

**REPORT DOCUMENTATION PAGE**

*Form Approved  
OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.**

<b>1. REPORT DATE (DD-MM-YYYY)</b> 08/26/2016		<b>2. REPORT TYPE</b> FINAL		<b>3. DATES COVERED (From - To)</b> 06/01/2013-05/31/2016	
<b>4. TITLE AND SUBTITLE</b> Kinetic self-assembly of DNA tiles and bricks				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b> N00014-13-1-0593	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Peng Yin				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> President and Fellows of Harvard College 1033 Massachusetts Avenue, 5th Floor Cambridge, MA 02138- 5369				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  167868	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> ONR REG BOSTON N62879 495 SUMMER STREET ROOM 627 BOSTON, MA 02210-2109				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>  ONR	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>  N62879	
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> Approved for Public Release; distribution is Unlimited					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b> The objective of our work is to engineer synthetic developmental self-assembly, where a synthetic molecular structure grows isothermally in a kinetically controlled fashion. We combine the structural insight gleaned from building complex structures from single stranded DNA bricks with the kinetic insight on programming assembly pathway using hairpin motifs and strand displacement interactions. This paradigm is fundamentally different from and conceptually more powerful than the dominant thermal annealing paradigm for assembling synthetic molecular structures. We are using this developmental self-assembly framework for triggered isothermal self-assembly of highly complex single-stranded tile and bricks based nanostructures. The successful implementation of this developmental self-assembly framework will introduce a profound paradigm shift in the field of synthetic self-assembly, and promises numerous technological applications with transformative impacts. As a parallel and supplemental effort we develop software tools to automate the design and in silico testing of these structures.					
<b>15. SUBJECT TERMS</b> self-assembly, DNA, molecular programming					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b> Peng Yin
<b>a. REPORT</b> U	<b>b. ABSTRACT</b> U	<b>c. THIS PAGE</b> U			<b>19b. TELEPHONE NUMBER (Include area code)</b> 617-432-7731

## INSTRUCTIONS FOR COMPLETING SF 298

**1. REPORT DATE.** Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

**2. REPORT TYPE.** State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

**3. DATES COVERED.** Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

**4. TITLE.** Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

**5a. CONTRACT NUMBER.** Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

**5b. GRANT NUMBER.** Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

**5c. PROGRAM ELEMENT NUMBER.** Enter all program element numbers as they appear in the report, e.g. 61101A.

**5d. PROJECT NUMBER.** Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

**5e. TASK NUMBER.** Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

**5f. WORK UNIT NUMBER.** Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

**6. AUTHOR(S).** Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES).** Self-explanatory.

**8. PERFORMING ORGANIZATION REPORT NUMBER.** Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES).** Enter the name and address of the organization(s) financially responsible for and monitoring the work.

**10. SPONSOR/MONITOR'S ACRONYM(S).** Enter, if available, e.g. BRL, ARDEC, NADC.

**11. SPONSOR/MONITOR'S REPORT NUMBER(S).** Enter report number as assigned by the sponsoring/monitoring agency, if available, e.g. BRL-TR-829; -215.

**12. DISTRIBUTION/AVAILABILITY STATEMENT.** Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

**13. SUPPLEMENTARY NOTES.** Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

**14. ABSTRACT.** A brief (approximately 200 words) factual summary of the most significant information.

**15. SUBJECT TERMS.** Key words or phrases identifying major concepts in the report.

**16. SECURITY CLASSIFICATION.** Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

**17. LIMITATION OF ABSTRACT.** This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Final Technical Report for N00014-13-1-0593

**Kinetic Self-assembly of Complex Structures  
from DNA Tiles and Bricks**

Principle Investigator: Peng Yin, Harvard University

**Contents**

1	Heading	3
2	Scientific and technical objectives	3
3	Approach	3
4	Concise accomplishment	3
5	Expanded accomplishments	5
5.1	Fundamental Accomplishments in Kinetic Self-Assembly . . . . .	5
5.1.1	Barcode Extension for Analysis and Reconstruction of Structures (BEARS) .	5
5.1.2	Eliminating Kinetic Barriers through Sequence Design Allows Rapid As- sembly of DNA Origami at Room Temperature . . . . .	5
5.1.3	Complex reconfiguration of DNA nanostructures . . . . .	6
5.1.4	Developmental self-assembly of a DNA tetrahedron . . . . .	6
5.1.5	Design space for complex DNA structures . . . . .	6
5.1.6	Isothermal self-assembly of complex DNA structures under diverse and bio- compatible conditions . . . . .	7
5.1.7	A Biochemical Nanoscope via Auto-cycling Proximity Recording . . . . .	7
5.2	Applications of Kinetic Self-Assembly . . . . .	7
5.2.1	Quantitative superresolution imaging with qPAINT . . . . .	7
5.2.2	Optical imaging of individual biomolecules in densely packed clusters . . . .	8
5.2.3	Paper-based synthetic gene networks . . . . .	8
5.2.4	Toehold switches: de-novo-designed regulators of gene expression . . . . .	8
5.2.5	Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT . . . . .	9
5.3	Software Efforts in Kinetic Self-Assembly . . . . .	9
5.3.1	DyNAMiC Workbench: An Integrated Development Environment for Dy- namic DNA Nanotechnology . . . . .	9
6	Major problems/issues (if any)	9

<b>7</b>	<b>Technology transfer</b>	<b>10</b>
<b>8</b>	<b>Foreign collaborations and supported foreign nationals</b>	<b>10</b>
<b>9</b>	<b>Productivity</b>	<b>10</b>
9.1	Refereed journal article . . . . .	10
9.2	Non-refereed journal article . . . . .	11
9.3	Workshops and conferences . . . . .	11
9.4	Patents . . . . .	14
9.5	Awards/Honors . . . . .	15
<b>10</b>	<b>Award participants</b>	<b>16</b>
<b>11</b>	<b>Appendix</b>	<b>16</b>

## 1 Heading

- PI: Peng Yin
- Organization: Harvard University
- ONR Award Number: N00014-13-1-0593
- Award Title: Kinetic Self-assembly of Complex Structures from DNA Tiles and Bricks

## 2 Scientific and technical objectives

The objective of our work is to engineer synthetic developmental self-assembly, where a synthetic molecular structure grows isothermally in a kinetically controlled fashion. We combine the structural insight gleaned from building complex structures from single stranded DNA bricks with the kinetic insight on programming assembly pathway using hairpin motifs and strand displacement interactions. This paradigm is fundamentally different from and conceptually more powerful than the dominant thermal annealing paradigm for assembling synthetic molecular structures. We are using this developmental self-assembly framework for triggered isothermal self-assembly of highly complex single-stranded tile and bricks based nano-structures. The successful implementation of this developmental self-assembly framework will introduce a profound paradigm shift in the field of synthetic self-assembly, and promises numerous technological applications with transformative impacts. As a parallel and supplemental effort we develop software tools to automate the design and *in silico* testing of these structures.

## 3 Approach

We report herein on the following approaches:

1. Development of fundamental molecular mechanisms that enable the developmental self-assembly of complex DNA nano structures. See section 5.1 for details.
2. Development of transformative real world applications based on dynamic DNA nanotechnology. See section 5.2 for details.
3. Development of software tools to enable automated design and *in silico* testing of kinetic DNA nanosystems. See section 5.3.

## 4 Concise accomplishment

- *A Biochemical Nanoscope via Auto-cycling Proximity Recording*. Here, we present a microscope-free “biochemical nanoscopy” method that records sophisticated nanostructure features in situ for later readout. The method is based on a conceptually novel Auto-cycling Proximity Recording (APR) mechanism, which continuously and repeatedly produces proximity records of any nearby pairs of DNA-barcoded probes, at physiological temperature, without destroying the probes themselves.

- *Barcode Extension for Analysis and Reconstruction of Structures (BEARS)*. We introduce a new method of structure characterization for simultaneous quantification of every strand species incorporated into a structure. Our method uses barcode extension and next-generation DNA sequencing to quantitatively measure the relative incorporation of individual strands into a DNA nanostructure.
- *Eliminating Kinetic Barriers through Sequence Design Allows Rapid Assembly of DNA Origami at Room Temperature*. We use computational sequence design to eliminate prohibitive kinetic barriers and thus assemble DNA origami at room temperature in as little as five minutes.
- *Quantitative superresolution imaging with qPAINT*. Here, we use the programmable and specific binding of dye-labeled DNA probes to count integer numbers of targets. This method, called quantitative points accumulation in nanoscale topography (qPAINT), works independently of dye photophysics for robust counting with high precision and accuracy over a wide dynamic range.
- *Optical imaging of individual biomolecules in densely packed clusters*. Here, we show that discrete molecular imaging is possible using DNA-PAINT (points accumulation for imaging in nanoscale topography) - a super-resolution fluorescence microscopy technique that exploits programmable transient oligonucleotide hybridization. Non synthetic DNA nanostructures.
- *DyNAMiC Workbench: An Integrated Development Environment for Dynamic DNA Nanotechnology*. Here, we present software implementing a three-tier design process: a high-level visual programming language is used to describe systems, a molecular compiler builds a DNA implementation and nucleotide sequences are generated and optimized. Additionally, our software includes tools for analysing and debugging the designs in silico, and for importing/exporting designs to other commonly used software systems.
- *Paper-based synthetic gene networks* We present an in vitro paper-based platform that provides an alternate, versatile venue for synthetic biologists to operate and a much-needed medium for the safe deployment of engineered gene circuits beyond the lab. We demonstrate this technology with small-molecule and RNA actuation of genetic switches, rapid prototyping of complex gene circuits, and programmable in vitro diagnostics, including glucose sensors and strain-specific Ebola virus sensors.
- *Toehold switches: de-novo-designed regulators of gene expression* we report a class of de-novo-designed prokaryotic riboregulators called toehold switches that activate gene expression in response to cognate RNAs with arbitrary sequences. Toehold switches provide a high level of orthogonality and can be forward engineered to provide average dynamic range above 400.
- *Complex reconfiguration of DNA nanostructures* We have developed a general structural-reconfiguration method that utilizes the modularly interconnected architecture of single-stranded DNA tile and brick structures. By using this method, we reconfigured a two-dimensional rectangular DNA canvas and a three-dimensional DNA cuboid into diverse prescribed shapes.
- *Developmental self-assembly of a DNA tetrahedron* By integrating temporal with spatial control, here we demonstrate the “developmental” self-assembly of a DNA tetrahedron, where a

prescriptive molecular program orchestrates the kinetic pathways by which DNA molecules isothermally self-assemble into a well-defined three-dimensional wireframe geometry.

- *Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT* Here we use the transient binding of short fluorescently labeled oligonucleotides (DNA-PAINT, a variation of point accumulation for imaging in nanoscale topography) for simple and easy-to-implement multiplexed super-resolution imaging that achieves sub-10-nm spatial resolution in vitro on synthetic DNA structures.
- *Design space for complex DNA structures* We perform a comprehensive study of the design space for complex DNA structures, using more than 30 distinct motifs derived from single-stranded tiles. These motifs self-assemble to form structures with diverse strand weaving patterns and specific geometric properties, such as curvature and twist.
- *Isothermal self-assembly of complex DNA structures under diverse and biocompatible conditions* We describe a DNA self-assembly system that can be tuned to form a complex target structure isothermally at any prescribed temperature or homogeneous condition within a wide range. We were able to achieve isothermal assembly between 15 and 69 °C in a predictable fashion.

## 5 Expanded accomplishments

### 5.1 Fundamental Accomplishments in Kinetic Self-Assembly

#### 5.1.1 Barcode Extension for Analysis and Reconstruction of Structures (BEARS)

Collections of DNA sequences can be rationally designed to self-assemble into predictable three-dimensional structures. The geometric and functional diversity of DNA nanostructures created to date has been enhanced by improvements in DNA synthesis and computational design. However, existing methods for structure characterization typically image the final product or laboriously determine the presence of individual, labeled strands using gel electrophoresis. Here, we introduce a new method of structure characterization for simultaneous quantification of every strand species incorporated into a structure. Our method uses barcode extension and next-generation DNA sequencing to quantitatively measure the relative incorporation of individual strands into a DNA nanostructure. By quantifying the relative abundances of distinct DNA species in product and monomer bands, we can study the influence of geometry and sequence on assembly. We have tested our method using 2D and 3D DNA brick and DNA origami structures. Our method is general and should be extensible to a wide variety of synthetic DNA structures.

#### 5.1.2 Eliminating Kinetic Barriers through Sequence Design Allows Rapid Assembly of DNA Origami at Room Temperature

DNA Origami [3] is a robust and programmable technique to make DNA nanostructures. A key use of DNA origami structures has been as nanoscale breadboards that organize components like small proteins, nanoparticles, small molecules and other DNA strands. Typical DNA origami assembly requires that the scaffold strand be annealed with staple strands in the appropriate buffer. The high

temperatures ( $> 80\text{C}$ ) used in this process are not always compatible with temperature-sensitive components or contexts. In response, there have been attempts to isothermally fold DNA. Studies have shown that DNA origami can assemble isothermally if held at temperatures around  $60\text{C}$ , or at room temperature by using denaturing agents. However, rapid assembly of DNA origami at room temperature without the use of denaturing agents, which can be toxic or have other unintended consequences, has not been achieved. We assembled a rectangular DNA origami from a 1000 base long DeBruijn derived scaffold sequence and a corresponding staple set. The origami assembles isothermally in as little as five minutes at temperatures ranging from  $25\text{C}$  to  $50\text{C}$ . Refer to appendix for more details.

