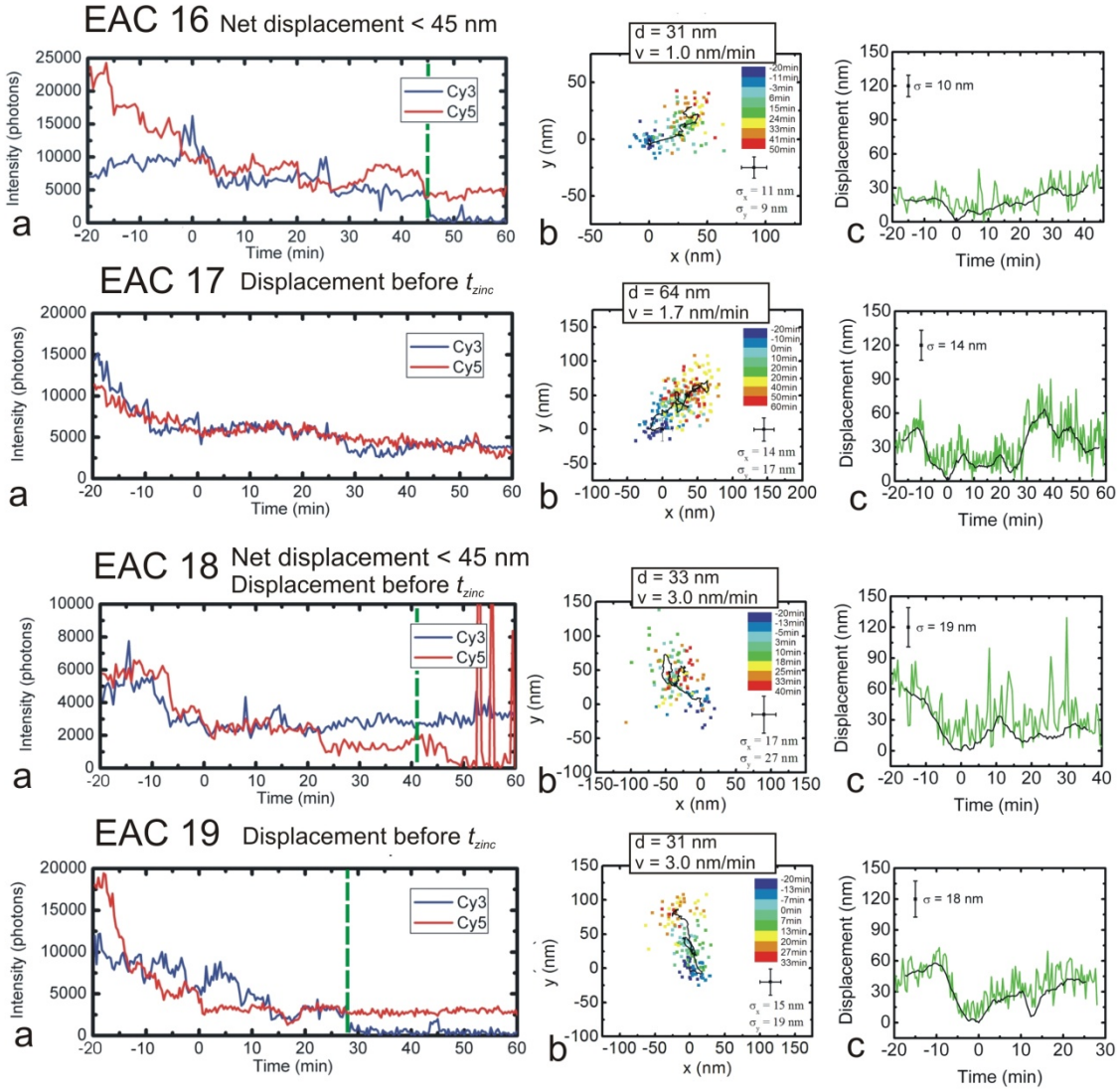
EAC Tier 1


EAC 1
d = 130 nm
v = 4.5 nm/min
σ = 23 nm
EAC 2
d = 99 nm
v = 3.1 nm/min
σ = 24 nm
EAC 3
d = 65 nm
v = 3.6 nm/min
σ = 16 nm
EAC 4
d = 88 nm
v = 1.5 nm/min
σ = 19 nm
EAC 5
d = 73 nm
v = 5.0 nm/min
σ = 10 nm
Cy3
Cy5
Intensity (photons)
Time (min)
x (nm)
y (nm)
Displacement (nm)
a
b
c

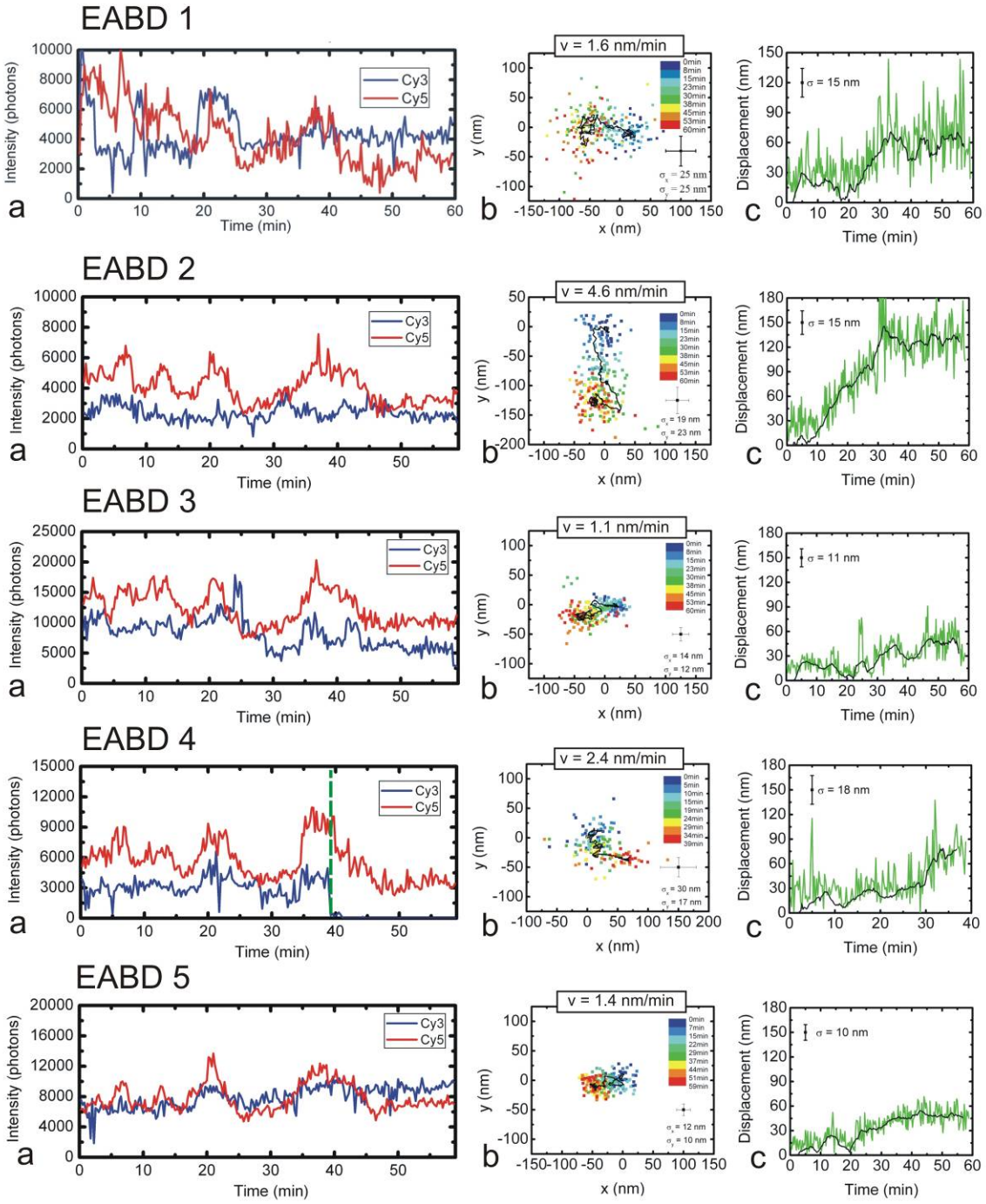

EAC 6
Intensity (photons)
Time (min)
Cy3
Cy5
a
d = 64 nm
v = 2.9 nm/min
y (nm)
x (nm)
σx = 27 nm
σy = 23 nm
b
Displacement (nm)
σ = 20 nm
Time (min)
c
EAC 7
d = 68 nm
v = 2.0 nm/min
σx = 14 nm
σy = 19 nm
σ =11 nm
EAC 8
d = 62 nm
v = 2.2 nm/min
σx = 17 nm
σy = 19 nm
σ = 12 nm
EAC 9
d = 96 nm
v = 3.6 nm/min
σx = 19 nm
σy = 23 nm
σ = 18 nm
EAC 10 displacement >> 115 nm
d = 190 nm
v = 5.4 nm/min
σx = 24 nm
σy = 28 nm
σ = 26 nm


EAC 11
Intensity (photons)
Time (min)
Cy3
Cy5
a
d = 56 nm
v = 3.0 nm/min
y (nm)
x (nm)
σx = 24 nm
σy = 24 nm
b
Displacement (nm)
σ = 18 nm
Time (min)
c
EAC 12
d = 98 nm
v = 5.1 nm/min
σx = 22.3
σy = 16.4
σ = 11 nm
EAC 13
d = 130 nm
v = 3.5 nm/min
σx = 19 nm
σy = 18 nm
σ = 13 nm
EAC 14
d = 46 nm
v = 2.0 nm/min
σx = 17 nm
σy = 20 nm
σ = 10 nm
EAC 15
d = 72 nm
v = 1.5 nm/min
σx = 28 nm
σy = 22 nm
σ = 20 nm

EAC Tier 2


EAC 16 Net displacement < 45 nm
Intensity (photons)
Cy3
Cy5
Time (min)
a
d = 31 nm
v = 1.0 nm/min
y (nm)
x (nm)
b
Displacement (nm)
σ = 10 nm
Time (min)
c
EAC 17 Displacement before t_{zinc}
d = 64 nm
v = 1.7 nm/min
σ = 14 nm
EAC 18 Net displacement < 45 nm
Displacement before t_{zinc}
d = 33 nm
v = 3.0 nm/min
σ = 19 nm
EAC 19 Displacement before t_{zinc}
d = 31 nm
v = 3.0 nm/min
σ = 18 nm

EABD


EABD 1
Intensity (photons)
Cy3
Cy5
a
Time (min)
v = 1.6 nm/min
y (nm)
x (nm)
b
Displacement (nm)
σ = 15 nm
c
Time (min)
EABD 2
v = 4.6 nm/min
σ = 15 nm
EABD 3
v = 1.1 nm/min
σ = 11 nm
EABD 4
v = 2.4 nm/min
σ = 18 nm
EABD 5
v = 1.4 nm/min
σ = 10 nm