### **5.1.3 Complex reconfiguration of DNA nanostructures**

Nucleic acids have been used to create diverse synthetic structural and dynamic systems. Toehold-mediated strand displacement has enabled the construction of sophisticated circuits, motors, and molecular computers. Yet it remains challenging to demonstrate complex structural reconfiguration in which a structure changes from a starting shape to another arbitrarily prescribed shape. To address this challenge, we have developed a general structural-reconfiguration method that utilizes the modularly interconnected architecture of singlestranded DNA tile and brick structures. The removal of one component strand reveals a newly exposed toehold on a neighboring strand, thus enabling us to remove regions of connected component strands without the need to modify the strands with pre-designed external toeholds. By using this method, we reconfigured a two-dimensional rectangular DNA canvas into diverse prescribed shapes. We also used this method to reconfigure a three-dimensional DNA cuboid.

### **5.1.4 Developmental self-assembly of a DNA tetrahedron**

Kinetically controlled isothermal growth is fundamental to biological development, yet it remains challenging to rationally design molecular systems that self-assemble isothermally into complex geometries via prescribed assembly and disassembly pathways. By exploiting the programmable chemistry of base pairing, sophisticated spatial and temporal control have been demonstrated in DNA self-assembly, but largely as separate pursuits. By integrating temporal with spatial control, here we demonstrate the “developmental” self-assembly of a DNA tetrahedron, where a prescriptive molecular program orchestrates the kinetic pathways by which DNA molecules isothermally self-assemble into a well-defined three-dimensional wireframe geometry. In this reaction, nine DNA reactants initially coexist metastably, but upon catalysis by a DNA initiator molecule, navigate 24 individually characterizable intermediate states via prescribed assembly pathways, organized both in series and in parallel, to arrive at the tetrahedral final product. In contrast to previous work on dynamic DNA nanotechnology, this developmental program coordinates growth of ringed substructures into a three-dimensional wireframe superstructure, taking a step toward the goal of kinetically controlled isothermal growth of complex three-dimensional geometries.

### **5.1.5 Design space for complex DNA structures**

Nucleic acids have emerged as effective materials for assembling complex nanoscale structures. To tailor the structures to function optimally for particular applications, a broad structural design space is desired. Despite the many discrete and extended structures demonstrated in the past few decades,

the design space remains to be fully explored. In particular, the complex finite-sized structures produced to date have been typically based on a small number of structural motifs. Here, we perform a comprehensive study of the design space for complex DNA structures, using more than 30 distinct motifs derived from single-stranded tiles. These motifs self-assemble to form structures with diverse strand weaving patterns and specific geometric properties, such as curvature and twist. We performed a systematic study to control and characterize the curvature of the structures, and constructed a flat structure with a corrugated strand pattern. The work here reveals the broadness of the design space for complex DNA nanostructures.

#### **5.1.6 Isothermal self-assembly of complex DNA structures under diverse and biocompatible conditions**

Nucleic acid nanotechnology has enabled researchers to construct a wide range of multidimensional structures in vitro. Until recently, most DNA-based structures were assembled by thermal annealing using high magnesium concentrations and non-physiological environments. Here, we describe a DNA self-assembly system that can be tuned to form a complex target structure isothermally at any prescribed temperature or homogeneous condition within a wide range. We were able to achieve isothermal assembly between 15 and 69 ° C in a predictable fashion by altering the strength of strand-strand interactions in several different ways, for example, domain length, GC content, and linker regions between domains. We also observed the assembly of certain structures under biocompatible conditions, that is, at physiological pH, temperature, and salinity in the presence of the molecular crowding agent polyethylene glycol (PEG) mimicking the cellular environment. This represents an important step toward the self-assembly of geometrically precise DNA or RNA structures in vivo.

#### **5.1.7 A Biochemical Nanoscope via Auto-cycling Proximity Recording**

Analysis of the spatial arrangement of molecular features enables the engineering of synthetic nanostructures and the understanding of natural ones. Microscopy methods can provide direct visualization of complex spatial features but typically require sophisticated equipment with limited throughput. Here, we present a microscope-free “biochemical nanoscopy” method that records sophisticated nanostructure features in situ for later readout. The method is based on a conceptually novel Auto-cycling Proximity Recording (APR) mechanism, which continuously and repeatedly produces proximity records of any nearby pairs of DNA-barcoded probes, at physiological temperature, without destroying the probes themselves. We demonstrate the proximity and auto-cycling characteristics, and then apply the system as a biochemical nanoscope to accurately decode various spatial arrangements of 7 unique probes in homogeneous test samples. We further demonstrate the ability of APR to re-sample the same dynamic system repeatedly. Future development is expected to enable readout with massively parallel sequencing methods, creating a truly single-molecule technique. Refer to appendix for more details

## **5.2 Applications of Kinetic Self-Assembly**

### **5.2.1 Quantitative superresolution imaging with qPAINT**

Counting molecules in complexes is challenging, even with super-resolution microscopy. Here, we use the programmable and specific binding of dye-labeled DNA probes to count integer numbers of

targets. This method, called quantitative points accumulation in nanoscale topography (qPAINT), works independently of dye photophysics for robust counting with high precision and accuracy over a wide dynamic range. qPAINT was benchmarked on DNA nanostructures and demonstrated for cellular applications by quantifying proteins in situ and the number of single-molecule FISH probes bound to an mRNA target.

### **5.2.2 Optical imaging of individual biomolecules in densely packed clusters**

Recent advances in fluorescence super-resolution microscopy have allowed subcellular features and synthetic nanostructures down to 10 – 20 nm in size to be imaged. However, the direct optical observation of individual molecular targets ( $\approx 5$  nm) in a densely packed biomolecular cluster remains a challenge. Here, we show that such discrete molecular imaging is possible using DNA-PAINT (points accumulation for imaging in nanoscale topography) - a super-resolution fluorescence microscopy technique that exploits programmable transient oligonucleotide hybridization - on synthetic DNA nanostructures. We examined the effects of a high photon count, high blinking statistics and an appropriate blinking duty cycle on imaging quality, and developed a software-based drift correction method that achieves  $< 1$  nm residual drift (root mean squared) over hours. This allowed us to image a densely packed triangular lattice pattern with  $\approx 5$  nm point-to-point distance and to analyse the DNA origami structural offset with angstrom-level precision ( $2 \text{ \AA}$ ) from single-molecule studies. By combining the approach with multiplexed exchange-PAINT imaging, we further demonstrated an optical nanodisplay with  $5 \times 5$  nm pixel size and three distinct colours with  $< 1$  nm cross-channel registration accuracy.

### **5.2.3 Paper-based synthetic gene networks**

Synthetic gene networks have wide-ranging uses in reprogramming and rewiring organisms. To date, there has not been a way to harness the vast potential of these networks beyond the constraints of a laboratory or in vivo environment. Here, we present an in vitro paper-based platform that provides an alternate, versatile venue for synthetic biologists to operate and a much-needed medium for the safe deployment of engineered gene circuits beyond the lab. Commercially available cell-free systems are freeze dried onto paper, enabling the inexpensive, sterile, and abiotic distribution of synthetic-biology-based technologies for the clinic, global health, industry, research, and education. For field use, we create circuits with colorimetric outputs for detection by eye and fabricate a low-cost, electronic optical interface. We demonstrate this technology with small-molecule and RNA actuation of genetic switches, rapid prototyping of complex gene circuits, and programmable in vitro diagnostics, including glucose sensors and strain-specific Ebola virus sensors.

### **5.2.4 Toehold switches: de-novo-designed regulators of gene expression**

Efforts to construct synthetic networks in living cells have been hindered by the limited number of regulatory components that provide wide dynamic range and low crosstalk. Here, we report a class of de-novo-designed prokaryotic riboregulators called toehold switches that activate gene expression in response to cognate RNAs with arbitrary sequences. Toehold switches provide a high level of orthogonality and can be forward engineered to provide average dynamic range above 400. We show that switches can be integrated into the genome to regulate endogenous genes and use them as sensors that respond to endogenous RNAs. We exploit the orthogonality of toehold switches to

regulate 12 genes independently and to construct a genetic circuit that evaluates 4-input AND logic. Toehold switches, with their wide dynamic range, orthogonality, and programmability, represent a versatile and powerful platform for regulation of translation, offering diverse applications in molecular biology, synthetic biology, and biotechnology.

### **5.2.5 Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT**

Super-resolution fluorescence microscopy is a powerful tool for biological research, but obtaining multiplexed images for a large number of distinct target species remains challenging. Here we use the transient binding of short fluorescently labeled oligonucleotides (DNDNA-PAINTINTINT, a variation of point accumulation for imaging in nanoscale topography) for simple and easy-to-implement multiplexed super-resolution imaging that achieves sub-10-nm spatial resolution in vitro on synthetic DNA structures. We also report a multiplexing approach (Exchange-PAINTINTINT) that allows sequential imaging of multiple targets using only a single dye and a single laser source. We experimentally demonstrate ten-color super-resolution imaging in vitro on synthetic DNDNA structures as well as four-color two-dimensional (2D) imaging and three-color 3D imaging of proteins in fixed cells.

## **5.3 Software Efforts in Kinetic Self-Assembly**

### **5.3.1 DyNAMiC Workbench: An Integrated Development Environment for Dynamic DNA Nanotechnology**

DynamicDNA nanotechnology provides a promising avenue for implementing sophisticated assembly processes, mechanical behaviours, sensing and computation at the nanoscale. However, design of these systems is complex and error-prone, because the need to control the kinetic pathway of a system greatly increases the number of design constraints and possible failure modes for the system. Previous tools have automated some parts of the design workflow, but an integrated solution is lacking. Here, we present software implementing a three-tier design process: a high-level visual programming language is used to describe systems, a molecular compiler builds a DNA implementation and nucleotide sequences are generated and optimized. Additionally, our software includes tools for analysing and debugging the designs in silico, and for importing/exporting designs to other commonly used software systems. The software we present is built on many existing pieces of software, but is integrated into a single package accessible using a Web-based interface at <http://molecular-systems.net/workbench>. We hope that the deep integration between tools and the flexibility of this design process will lead to better experimental results, fewer experimental design iterations and the development of more complex DNA nanosystems.

## **6 Major problems/issues (if any)**

None.

## 7 Technology transfer

See Sect. 9.4 "Patents".

## 8 Foreign collaborations and supported foreign nationals

**Foreign Nationals:** Nikhil Gopalkrishnan, Alexander Green, Dongran Han, Ralf Jungmann, Jie Shen, Yonggang Ke, Feng Xuan, Mingjie Dai, Johannes Woehrstein, Maximilian Strauss and Juanita Lara

**Foreign Collaborations:** N/A

## 9 Productivity

### 9.1 Refereed journal article

1. Thomas E. Schaus, Sungwook Woo, Feng Xuan, Xi Chen and Peng Yin. *A Biochemical Nanoscope via Auto-cycling Proximity Recording*. Submitted
2. Cameron Myhrvold, Michael Baym, Nikita Hanikel, Luvena Ong, Jonathan Gootenberg and Peng Yin. *Barcode Extension for Analysis and Reconstruction of Structures (BEARS)*. In preparation
3. Nikhil Gopalkrishnan, Thomas Schaus and Peng Yin. *Eliminating Kinetic Barriers through Sequence Design Allows Rapid Assembly of DNA Origami at Room Temperature*. In preparation
4. R. Jungmannă, M. S. Avendanoă, M. Dai, J. B. Woehrstein, S. S. Agasti, Z. Feiger, A. Rodal and P. Yin. *Quantitative super-resolution imaging with qPAINT*, Nature Methods, 13, 439-442 (2016)
5. M. Dai, R. Jungmann and P. Yin. *Optical imaging of individual biomolecules in densely packed clusters*. Nature Nanotechnology (2016).
6. C. Grun, J. Werfel, D. Y. Zhang, and P. Yin. *DyNAMiC Workbench: an integrated development environment for dynamic DNA nanotechnology*. J. R. Soc. Interface. 12: 20150580 (2015)
7. K. Pardee, A.A. Green, T. Ferrante, E. Cameron, A. DaleyKeyser, P. Yin, and J.J. Collins. *Paper-based synthetic gene networks*. Cell159:940-954
8. A.A. Green, P.A. Silver, J.J. Collins, P. Yin. *Toehold switches: de-novo-designed regulators of gene expression*. Cell 159: 925-939 (2014)
9. B. Wei, L.L. Ong, J. Chen, A. Jaffe, P. Yin. *Complex reconfiguration of DNA nanostructures*. Angewandte Chemie 126:7605-7609 (2014).
10. J. Sadowski, C.R. Calvert, D.Y. Zhang, N.A. Pierce, P. Yin. *Developmental self-assembly of a DNA tetrahedron*. ACS Nano DOI: 10.1021/nn4038223 (2014).