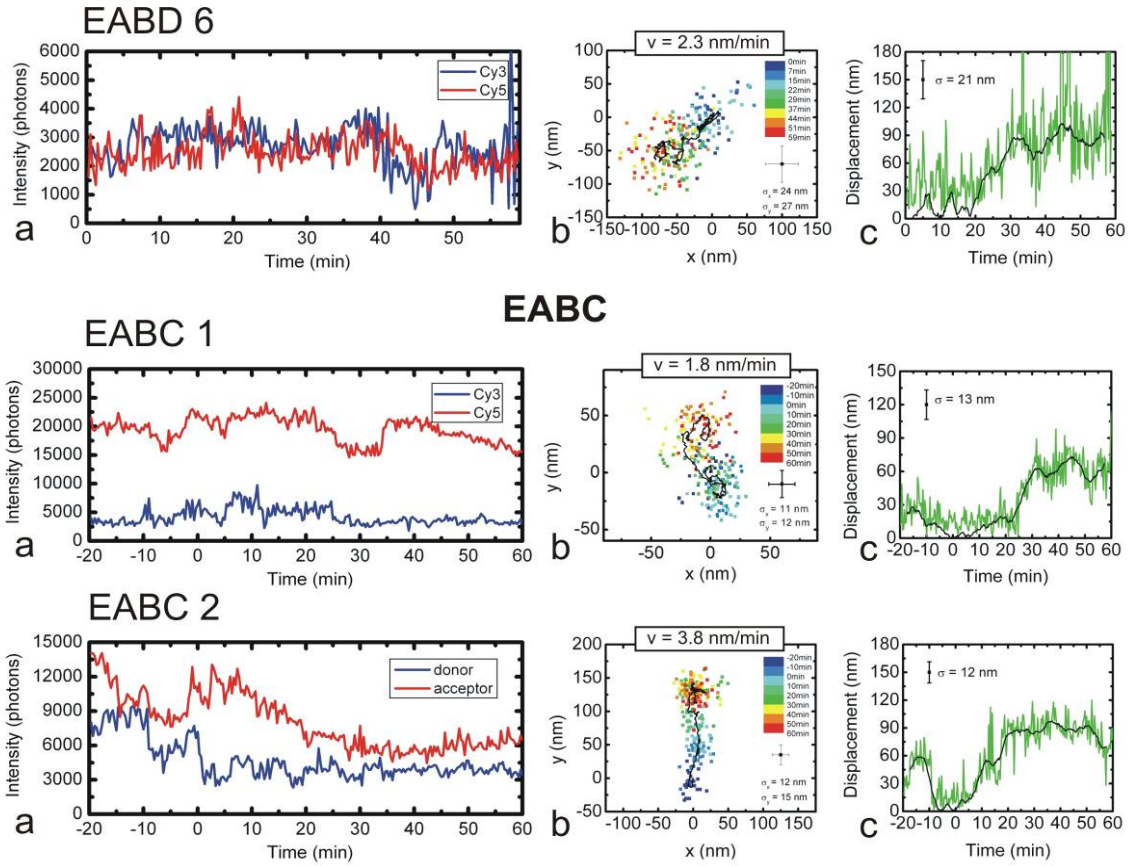
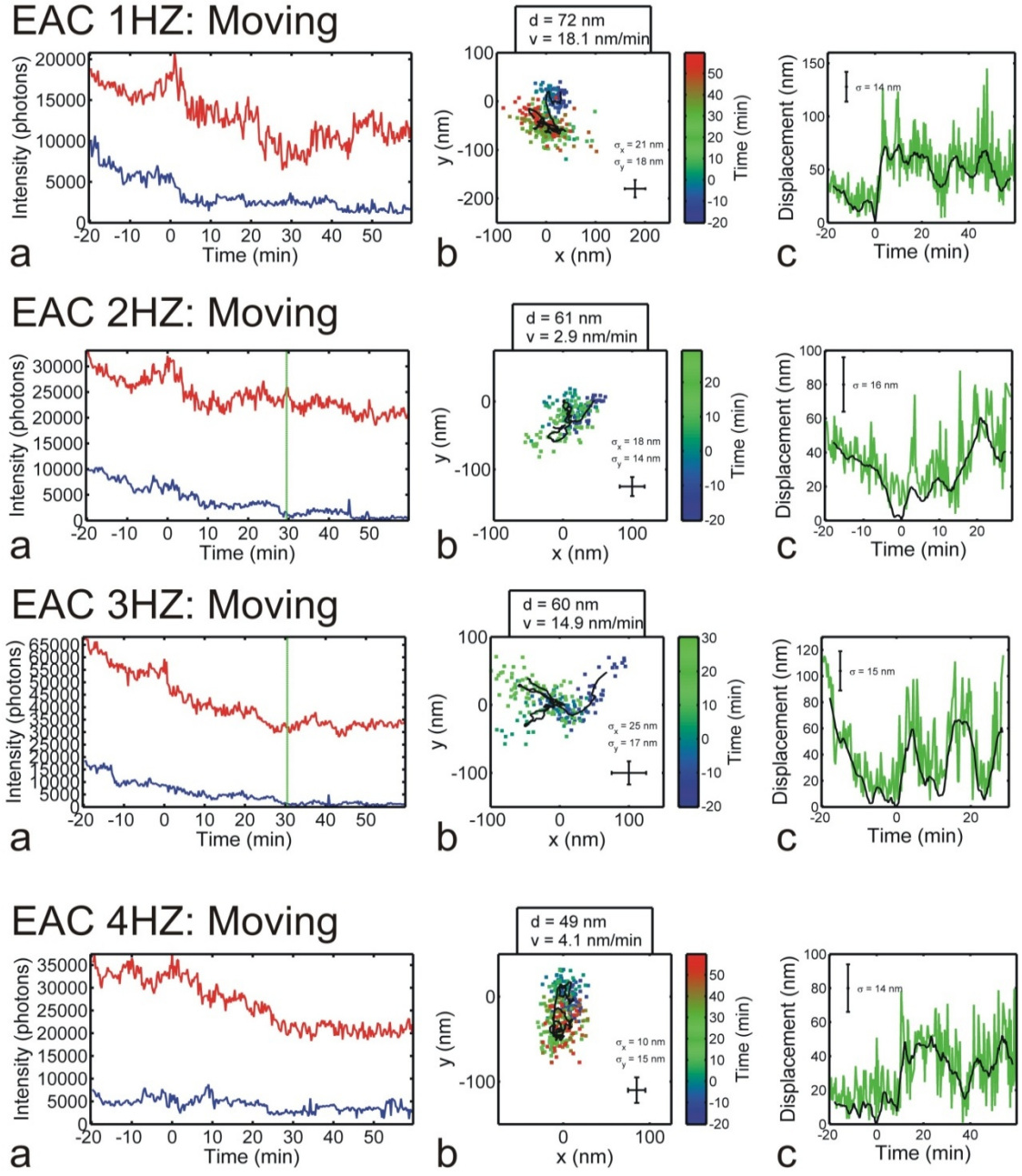
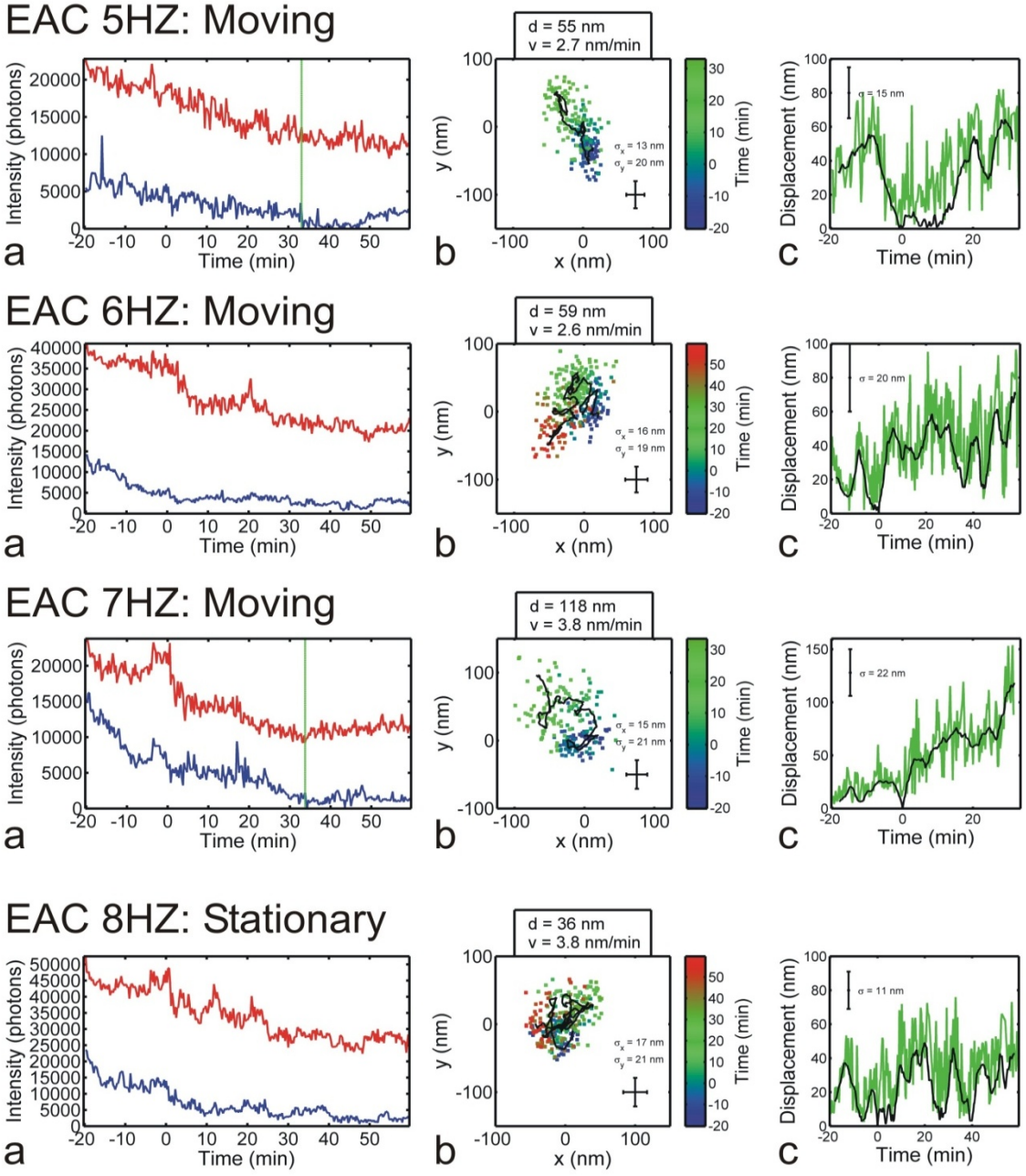
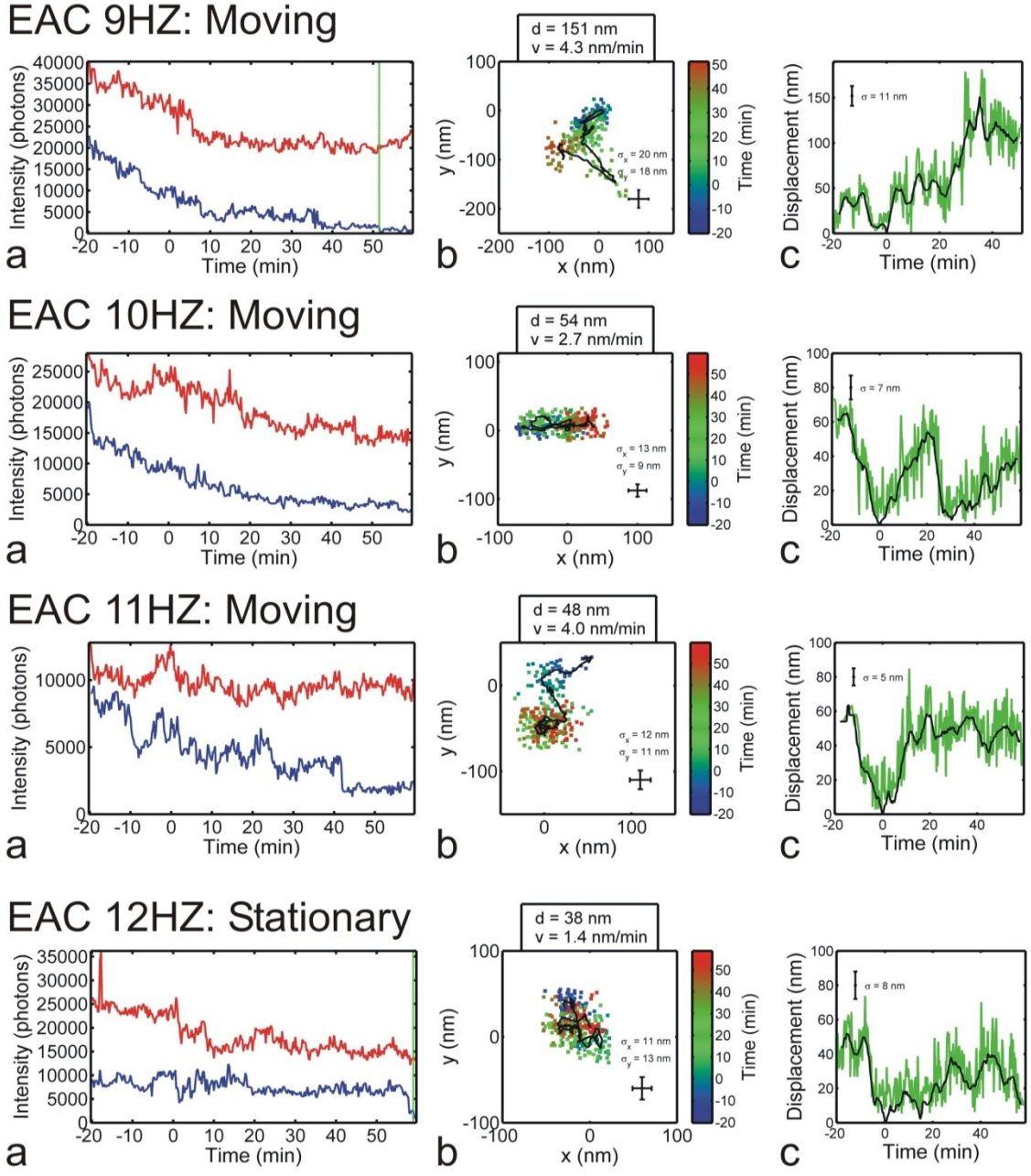

EABD 6
Intensity (photons)
Cy3
Cy5
Time (min)
a
v = 2.3 nm/min
y (nm)
x (nm)
b
Displacement (nm)
σ = 21 nm
Time (min)
c
EABC
EABC 1
Intensity (photons)
Cy3
Cy5
Time (min)
a
v = 1.8 nm/min
y (nm)
x (nm)
b
Displacement (nm)
σ = 13 nm
Time (min)
c
EABC 2
Intensity (photons)
donor
acceptor
Time (min)
a
v = 3.8 nm/min
y (nm)
x (nm)
b
Displacement (nm)
σ = 12 nm
Time (min)
c

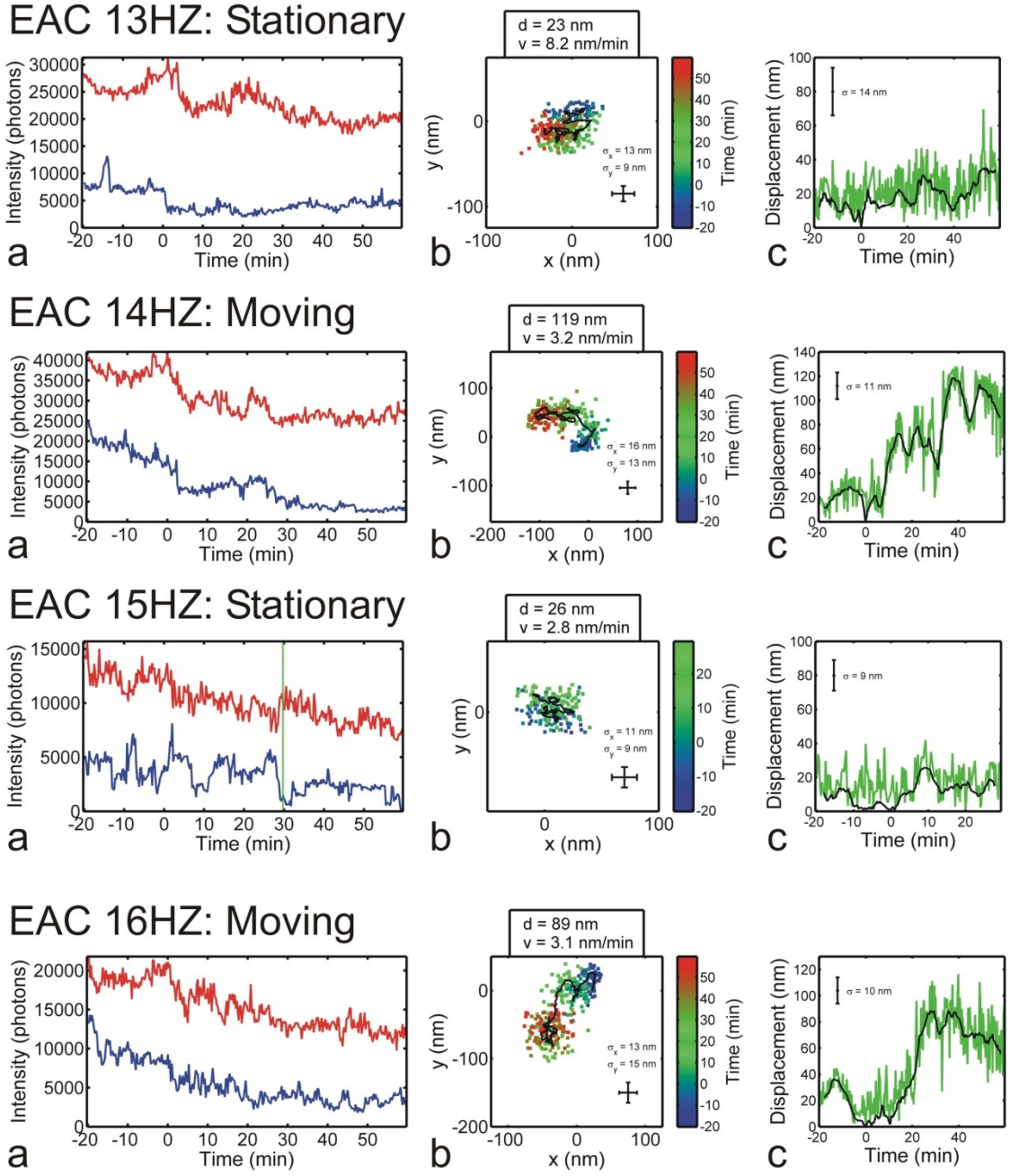
Figure A1.24. Fluorophore emission intensity over time (**a**), (non-averaged) 2-dimensional trajectories of the motion of individual spiders relative to the Cy5 PSF (**b**), and displacement versus time plots of individual spiders (**c**) on the EAC substrate track as imaged by fluorescence microscopy in 1× HBS buffer with 1× HBS buffer containing 5 mM (EAC HZ 1-16) or 0 mM $ZnSO_4$ (EAC H 1-21) added after 20 minutes of imaging. In (**a**), the vertical green line represents the point after which the trace is no longer analyzed due to photobleaching. The black line in (**b**) represents the smoothed trajectory obtained by applying a 16-frame rolling average as described in the text. The origin is chosen to coincide with each spider's coordinates at the time of adding 1× HBS buffer containing 0 or 5 mM $ZnSO_4$. Plots of displacement versus time for the raw trajectory (**c**, green line) and smoothed trajectory (**c,** black line) are also shown for each trace. The addition of 5 mM $ZnSO_4$ occurred at t = 0 min. Also shown in panel b are values of net displacement (d) and mean velocity (v) calculated as described in the fluorescence microscopy analysis section. For comparison between experiments performed in 0 and 5 mM $ZnSO_4$, all stationary spiders (those with net displacements less than or equal to 45 nm after zinc addition) observed under each set of conditions are also shown. In the presence of 5 mM $ZnSO_4$, 12 of 16 trajectories move > 45 nm, while only 3 of 21 trajectories collected in absence of Zn^{2+} ions appear to move > 45 nm.

HBS, 5 mM Zn^{2+} (EAC HZ)

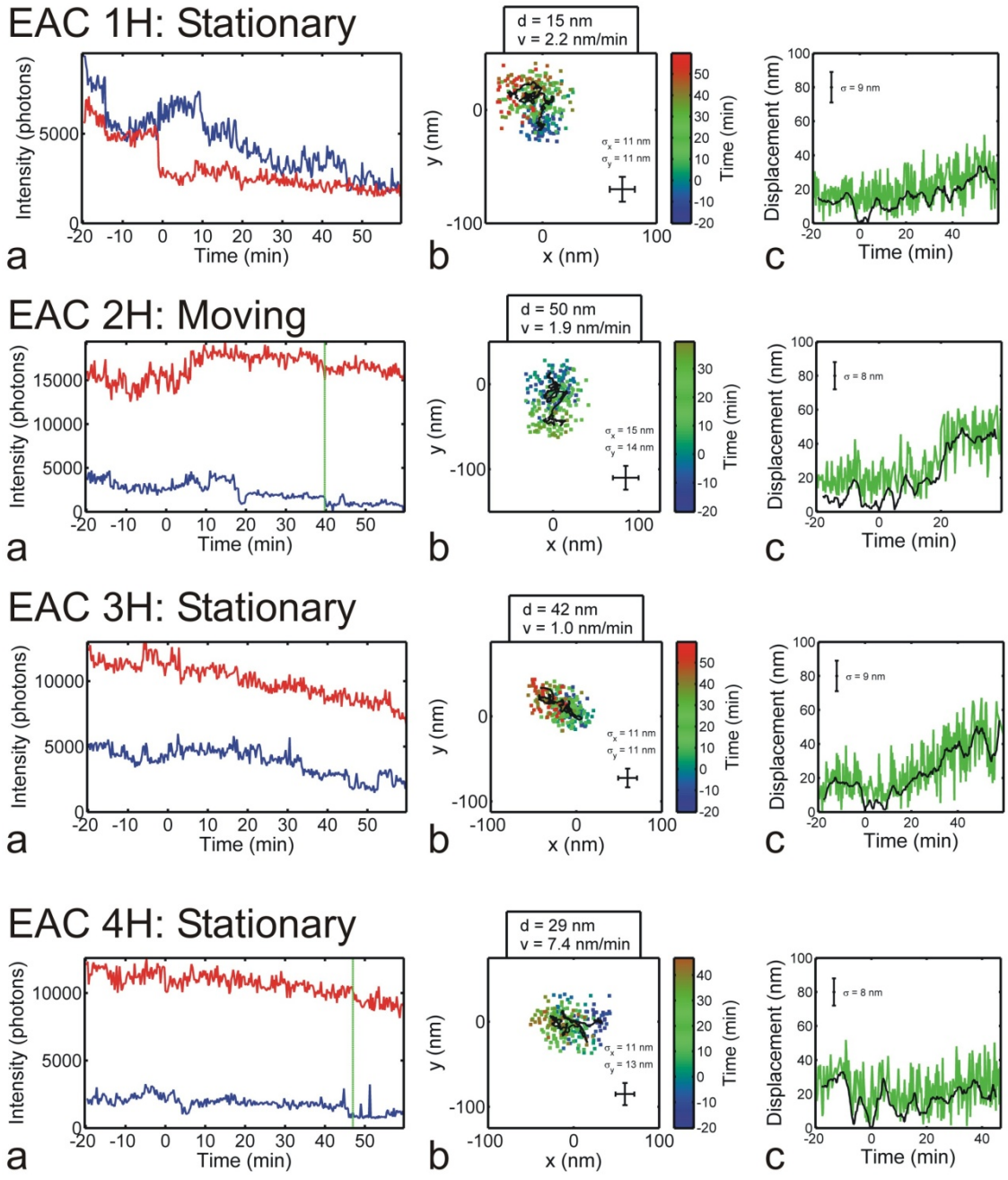


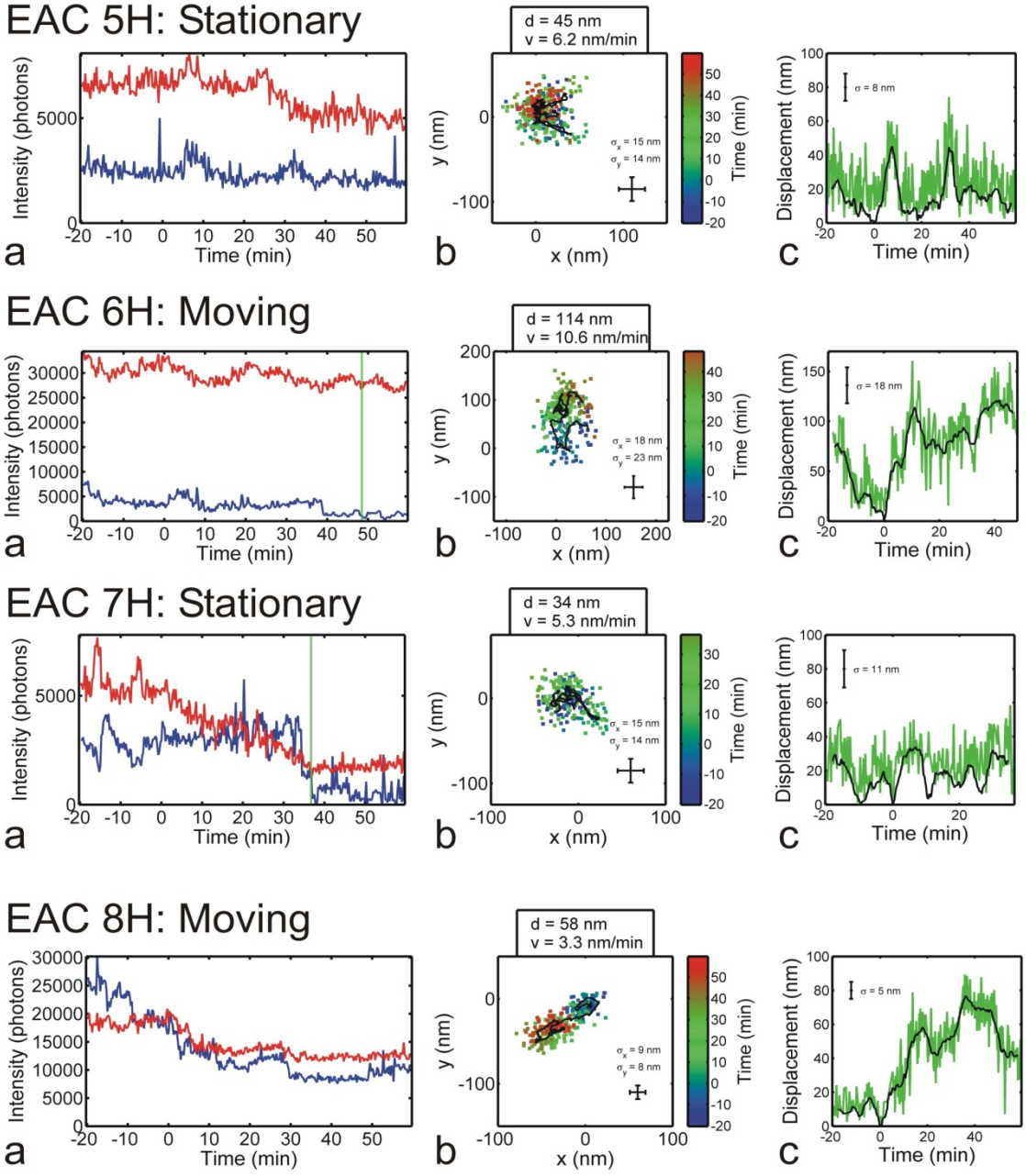

EAC 5HZ: Moving
d = 55 nm
v = 2.7 nm/min
σ = 15 nm
EAC 6HZ: Moving
d = 59 nm
v = 2.6 nm/min
σ = 20 nm
EAC 7HZ: Moving
d = 118 nm
v = 3.8 nm/min
σ = 22 nm
EAC 8HZ: Stationary
d = 36 nm
v = 3.8 nm/min
σ = 11 nm
Intensity (photons)
Time (min)
y (nm)
x (nm)
Time (min)
Displacement (nm)
Time (min)
a
b
c

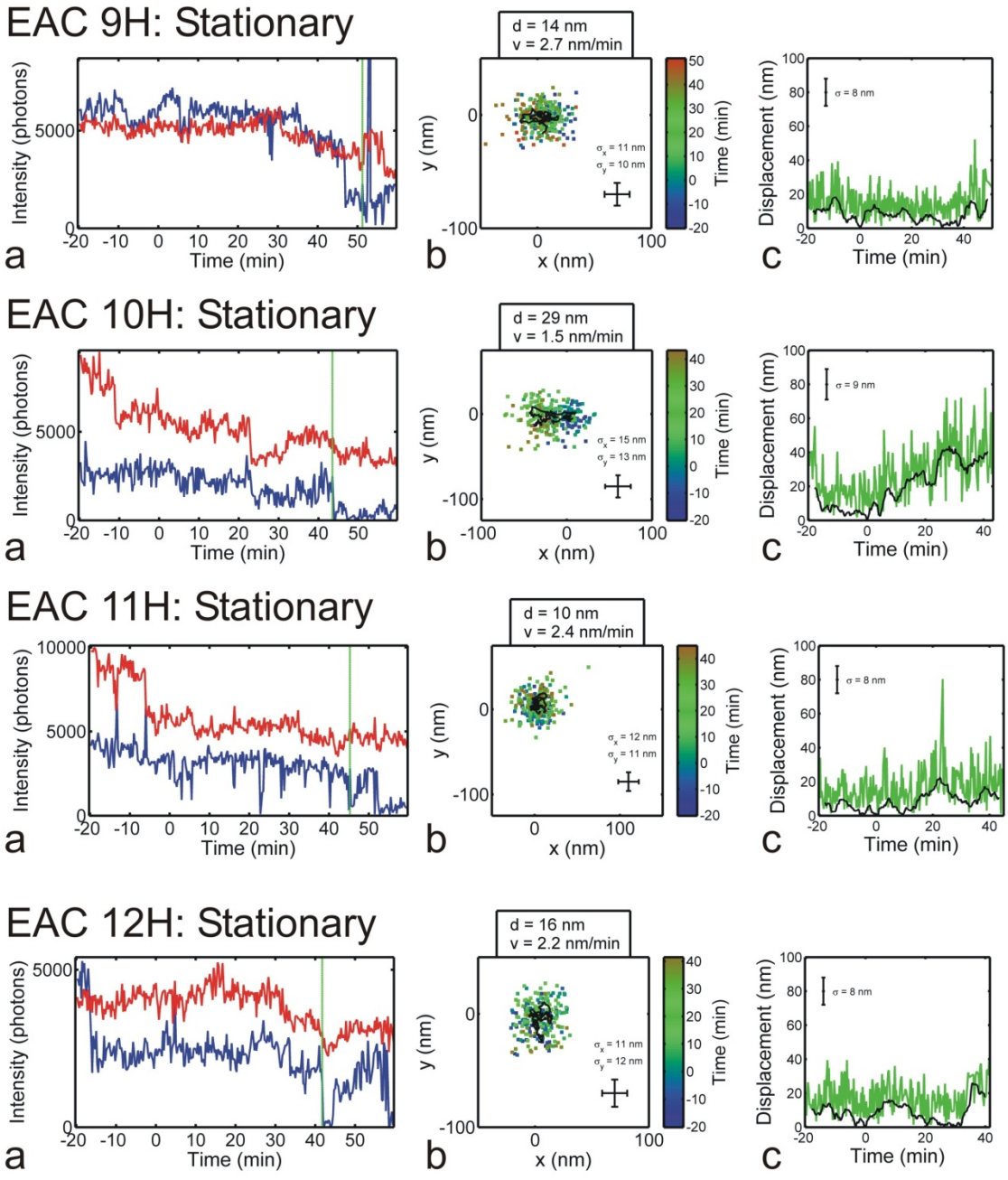

EAC 9HZ: Moving
d = 151 nm
v = 4.3 nm/min
σ = 11 nm
EAC 10HZ: Moving
d = 54 nm
v = 2.7 nm/min
σ = 7 nm
EAC 11HZ: Moving
d = 48 nm
v = 4.0 nm/min
σ = 5 nm
EAC 12HZ: Stationary
d = 38 nm
v = 1.4 nm/min
σ = 8 nm
Intensity (photons)
Time (min)
y (nm)
x (nm)
Time (min)
Displacement (nm)
Time (min)
a
b
c