11. R. Jungmannă, M.S. Avendanoă, J.B. Woehrsteină, M. Dai, W.M. Shih, P. Yin. *Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT*. *Nature Methods* 11(3):313-318 (2014)
12. B. Wei, M. Dai, C. Myhrvold, Y. Ke, R. Jungmann, P. Yin. *Design space for complex DNA structures*. *J. Am. Chem. Soc.* 135 (48):18080-8 (2013)
13. C. Myhrvold, M. Dai, P.A. Silver, P. Yin. *Isothermal self-assembly of complex DNA structures under diverse and biocompatible conditions*. *Nano Letters* 13(9): 4242-8 (2013)

## 9.2 Non-refereed journal article

None.

## 9.3 Workshops and conferences

1. 7th international symposium on Bioanalysis, Biomedical Engineering and Nanotechnology (ISBBN 2016), Changsha, China, May 27th, 2016.
2. 3rd Materials Beyond Symposium, Shanghai, China, May 26th, 2016.
3. Worcester annual nanotech symposium (NanoWorcester), Worcester, MA, Apr. 14th, 2016.
4. Workshop on 10 years of DNA origami, Pasadena, CA, March 14th, 2016.
5. Biophysical Society 60th Annual Meeting, Los Angeles, CA, Feb. 27th, 2016.
6. Department of Bioengineering, Rice University, Houston, TX, Feb. 9th, 2016.
7. Amgen, Inc., Cambridge, MA, Nov. 10th, 2015.
8. Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA, Nov. 3rd, 2015.
9. Tenth International Meeting of Institute for Translational Medicine and Therapeutics at Penn, Pennsylvania, PA, Oct. 11th, 2015.
10. Physical Sciences Symposia-2015 on Quantum Science Symposium and Crystal/Graphene Science Symposium, Cambridge, MA, Sep. 22, 2015
11. 2nd ACM International Conference on Nanoscale Computing and Communication, Boston, MA, Sep. 21st, 2015 [Keynote].
12. DARPA Workshop on Transient Materials, Arlington, VA, Sep. 2nd, 2015.
13. The Second Blavatnik Science Symposium, New York City, NY, Aug. 6th, 2015.
14. The Eleventh International Fab Lab Conference (Fab11), Boston, MA, Aug. 5th, 2015.
15. The Third Annual Workshop on Micro- and Nanotechnologies in Medicine, Cambridge, MA, July 28th, 2015.

16. Gordon Conference on Nucleosides, Nucleotides and Oligonucleotides, Newport, RI, July 1st, 2015.
17. Gordon Conference on Synthetic Biology, Newry, ME, June 30th, 2015.
18. Albany 2015: The 19th Conversation, Albany, NY, June 11th, 2015.
19. ECI Conference on Photonics for Biology, Medicine and Surgery, Vail, CO, June, 7th, 2015.
20. 1st Synthetic Biology Young Scholar Forum (SynBioYSF), Beijing, May 23rd, 2015 [Keynote].
21. College of Chemistry and Molecular Engineering, Beijing, May 22nd, 2015.
22. Yale Chemical Biology Symposium, New Haven, CT, May, 15th, 2015.
23. Department of Materials Science and Engineering, Northwestern University, Evanston, IL, May 12, 2015
24. First Active Matter Summit, Cambridge, MA, Apr. 24, 2015.
25. Harvard Chinese Life Sciences Annual Research Symposium, Boston, MA, March 28, 2015.
26. SPIE Photonics West BiOS, San Francisco, CA, Feb. 7th, 2015 [Keynote].
27. Frontiers in Quantitative Biology Seminar Series, Stanford University, Stanford, CA, Jan 15th, 2015.
28. Joint seminar for the Graduate Programs in Bioengineering and Pharmaceutical Sciences and Pharmacogenomics, UCSF, San Francisco, CA, Jan 13th, 2015.
29. Department of Pathology, Brigham and Women's Hospital, Boston, Dec. 16th, 2014.
30. IEEE EMBS Micro and Nanotechnology in Medicine Conference, Oahu, HI, Dec. 9th, 2014.
31. Sino-US Synthetic Biology Workshop, Tianjin, China, Dec. 6th, 2014.
32. School of Life Sciences, Tsinghua University, Beijing, China, Dec. 5th, 2014.
33. Biodynamic Optical Imaging Center, Peking University, Beijing, China, Dec. 5th, 2014.
34. Fall Symposium of the New England Society for Microscopy, Cambridge, MA, Nov. 20th, 2014.
35. Department of Physics, Columbia University, New York, NY, Nov. 10th, 2014.
36. Department of Materials Science and Engineering, Boise State University, Boise, ID, Nov. 7th, 2014.
37. Department of Chemical and Biomolecular Engineering, Lehigh University, Bethlehem, PA, Oct. 29th, 2014.
38. Naval Research Laboratory, Washington, DC, Oct. 20th, 2014.

39. Fourth U.S.-China Symposium on Nanobiology and Nanomedicine, Washington, DC, Oct. 17th, 2014.
40. Twenty Ninth Chinese Chemical Society Annual Conference, Beijing, China, Aug. 5th, 2014.
41. EITA-New Media and Bio, Cambridge, MA, July 31st, 2014.
42. Second Annual workshop on Micro- and Nanotechnologies in Medicine, Cambridge, MA, July 31st, 2014.
43. SEED2014 (Synthetic Biology, Engineering, Evolution and Design), Manhattan Beach, CA, July 14th, 2014 [Young Investigator Award Lecture].
44. Blavatnik Science Symposium, New York City, NY, July 8th, 2014.
45. Seventh world congress on biomechanics (DNA mechanics and assembly session), Boston, MA, July 6th, 2014.
46. Seventh world congress on biomechanics (Nucleic acid nanostructures session), Boston, MA, July 6th, 2014.
47. BIO International Convention, San Diego, CA, June 25th, 2014.
48. International Workshop on DNA-Based Nanotechnology: Digital Chemistry (DNATEC14), Dresden, Germany, May 5th, 2014.
49. Oncology Research Seminars, Koch Institute for Integrative Cancer Research, MIT, Cambridge, MA, Apr. 28th, 2014.
50. Foundation of Nanosciences (FNANO2014), Snowbird, Apr. 16th, 2014.
51. Biological Imaging Meeting, Northwestern University, Evanston, March 28th, 2014.
52. National Academy of Sciences Frontiers of Engineering and Sciences Program, Rio de Janeiro, Brazil, March 17th, 2014.
53. Biophysics seminar, MIT, Cambridge, MA, Feb. 12th, 2014.
54. NSF Nanoscale Science and Engineering Grantees Conference: Current and Future Trends in Nanotechnology and Environment and Nanomanufacturing, Arlington, VA, Dec. 6th, 2013.
55. The Nano Area of Advance at Chalmers University of Technology, Goteborg, Sweden, Nov. 28th, 2013.
56. Fourth International Congress on Biotechnology and Genomics, Monterrey, Mexico, Nov. 21st, 2013.
57. NIH Common Fund High-Risk High-Reward Symposium, Bethesda, MD, Nov. 18th, 2013.
58. Department of Chemistry and Biotechnology, Tokyo University, Tokyo, Japan, Oct. 30th, 2013.

59. Annual Meeting of Chem-Bio Informatics Society, Tokyo, Japan, Oct. 30th, 2013.
60. Department of Chemistry, Kyoto University, Kyoto, Japan, Oct. 29th, 2013.
61. Annual Meeting of the Biophysical Society of Japan, Kyoto, Japan, Oct. 29th, 2013.
62. U.S. Army Research Office Workshop: Challenges and Opportunities in Nanomanufacturing, Raleigh, NC, Oct. 1st, 2013.
63. International Conference on DNA Computing and Molecular Programming, Tempe, Sep. 27th, 2013.
64. Symposium on Advanced Nano/Biosystems: Design, Fabrication, and Characterization, University of Illinois at Urbana-Champaign, Urbana, IL, Sep. 25th, 2013.
65. The Ninth International Fab Lab Conference, Yokohama, Japan, Aug. 21st, 2013.
66. NSF Workshop on BioMEMS and Tissue Engineering, Cambridge, July 30th, 2013.
67. NSF Workshop: Advanced Biomanufacturing, Arlington, VA. July 14, 2013.
68. The Sixth International Meeting on Synthetic Biology (SB 6.0), London, England, July 10th, 2013
69. Sino-US Synthetic Biology Workshop, Tianjin, China, June 21st, 2013
70. Fourth International Wyss Symposium: Nanotherapeutics and Diagnostics, Boston, June 6th, 2013

#### 9.4 Patents

- Yin, Peng; Jungmann, Ralf; Dai, Mingjie; Avendano, Maier; Woehrstein, Johannes. Quantitative DNA-Based Super-Resolution Imaging US 14/908,333. *Pending*
- Yin, Peng; Sun, Wei; Shen, Jie. Using DNA Nano Structure as Reactive Ion Etching Mask for Nanoelectronics Fabrication PCT/US15/32198. *Pending*
- Yin, Peng; Agasti, Sarit; Chen, Xi; Jungmann, Ralf. High Throughput and Highly Multiplexed Imaging with Programmable DNA Probe US 15/108,911. *Pending*
- Yin, Peng; Han, Dongran; Myhrvold, Cameron. Single-Stranded DNA Nanostructures PCT/US16/20893. *Pending*
- Yin, Peng; Chen, Xi; Schaus, Thomas. Microscope-Free Imaging PCT/US16/15503. *Pending*
- Yin, Peng; Agasti, Sarit; Gopalkrishnan, Nikhil; Yaghi, Omar Target Detection Using Barcode Conjugates and Nucleic Acid Nanoarrays 62/264,821. *Pending*
- Yin, Peng; Chen, Xi; Zhang, Mengmeng. Nucleic Acid Retro-Activated Primers PCT/US16/25480. *Pending*

- Yin, Peng; Silverberg, Jesse. Molecular Atlas Project - Software and Netware. *Disclosure*
- Yin, Peng; Gutierrez, Juanita Maira Lara; Amado, Maier S. Avendano. Proximity PAINT: DNA-Based Super Resolution Imaging of Protein-Protein Interactions 62/202,327. *Pending*
- Yin, Peng; Jungmann, Ralf; Wohrstein, Johannes. Kinetic Barcoding. *Disclosure*
- Yin, Peng; Ong, Luvena; Simmel, Friedrich; Scheible, Max. Compact DNA Cubes. *Disclosure*
- Yin, Peng; Kim, Jongmin. Synthetic biological circuits for transient materials 62/213,217. *Pending*
- Yin, Peng; Kishi, Jocelyn; Schaus, Thomas; Xuan, Feng; Gopalkrishnan, Nikhil; Woo Sungwook. Primer exchange reactions (PER) 62/296,310; 62/299,206. *Pending*
- Yin, Peng; Beliveau, Brian; Kishi, Jocelyn. Probe Miner: A rapid, genome-scale oligo FISH probe discovery pipeline. *Disclosure*
- Yin, Peng; Schaus, Thomas; Gopalkrishnan, Nikhil. Sequence design for efficient assembly of nucleic acid structures 62/321,146. *Pending*
- Yin, Peng; Xuan, Feng; Dai, Mingjie; Chen, Xi. Conditional Primer Extension for Single-Molecule Protein Detection 62/342,401. *Pending*
- Yin, Peng; Kishi, Jocelyn; Xuan, Feng. Primer Exchange Reactions (PER) for Isothermal PCR 62/342,295. *Pending*

## 9.5 Awards/Honors

- World Economic Forum Young Scientist Award, 2015 Press:<https://agenda.weforum.org/news/world-economic-forum-honours-its-2015-young-scientists-community-at-annual-meeting-of-the-new-champions/>
- Blavatnik National Awards for Young Scientists, Finalist, 2015 Press:<http://www.prnewswire.com/news-releases/blavatnik-national-awards-for-young-scientists-announce-2015-finalists-300087851.html>
- World Economic Forum Young Scientist Award, 2014 Press:<http://www.weforum.org/news/world-economic-forum-honours-its-2014-young-scientists-community-annualmeeting-new-champions>
- Inaugural ACS Synthetic Biology Young Scientist Award, 2014 Press:<http://wyss.harvard.edu/viewpage/540/>
- Blavatnik National Awards for Young Scientists, Finalist, 2014 Press:<http://blavatnikawards.org/news/items/blavatnik-national-awards-announce-30-finalists/>
- NIH Transformative Research Award, 2013 Press:<http://wyss.harvard.edu/viewPressrelease/126/>

- NSF Expedition in Computing Award, 2013 Press:[http://www.nsf.gov/news/news\\_summ.jsp?cntn\\_id=128886](http://www.nsf.gov/news/news_summ.jsp?cntn_id=128886)

## **10 Award participants**

**Military Personnel:** N/A

**Salary Support:** Peng Yin, Nikhil Gopalkrishnan, Alexander Green, Dongran Han, Ralf Jungmann, Jie Shen, Yonggang Ke, Feng Xuan, Luvena Ong, Mingjie Dai, Cameron Myhrvold, John Paul Sadowski, Johannes Woehrstein, Maximilian Strauss, Juanita Lara and Anita Chandrahas.

## **11 Appendix**

The relevant reprints, preprints and technical reports are included in this appendix.