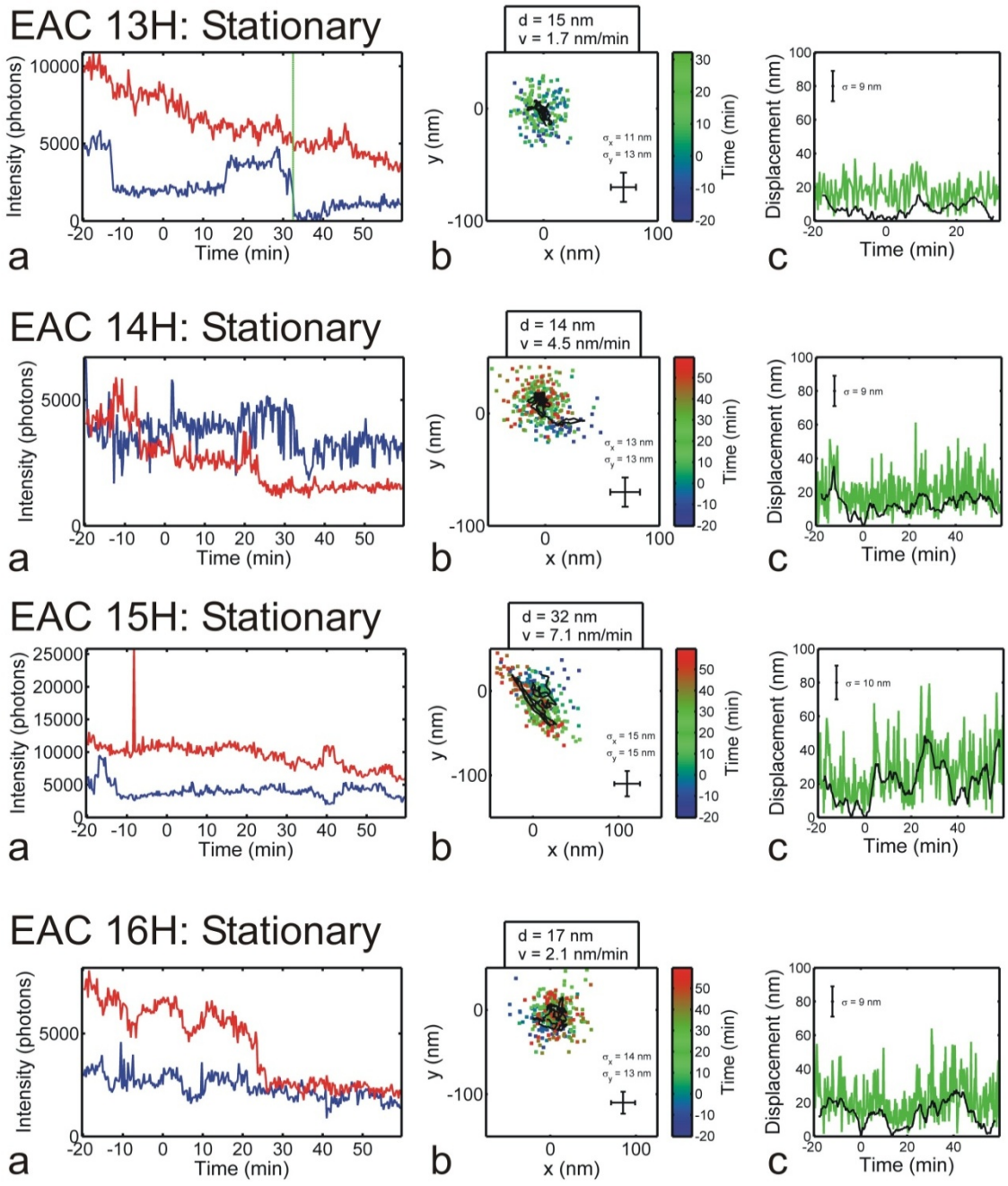

EAC 13HZ: Stationary
Intensity (photons)
Time (min)
a
d = 23 nm
v = 8.2 nm/min
σx = 13 nm
σy = 9 nm
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
σ = 14 nm
Time (min)
c
EAC 14HZ: Moving
Intensity (photons)
Time (min)
a
d = 119 nm
v = 3.2 nm/min
σx = 16 nm
σy = 13 nm
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
σ = 11 nm
Time (min)
c
EAC 15HZ: Stationary
Intensity (photons)
Time (min)
a
d = 26 nm
v = 2.8 nm/min
σx = 11 nm
σy = 9 nm
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
σ = 9 nm
Time (min)
c
EAC 16HZ: Moving
Intensity (photons)
Time (min)
a
d = 89 nm
v = 3.1 nm/min
σx = 13 nm
σy = 15 nm
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
σ = 10 nm
Time (min)
c

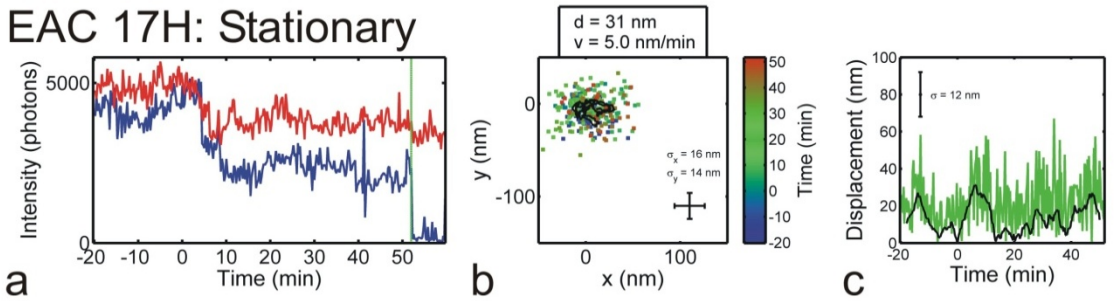
HBS, 0 mM Zn^{2+} Control (EAC H)

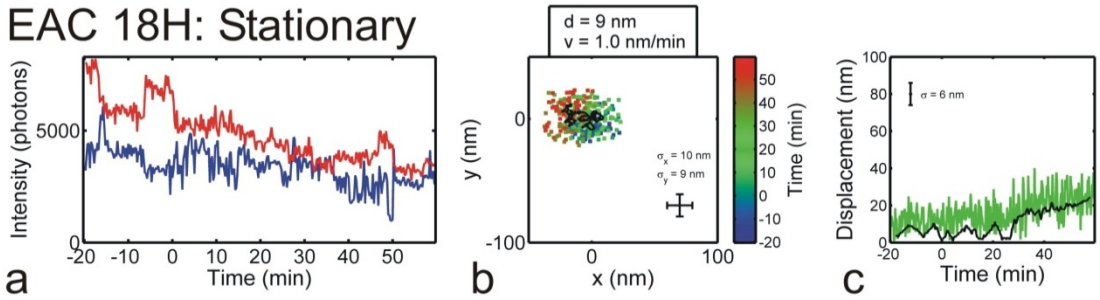


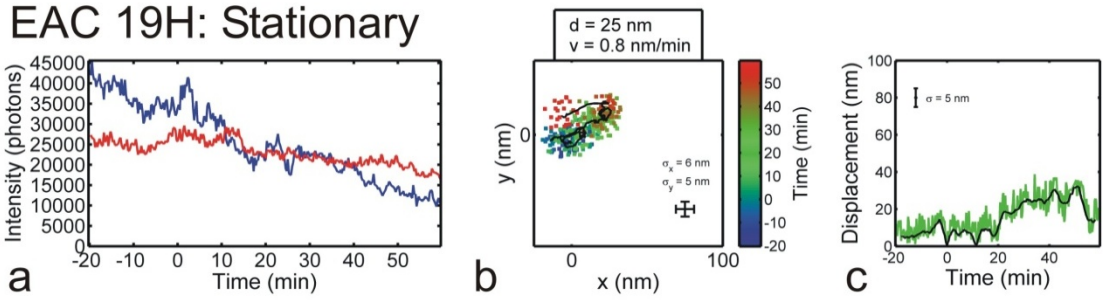

EAC 5H: Stationary
Intensity (photons)
Time (min)
a
d = 45 nm
v = 6.2 nm/min
σx = 15 nm
σy = 14 nm
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
σ = 8 nm
Time (min)
c
EAC 6H: Moving
Intensity (photons)
Time (min)
a
d = 114 nm
v = 10.6 nm/min
σx = 18 nm
σy = 23 nm
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
σ = 18 nm
Time (min)
c
EAC 7H: Stationary
Intensity (photons)
Time (min)
a
d = 34 nm
v = 5.3 nm/min
σx = 15 nm
σy = 14 nm
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
σ = 11 nm
Time (min)
c
EAC 8H: Moving
Intensity (photons)
Time (min)
a
d = 58 nm
v = 3.3 nm/min
σx = 9 nm
σy = 8 nm
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
σ = 5 nm
Time (min)
c

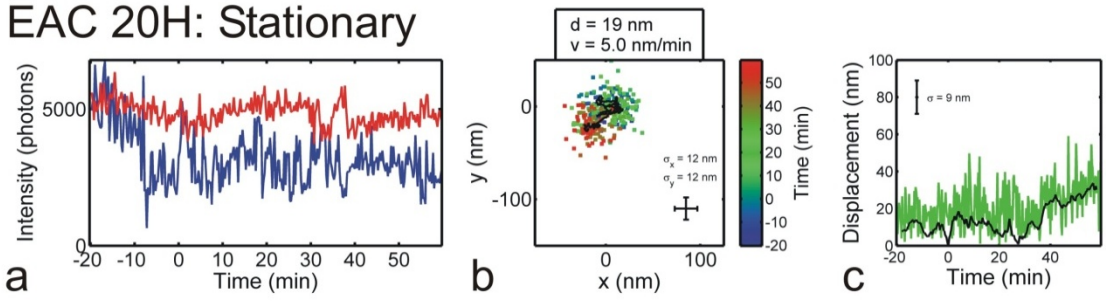

EAC 9H: Stationary
d = 14 nm
v = 2.7 nm/min
σx = 11 nm
σy = 10 nm
σ = 8 nm
EAC 10H: Stationary
d = 29 nm
v = 1.5 nm/min
σx = 15 nm
σy = 13 nm
σ = 9 nm
EAC 11H: Stationary
d = 10 nm
v = 2.4 nm/min
σx = 12 nm
σy = 11 nm
σ = 8 nm
EAC 12H: Stationary
d = 16 nm
v = 2.2 nm/min
σx = 11 nm
σy = 12 nm
σ = 8 nm
Intensity (photons)
Time (min)
y (nm)
x (nm)
Time (min)
Displacement (nm)
Time (min)
a
b
c

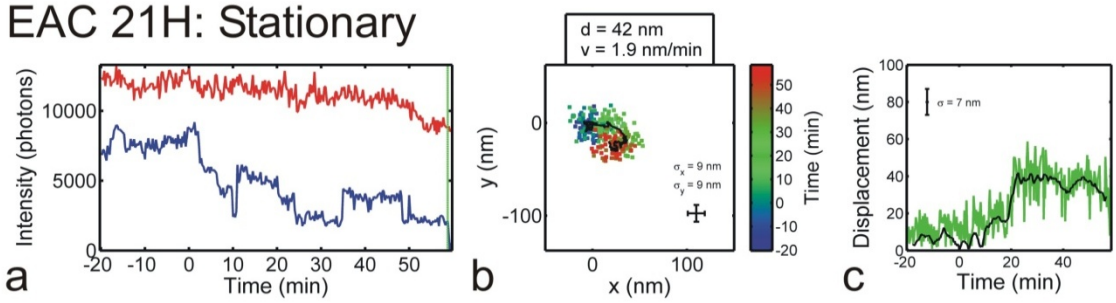

EAC 13H: Stationary
d = 15 nm
v = 1.7 nm/min
σx = 11 nm
σy = 13 nm
σ = 9 nm
EAC 14H: Stationary
d = 14 nm
v = 4.5 nm/min
σx = 13 nm
σy = 13 nm
σ = 9 nm
EAC 15H: Stationary
d = 32 nm
v = 7.1 nm/min
σx = 15 nm
σy = 15 nm
σ = 10 nm
EAC 16H: Stationary
d = 17 nm
v = 2.1 nm/min
σx = 14 nm
σy = 13 nm
σ = 9 nm
Intensity (photons)
Time (min)
y (nm)
x (nm)
Time (min)
Displacement (nm)
Time (min)
a
b
c


EAC 17H: Stationary
Intensity (photons)
Time (min)
a
d = 31 nm
v = 5.0 nm/min
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
Time (min)
c


EAC 18H: Stationary
Intensity (photons)
Time (min)
a
d = 9 nm
v = 1.0 nm/min
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
Time (min)
c


EAC 19H: Stationary
Intensity (photons)
Time (min)
a
d = 25 nm
v = 0.8 nm/min
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
Time (min)
c


EAC 20H: Stationary
Intensity (photons)
Time (min)
a
d = 19 nm
v = 5.0 nm/min
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
Time (min)
c


EAC 21H: Stationary
Intensity (photons)
Time (min)
a
d = 42 nm
v = 1.9 nm/min
y (nm)
x (nm)
Time (min)
b
Displacement (nm)
Time (min)
c

	Total PSF Pair Candidates	PSF Pairs Fit to Gaussians	PSF Pairs with Satisfactory Fitting	Putative Moving Spiders
EAC SSC + Zn^{2+}	3,821	139	83	21
EAC SSC - Zn^{2+}	303	15	7	0
EAC HBS + Zn^{2+}	384	22	15	11
EAC HBS - Zn^{2+}	127	22	21	3
EABD TA-Mg + Zn^{2+}	477	28	11	6
EABC TA-Mg + Zn^{2+}	227	9	7	2

Table A1.3. Trajectory filtering statistics for spiders imaged by fluorescence microscopy on substrate tracks. These statistics reflect the filtering of raw fluorescence microscopy data to yield the spider trajectories shown in Figures A1.23 and A1.24. Total PSF pair candidates (column 1) include PSFs identified as possible signal by an automated image analysis routine. Most of these contain detectible signal from only Cy3, only Cy5, or neither. This is highly variable from experiment to experiment and may depend on how free a particular slide is of fluorescent contaminants. PSF pairs were fit to Gaussians (column 2) if both Cy3 and Cy5 were present for at least 25 minutes with at least 1,000 photon counts per movie frame. PSF pairs with satisfactory fitting (column 3) lacked excessive blinking or interference from other nearby PSFs. Finally, putative moving spiders (column 4) satisfy the selection criteria listed in the Materials and Methods and are also shown in Figures A1.23 and A1.24.

	Total PSF Pair Candidates	Spiders Included in Figure 2.9b
EAC substrate track	3,821	85
EAC product track, Long TRIGGER pre-incubation	276	29
EAC product track, Short TRIGGER pre-incubation	74	18

Table A1.4. Trajectory filtering statistics for spiders imaged by fluorescence microscopy on the EAC track. These statistics reflect the filtering performed to produce the MSD plot in Figure 2.9b. The only criterion these spider-origami pairs needed to satisfy is to have detectable Cy3 and Cy5 for at least 1/3 of the experiment's duration (20-30 min). All are imaged in 1× SSC buffer in the presence of Zn^{2+}. The EAC substrate track was in all cases incubated with TRIGGER for 30-60 min prior to imaging, while TRIGGER was added to the EAC product track either 30-60 min (row 2) or 10-15 min (row 3) prior to imaging by fluorescence microscopy.

APPENDIX 2

EVIDENCE OF MULTIPLE NATIVE STATES IN A VARKUD SATELLITE RIBOZYME[8]

A2.1 Introduction

The Varkud satellite (VS) ribozyme is a catalytic motif embedded in certain satellite RNAs isolated from the mitochondria of *Neurospora*, where it mediates the self-cleavage and ligation reactions thought to be necessary for the replication cycle of the satellite RNA[151]. It is classified as a small nucleolytic RNA, along with the hairpin, hepatitis delta virus (HDV), and hammerhead ribozymes, all of which catalyze a transesterification reaction converting a 3′-5′ phosphodiester into two products bearing a 2′-3′-cyclic phosphate and a 5′-terminal hydroxyl group, respectivley[152]. While the VS ribozyme motif has not yet been found in other organisms, the functionally analogous HDV and hammerhead ribozymes have recently been discovered in intergenic regions of organisms from diverse kingdoms of life[153,154], providing further motivation to understand the biophysical and biochemical principles that govern the function of small nucleolytic RNAs.

The VS ribozyme, at approximately 160 nucleotides in length, is the largest motif yet discovered in its class, and is the only known small nucleolytic ribozyme for which no crystal structure has yet been reported in the literature. However, biochemical[155–157], steady-state FRET[158], and small-angle X-ray scattering[159] studies have given rise to a consistent and increasingly detailed understanding of the secondary and tertiary structure of this RNA. Furthermore, recent single-molecule FRET (smFRET) studies showed the VS ribozyme to exhibit dynamic, hierarchical three-state folding into its catalytically active conformation[12].

[8] Alexander Johnson-Buck performed all smFRET assays and docking thermodynamics analysis, and the activity assay at 10 mM Mg^{2+}. Miguel J. Pereira performed the cleavage assay at 200 mM Mg^{2+}. Richard A. Collins, Shawna Hiley, and Dominic Jaikaran designed the FR3 VS ribozyme, and Shawna Hiley and Dominic Jaikaran synthesized it.

Here, we use smFRET[13] to investigate a different variant of the VS ribozyme called FR3. This variant is derived from the fast-cleaving RS19 family of VS ribozymes[160], in which stem-loop I, which contains the site of cleavage, is attached via a linker of arbitrary sequence to the 3′ end of the ribozyme core (Figure A2.1a). We show that the FR3 ribozyme exhibits similar conformational dynamics as the variant previously characterized by smFRET, but with only two states instead of three. We also show that individual VS ribozymes show patterns of conformational transitions that persist for several times longer than the timescale of catalysis, yet slowly interconvert, suggesting the side-by-side existence of multiple catalytically competent native states.

A2.2 Materials and Methods

Preparation of FR3 RNA. Synthesis of FR3 was accomplished by VS ribozyme-mediated ligation using a strategy similar to that used previously to incorporate a site-specific 4-thio-uridine nucleotide into stem-loop I[161]. The ribozyme portion was obtained by in vitro transcription by T7 RNA polymerase of a linearized plasmid template, followed by self-cleavage and gel-purification of the upstream cleavage product which ends at G620 and contains a 2′3′ cyclic phosphate terminus; the transcription mixture included 4 mM ApG dinucleotide in which the adenosine contained an amino group at the end of a six-carbon linker attached to the 5′ phosphate (Dharmacon, Inc) to allow for subsequent labeling of the 5′ end of the RNA with Cy5 mono-reactive dye (GE Healthcare). A second RNA beginning at position 621 (with a 5′ hydroxyl) was chemically-synthesized (Dharmacon, Inc.) and contains a 5-amino-allyl-uridine for subsequent labeling with Cy3 mono-reactive dye (GE Healthcare) at the position indicated in Figure A2.1a and a 3′ terminal biotin. Incubation of these two Cy-labeled RNAs in the presence of 200 mM Mg^{2+} results in ligation to form the full-length FR3 RNA, which was gel-purified and ethanol-precipitated.

Activity Assay of FR3: A 150 µl solution containing 45 picomoles of FR3 in 20 mM HEPES-KOH pH 7.4 + 100 mM KCl was heated to 70°C for 2 min, then cooled to room temperature for 5 min. Then 25 µl of it were removed into 25 µl stop mix (formamide + 200 mM EDTA). The remainder was added to 25 µl of a $MgCl_2$ solution to achieve a 200

mM final Mg^{2+} concentration. At time points of 15 sec, 30 sec, 1 min, 2 min, 5 min, and 10 min, 25 µl of the reaction mix were removed into 25 µl of stop mix. The cleaved and uncleaved material were separated on a 10% denaturing urea-PAGE gel and the intensity of Cy5 fluorescence on the gel was quantified for each band and analyzed using the Amersham Biosciences Typhoon 9410 Variable Mode Imager instrument and ImageQuant 5.2 analysis software.

Single-molecule FRET measurements. All single-molecule experiments and preparation steps were carried out in the presence of imaging buffer ≡ 20 mM HEPES-KOH, pH 7.4 and 100 mM KCl. FR3 molecules were diluted to 50 pM in Imaging Buffer and heated to 70°C for 2 min before cooling to room temperature over 5 min. Molecules were immobilized on a biotinylated BSA/streptavidin-coated microscope slide as described for DNA origami in Chapter 4.

Single-molecule imaging was performed on the prism-type TIRF microscope described in Chapters 2-4, with excitation at 532 nm (9 W/cm^2). At the beginning and end of each experiment, the sample was illuminated briefly at 640 nm (8 W/cm^2) to verify the presence of the acceptor fluorophore, Cy5. Only molecules with an active acceptor were analyzed further. All smFRET measurements were performed in the presence of imaging buffer containing an enzymatic oxygen scavenging system[100] and 35 mM $MgCl_2$. In some experiments, the same molecules were observed before and after periods without excitation. Single-molecule FRET-versus-time trajectories were generated using the formula $FRET = I_{Cy5}/(I_{Cy3} + I_{Cy5})$ as described[12].

Analysis of FR3 docking energetics. For each molecule observable for at least 100 s without either fluorophore photobleaching, the apparent equilibrium constant of undocking was calculated as $K_{undock} = t_{undocked}/t_{docked}$, where $t_{undocked}$ and t_{docked} are the time the molecule exhibits FRET values less than and greater than 0.7, respectively. The apparent free energy of undocking was calculated as $\Delta G_{undock} = -RT\ ln\ K_{undock}$.

A2.3 Results

The FR3 VS ribozyme is based on a fast-cleaving family of VS ribozymes known as RS19 that are related to the canonical sequence *via* circular permutation[160]. It is most closely related to RS19ΔL, which bears an extended linker between stem-loop I (SLI) and the rest of the ribozyme, as well the canonical sequence of the substrate internal loop in SLI that is ligated about ten times more rapidly than it is cleaved (Figure A2.1a)[160]. Its mode of synthesis, which involves the self-ligation of pre-cleaved VS ribozymes to a synthetic 3′ segment followed by PAGE purification, ensures a relatively homogeneous population of molecules that all possess the same, catalytically competent sequence. Indeed, despite the fluorophore modifications, the activity of FR3 is very similar to RS19ΔL in 200 mM Mg^{2+}, reaching the cleavage-ligation equilibrium in < 1 min (k_{obs} = 3.3 min^{-1}, Figure A2.1b). Even in the presence of the lower concentration of 10 mM Mg^{2+}, the observed rate constant is 1.1 min^{-1}. However, since ligation is favored, and since even the cleaved form of SLI can form several base pairs with the rest of the ribozyme, including the 3-bp tertiary “kissing” interaction between the terminal loops of stems I and V (Figure A2.1a), the substrate-ribozyme complex is expected to remain intact for several minutes to hours of observation, even at the low concentrations used in single molecule experiments.

As previously reported for a different VS ribozyme[12], FR3 exhibits transitions between at least two resolvable FRET states. These transitions report on docking of SLI into the active site, as docking juxtaposes the Cy3 label on SLI with the Cy5 label at the 5′-end of stem II. Strikingly, individual FR3 molecules exhibit drastically different conformational behaviors. For instance, while some molecules are very dynamic, sampling the FRET ≈ 0.64 undocked state as much as, or more than, the FRET ≈ 0.82 docked state (Figure A2.2a), other molecules undock only rarely (Figure A2.2b). A histogram of FRET values for many molecules reveals two apparent distributions (Figure A2.2c), as expected based on individual FRET trajectories. The observation of a highly populated high-FRET (docked) state is consistent with the faster rate constant of cleavage of FR3 compared to the variant previously studied by smFRET[12].

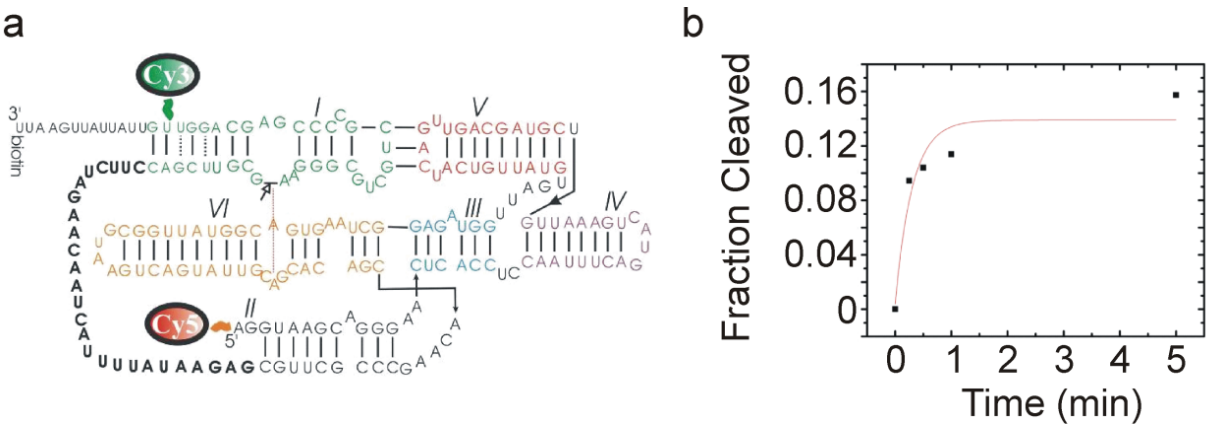


Figure A2.1. (a) Structure of the FR3 VS ribozyme[12], a fluorescently labeled RNA based on the RS19ΔL variant[160]. The open arrow denotes the site of self-cleavage. (b) Results of a urea-PAGE assay of FR3 self-cleavage in the presence of 200 mM Mg^{2+}. Fitting to a single exponential (red curve) yields an apparent rate constant of 3.3 min^{-1} and a final fraction cleaved of 0.139, compared to 1.9 min^{-1} and 0.12 for RS19ΔL[160].

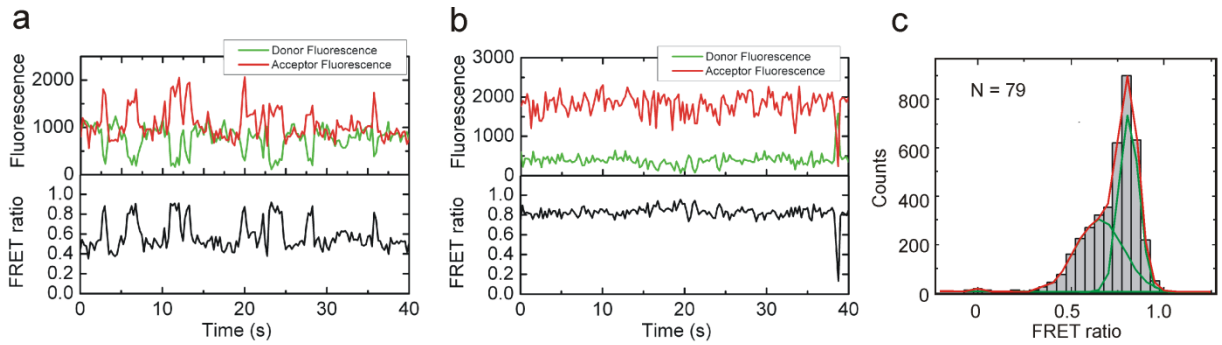


Figure A2.2. Heterogeneity of conformational dynamics in FR3 ribozymes as detected by smFRET. (a),(b) smFRET traces of representative single molecules exhibiting fast (a) and slow (b) conformational dynamics. (c) Histogram of FRET ratios for 79 FR3 molecules. Modeling the distribution with three Gaussian functions (green curves) yields mean FRET values of 0.00 (2%), 0.64 (26%), and 0.82 (72%).

To further characterize the heterogeneity of docking behavior between different FR3 molecules, we calculated the apparent Gibbs free energy of undocking, ΔG_{undock}. Individual molecules display a wide range of ΔG_{undock} values, spanning from approximately -10 to 10 kJ/mol (Figure A2.3a). The majority of molecules have ΔG_{undock} near zero, with a slight skew towards positive values (mean = 0.7 kJ/mol), consistent with the high population of the docked state. Furthermore, the ΔG_{undock} of an individual molecule is relatively stable over several minutes, as is attested by the positive correlation (R = 0.83) between the calculated free energy before and after a 5-min dark period (Figure A2.3b). Intriguingly, a small fraction (~5%) of molecules do change their apparent docking energetics over a 5-minute window (Figure A2.3b,c), suggesting that a reversible conformational or chemical change is responsible for at least some of the diversity in docking energetics.