# Kinetic Self-assembly of DNA Tiles and Bricks

Peng Yin, Harvard University

## 2016

- Thomas E. Schaus, Sungwook Woo, Feng Xuan, Xi Chen and Peng Yin. A Biochemical Nanoscope via Auto-cycling Proximity Recording. Submitted
- Cameron Myhrvold, Michael Baym, Nikita Hanikel, Luvena Ong, Jonathan Gootenberg and Peng Yin. Barcode Extension for Analysis and Reconstruction of Structures (BEARS). In preparation
- Nikhil Gopalkrishnan, Thomas Schaus and Peng Yin. Eliminating Kinetic Barriers through Sequence Design Allows Rapid Assembly of DNA Origami at Room Temperature. In preparation
- R. Jungmann†, M. S. Avendano†, M. Dai, J. B. Woehrstein, S. S. Agasti, Z. Feiger, A. Rodal and P. Yin. Quantitative super-resolution imaging with qPAINT, *Nature Methods*, 13, 439-442 (2016)
- M. Dai, R. Jungmann and P. Yin. Optical imaging of individual biomolecules in densely packed clusters. *Nature Nanotechnology* (2016).

## 2015

- C. Grun, J. Werfel, D. Y. Zhang, and P. Yin. DyNAMiC Workbench: an integrated development environment for dynamic DNA nanotechnology. *J. R. Soc. Interface*. 12: 20150580 (2015)

## 2014

- K. Pardee, A.A. Green, T. Ferrante, E. Cameron, A. DaleyKeyser, P. Yin, and J.J. Collins. Paper-based synthetic gene networks. *Cell*159:940-954
- A.A. Green, P.A. Silver, J.J. Collins, P. Yin. Toehold switches: de-novo-designed regulators of gene expression. *Cell* 159: 925-939 (2014)
- B. Wei, L.L. Ong, J. Chen, A. Jaffe, P. Yin. Complex reconfiguration of DNA nanostructures. *Angewandte Chemie* 126:7605-7609 (2014).
- J. Sadowski, C.R. Calvert, D.Y. Zhang, N.A. Pierce, P. Yin. Developmental self-assembly of a DNA tetrahedron. *ACS Nano* DOI: 10.1021/nn4038223 (2014).
- R. Jungmann†, M.S. Avendano†, J.B. Woehrstein†, M. Dai, W.M. Shih, P. Yin. Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nature Methods*11(3):313-318 (2014)

## 2013

- B. Wei, M. Dai, C. Myhrvold, Y. Ke, R. Jungmann, P. Yin. Design space for complex DNA structures. *J. Am. Chem. Soc.* 135 (48):18080-8 (2013)
- C. Myhrvold, M. Dai, P.A. Silver, P. Yin. Isothermal self-assembly of complex DNA structures under diverse and biocompatible conditions. *Nano Letters* 13(9): 4242-8 (2013)

# A Biochemical Nanoscope via Auto-cycling Proximity Recording

Thomas E. Schaus<sup>1,2</sup> Sungwook Woo<sup>1,2,3</sup> Feng Xuan<sup>1,2,3</sup> Xi Chen<sup>1,2</sup> Peng Yin<sup>1,2\*</sup>

Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115<sup>1</sup>

Department of Systems Biology, Harvard Medical School, Boston, MA 02115<sup>2</sup>,

Authors contributed equally to this work<sup>3</sup>,

E-mail address\*: [py@hms.harvard.edu](mailto:py@hms.harvard.edu)

**Abstract.** Analysis of the spatial arrangement of molecular features enables the engineering of synthetic nanostructures and the understanding of natural ones. Microscopy methods can provide direct visualization of complex spatial features but typically require sophisticated equipment with limited throughput. Here, we present a microscope-free “biochemical nanoscopy” method that records sophisticated nanostructure features in situ for later readout. The method is based on a conceptually novel Auto-cycling Proximity Recording (APR) mechanism, which continuously and repeatedly produces proximity records of any nearby pairs of DNA-barcoded probes, at physiological temperature, without destroying the probes themselves. We demonstrate the proximity and auto-cycling characteristics, and then apply the system as a biochemical nanoscope to accurately decode various spatial arrangements of 7 unique probes in homogeneous test samples. We further demonstrate the ability of APR to re-sample the same dynamic system repeatedly. Future development is expected to enable readout with massively parallel sequencing methods, creating a truly single-molecule technique.

## 17 A foundational tool for studying nanostructure

18 Spatial organization is fundamental to the function of syn-  
 19 thetic<sup>1-5</sup> and biological<sup>6-8</sup> nanostructures. The organization of  
 20 such nanostructures can be described as a set of pairwise prox-  
 21 imities between molecular components. Single-molecule meth-  
 22 ods to study proximity and organization must examine individ-  
 23 ual nanostructures with molecular-scale precision, and are foun-  
 24 dational to advancing nanoscience.<sup>9</sup>

25 Though limited to multiplexing a modest number of simul-  
 26 taneous species, direct visualization by electron, atomic force,  
 27 and optical microscopy have identified individual macromolec-  
 28 ular associations in synthetic<sup>4,5,10</sup> and biological<sup>11-13</sup> systems,  
 29 and achieve molecular resolution in controlled environments.<sup>14</sup>  
 30 Resonance energy transfer (FRET) techniques further allow for  
 31 dynamic measurement of pairwise proximity in solution.<sup>15,16</sup>

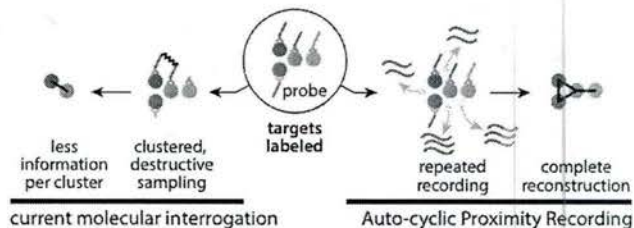
32 Complementary to the sophisticated microscopy platforms  
 33 are biochemical techniques that have long enabled molecular-  
 34 scale precision, but typically make only ensemble measurements  
 35 of simple pairwise interactions. Techniques such as Yeast Two-  
 36 Hybrid<sup>17-19</sup> and related<sup>20</sup> assays, affinity purification / mass  
 37 spectroscopy,<sup>21</sup> and Co-Immuno Precipitation<sup>22</sup> are common in  
 38 elucidating protein interaction networks, and are relatively eas-  
 39 ily parallelized, executed, and even automated.

40 Barcoding with DNA has recently given biochemical tech-  
 41 niques the potential for enormous multiplexing. Proximity Liga-  
 42 tion (PLA)<sup>23</sup> or Extension Assays (PEA),<sup>24</sup> which log the co-  
 43 localization of two probes by ligating or extending probe-bound  
 44 DNA sequences, now routinely multiplex ~100 signals with or-  
 45 thogonal sequences. The information content of DNA itself  
 46 is much larger, however. There are  $4^N$  combinations of  $N$   
 47 nucleotides, enabling 1 million-plex with only 10-nucleotide  
 48 strings. The destructive nature of PLA and PEA, wherein only  
 49 a single association can be measured for each probe pair, limits  
 50 the utility of unique barcoding in these methods; it is possible to  
 51 measure pairings uniquely, but not to connect those unique pairs  
 52 into large structures on the single molecule level.

53 Here, we present a novel biochemical interrogation technique,  
 54 termed "Auto-cycling Proximity Recording" (APR), for the con-  
 55 tinuous generation of molecular records of target proximity  
 56 (Fig. 1). In contrast to existing methods, proximity data are gen-  
 57 erated continuously and repeatedly, at tunable distances and un-  
 58 der physiological conditions, by non-destructively copying pairs  
 59 of target-bound, DNA-barcoded probes into each new molecu-  
 60 lar record. APR thus allows for re-sampling of altered molecu-  
 61 lar arrangement, a high signal-to-noise ratio, and the potential  
 62 for developing single molecule techniques. The resulting set of  
 63 proximity pairs represents a rich set of information from which to  
 64 reconstruct spatial organization, and allows us to demonstrate  
 65 a "biochemical nanoscope" that decodes the complex spatial ar-  
 66 rangements of 7 unique APR probes on a homogeneous test  
 67 sample. Like other DNA barcoded methods, APR is in principle  
 68 compatible with massively-parallel sequencing readout. Future  
 69 development will enable the APR nanoscope to be a truly single-  
 70 molecule technique, bridging microscopy's capacity for analyz-  
 71 ing individual nanostructures with the molecular-scale precision  
 72 and broad applicability of biochemical techniques.

### 73 The APR mechanism

74 The APR reaction proceeds in an autonomous, cyclic fash-

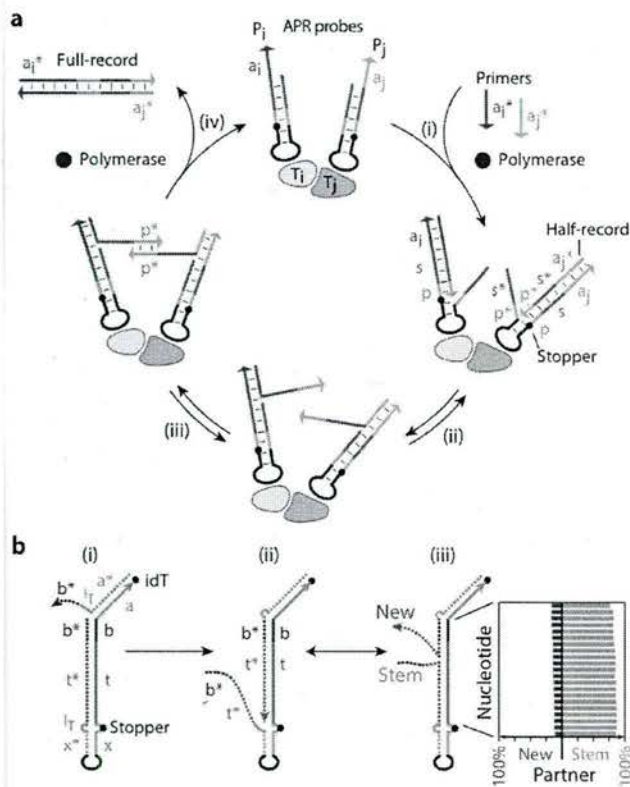


**Figure 1: Concept.** From a cluster of labeled targets (center), Auto-cycling Proximity Recording (APR, right) generates repeated records of any nearby probe pairs by copying sequences encoded in the probes. This allows re-sampling of the same targets in different organization or, in principle, full reconstruction of an individual cluster. In contrast, current analysis techniques (left) provide limited macromolecular geometry information and typically destroy associations as they are measured, precluding repeated sampling.

503 ion to convert soluble primers into molecular "records" of probe  
 504 pair proximity (Fig. 2a). DNA Hairpin probes  $P_i$  and  $P_j$  are  
 505 first attached to target molecules  $T_i$  and  $T_j$  via antibodies, ap-  
 506 tamers, direct nucleic acid hybridization, or other means, and  
 507 encode the identity of the target type within unique primer bind-  
 508 ing sequences (labeled as sequence domains  $a_i$  and  $a_j$ ). In step  
 509 (i), the soluble primers (sequences  $a_i^*$  and  $a_j^*$ ) stably bind their  
 510 respective hairpin probe type and are extended by a polymerase  
 511 through the spacer ( $s$ ) and palindromic ( $p$ ) domains up to the  
 512 "stopper" site, thereby producing "Half-records" ( $a_i^* - s^* - p^*$   
 513 and  $a_j^* - s^* - p^*$ ). (ii) Half-records are partially displaced by  
 514 spontaneously re-forming hairpin stems ( $s^* - p^* / p - s$ ) and (iii)  
 515 bind any nearby Half-records at their 3' palindromic ( $p^*$ ) seg-  
 516 ments. (iv) These sequences are again extended, this time over  
 517 the partner spacer  $s^*$  and primer  $a^*$  segments, effecting the re-  
 518 lease of soluble, double-stranded "Full-records" that carry both  
 519 probe identities as sequences  $a_i^*$  and  $a_j^*$ . Probes, thus regene-  
 520 rated to their initial state, may undergo additional cycles in the  
 521 same or other pairings. Upon termination of this cyclic record-  
 522 ing reaction, a specific Full-record is tested for by sample PCR  
 523 amplification with the respective primer pair ( $a_i, a_j$ ) and subse-  
 524 quent detection of Full-record copies by gel electrophoresis. In  
 525 this manner, for example, an isolated sample with PCR primers  
 526 for  $a_i$  and  $a_j$  would yield a positive result in the presence of  
 527  $T_i-T_j$  proximity.

528 This cycle solves two main challenges in generating multi-  
 529 ple records between any nearby pairs: (1) a template sequence  
 530 (in this case,  $s^* - p^*$ ) must be copied from the probe and then  
 531 spontaneously made single-stranded and available for further re-  
 532 action, and (2) in the context of probe proximity, pairs of copied  
 533 sequences must be combined into a single molecular record and  
 534 then released.