A2.4 Discussion

Like the hairpin, VS, and *Tetrahymena* group I ribozymes previously studied by smFRET[11,12,15,16], FR3 undergoes global conformational changes on a timescale of seconds to minutes, with dramatic variation between individual molecules in the stability of the docked, catalytically active conformation. Furthermore, single FR3 molecules persist in their docking behavior for at least 5 minute in general, which is far longer than the time required to reach the cleavage/ligation equilibrium. Only ~5% of FR3 molecules are observed to change their docking behavior over 5 min (Figure A2.3b), whereas the similarity between the bulk kinetic assays of FR3 (Figure A2.1b) and RS19ΔL are consistent with the majority of the ribozymes being catalytically active. Together, these data suggest that all or most of the diverse patterns of smFRET behavior represent catalytically active FR3 molecules, recalling the multiple native states reported recently for the *Tetrahymena* group I ribozyme[16]. We therefore propose a tentative model in which there exist side-by-side two or more native populations of FR3 molecules with different docking thermodynamics (Figure A2.4). The populations interconvert, but on a timescale slower than catalysis, so that they effectively constitute isolated populations of natively folded ribozymes. This implies that there is a steep energetic barrier separating

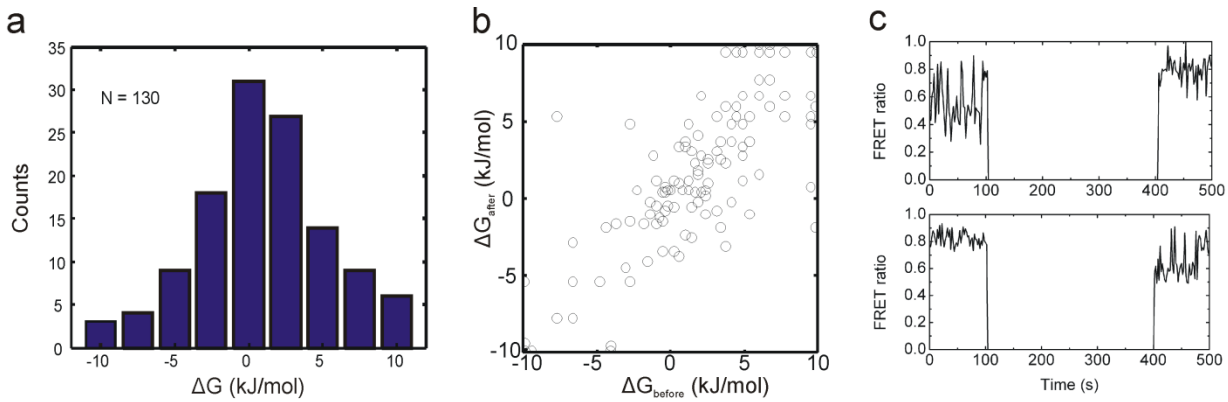


Figure A2.3. Free energy of undocking for FR3 molecules. (a) Distribution of ΔG_{undock} for 130 molecules. Molecules with $\Delta G \leq -10$ or $\Delta G \geq 10$ are combined into the leftmost and rightmost bins, respectively. (b) Free energy of undocking of the molecules in panel (a) before and after a 5-min dark period. Correlation coefficient R = 0.83. Only 6 molecules (4.6%) exhibit significant changes in ΔG_{undock} in this period. (c) Examples of individual molecules that switch between dynamic and non-dynamic behavior. From t = 100-400s, the laser shutter is closed to avoid photobleaching of Cy3 and Cy5.

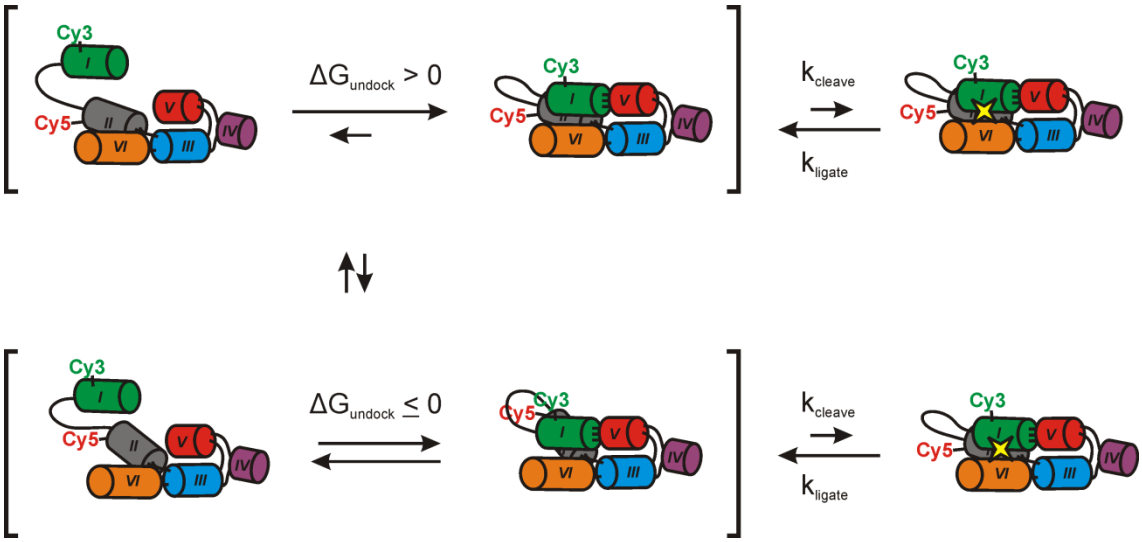


Figure A2.4. Proposed model of heterogeneous docking behavior in the FR3 VS ribozyme. Different populations (brackets, top and bottom) exhibit distinct kinetics of docking and undocking of stem-loop I into the catalytically active conformation. The populations can interconvert only slowly, but both can cleave with a collective observed rate constant of ~1 min^{-1}.

the different populations of FR3 molecules, but that it is not a permanent chemical modification such as UV- or heat-induced damage that can also cause heterogeneous folding. Rather, it is likely to result from steep conformational barriers that are very common in RNA, yet whose molecular basis remains elusive[7].

There are some caveats to consider when interpreting these results. First, while it is likely, based on the activity comparison with RS19ΔL, that the majority of FR3 molecules are active, we have not yet been able to confirm that this is the case for molecules at the surface of a microscope slide. The FR3 sequence was designed with prolonged single-molecule observation in mind, with extensive interactions between the 3'- and 5'-segments of SLI, and strongly favors ligation. As such, it is difficult to directly determine which individual molecules are active, for example, by disappearance of Cy5 from the surface upon cleavage. While we have ventured to design single-molecule activity assays, the conditions needed to reliably disrupt the interactions linking the cleaved product to the ribozyme – conditions such as 60-80% formamide or 6 M urea – are also prone to strip molecules from the surface of the microscope slide.

Nevertheless, these results are strongly suggestive of multiple native states in the FR3 variant of the VS ribozyme. Interestingly, a procedure involving co-transcriptional folding and native purification of a VS ribozyme has been reported to remove severe heterogeneity observed as smearing in a nondenaturing polyacrylamide gel[41]. Since FR3 was purified by denaturing PAGE and subsequently refolded by heating, it is likely this produced other kinetically trapped folds that may not have existed in a co-transcriptionally folded ribozyme. It remains to be seen whether co-transcriptionally folding the FR3 ribozyme would also reduce the variety of conformational behaviors observed.

APPENDIX 3

NO EVIDENCE OF PRE-CLEAVAGE DYNAMICS IN THE 5′-FLANKING REGION OF A HEPATITIS DELTA VIRUS RIBOZYME[9]

Of the known naturally occurring RNA enzymes (ribozymes), the hepatitis delta virus ribozyme (HDVr) is the only one found within a human pathogen[162]. This ~85-nt self-cleaving RNA motif plays a critical role in replication of the hepatitis delta virus, and structurally homologous motifs have been discovered in diverse phyla of the kingdom *Animalia*[153,163]. HDVr has a complex secondary and tertiary structure comprising a double-nested pseudoknot wrapped into two adjacent helical stacks with multiple helical crossovers (Figure A3.1a)[164–167]. Cytosine 75 (C75) has been implicated as the general acid or base that catalyzes the cleavage immediately upstream of G1 within the ribozyme core[164,167–169]. Using ensemble fluorescence measurements, our group has discovered simultaneous global[170] and local[171] conformational changes that accompany catalysis in a *trans*-cleaving HDV ribozyme, suggesting a close relationship between dynamics and catalytic function. Furthermore, a sharply kinked uridine turn, or U-turn, motif was found within the 5'-sequence flanking the cleavage site which appears to position the scissile phosphate for cleavage within the active site[172].

To investigate experimentally whether pre-cleavage dynamics in the 5'-flanking sequence play a role in catalysis, we designed a version of HDVr bearing the following modifications: (1) a 2′-*O*-methyl modification at the U-1 position to prevent cleavage; (2) a donor fluorophore, Cy3, at the end of the P2 helix; and (3) an acceptor fluorophore, Cy5, at the fourth nucleotide upstream of the cleavage site, attached via conjugation to a synthetic 5-aminoallyl uridine. The HDVr molecule was synthesized via a splinted ligation strategy using T4 DNA ligase[173] starting with two synthetic RNAs (purchased

[9] The HDVr construct for smFRET was based on a design by Chamaree de Silva. It was synthesized by Alexander Johnson-Buck and Kamali Sripathi. Single-molecule experiments were carried out by Alexander Johnson-Buck.

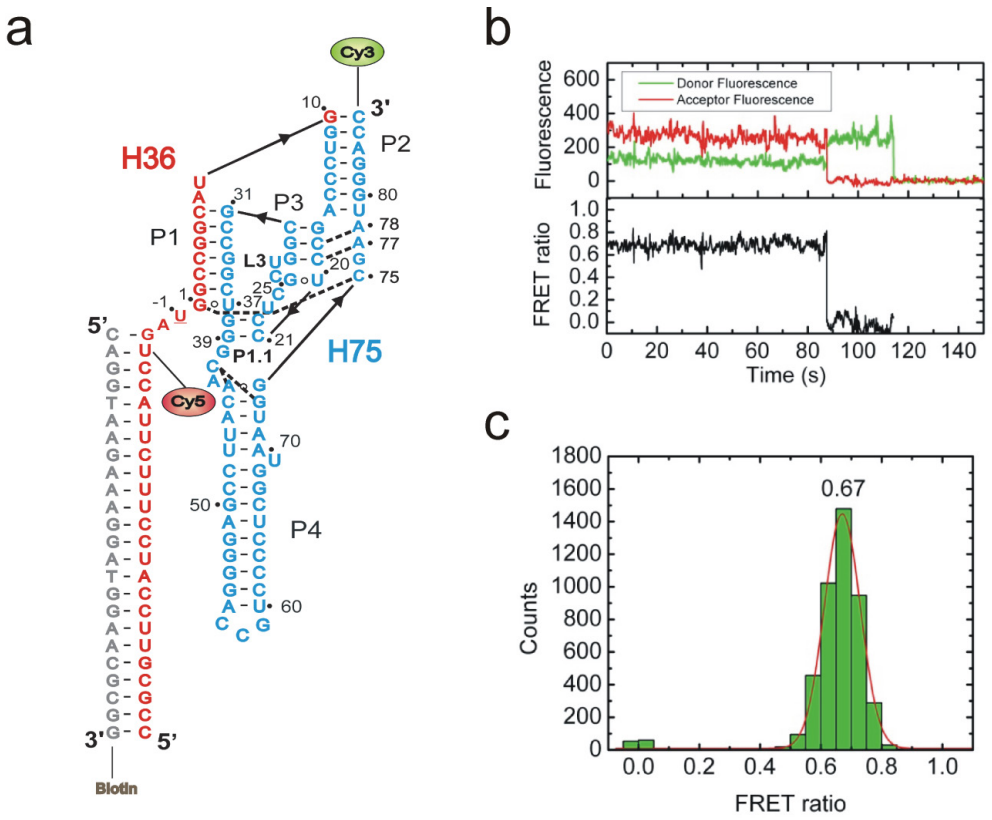


Figure A3.1. (a) Construct of HDVr for smFRET studies. The strands H36 (red) and H75 (blue) were ligated to form a single-stranded 111-mer ribozyme that was surface-immobilized via hybridization to a biotinylated capture strand (gray). U-1 (underlined) was modified with a 2′-*O*-methyl group to prevent cleavage during single-molecule measurements. (b) Representative FRET trace of a single HDVr molecule. Behavior was the same in 0 or 10 mM $MgCl_2$. (c) FRET histogram of 11 molecules in 10 mM $MgCl_2$.

from Dharmacon): a 36-nt 5′ segment (H36) bearing a 5-aminoallyl uridine for labeling with Cy5 monoreactive NHS ester (GE Healthcare), and a 75-nt 3′ segment (H75) bearing a 3′-amine modification for labeling with Cy3 monoreactive NHS ester (GE Healthcare). H36 and H75 were labeled according to the protocol in Chapter 4, ethanol precipitated, and purified by denaturing PAGE prior to ligation. After ligation, the full-length ribozyme was again purified by denaturing PAGE. A biotinylated DNA tether for surface immobilization (Figure A3.1a) was ordered from integrated DNA technologies and used as supplied.

The ligated HDV ribozyme was annealed with a 10-fold excess of the DNA tether in standard buffer (50 mM Tris-HCl pH 7.5, 100 mM MgCl2) at 90 °C for 2 min, then cooled to room temperature for 5 minutes. The HDVr solution was diluted to 50 pM and immobilized on a biotinylated BSA/streptavidin coated microscope slide. Single-molecule FRET was carried out according to the imaging protocol in Appendix 2.

HDVr molecules exhibited a stable FRET ratio of 0.67, with few or no transitions (Figure A3.1b, c). Occasional excursions to a FRET value of 0 occurred, but these were rare and could not be distinguished from blinking of Cy5[174]. This suggests that either (1) the HDV ribozyme does not exhibit conformational dynamics measurable by smFRET along the Cy3-Cy5 axis chosen here, or (2) such dynamics are faster than our time resolution of 10 Hz. Although a negative result, this study serves as a control for other single-molecule studies of ribozymes in which dynamic and heterogeneous transitions between FRET states are observed[12,14,15] (see also Appendix 2).

REFERENCES

1. Watson, J. D. & Crick, F. H. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
2. Palade, G. E. A Small Particulate Component of the Cytoplasm. *J. Biophys. Biochem. Cytol.* **1**, 59–68 (1955).
3. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science* **289**, 905–920 (2000).
4. Schluenzen, F. *et al.* Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell* **102**, 615–623 (2000).
5. Cech, T. R. Structural Biology. The Ribosome Is a Ribozyme. *Science* **289**, 878–879 (2000).
6. Munro, J. B., Sanbonmatsu, K. Y., Spahn, C. M. T. & Blanchard, S. C. Navigating the Ribosome's Metastable Energy Landscape. *Trends in Biochemical Sciences* **34**, 390–400 (2009).
7. Marek, M. S., Johnson-Buck, A. & Walter, N. G. The Shape-shifting Quasispecies of RNA: One Sequence, Many Functional Folds. *Phys. Chem. Chem. Phys.* **13**, 11524 (2011).
8. Lloyd, S. Quantum Thermodynamics: Excuse Our Ignorance. *Nat. Phys.* **2**, 727–728 (2006).
9. Cleland, W. W. Enzyme Kinetics. *Annu. Rev. Biochem.* **36**, 77–112 (1967).
10. Xie, X. S. & Lu, H. P. Single-Molecule Enzymology. *J. Biol. Chem.* **274**, 15967–15970 (1999).
11. Zhuang, X. *et al.* A Single-Molecule Study of RNA Catalysis and Folding. *Science* **288**, 2048–2051 (2000).
12. Pereira, M. J. B. *et al.* Single VS Ribozyme Molecules Reveal Dynamic and Hierarchical Folding Toward Catalysis. *J. Mol. Biol.* **382**, 496–509 (2008).
13. Johnson-Buck, A., Walter, N. G. & Blanco, M. R. Single-Molecule Fluorescence Resonance Energy Transfer. *Encyclopedia of Biophysics* **1**, (2012).
14. Zhuang, X. *et al.* Correlating Structural Dynamics and Function in Single Ribozyme Molecules. *Science* **296**, 1473–1476 (2002).
15. Ditzler, M. A., Rueda, D., Mo, J., Håkansson, K. & Walter, N. G. A Rugged Free Energy Landscape Separates Multiple Functional RNA Folds Throughout Denaturation. *Nucleic Acids Res.* **36**, 7088–7099 (2008).
16. Solomatin, S. V., Greenfeld, M., Chu, S. & Herschlag, D. Multiple Native States Reveal Persistent Ruggedness of an RNA Folding Landscape. *Nature* **463**, 681–684 (2010).

17. Walstrum, S. A. & Uhlenbeck, O. C. The Self-splicing RNA of Tetrahymena Is Trapped in a Less Active Conformation by Gel Purification. *Biochemistry* **29**, 10573–10576 (1990).
18. Esteban, J. A., Banerjee, A. R. & Burke, J. M. Kinetic Mechanism of the Hairpin Ribozyme: Identification and Characterization of Two Nonexchangeable Conformations. *J. Biol. Chem.* **272**, 13629–13639 (1997).
19. Pan, T. & Sosnick, T. R. Intermediates and Kinetic Traps in the Folding of a Large Ribozyme Revealed by Circular Dichroism and UV Absorbance Spectroscopies and Catalytic Activity. *Nat. Struct. Mol. Biol.* **4**, 931–938 (1997).
20. Treiber, D. K., Rook, M. S., Zarrinkar, P. P. & Williamson, J. R. Kinetic Intermediates Trapped by Native Interactions in RNA Folding. *Science* **279**, 1943–1946 (1998).
21. Chadalavada, D. M., Knudsen, S. M., Nakano, S. & Bevilacqua, P. C. A Role for Upstream RNA Structure in Facilitating the Catalytic Fold of the Genomic Hepatitis Delta Virus Ribozyme. *J. Mol. Biol.* **301**, 349–367 (2000).
22. Huang, Z. *et al.* One RNA Aptamer Sequence, Two Structures: A Collaborating Pair That Inhibits AMPA Receptors. *Nucleic Acids Res.* **37**, 4022–4032 (2009).
23. Choi, P. J., Cai, L., Frieda, K. & Xie, X. S. A Stochastic Single-Molecule Event Triggers Phenotype Switching of a Bacterial Cell. *Science* **322**, 442–446 (2008).
24. Wolffe, A. P. & Matzke, M. A. Epigenetics: Regulation Through Repression. *Science* **286**, 481–486 (1999).
25. Hannon, G. J. RNA Interference. *Nature* **418**, 244–251 (2002).
26. Nahvi, A. *et al.* Genetic Control by a Metabolite Binding mRNA. *Chemistry & Biology* **9**, 1043–1049 (2002).
27. Mironov, A. S. *et al.* Sensing Small Molecules by Nascent RNA: A Mechanism to Control Transcription in Bacteria. *Cell* **111**, 747–756 (2002).
28. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine Derivatives Bind Messenger RNAs Directly to Regulate Bacterial Gene Expression. *Nature* **419**, 952–956 (2002).
29. Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. An mRNA Structure That Controls Gene Expression by Binding FMN. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15908–15913 (2002).
30. Graveley, B. R. Alternative Splicing: Increasing Diversity in the Proteomic World. *Trends Genet.* **17**, 100–107 (2001).
31. Blencowe, B. J. Alternative Splicing: New Insights from Global Analyses. *Cell* **126**, 37–47 (2006).
32. Cai, L., Friedman, N. & Xie, X. S. Stochastic Protein Expression in Individual Cells at the Single Molecule Level. *Nature* **440**, 358–362 (2006).
33. Taniguchi, Y. *et al.* Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **329**, 533–538 (2010).
34. Reed, D. H. & Frankham, R. Correlation between Fitness and Genetic Diversity. *Conservation Biology* **17**, 230–237 (2003).
35. Allison, A. C. Malaria in Carriers of the Sickle-cell Trait and in Newborn Children. *Exp. Parasitol.* **6**, 418–447 (1957).
36. Uenoyama, M. *Pesticide Resistance: Strategies and Tactics for Management*. (National Academies, 1986).

37. Lyon, B. R. & Skurray, R. Antimicrobial Resistance of Staphylococcus Aureus: Genetic Basis. *Microbiol. Rev.* **51**, 88–134 (1987).
38. Kuwahara, H. & Soyer, O. S. Bistability in Feedback Circuits as a Byproduct of Evolution of Evolvability. *Mol. Syst. Biol.* **8**, (2012).
39. Mahen, E. M., Harger, J. W., Calderon, E. M. & Fedor, M. J. Kinetics and Thermodynamics Make Different Contributions to RNA Folding In Vitro and in Yeast. *Mol. Cell* **19**, 27–37 (2005).
40. Mahen, E. M., Watson, P. Y., Cottrell, J. W. & Fedor, M. J. mRNA Secondary Structures Fold Sequentially but Exchange Rapidly in Vivo. *PLoS Biol.* **8**, e1000307 (2010).
41. Pereira, M. J. B., Behera, V. & Walter, N. G. Nondenaturing Purification of Co-Transcriptionally Folded RNA Avoids Common Folding Heterogeneity. *PLoS One* **5**, e12953 (2010).
42. Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in Bacterial Chemotaxis. *Nature* **397**, 168–171 (1999).
43. Barkai, N. & Leibler, S. Robustness in Simple Biochemical Networks. *Nature* **387**, 913–917 (1997).
44. Li, X., Cassidy, J. J., Reinke, C. A., Fischboeck, S. & Carthew, R. W. A MicroRNA Imparts Robustness against Environmental Fluctuation during Development. *Cell* **137**, 273–282 (2009).
45. Michelotti, N., Johnson-Buck, A., Manzo, A. J. & Walter, N. G. Beyond DNA Origami: The Unfolding Prospects of Nucleic Acid Nanotechnology. *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.* **4**, 139–152 (2012).
46. Petrillo, M. L. *et al.* The Ligation and Flexibility of Four-Arm DNA Junctions. *Biopolymers* **27**, 1337–1352 (1988).
47. Holliday, R. A Mechanism for Gene Conversion in Fungi. *Genet. Res.* **5**, 282–304 (1964).
48. Grindley, N. D. F., Whiteson, K. L. & Rice, P. A. Mechanisms of Site-Specific Recombination. *Annu. Rev. Biochem.* **75**, 567–605 (2006).
49. Fu, T. J. & Seeman, N. C. DNA Double-Crossover Molecules. *Biochemistry* **32**, 3211–3220 (1993).
50. LaBean, T. H. *et al.* Construction, Analysis, Ligation, and Self-Assembly of DNA Triple Crossover Complexes. *J. Am. Chem. Soc.* **122**, 1848–1860 (2000).
51. Wei, B., Dai, M. & Yin, P. Complex Shapes Self-Assembled from Single-stranded DNA Tiles. *Nature* **485**, 623–626 (2012).
52. Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* **338**, 1177–1183 (2012).
53. Rothemund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **440**, 297–302 (2006).
54. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
55. Han, D. *et al.* DNA Origami with Complex Curvatures in Three-Dimensional Space. *Science* **332**, 342–346 (2011).
56. Fu, J., Liu, M., Liu, Y., Woodbury, N. W. & Yan, H. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *J. Am. Chem. Soc.* **134**, 5516–5519 (2012).

57. Wu, N. *et al.* In Situ Monitoring of Single Molecule Binding Reactions with Time-lapse Atomic Force Microscopy on Functionalized DNA Origami. *Nanoscale* **3**, 2481 (2011).
58. Maune, H. T. *et al.* Self-Assembly of Carbon Nanotubes into Two-dimensional Geometries Using DNA Origami Templates. *Nat. Nanotechnol.* **5**, 61–66 (2009).
59. Gu, H., Chao, J., Xiao, S.-J. & Seeman, N. C. A Proximity-Based Programmable DNA Nanoscale Assembly Line. *Nature* **465**, 202–205
60. Kuzyk, A. *et al.* DNA-Based Self-Assembly of Chiral Plasmonic Nanostructures with Tailored Optical Response. *Nature* **483**, 311–314 (2012).
61. He, Y. & Liu, D. R. Autonomous Multistep Organic Synthesis in a Single Isothermal Solution Mediated by a DNA Walker. *Nat. Nanotechnol.* **5**, 778–782 (2010).
62. Qian, L. & Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **332**, 1196–1201 (2011).
63. Stojanovic, M. & Stefanovic, D. A Deoxyribozyme-Based Molecular Automaton. *Nat. Biotechnol.* **21**, 1069–1074 (2003).
64. Von Delius, M. & Leigh, D. A. Walking Molecules. *Chem. Soc. Rev.* **40**, 3656 (2011).
65. Stojanovic, M. N. & Stefanovic, D. Chemistry at a Higher Level of Abstraction. *J. Comput. Theor. Nanosci.* **8**, 434–440 (2011).
66. Tian, Y., He, Y., Chen, Y., Yin, P. & Mao, C. A DNAzyme That Walks Processively and Autonomously along a One-Dimensional Track. *Angew. Chem., Int. Ed.* **44**, 4355–4358 (2005).
67. Wickham, S. F. J. *et al.* A DNA-Based Molecular Motor That Can Navigate a Network of Tracks. *Nat. Nanotechnol.* **advance online publication**, (2012).
68. Li, Z. *et al.* Molecular behavior of DNA origami in higher-order self-assembly. *J. Am. Chem. Soc.* **132**, 13545–13552 (2010).
69. Ke, Y., Lindsay, S., Chang, Y., Liu, Y. & Yan, H. Self-Assembled Water-soluble Nucleic Acid Probe Tiles for Label-Free RNA Hybridization Assays. *Science* **319**, 180–183 (2008).
70. Li, Z., Wang, L., Yan, H. & Liu, Y. Effect of DNA Hairpin Loops on the Twist of Planar DNA Origami Tiles. *Langmuir* **28**, 1959–1965 (2011).
71. Meyer, E. Atomic Force Microscopy. *Prog. Surf. Sci.* **41**, 3–49 (1992).
72. Thomas, J. M. & Ducati, C.in *Characterization of Solid Materials and Heterogeneous Catalysts* (Che, M. & Védrine, J. C.) 655–701 (Wiley-VCH Verlag GmbH & Co. KGaA, 2012).
73. Henderson, E. Imaging of Living Cells by Atomic Force Microscopy. *Prog. Surf. Sci.* **46**, 39–60 (1994).
74. Agard, D. A., Hiraoka, Y., Shaw, P. & Sedat, J. W. Fluorescence Microscopy in Three Dimensions. *Methods Cell Biol.* **30**, 353–377 (1989).
75. Gonçalves, M. S. T. Fluorescent Labeling of Biomolecules with Organic Probes. *Chem. Rev.* **109**, 190–212 (2009).
76. Wood, B. T., Thompson, S. H. & Goldstein, G. Fluorescent Antibody Staining III. Preparation of Fluorescein-Isothiocyanate-Labeled Antibodies. *J. Immunol.* **95**, 225–229 (1965).

77. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green Fluorescent Protein as a Marker for Gene Expression. *Science* **263**, 802–805 (1994).
78. Walter, N. G., Huang, C.-Y., Manzo, A. J. & Sobhy, M. A. Do-It-Yourself Guide: How to Use the Modern Single-Molecule Toolkit. *Nat. Methods* **5**, 475–489 (2008).
79. Lipson, A., Lipson, S. G. & Lipson, H. *Optical Physics*. (Cambridge University Press, 2010).
80. Thompson, R. E., Larson, D. R. & Webb, W. W. Precise Nanometer Localization Analysis for Individual Fluorescent Probes. *Biophys. J.* **82**, 2775–2783 (2002).
81. Yildiz, A. & Selvin, P. R. Fluorescence Imaging with One Nanometer Accuracy: Application to Molecular Motors. *Acc. Chem. Res.* **38**, 574–582 (2005).
82. Rueda, D. & Walter, N. G. Single Molecule Fluorescence Control for Nanotechnology. *J. Nanosci. Nanotechnol.* **5**, 1990–2000 (2005).
83. Voigt, N. V. *et al.* Single-molecule Chemical Reactions on DNA Origami. *J. Nanosci. Nanotechnol.* **5**, 200–203 (2010).
84. Rust, M. J., Bates, M. & Zhuang, X. Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM). *Nat. Methods* **3**, 793–796 (2006).
85. Hess, S. T., Girirajan, T. P. K. & Mason, M. D. Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy. *Biophys. J.* **91**, 4258–4272 (2006).
86. Sharonov, A. & Hochstrasser, R. M. Wide-Field Subdiffraction Imaging by Accumulated Binding of Diffusing Probes. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18911–18916 (2006).
87. Cordes, T. *et al.* Resolving Single-Molecule Assembled Patterns with Superresolution Blink-Microscopy. *Nano Lett.* **10**, 645–651 (2009).
88. Jungmann, R. *et al.* Single-Molecule Kinetics and Super-Resolution Microscopy by Fluorescence Imaging of Transient Binding on DNA Origami. *Nano Lett.* **10**, 4756–4761 (2010).
89. Huang, B., Bates, M. & Zhuang, X. Super-Resolution Fluorescence Microscopy. *Annu. Rev. Biochem.* **78**, 993–1016 (2009).
90. Vale, R. D. The Molecular Motor Toolbox for Intracellular Transport. *Cell* **112**, 467–480 (2003).
91. Sawin, K. E. & Scholey, J. M. Motor Proteins in Cell Division. *Trends Cell Biol.* **1**, 122–129 (1991).
92. Pierobon, P. *et al.* Velocity, Processivity, and Individual Steps of Single Myosin V Molecules in Live Cells. *Biophys. J.* **96**, 4268–4275 (2009).
93. Ovchinnikov, V., Trout, B. L. & Karplus, M. Mechanical Coupling in Myosin V: A Simulation Study. *J. Mol. Biol.* **395**, 815–833 (2010).
94. Shin, J.-S. & Pierce, N. A. A Synthetic DNA Walker for Molecular Transport. *J. Am. Chem. Soc.* **126**, 10834–10835 (2004).
95. Sherman, W. B. & Seeman, N. C. A Precisely Controlled DNA Biped Walking Device. *Nano Lett.* **4**, 1203–1207 (2004).
96. Zhang, D. Y. & Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *J. Am. Chem. Soc.* **131**, 17303–17314 (2009).
97. Bath, J., Green, S. J. & Turberfield, A. J. A Free-Running DNA Motor Powered by a Nicking Enzyme. *Angew. Chem.* **117**, 4432–4435 (2005).