535 The first challenge is solved by a modified, DNA-based  
 536 "Copy-and-Release Hairpin" (CRH) that enables the isother-  
 537 mal copy and release of an arbitrary sequence onto each primer  
 538 strand. The detailed mechanism is as follows (Fig. 2b): (i) DNA  
 539 primer strand  $a^* - l_T - b^*$  (i.e., of domains  $a^*$  and  $b^*$  with  
 540 intervening single T nucleotide linker  $l_T$ ) binds the 3' end of  
 541 the hairpin at primer binding domain  $a$ . (ii) The primer and  
 542 hairpin  $b^*$  domains compete for short probe domain  $b$ . Primer  
 543 domain  $b^*$  occasionally binds, despite the resulting single  $T$  nu-  
 544 cleotide  $l_T$  bulge, and is extended by a displacing polymerase  
 545



**Figure 2: Auto-cycling Proximity Recording (APR) mechanisms.** (a) The APR cycle creates molecular records of proximity from pairs of hairpin probes. Probe-specific primers are (i) extended to “Half-records” and (ii) reversibly displaced, (iii) bind palindromic domains of nearby Half-records, and (iv) are extended on each other to create and displace “Full-records,” regenerating the probes. See text for detailed description and SI Note 1 for other considerations. (b) Copy-and-Release Hairpin (CRH) detail. Shown is the mechanism of (i) initial primer binding, (ii) extension, and (iii) random walk of the strand displacement branch. The hairpin template strand  $t - b$  is computationally predicted to pair predominantly with the hairpin stem complement (iii, inset).

(here, Bst, New England Biolabs) through stem template domain  $t$ . When primer extension reaches a “stopper” modification near the loop (here, covalent linker triethylene glycol “Spacer 9,” Integrated DNA Technologies, opposite single T  $l_T$ ), the polymerase dissociates. (iii) The new extension and the displaced hairpin stem, being identical in sequence ( $b^* - t^*$ ), rapidly compete for stem template domains  $t - b$  via the strand displacement mechanism<sup>25</sup> with millisecond transit times.<sup>26,27</sup> Computer models (NUPACK<sup>28</sup>) predict that the stem template  $t - b$  is more likely to be bound to its stem-based complement than to the new primer extension (iii, plot, as well as SI Note 1), ensuring the Half-record is largely single-stranded and available for further reaction. In contrast, because sequence  $a^*$  is longer than  $\sim 11$  nt, the Half-record remains stably bound at the primer binding site  $a$  until forcibly displaced by a polymerase.

The advance of this copy-and-release mechanism is not in copying a new sequence, of course, but in autonomously displacing that stably-bound sequence to reveal a single-stranded product without external thermal or chemical influence. Simply extending a primer on a single-stranded template, as in PCR,

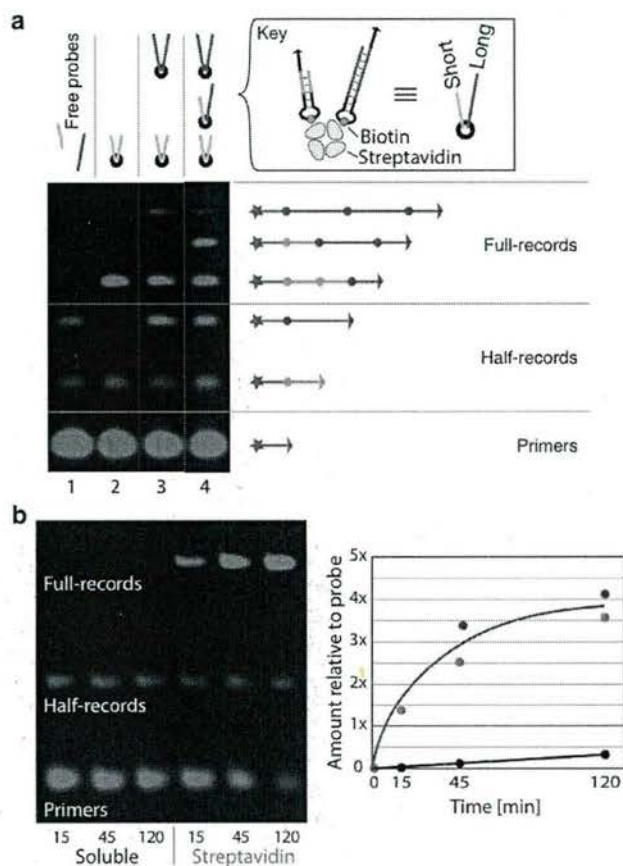
would leave the product stably bound without external manipulation, and therefore unavailable for further reaction. Here, template strand  $t - b$  is bound to a complementary hairpin competitor strand  $b^* - t^*$ . The free energy released from dNTP additions, in the presence of a displacing polymerase, allows the otherwise unfavorable hairpin opening and template copying. Once the polymerase is released at the stopper, the hairpin stem can close again to displace the new strand.

The details of the CRH probe are important to optimal functioning. The Spacer9 stopper modification, opposite an unpaired “T” ( $l_T$ ) to create a small internal loop adjacent to  $x$ , was chosen empirically for its short,  $\sim 1$  nt length and its ability to consistently stop the polymerase in this context. Even at this short length, however, the internal loop slows the ability of the hairpin competitor strand  $b^* - t^*$  to re-bind template strand  $t - b$  and displace new sequence  $b^* - t^*$ . Thermodynamic calculations (NUPACK, see SI Fig. 1) and experimental evidence (not shown) confirm that this leads to new strand extension  $b^* - t^*$  preferentially binding template  $t - b$  and unavailable for further reaction. The unpaired nucleotide within the  $a^* - l_T - b^*$  primer sequence was added to balance this preference (Fig. 2b, ii), adding an even weaker bulge loop to the other end.<sup>29</sup> Furthermore, to minimize secondary structure and prevent spurious binding and extension of the single-stranded new sequence  $b^* - t^*$  (and other segments), the primer-binding ( $b - a$ ) and template ( $t$ ) domains utilize only three base types (A, T, and G). See SI Note 1 for other considerations.

The second challenge, linking these Half-records together in the context of proximity, and then releasing a single Full-record, is solved by copying a short, 6 nucleotide palindromic sequence  $p$  (AGGCCT) to the 3' end of the Half-record. Here, a palindrome is defined such that  $p = p^*$ , enabling any Half-record to bind to any other Half-record by their 3' ends. Subsequent extension by the polymerase in solution copies the information on each Half-record onto the other, as well as ultimately displaces and releases the nascent Full-records from the probes. Note that the non-specific reaction of unbound Half-record is prevented by two features: the  $p$  domain is not part of the soluble primer but only added upon Half-record formation, and these Half-record are only released once they pair. Both features limit reactive,  $p$ -containing strands free-floating in solution. See SI Fig. 2 for further probe detail and sequences.

### Proof of principle and characterization

Figure 3 presents the proof of principle experiments, demonstrating proximity-based, auto-cycling, Full-record generation. Fig. 3a focuses on the proximity dependence of reactions. Lanes 1-4 each contained combinations of 2 probe types, made different only in spacer length for the purpose of differentiating products on gel electrophoresis, together with a common Cy5 fluorophore-labeled primer and polymerase (not depicted at top). In lane 1, short and long barcode probes existed unbound in solution. The primers (red, in excess) were extended to short or long Half-records, respectively, but no Full-records were generated because probes were not in close proximity of each other. In lane 2, short barcode probes were held in proximity on streptavidin molecules, and the system produced short Half-records and, in turn, short Full-records, as expected. When streptavidin was prepared in separate vessels with either short or



**Figure 3: Proof of principle experiments.** (a) Generation of Full-records requires the co-localization of probes, here by biotinylated hairpin loops bound to streptavidin, whereas isolated probes can always generate Half-records. Cropped denaturing PAGE gel depicting 10  $\mu$ l reactions (40 min at 37°C) with biotin-streptavidin association and 4:1 overall probe:streptavidin stoichiometry (inset), 8 and 22 nt barcodes (19 and 33 nt stem lengths copied), 10:1 primer:probe, and 40 nM total probe concentration. A single primer sequence was used and no secondary amplification was performed. (b) Auto-cycling is also demonstrated by quantification of Cy5-labeled probes on cropped denaturing PAGE gels. Reaction and gel conditions identical to those of (a) except: 10 nt spacer (21 nt stem length copied), 5:1 primer:probe with probes still at 40 nM total. Light green dots are from gel shown, dark green from repeat experiment. See probe details and sequences in SI Fig. 2 and full gels in SI Fig. 3.

long probes, and then mixed together for reaction (lane 3), both Half-records and Full-records of short and long lengths were generated. Interaction among probes from different streptavidin molecules would have also produced Full-records of intermediate length (i.e., of long + short Half-records), but these were instead only present when short and long probes were randomly arranged on the same streptavidin molecules, as in lane 4.

Figure 3b demonstrates the cycling ability of the probes by observing product formation over time. Because primers were each labeled with a single copy of Cy5, the amount of product in each band could be quantified. Beginning with a 5-fold abundance of primers with respect to probes, probes unbound in solution generated few Full-records of background signal. When held on streptavidin, however, record quantification reveals that

each probe was read multiple times on average, with an initial record generation rate of one per 7 min. Conclusive proof that a given individual probe sequentially generates records with multiple different partner probes was provided by massively parallel sequencing (see SI Fig. 4).

Probe design dictates the “reach” of a probe pair, or probe-probe proximity within which Full-records can be produced. To characterize the rates of Full-record production as a function of probe-probe distance, pairs of probes were fixed to programmed positions on 2D DNA origami structures<sup>3</sup> (Fig. 4a). Three types of probes were tested, encompassing three probe spacer lengths and two attachment methods. Two types were attached as extensions of intrinsic origami “staple” strands (Fig. 4a, left inset), while a third was held by a click-chemistry azide-alkyne linkage to an intermediate strand (right inset) to demonstrate how probes may be attached to arbitrary moieties. Both methods incorporated single-stranded DNA as flexible linkers at the base of each probe. Figure 4b diagrams the flexible, single-stranded and the rigid, double-stranded components contributing to the reach between two probes.

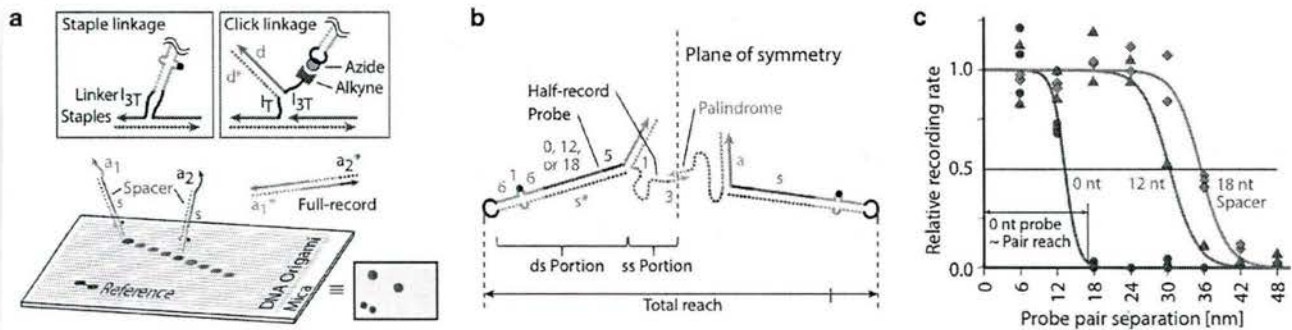
Recording reactions for each probe and separation distance were carried out separately. Many copies of a given origami structure were held flat and immobile on a mica surface, as is done for atomic force microscopy of DNA structures (see SI Note 2 and SI Fig. 7 for details of how this stabilizes structures). Records were then amplified by PCR, and products quantified by gel electrophoresis. To account for variation in experimental conditions, especially in the number of origami, a second probe pair with orthogonal PCR sequences was present on each origami type, but always at the same fixed separation distance (Fig. 4a). The rate of production was calculated with respect to this reference pair.

Figure 4c shows the relative recording rate for three probe designs, containing 0, 12, and 18 nt spacer domains  $s$ . Each probe pair was tested every 6 nm for separations of 6 to 48 nm, and mathematical fits to record generation rates were normalized to a common maximum. Zero, 12, and 18 nt probes all produced records at near maximum rates when pairs were closest to one another. Rates were reduced to half at 13, 20, and 25 nm, respectively, and had a maximum reach (1% of maximum recording rate) at 17, 39, and 44 nm. In absolute terms, the 0 nt spacer probe produced records at the slowest rate, approximately half as fast as longer probes (not shown).

All three maximum-reach distances correspond well to the values expected when DNA probes and attached Half-records are oriented optimally, as in a straight chain (Fig. 4b). For the 0 nt spacer probe, the maximum expected distance is approximately twice the sum of the probe length (19 bp of double-stranded DNA, at 0.34 nm per base pair,<sup>32</sup> ~6.5 nm) and Half-record length (11 nt of single-stranded DNA minus 3 nt palindrome overlap, at the maximum 0.58 nm per base,<sup>32</sup> ~4.6 nm), totaling ~22 nm. Similarly, the 12 nt and 18 nt spacer probes have a maximum expected reach of ~36 and ~43 nm, respectively. See SI Note 3 and SI Fig. 8 for estimates of reach profiles using more sophisticated Worm-Like Chain models.