98. Green, S. J., Bath, J. & Turberfield, A. J. Coordinated Chemomechanical Cycles: A Mechanism for Autonomous Molecular Motion. *Phys. Rev. Lett.* **101**, 238101 (2008).
99. Pei, R. *et al.* Behavior of Polycatalytic Assemblies in a Substrate-Displaying Matrix. *J. Am. Chem. Soc.* **128**, 12693–12699 (2006).
100. Aitken, C. E., Marshall, R. A. & Puglisi, J. D. An Oxygen Scavenging System for Improvement of Dye Stability in Single-Molecule Fluorescence Experiments. *Biophys. J.* **94**, 1826–1835 (2008).
101. Churchman, L. S., Ökten, Z., Rock, R. S., Dawson, J. F. & Spudich, J. A. Single Molecule High-Resolution Colocalization of Cy3 and Cy5 Attached to Macromolecules Measures Intramolecular Distances Through Time. *Proc. Nat. Acad. Sci. U.S.A.* **102**, 1419–1423 (2005).
102. Li, J., Zheng, W., Kwon, A. H. & Lu, Y. In Vitro Selection and Characterization of a Highly Efficient Zn(II)-Dependent RNA-Cleaving Deoxyribozyme. *Nucleic Acids Res.* **28**, 481–488 (2000).
103. Schütz, G. J., Trabesinger, W. & Schmidt, T. Direct Observation of Ligand Colocalization on Individual Receptor Molecules. *Biophys. J.* **74**, 2223–2226 (1998).
104. N. Schellartin *Image Analysis: Methods and Applications, Second Edition* (Donat P. Hader) 359–361 (CRC Press, 2001).
105. Wetmur, J. G. & Fresco, J. DNA Probes: Applications of the Principles of Nucleic Acid Hybridization. *Crit. Rev. Biochem. Mol. Biol.* **26**, 227–259 (1991).
106. Santoro, S. W. & Joyce, G. F. A General Purpose RNA-Cleaving DNA Enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4262–4266 (1997).
107. Antal, T. & Krapivsky, P. L. Molecular Spiders with Memory. *Phys. Rev. E* **76**, 021121 (2007).
108. Omabegho, T., Sha, R. & Seeman, N. C. A Bipedal DNA Brownian Motor with Coordinated Legs. *Science* **324**, 67–71 (2009).
109. Saffarian, S., Collier, I. E., Marmer, B. L., Elson, E. L. & Goldberg, G. Interstitial Collagenase Is a Brownian Ratchet Driven by Proteolysis of Collagen. *Science* **306**, 108–111 (2004).
110. Yurke, B., Turberfield, A. J., Mills, A. P., Simmel, F. C. & Neumann, J. L. A DNA-Fuelled Molecular Machine Made of DNA. *Nature* **406**, 605–608 (2000).
111. Bouchaud, J.-P., Georges, A., Koplik, J., Provata, A. & Redner, S. Superdiffusion in Random Velocity Fields. *Phys. Rev. Lett.* **64**, 2503–2506 (1990).
112. Samii, L., Linke, H., Zuckermann, M. J. & Forde, N. R. Biased Motion and Molecular Motor Properties of Bipedal Spiders. *Phys. Rev. E* **81**, 021106 (2010).
113. Kube, C. R. & Zhang, H. Collective Robotics: From Social Insects to Robots. *Adaptive Behavior* **2**, 189–218 (1993).
114. Rus, D., Butler, Z., Kotay, K. & Vona, M. Self-Reconfiguring Robots. *Commun. ACM* **45**, 39–45 (2002).
115. Dorigo, M., Birattari, M. & Stutzle, T. *Ant Colony Optimization*. **1**, (MIT Press, 2004).
116. Semenov, O. & Stefanovic, D. Multiple Molecular Spiders with a Single Localized Source—The One-Dimensional Case. *Lecture Notes in Computer Science* **6937**, 204–216 (2011).

117. Kim, J. & Winfree, E. Synthetic in Vitro Transcriptional Oscillators. *Mol. Syst. Biol.* **7**, (2011).
118. Zhang, D. Y. Towards Domain-Based Sequence Design for DNA Strand Displacement Reactions. *Lecture Notes in Computer Science* **6518**, 162–175 (2011).
119. Douglas, S. M., Bachelet, I. & Church, G. M. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science* **335**, 831–834 (2012).
120. Lund, K. *et al.* Molecular Robots Guided by Prescriptive Landscapes. *Nature* **465**, 206–210
121. Zhou, X. *et al.* Quantitative Super-Resolution Imaging Uncovers Reactivity Patterns on Single Nanocatalysts. *Nat. Nanotechnol.* **7**, 237–241 (2012).
122. Cang, H. *et al.* Probing the Electromagnetic Field of a 15-Nanometre Hotspot by Single Molecule Imaging. *Nature* **469**, 385–388 (2011).
123. Huang, B., Babcock, H. & Zhuang, X. Breaking the Diffraction Barrier: Super-Resolution Imaging of Cells. *Cell* **143**, 1047–1058 (2010).
124. Zürner, A., Kirstein, J., Döblinger, M., Bräuchle, C. & Bein, T. Visualizing Single-Molecule Diffusion in Mesoporous Materials. *Nature* **450**, 705–708 (2007).
125. Kirstein, J. *et al.* Exploration of Nanostructured Channel Systems with Single-Molecule Probes. *Nat. Mater.* **6**, 303–310 (2007).
126. Brown, T. A. *et al.* Superresolution Fluorescence Imaging of Mitochondrial Nucleoids Reveals Their Spatial Range, Limits, and Membrane Interaction. *Mol. Cell. Biol.* **31**, 4994–5010 (2011).
127. Walder, R., Nelson, N. & Schwartz, D. K. Super-Resolution Surface Mapping Using the Trajectories of Molecular Probes. *Nat. Commun.* **2**, 515 (2011).
128. Michelotti, N., De Silva, C., Johnson-Buck, A. E., Manzo, A. J. & Walter, N. G. A Bird's Eye View: Tracking Slow Nanometer-Scale Movements of Single Molecular Nano-Assemblies. *Methods Enzymol.* **475**, 121–148 (2010).
129. Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez Jr., R. L. & Wiggins, C. H. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. *Biophys. J.* **97**, 3196–3205 (2009).
130. Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. Efficient Subpixel Image Registration Algorithms. *Opt. Lett.* **33**, 156–158 (2008).
131. Potthoff, R. F. & Whittinghill, M. Testing for Homogeneity: II. The Poisson Distribution. *Biometrika* **53**, 183–190 (1966).
132. Kim, D.-N., Kilchherr, F., Dietz, H. & Bathe, M. Quantitative Prediction of 3D Solution Shape and Flexibility of Nucleic Acid Nanostructures. *Nucleic Acids Res.* **40**, 2862–2868 (2012).
133. Nangreave, J., Han, D., Liu, Y. & Yan, H. DNA Origami: A History and Current Perspective. *Curr. Opin. Chem. Biol.* **14**, 608–615 (2010).
134. Castro, C. E. *et al.* A Primer to Scaffolded DNA Origami. *Nat. Methods* **8**, 221–229 (2011).
135. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **34**, 5402–5415 (2006).
136. Stein, I. H., Schüller, V., Böhm, P., Tinnefeld, P. & Liedl, T. Single-Molecule FRET Ruler Based on Rigid DNA Origami Blocks. *ChemPhysChem* **12**, 689–695 (2011).

137. Langecker, M. *et al.* Synthetic Lipid Membrane Channels Formed by Designed DNA Nanostructures. *Science* **338**, 932–936 (2012).
138. Ha, T. Single-Molecule Fluorescence Resonance Energy Transfer. *Methods* **25**, 78–86 (2001).
139. Wilkins Stevens, P., Henry, M. R. & Kelso, D. M. DNA Hybridization on Microparticles: Determining Capture-Probe Density and Equilibrium Dissociation Constants. *Nucleic Acids Res.* **27**, 1719–1727 (1999).
140. Watterson, J. H., Piunno, P. A. E., Wust, C. C. & Krull, U. J. Effects of Oligonucleotide Immobilization Density on Selectivity of Quantitative Transduction of Hybridization of Immobilized DNA. *Langmuir* **16**, 4984–4992 (2000).
141. Johnson-Buck, A. E., McDowell, S. E. & Walter, N. G. Metal Ions: Supporting Actors in the Playbook of Small Ribozymes. *Met. Ions Life Sci.* **9**, 175–196 (2011).
142. Rueda, D., Wick, K., McDowell, S. E. & Walter, N. G. Diffusely Bound Mg2+ Ions Slightly Reorient Stems I and II of the Hammerhead Ribozyme to Increase the Probability of Formation of the Catalytic Core. *Biochemistry* **42**, 9924–9936 (2003).
143. Duckett, D. R., Murchie, A. I. & Lilley, D. M. The Role of Metal Ions in the Conformation of the Four-Way DNA Junction. *EMBO J.* **9**, 583–590 (1990).
144. Kershner, R. J. *et al.* Placement and Orientation of Individual DNA Shapes on Lithographically Patterned Surfaces. *Nat. Nanotechnol.* **4**, 557–561 (2009).
145. Pastré, D. *et al.* Anionic Polyelectrolyte Adsorption on Mica Mediated by Multivalent Cations: A Solution to DNA Imaging by Atomic Force Microscopy under High Ionic Strengths. *Langmuir* **22**, 6651–6660 (2006).
146. Derr, N. D. *et al.* Tug-of-War in Motor Protein Ensembles Revealed with a Programmable DNA Origami Scaffold. *Science* **338**, 662–665 (2012).
147. Acuna, G. P. *et al.* Fluorescence Enhancement at Docking Sites of DNA-Directed Self-Assembled Nanoantennas. *Science* **338**, 506–510 (2012).
148. Tyagi, S. & Kramer, F. R. Molecular Beacons: Probes That Fluoresce upon Hybridization. *Nat. Biotechnol.* **14**, 303–308 (1996).
149. Levene, M. J. *et al.* Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations. *Science* **299**, 682–686 (2003).
150. Castronovo, M. *et al.* Two-Dimensional Enzyme Diffusion in Laterally Confined DNA Monolayers. *Nat. Commun.* **2**, 297 (2011).
151. Saville, B. J. & Collins, R. A. A Site-Specific Self-Cleavage Reaction Performed by a Novel RNA in Neurospora Mitochondria. *Cell* **61**, 685–696 (1990).
152. Cochrane, J. C. & Strobel, S. A. Catalytic Strategies of Self-Cleaving Ribozymes. *Acc. Chem. Res.* **41**, 1027–1035 (2008).
153. Salehi-Ashtiani, K., Lupták, A., Litovchick, A. & Szostak, J. W. A Genomewide Search for Ribozymes Reveals an HDV-Like Sequence in the Human CPEB3 Gene. *Science* **313**, 1788–1792 (2006).
154. Hammann, C., Luptak, A., Perreault, J. & Peña, M. de la The Ubiquitous Hammerhead Ribozyme. *RNA* **18**, 871–885 (2012).
155. Beattie, T. L., Olive, J. E. & Collins, R. A. A Secondary-Structure Model for the Self-Cleaving Region of Neurospora VS RNA. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4686–4690 (1995).

156. Rastogi, T., Beattie, T. L., Olive, J. E. & Collins, R. A. A Long-Range Pseudoknot Is Required for Activity of the Neurospora VS Ribozyme. *EMBO J.* **15**, 2820–2825 (1996).
157. Hiley, S. L. & Collins, R. A. Rapid Formation of a Solvent-Inaccessible Core in the Neurospora Varkud Satellite Ribozyme. *EMBO J.* **20**, 5461–5469 (2001).
158. Lafontaine, D. A., Norman, D. G. & Lilley, D. M. J. Structure, Folding and Activity of the VS Ribozyme: Importance of the 2-3-6 Helical Junction. *EMBO J.* **20**, 1415–1424 (2001).
159. Lipfert, J., Ouellet, J., Norman, D. G., Doniach, S. & Lilley, D. M. J. The Complete VS Ribozyme in Solution Studied by Small-Angle X-Ray Scattering. *Structure* **16**, 1357–1367 (2008).
160. Zamel, R. *et al.* Exceptionally Fast Self-Cleavage by a Neurospora Varkud Satellite Ribozyme. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1467–1472 (2004).
161. Hiley, S. L., Sood, V. D., Fan, J. & Collins, R. A. 4-thio-U Cross-linking Identifies the Active Site of the VS Ribozyme. *EMBO J.* **21**, 4691–4698 (2002).
162. Lai, M. M. The Molecular Biology of Hepatitis Delta Virus. *Annu. Rev. Biochem.* **64**, 259–286 (1995).
163. Webb, C.-H. T., Riccitelli, N. J., Ruminski, D. J. & Lupták, A. Widespread Occurrence of Self-Cleaving Ribozymes. *Science* **326**, 953–953 (2009).
164. Ferré-D'Amaré, A. R., Zhou, K. & Doudna, J. A. Crystal Structure of a Hepatitis Delta Virus Ribozyme. *Nature* **395**, 567–574 (1998).
165. Wadkins, T. S., Perrotta, A. T., Ferré-D'Amaré, A. R., Doudna, J. A. & Been, M. D. A Nested Double Pseudoknot Is Required for Self-Cleavage Activity of Both the Genomic and Antigenomic Hepatitis Delta Virus Ribozymes. *RNA* **5**, 720–727 (1999).
166. Ke, A., Zhou, K., Ding, F., Cate, J. H. D. & Doudna, J. A. A Conformational Switch Controls Hepatitis Delta Virus Ribozyme Catalysis. *Nature* **429**, 201–205 (2004).
167. Chen, J.-H. *et al.* A 1.9 Å Crystal Structure of the HDV Ribozyme Precleavage Suggests both Lewis Acid and General Acid Mechanisms Contribute to Phosphodiester Cleavage. *Biochemistry* **49**, 6508–6518 (2010).
168. Nakano, S., Chadalavada, D. M. & Bevilacqua, P. C. General Acid-Base Catalysis in the Mechanism of a Hepatitis Delta Virus Ribozyme. *Science* **287**, 1493–1497 (2000).
169. Das, S. R. & Piccirilli, J. A. General Acid Catalysis by the Hepatitis Delta Virus Ribozyme. *Nat. Chem. Biol.* **1**, 45–52 (2005).
170. Pereira, M. J. B., Harris, D. A., Rueda, D. & Walter, N. G. Reaction Pathway of the Trans-Acting Hepatitis Delta Virus Ribozyme: A Conformational Change Accompanies Catalysis. *Biochemistry* **41**, 730–740 (2002).
171. Harris, D. A., Rueda, D. & Walter, N. G. Local Conformational Changes in the Catalytic Core of the Trans-Acting Hepatitis Delta Virus Ribozyme Accompany Catalysis. *Biochemistry* **41**, 12051–12061 (2002).
172. Sefcikova, J., Krasovska, M. V., Šponer, J. & Walter, N. G. The Genomic HDV Ribozyme Utilizes a Previously Unnoticed U-turn Motif to Accomplish Fast Site-Specific Catalysis. *Nucleic Acids Res.* **35**, 1933–1946 (2007).

173. Lang, K. & Micura, R. The Preparation of Site-Specifically Modified Riboswitch Domains as an Example for Enzymatic Ligation of Chemically Synthesized RNA Fragments. *Nat. Protoc.* **3**, 1457–1466 (2008).
174. Rasnik, I., McKinney, S. A. & Ha, T. Nonblinking and Long-lasting Single-Molecule Fluorescence Imaging. *Nat. Methods* **3**, 891–893 (2006).