#### Application: Interrogation of complex geometries

Auto-cycling Proximity Recording can be used to interrogate the relative proximities of a group of targets, enabling recon-



**Figure 4: Probe reach.** (a) To characterize reach, relative probe positioning is controlled precisely. Probes were attached to 2D, rectangular, twist-corrected DNA origami structures measuring  $\sim 70$  by  $100$  nm (origami shown with all possible Fig. 4 probe positions, and equivalent cartoon with 4 probes; see SI Fig. 5 for geometry, SI Table 1 for staple sequences, and SI Fig. 6 for precise probe positions). After complete assembly, origami were deposited randomly on mica surfaces in the presence of  $12.5$  mM  $Mg^{2+}$  for firm adhesion, a condition we find protects origami from degradation by polymerase (see SI Note 2). Each probe was held by direct extension of two origami staples (inset, left) or covalent attachment to an intermediary oligonucleotide (inset, right; azide on probe loop covalently attached to alkyne on  $TTT-L$  intermediary via DIBO-based click chemistry and gel-purified). Separation of probes was measured at the origami attachment point. Records produced were collected from the supernatant without disturbing the underlying origami, and PCR-amplified with, for example,  $a_1$  and  $a_2^*$  primers. (b) The maximum “reach” across which a probe pair can make a Full-record is determined by the cumulative lengths of their components, including double-stranded probe components and single-stranded Half-record components. These are shown as rigid and flexible, respectively, corresponding to their persistence lengths of  $\sim 50$  nm<sup>30</sup> and  $\sim 2$  nm.<sup>31</sup> Numbers indicate length in nucleotides. (c) Otherwise identical probes with spacer lengths of 0 or 12 nt (attached by staple extension), or 18 nt (covalently attached), were held in pairs separated from 6 to 48 nm by 6 nm increments (red positions in (a)), recorded (1 hour at room temperature, 100 nM primers), and Log-phase PCR amplified (20 cycles, 500 nM primers) to gel-detectable levels. Denaturing PAGE band quantification was normalized to a constant reference pair for each well. The entire series for a given probe pair type was then fit by least squares recursion to a simple sigmoidal curve  $c1/(1 + Exp[c2(dist - c3)])$ , where  $dist$  represents the separation distance and  $c1$ - $c3$  were fit, and normalized to a maximum rate of 1. See probe sequences in SI Table 2.

267 structure of overall geometry (Fig. 5). Using a series of PCR  
 268 assays to test if each pairwise combination of probes produced  
 269 records (i.e., is within reach, per Fig. 4c), a list of individual  
 270 proximities was generated from which to reconstruct the under-  
 271 lying geometry. The longest probe tested in Fig. 4c, with an 18  
 272 nt spacer ( $s$ ), was first used to determine the relative proxim-  
 273 ities of three targets. Three probes, programmed with different  
 274 primer binding sequences ( $a$ ) and therefore generating records  
 275 with unique ends (e.g., ends  $a_i^*$  and  $a_j$  in Fig. 2a), were again  
 276 fixed by 2D origami in a triangular configuration with 30 nm  
 277 probe separation (Fig. 5a). This made each probe within the  
 278 reach of the other two. A recording was performed, as before,  
 279 and the sample was split into 3 volumes, reflecting the number  
 280 of pairwise probe combinations. Each subsample was combined  
 281 with a different pair of PCR primers, and records, if any, were  
 282 amplified. Gel electrophoresis and band quantification indicated  
 283 if records were produced, and a “graph” depicting relative pos-  
 284 itioning was computationally reconstructed. As expected, the  
 285 triangular arrangement yielded all three record types, suggest-  
 286 ing that very geometry. In contrast, when the same probes were  
 287 placed in a straight line (Fig. 5b), only two record types were  
 288 produced, as the distance between P1 and P2 was beyond probe  
 289 reach.

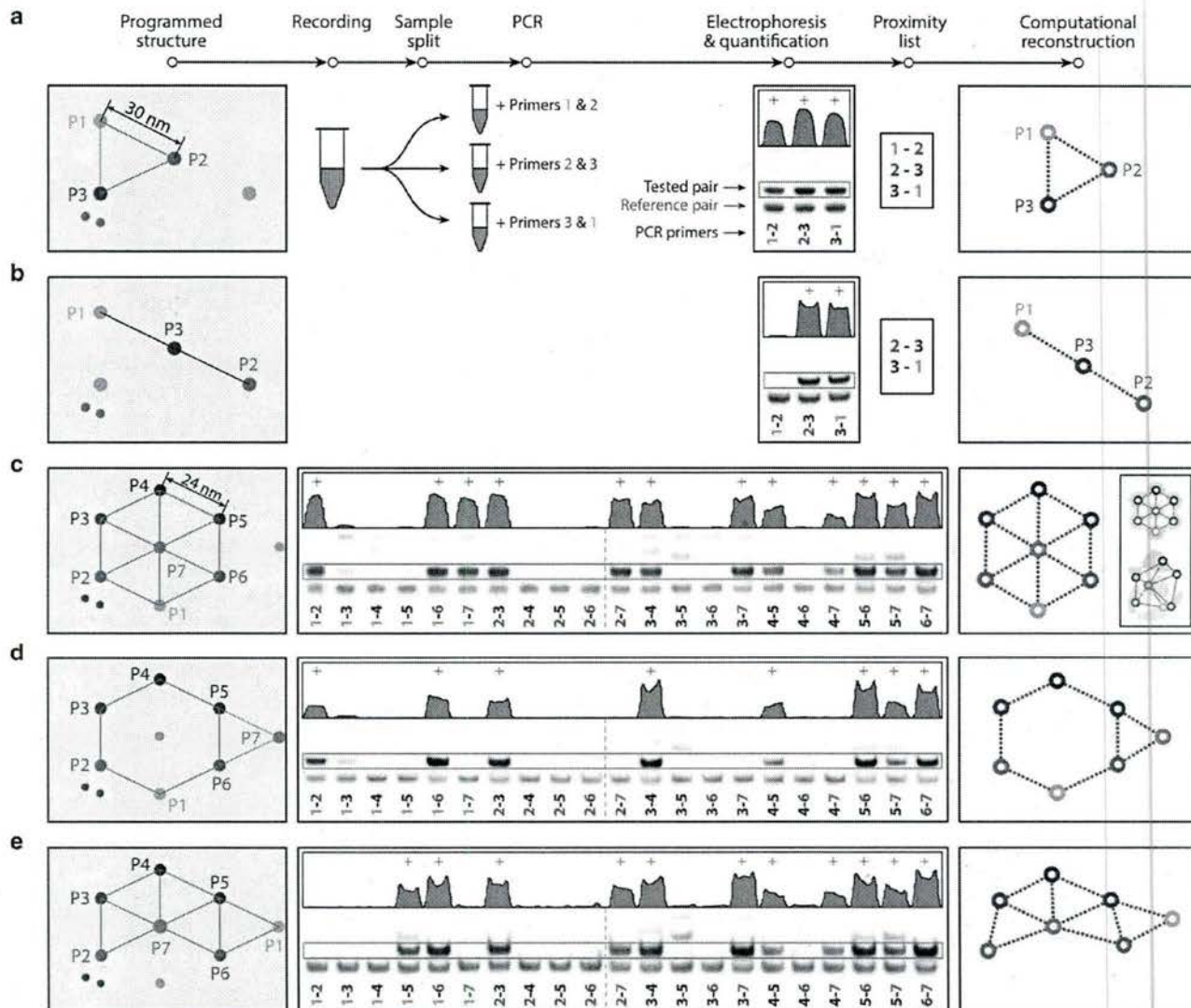
290 More complex, 7-point geometries were then tested, each  
 291 using shorter, 12 nt-spacer ( $s$ ) probes arranged on a smaller,  
 292 24 nm hexagonal pattern (Fig. 5c-e). This made non-adjacent  
 293 probes at least 42 nm apart, beyond the reach of these probe  
 294 pairs ( $\leq 39$  nm). Each 7-probe geometry required testing of  
 295 21 ( $6+5+4+3+2+1$ ) possible probe pairings, yielding one of  
 296  $2^{21} > 2 \times 10^6$  possible binary combinations, any potentially  
 297 valid. As in Fig. 5a, records were produced from these origami,  
 298 split, and amplified by PCR in the 21 possible primer pairings.

The hexagonal geometry of Fig. 5c yielded 12 of 21 positive  
 299 proximity pairs, as seen by gel electrophoresis, in addition to the  
 300 21 reference pairs that served as positive PCR controls. Simple  
 301 graphing of these 12 pairs (Wolfram Mathematica, “GraphPlot”  
 302 function, see SI Note 4) resulted in the correct geometry.  
 303

The exceptional fidelity of reconstruction in this case is due  
 304 to a combination of experimental and reconstruction details.  
 305 The graphing algorithm used here is based on a physical model  
 306 (“spring-electrical”) that separates the final probe positions rel-  
 307 atively uniformly, and is therefore particularly congruent with the  
 308 regular hexagonal organization of the origami probes. A sense  
 309 of the precision of this reconstruction is shown in Fig. 5c, inset.  
 310 Given the positions shown for any 6 probes, the 7th probe may  
 311 be anywhere within the shaded region and still satisfy the gel  
 312 data shown. In addition, a series of graphs with random probe  
 313 positions were generated and tested for compatibility with these  
 314 same data. One compatible example is shown in the Fig. 5c in-  
 315 set, lower graph, again with shaded regions showing the preci-  
 316 sion of a given probe position. See SI Fig. 10 for more examples.  
 317

Two further examples are demonstrated. Figure 5d, where  
 318 the center probe was moved to the outside, yielded the 8 correct  
 319 proximity pairings and geometry. Figure 5e, where the lower  
 320 probe was moved to the outside, yielded the 11 correct proxim-  
 321 ities and geometry.  
 322

Note that the relative quantity of Full-records created should  
 323 not be taken as a measure of relative distance between probes,  
 324 but only that the appearance of records should be taken to indi-  
 325 cate probes within reach (proximity); the non-linear record gen-  
 326 eration rate as a function of distance (Fig. 4c), variable probe  
 327 attachment to target, and possible sequence-dependence make  
 328 distance approximation unreliable from this simple measure. In-  
 329 stead, further refinement of target-target distances may be per-  
 330

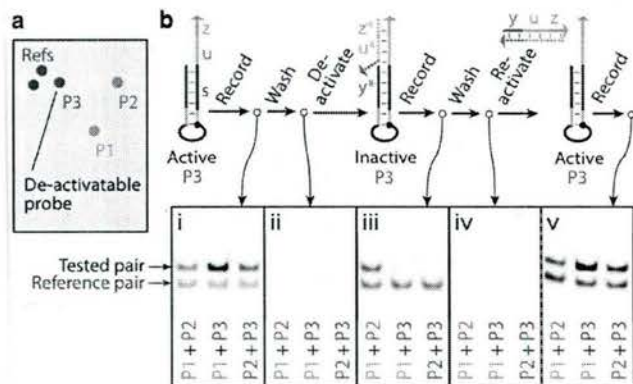


**Figure 5: Complex geometry interrogation.** (a) The same rectangular origami base used in Fig. 4b was adapted to hold three probes, of 18 nt spacer design, at 30 nm spacing (see SI Fig. 6). In the same conditions as in Fig. 4, probes were first placed in a triangular arrangement and recording was carried out. The sample was split, records PCR-amplified with separate pairs of probe primers (and reference primers), and products separated by gel electrophoresis and quantified (green plot profiles, marked (+) when judged to qualify for the proximity list). As expected, all three probe pairs yielded records, suggesting a triangular shape in reconstruction. (b) When those same probes were attached in a line, with P1 and P2 out of reach from one another, only two record types were amplified. This suggests a relatively linear arrangement. (c) More complex, 7-probe geometries were also interrogated, this time with 24 nm hexagonal grid spacing and 12 nt spacer probes. The full set of 21 primer pair amplifications yielded the gel shown in the center, with 12 pair in proximity. Computational reconstruction yielded the hexagonal pattern on the right. Inset shows positional precision of any single point, with respect to others shown, as shaded region that satisfies this gel. Top shows precision for computed hex pattern, bottom shows precision for random pattern that also satisfies gel result. (d) When center probe P7 was moved outside the hexagon, APR yielded the 8 expected proximities and correct reconstruction. (e) Similarly, when P1 was moved, the 11 expected proximities and correct reconstruction resulted. Throughout, computational reconstruction images were reflected and rotated, but not distorted, for best comparison to the programmed structure diagrams. See probe sequences in SI Table 2 and full gels in SI Fig. 9.

331 formed by sequential experiments with probes of different reach.  
 332 For example, a pair of 12 nt-spacer probes may produce records,  
 333 indicating targets closer than 36 nm, while a pair of 0 nt-spacer  
 334 probes at those same positions do not, indicating those same  
 335 targets are also farther than 18 nm. Alternatively, a single experi-  
 336 ment with a mixture of probes with unique primers and variable  
 337 reach could be used. See SI Fig. 11 for examples of how recon-  
 338 struction precision could be further refined with this strategy.

#### Application: Re-sampling of a dynamic system

339  
 340 In addition to the successful interrogation of complex geometries,  
 341 we demonstrate the advantages of non-destructive recording by  
 342 repeatedly sampling the same probes in a changing system. The  
 343 same origami-based, triangular arrangement of probes of Fig. 5a-b  
 344 was constructed, with a mechanism for inactivating probe P3 (Fig. 6a).  
 345 At each sampling point, the supernatant was split, PCR-amplified,  
 346 and observed by gel electrophoresis. First, the triangular arrange-  
 347 ment was recorded (with primer



**Figure 6: State change.** (a) The same origami probe triangle was used as in Fig. 5c(i), but probe P3 could be de-activated. (see SI Fig. 6 for precise probe positions.) (b) PCR amplification and denaturing PAGE of recording from active probe (i), after intermediate wash (ii), from de-activated probe (iii), after second intermediate wash (iv), and from re-activated probe (v). The intermediate wash results demonstrate that there were no leftover records. A single set of origami was used for all steps and re-sampled. Origami and PAGE otherwise treated as in Fig. 5. See probe sequences in SI Table 2 and full gels in SI Fig. 12.

348  $u^*$  on P3) and analyzed, indicating the co-localization of three  
 349 probes as expected (Fig. 6b, i). After a wash with buffer, the  
 350 supernatant showed no residual records (Fig. 6b, ii). Then, by  
 351 applying an inextensible (inverted dT at 3' end to prevent exten-  
 352 sion) P3 primer blocker  $z^* - u^* - y^*$  to the system, any leftover  
 353 P3 primers and Half-records were displaced and probe P3 was  
 354 made unable to bind primer, rendering it de-activated. Record-  
 355 ing then indicated co-localization of only P1 and P2 (Fig. 6b, iii).  
 356 To sample a third state, the same origami were again washed  
 357 and shown record-free (Fig. 6b, iv). Probes P3 were then  
 358 re-activated by addition of strand  $y - u - z$ , which stripped off the  
 359 primer blockers. Recording and analysis once again indicated  
 360 the full triangular arrangement (Fig. 6b, v), completing three  
 361 correct samplings of the same targets.

## 362 Discussion

363 Given that structure dictates function in synthetic and natu-  
 364 ral nanostructures, we sought to record nanoscale proximity and  
 365 organization in as precise and complete a manner as possible.  
 366 To this end, we have developed a biochemical technique termed  
 367 Auto-cycling Proximity Recording (APR). The two key mech-  
 368 anisms requiring development were a non-destructive, contin-  
 369 uous (i.e., “auto-cycling”) method for copying DNA strands from  
 370 probes, and a method for pairing these strands only in the pres-  
 371 ence of probe proximity (“proximity recording”). The first was  
 372 solved by the development of the Copy-and-Release Hairpin  
 373 (CRH), which allows a probe-specific primer to be extended by  
 374 a short template, and then isothermally released for further re-  
 375 action. The second, proximity requirement was enforced by en-  
 376 coding a short palindromic site at the 3' end of the extension.  
 377 The extended primers represent “Half-records” that must pair  
 378 by their palindromes with other nearby Half-records before they  
 379 can extend over each other, through their primer domains, and  
 380 be released into solution as “Full-records” of probe pair proxim-  
 381 ity. Subsequent identification of Full-record sequences reveals  
 382 proximity or putative interactions.

The APR method shares a superficial resemblance to Proxi- 382  
 mity Ligation (PEA) and Extension (PEA) techniques in that, 383  
 fundamentally, it records pairwise interactions in DNA. A key 384  
 difference is that, because the probes are not consumed in de- 385  
 tecting each proximity, many pairwise recordings can be made 386  
 from the same probe. This critical factor allows for not only bet- 387  
 ter signal-to-noise ratios, but also re-sampling of dynamic sys- 388  
 tems, and ultimately the interrogation of true single-molecule 389  
 networks. 390

We have taken advantage of the performance of APR to 391  
 demonstrate the detection and reconstruction of complex ge- 392  
 ometries and the re-sampling of dynamic interactions among the 393  
 same targets, all with simple PCR and gel-based readouts. Ge- 394  
 ometry reconstructions put 7 probes in the correct relative posi- 395  
 tion, a  $1\text{-in-}2 \times 10^6$  event by chance alone, and reconstruction 396  
 precision is tunable by using probes of different length. Re- 397  
 sampling allowed the same probes to be recorded 3 times in 398  
 different molecular-scale states, a task impossible with most as- 399  
 says. 400

Several simple modifications to probe design are apparent. 401  
 Probes can be attached to targeting moieties in a variety of 402  
 ways, and spacer length within the probe allows for reach to 403  
 be tuned. Within the probe stem sequence, the choice of a palin- 404  
 drome allows any probe to potentially produce Full-records with 405  
 any other probe. If target interactions are simple heterodimers, 406  
 however, stronger and more specific Half-Record binding can 407  
 be designed by using longer, probe-specific, non-palindromic 408  
 sequences. Cycling rates can perhaps be increased with these 409  
 longer, non-palindromic 3' sequences, by keeping probe spacers 410  
 short, or by using alternative means of biasing the Half-record 411  
 strand to remain single-stranded and available for pairing. Fi- 412  
 nally, a variety of stopper modifications can be used for optimal 413  
 performance under different conditions. 414

An assumption in describing the geometry of a complex de- 415  
 tected in bulk is that all structures are the same or similar. In 416  
 general, detection in bulk cannot differentiate between a single 417  
 structure and several overlapping pieces in the system. Be- 418  
 yond bulk geometry and re-sampling, however, the APR method 419  
 suggests the exciting prospect of single-molecule recording, 420  
 wherein globally unique barcodes, corresponding to “names” of 421  
 each individual target (rather than target species), are encoded in 422  
 the Full-records. Massively parallel Next Generation Sequenc- 423  
 ing could then be used as a readout to describe every interaction 424  
 on a per-molecule basis, allowing for protein counting, interac- 425  
 tion stoichiometry and turnover, real network connectivity (as 426  
 opposed to averaged or inferred connectivity), and other mea- 427  
 sures not possible with current techniques. We are actively pur- 428  
 suing these single-molecule capabilities. 429  
 430

## 431 References

- 432 Rajat Kanti Paul, Maziar Ghazinejad, Miroslav Penchev, Jian Lin,  
433 Mihrimah Ozkan, and Cengiz Sinan Ozkan. Synthesis of a pillared  
434 graphene nanostructure: A counterpart of three-dimensional carbon archi-  
435 tectures. *Small*, 6(20):2309–2313, 2010.
- 436 Tian Ming, Xiaoshan Kou, Huanjun Chen, Tao Wang, Hoi-Lam Tam,  
437 Kok-Wai Cheah, Ji-Yao Chen, and Jianfang Wang. Ordered gold nanos-  
438 tructure assemblies formed by droplet evaporation. *Angewandte Chemie*,  
439 120(50):9831–9836, 2008.
- 440 P.W.K. Rothemund. Folding DNA to create nanoscale shapes and patterns.  
441 *Nature*, 440(7082):297–302, 2006.

442 4. D. Wei, M. Dai, and P. Yin. Complex shapes self-assembled from modular  
443 molecular components. *Nature*, 485:623–626, 2012.

444 5. Y. Ke, L.L. Ong, W. M. Shih, and P. Yin. Three-dimensional structures  
445 self-assembled from dna bricks. *Science*, 2012.

446 6. Frank Alber, Svetlana Dokudovskaya, Liesbeth M Veenhoff, Wenzhu  
447 Zhang, Julia Kipper, Damien Devos, Adisetyantari Suprpto, Orit Kami-  
448 Schmidt, Rosemary Williams, Brian T Chait, et al. The molecular archite-  
449 cture of the nuclear pore complex. *Nature*, 450(7170):695–701, 2007.

450 7. Ronen Zaidel-Bar, Shalev Itzkovitz, Avi Ma'ayan, Ravi Iyengar, and Ben-  
451 jamin Geiger. Functional atlas of the integrin adhesome. *Nature cell biol-*  
452 *ogy*, 9(8):858–867, 2007.

453 8. Ryan Suderman and Eric J Deeds. Machines vs. ensembles: effective mapk  
454 signaling through heterogeneous sets of protein complexes. *PLoS Comput*  
455 *Biol*, 9(10):e1003278, 2013.

456 9. Albert-Laszlo Barabasi and Zoltan N. Oltvai. Network biology: un-  
457 derstanding the cell's functional organization. *Nature Reviews Genetics*,  
458 5:101–113, 2004.

459 10. Ralf Jungmann, Maier S Avendaño, Johannes B Woehrstein, Mingjie Dai,  
460 William M Shih, and Peng Yin. Multiplexed 3d cellular super-resolution  
461 imaging with dna-paint and exchange-paint. *Nature methods*, 11(3):313–  
462 318, 2014.

463 11. Tatyana M Svitkina and Gary G Borisy. Arp2/3 complex and actin depoly-  
464 merizing factor/cofilin in dendritic organization and treadmilling of actin  
465 filament array in lamellipodia. *The Journal of cell biology*, 145(5):1009–  
466 1026, 1999.

467 12. Peter Hinterdorfer and Yves F Dufrene. Detection and localization of single  
468 molecular recognition events using atomic force microscopy. *Nature*  
469 *methods*, 3(5):347–355, 2006.

470 13. Pakorn Kanchanawong, Gleb Shtengel, Ana M Pasapera, Ericka B Ramko,  
471 Michael W Davidson, Harald F Hess, and Clare M Waterman. Nanoscale  
472 architecture of integrin-based cell adhesions. *Nature*, 468(7323):580–584,  
473 2010.

474 14. M. Dai, R. Jungmann, and P. Yin. Optical imaging of individual  
475 biomolecules in densely packed clusters. *Nature Nanotechnology*, 2016.

476 15. P. Santangelo, N. Nitin, and G. Bao. Nanostructured probes for RNA detec-  
477 tion in living cells. *Annals of Biomedical Engineering*, 34(1):39–50, 2006.  
478 026PO Times Cited:6 Cited References Count:65.

479 16. W. M. Shih, Z. Gryczynski, J. R. Lakowicz, and J. A. Spudich. A fret-  
480 based sensor reveals atp hydrolysis-induced large conformational changes  
481 and three distinct states of the molecular motor myosin. *Cell*, 102:683–694,  
482 2000.

483 17. Ulrich Stelzl, Uwe Worm, Maciej Lalowski, Christian Haenig, Felix H  
484 Brembeck, Heike Goehler, Martin Stroedicke, Martina Zenkner, Anke  
485 Schoenherr, Susanne Koeppen, et al. A human protein-protein interaction  
486 network: a resource for annotating the proteome. *Cell*, 122(6):957–968,  
487 2005.

488 18. Haiyuan Yu, Pascal Braun, Muhammed A Yildirim, Irma Lemmens,  
489 Kavitha Venkatesan, Julie Sahalie, Tomoko Hirozane-Kishikawa, Fana Ge-  
490 breab, Na Li, Nicolas Simonis, et al. High-quality binary protein interaction  
491 map of the yeast interactome network. *Science*, 322(5898):104–110, 2008.

492 19. Takashi Ito, Tomoko Chiba, Ritsuko Ozawa, Mikio Yoshida, Masahira Hat-  
493 tori, and Yoshiyuki Sakaki. A comprehensive two-hybrid analysis to ex-  
494 plore the yeast protein interactome. *Proceedings of the National Academy*  
495 *of Sciences*, 98(8):4569–4574, 2001.

496 20. Kirill Tarasov, Vincent Messier, Christian R Landry, Stevo Radinovic,  
497 Mercedes M Serna Molina, Igor Shames, Yelena Malitskaya, Jackie Vogel,  
498 Howard Bussey, and Stephen W Michnick. An in vivo map of the yeast  
499 protein interactome. *Science*, 320(5882):1465–1470, 2008.

500 21. Anne-Claude Gavin, Markus Böschke, Roland Krause, Paola Grandi, Mar-  
501 tina Marzoch, Andreas Bauer, Jörg Schultz, Jens M Rick, Anne-Marie  
502 Michon, Cristina-Maria Cruciat, et al. Functional organization of the  
503 yeast proteome by systematic analysis of protein complexes. *Nature*,  
504 415(6868):141–147, 2002.

505 22. Jin Gao, Wen-Xi Li, Si-Qian Feng, Yun-Sheng Yuan, Da-Fang Wan, Wei  
506 Han, and Yan Yu. A protein-protein interaction network of transcription  
507 factors acting during liver cell proliferation. *Genomics*, 91(4):347–355,  
508 2008.

509 23. Simon Fredriksson, Mats Gullberg, Jonas Jarvius, Charlotta Olsson, Kris-  
510 tian Pietras, Sigrun Margret Gustafsdottir, Arne Ostman, and Ulf Lande-  
511 gren. Protein detection using proximity-dependent dna ligation assays. *Nat*  
512 *Biotech*, 20(5):473–477, 05 2002.

513 24. Erika Assarsson, Martin Lundberg, Göran Holmquist, Johan Björkesten,  
514 Stine Bucht Thorsen, Daniel Ekman, Anna Eriksson, Emma Rennel Dick-  
515 ens, Sandra Ohlsson, Gabriella Edfeldt, Ann-Catrin Andersson, Patrik  
Lindstedt, Jan Stenvang, Mats Gullberg, and Simon Fredriksson. Homoge-  
516 nous 96-plex pea immunoassay exhibiting high sensitivity, specificity, and  
517 excellent scalability. *PLoS ONE*, 9(4):e95192, 04 2014.

518 25. B. Yurke, A.J. Turberfield, A.P. Mills, Jr., F.C. Simmel, and J.L. Neumann.  
519 A DNA-fuelled molecular machine made of DNA. *Nature*, 406:605–608,  
520 2000.

521 26. C.M. Radding, K.L. Beattie, W.K. Holloman, and R.C. Wiegand. Uptake  
522 of homologous single-stranded fragments by superhelical DNA. IV. branch  
523 migration. *J. Mol. Biol.*, 166:825–839, 1977.

524 27. D. Y. Zhang and E. Winfree. Control of dna strand displacement kinetics  
525 using toehold exchange. *J Am Chem Soc*, 131:17303, 2009.

526 28. J. N. Zadeh, J. S. Bois, M. B. Pierce, R. M. Dirks, and N. A. Pierce. Nupack:  
527 a web-based tool for the analysis and design of nucleic acid systems. *In*  
528 *preparation*, 2007.

529 29. Jr. John SantaLucia and Donald Hicks. The thermodynamics of dna struc-  
530 tural motifs. *Annual Review of Biophysics and Biomolecular Structure*,  
531 33(1):415–440, 2004. PMID: 15139820.

532 30. Claudio Rivetti, Martin Guthold, and Carlos Bustamante. Scanning force  
533 microscopy of dna deposited onto mica: Equilibrationversuskinetic trap-  
534 ping studied by statistical polymer chain analysis. *Journal of molecular*  
535 *biology*, 264(5):919–932, 1996.

536 31. MC Murphy, Ivan Rasnik, Wei Cheng, Timothy M Lohman, and Taekjip  
537 Ha. Probing single-stranded dna conformational flexibility using fluores-  
538 cence spectroscopy. *Biophysical Journal*, 86(4):2530–2537, 2004.

539 32. S. B. Smith, C. Yujia, and C. Bustamante. Overstretching b-dna: the elastic  
540 response of individual double-stranded and single-stranded dna groups.  
541 *Science*, 271:795–799, 1996.

542 33. Sungwook Woo and Paul WK Rothmund. Programmable molecular recog-  
543 nition based on the geometry of dna nanostructures. *Nature chemistry*,  
544 3(8):620–627, 2011.

**Acknowledgments:** We thank Jocelyn Kishi, Reza Kalhor, and Mitchell Zakin for fruitful discussions. This work is supported by a National Institutes of Health (NIH) Director's New Innovator Award (1DP2OD007292), an NIH Transformative Research Award (1R01EB018659), an Office of Naval Research (ONR) Young Investigator Program Award (N000141110914), ONR grants (N000141010827 and N000141310593), a National Science Foundation (NSF) Faculty Early Career Development Award (CCF1054898), an NSF Expedition in Computing Award (CCF1317291), and a Wyss Institute for Biologically Engineering Faculty Startup Fund to P.Y. T.S. and X.C. acknowledge support from the Jane Coffin Childs Postdoctoral Fellowship. S.W. acknowledges support from the Damon Runyon Postdoctoral Fellowship. The authors have filed a provisional patent based on the work.

## Materials and methods

**Probes** Probes were ordered from Integrated DNA Technologies with possible modifications of internal biotin (*iBiodT*), internal azide (*iAzideN*), or internal "Spacer 9" (*iSp9*). See SI Fig. 2 and SI Tables. 1 and 2 for probe sequences. Strands were ordered at 100 nmol or 250 nmol scale, with HPLC purification.

Some probes were ordered with an internal biotin modification in the center of the 5 nt loop domain, in which case probes were held in proximity with streptavidin, as indicated in Fig. 3 and SI Fig. 2. Probes were attached to streptavidin (New England Biolabs (NEB)) in a 20  $\mu$ l reaction containing 25 nM streptavidin (New England Biolabs (NEB)) and 140 nM probe (for a total of more than 4:1 probe:streptavidin to ensure streptavidin saturation), held at 37°C for 1 hour.

Other probes were covalently linked to alkyne-modified strands by click chemistry. Azide modified raw probes were ordered directly from IDT and dibenzocyclooctyne (DBCO) group modified linker strands were purchased from Boston Open Labs. Both DNA strands were HPLC purified by the manufacturers. Typically, 180 pmol of azide modified probes firstly went through Bst-mediated extension and USER-mediated cleavage (for random barcodes/spacers), and then column purified by using QIAquick PCR Purification Kit (Qiagen). A volume of 9  $\mu$ l of 1 mM DBCO-linker strands and 1  $\mu$ l of 10 $\times$  PBS (pH 7.4) buffer were added to 20  $\mu$ l of the eluted azide modified probes and incubated at room temperature for 12 h. The linked probes were purified by 8% denaturing polyacrylamide gel and quantified by NanoDrop UV-Vis Spectrophotometer (Thermo Scientific).

**NUPACK simulations** Expected reaction products at thermodynamic equilibrium were predicted using NUPACK<sup>28</sup> web-based software ([nupack.org](http://nupack.org)), "analysis" option. Specified options were "DNA" nucleic acid type, 37°C temperature, 10 nM reactants, with maximum complex size of two. Advanced options 0.05 mM Na, 0.002 mM Mg, and "all" dangle treatments. Pair probability results were downloaded in text form and re-plotted using custom software (Wolfram Mathematica). See SI Fig. 1.

**APR recording** Recording in solution was carried out as follows: Probe strands were bound to streptavidin, as above. A mixture of 100  $\mu$ M each dNTP (NEB), 1 $\times$  ThermoPol Buffer (NEB), and 2 $\times$  Bst large fragment polymerase (NEB), 200-400 nM Cy5-labeled primer (IDT), and 40 nM probe:streptavidin was prepared at 4°C at a volume of 10  $\mu$ l per reaction. The temperature was raised to 37°C for 120 min (or other reaction time), followed by 80°C heat inactivation for 20 min.

For recording reactions on origami, flow chambers were created by sticking a freshly cleaved mica (Ted Pella) piece to a channel slide system (ibidi, Sticky-Slide V10.4, Cat. No. 80608), yielding a 30  $\mu$ l chamber volume. After washing the chamber three times with 60  $\mu$ l of TAE/Mg buffer by adding the buffer to one reservoir (inlet) and simultaneously taking out the same volume from the other side (outlet), a 30  $\mu$ l solution containing origami at 50 pM concentration was introduced to the chamber. (In the first washing round, ~30  $\mu$ l of buffer occupies the chamber and only ~30  $\mu$ l of extra buffer comes out.) After 30 min of incubation, unbound origami and extra staple strands were removed by washing three times with 60  $\mu$ l of fresh TAE/Mg buffer and three times with 60  $\mu$ l of 1 $\times$  ThermoPol buffer (NEB, Cat. No. B9004S, 10 $\times$ ; 1 $\times$  contains 20 mM Tris-HCl, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 10 mM KCl, 2 mM MgSO<sub>4</sub>, 0.1% Triton X-100). To the chamber, a 40  $\mu$ l solution containing 8 units/ $\mu$ l of Bst polymerase (NEB, Cat. No. M0275S), 100  $\mu$ M dNTP (NEB, Cat. No. N0447S), and the relevant primer mixes typically at 100 nM each, in 1 $\times$  ThermoPol reaction buffer, was added and incubated for 1 hr at room temperature. After the recording reaction, the supernatant solution containing produced records was collected and the polymerase was heat inactivated by incubating the solution at 80°C for 20 min. For samples used in the geometry studies (three point and hexagonal grid patterns) and the state change study, before quenching the reaction, extra recording primers contained in the product solution were removed by mixing the solution with Exonuclease I (NEB, Cat. No. M0293S) at 9:1 volume ratio and incubating 20 min at 37°C. This minimize any false-positive PCR amplification by the recording primers.

State change recordings of Fig. 6 were carried out following the standard origami procedure described just above, except that the product solutions were collected after first adding 40  $\mu$ l of buffer into the inlet, to prevent drying of the chamber between rounds. After the first and second state recording rounds, the chamber was washed six times with 1 $\times$  ThermoPol buffer before the "wash" samples were collected, followed by three times with TAE/Mg. For deactivating or reactivating the probe P3, a 40  $\mu$ l solution containing the de-activator strand at 10  $\mu$ M or the re-activator strand at 0.1  $\mu$ M respectively in TAE/Mg was added and incubated for 30 min at room temperature, after which the chamber was washed four times with TAE/Mg and four times with 1 $\times$  ThermoPol buffer before each recording reaction.

For the 7 point hexagonal-grid geometry recordings of Fig. 5, the experimental condition is similar to the ones for stage change recording except that the annealing temperature for PCR amplification was to 52°C, with 200 nM of each primer sequence used in the PCR mixture.

**Electrophoresis** Gels (8  $\times$  8 cm) for denaturing PAGE were cast in house at 8% acrylamide (J.T.Baker), with 7 M urea (Sigma) and 1 $\times$  TAE (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) buffer (GrowCells.com), in plastic cassettes (Life Technologies), and run for 30-35 min at 200 V and 65°C in 1 $\times$  TAE buffer. Gels were then removed from cassettes, stained in 1 $\times$  SybrGold (Life Technologies) for 15 min, and imaged with a Typhoon scanner (General Electric).

**DNA Origami preparation** The design and folding protocols of the 2D rectangular origami were adapted from the original rectangle<sup>3</sup> and the twist-corrected version.<sup>10,33</sup> Single-stranded M13mp18 DNA (scaffold strand) was purchased from Bayou Biolabs (Catalogue #P-107) and staple strands were obtained from Integrated DNA Technologies. Staple strands were purchased and used unpurified, except the staples with probe extension at the 5' end of the staples, which were purchased PAGE-purified. Scaffold and staples were mixed together at target concentrations of 5 nM and 40 nM, respectively, in TAE (40 mM Tris acetate, 1 mM EDTA) buffer with 12.5 mM magnesium acetate (TAE/Mg). For origami with click-coupled probes, the probe strands were added at 1.2-fold the concentration of the staple strands containing the anchor. For origami folding, the mixtures were kept at 90°C for 5 min and annealed from 90°C to 60°C over the course of 30 min, then from 60°C to 45°C over the course of 90 min, and finally from 45°C to 25°C over the course of 20 min. Folded origami solutions were used without further purification except where otherwise stated. See SI Fig. 5 for geometry and layout and SI Table 1 for staple sequences.

**Record PCR amplification** PCR amplification utilized 1 $\times$  VentR (exo-) polymerase in 1 $\times$  ThermoPol Buffer (New England Biolabs (NEB)), for 20 cycles, in 20  $\mu$ l reactions. Primers were initially present at 0.5  $\mu$ M, and dNTP mix (NEB) at 100  $\mu$ M each. Temperature cycling included a 95°C melt for 2 min, 20 cycles (varies) of 95°C for 20 s denaturing, 56°C for 45 s annealing, and 72°C for 15 s extension, and a final 72°C extension for 3 min.

## Barcode Extension for Analysis and Reconstruction of Structures (BEARS)

Cameron Myhrvold<sup>1,2</sup>, Michael Baym<sup>2</sup>, Nikita Hanikel<sup>1</sup>, Luvena Ong<sup>1</sup>, Jonathan Gootenberg<sup>2</sup>, and Peng Yin<sup>1,2,\*</sup>

<sup>1</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, <sup>2</sup>Department of Systems Biology, Harvard Medical School. \*email: py@hms.harvard.edu

### Abstract (154 words):

Collections of DNA sequences can be rationally designed to self-assemble into predictable three-dimensional structures. The geometric and functional diversity of DNA nanostructures created to date has been enhanced by improvements in DNA synthesis and computational design. However, existing methods for structure characterization typically image the final product or laboriously determine the presence of individual, labeled strands using gel electrophoresis. Here, we introduce a new method of structure characterization for simultaneous quantification of every strand species incorporated into a structure. Our method uses barcode extension and next-generation DNA sequencing to quantitatively measure the relative incorporation of individual strands into a DNA nanostructure. By quantifying the relative abundances of distinct DNA species in product and monomer bands, we can study the influence of geometry and sequence on assembly. We have tested our method using 2D and 3D DNA brick and DNA origami structures. Our method is general and should be extensible to a wide variety of synthetic DNA structures.

## Introduction

The discovery in 1982 that DNA can self-assemble into designed structures initiated the field of structural DNA nanotechnology<sup>1</sup>. Over the past few decades, the field of structural DNA nanotechnology has produced a stunning array of two- and three-dimensional structures<sup>2-13</sup>. These structures have been used for a variety of applications, such as protein structure determination<sup>14</sup>, enzyme scaffolding<sup>15-17</sup>, photonics<sup>18-20</sup>, and drug delivery<sup>21-23</sup>. The standard workflow is typically: structures are designed on a computer, component oligonucleotides are ordered and synthesized commercially, the structures are assembled in the lab, and then characterized using imaging or other analytical methods, including gel electrophoresis. This design-build-test process can be iterated several times if necessary to achieve a design with high performance.

Most aspects of the design-build-test cycle have been dramatically improved over the past few decades. Structure design is much easier now than it was in 1982, as evidenced by new design paradigms (DNA origami<sup>5</sup>, DNA bricks<sup>8</sup>, gridiron<sup>6</sup>, 3d polyhedral meshes<sup>10</sup>) and software packages for automating structure design and analysis (NUPACK<sup>24</sup>, caDNAo<sup>25</sup>, CanDo<sup>26</sup>). The synthesis of the oligonucleotides that form the staples or tiles for a structure is growing exponentially cheaper. Structure assembly is also faster and easier than ever before, with the recent demonstration of isothermal assembly protocols for DNA origami and DNA brick structures<sup>27,28</sup>. As a result of these combined advances, one can now design and assemble multiple structure designs in a single round of testing.

Despite the many advances in structure design, synthesis, and assembly, structure imaging remains time-consuming and requires dedicated, expensive machines. In recent years, new technologies such as fast-scan atomic force microscopy (fast AFM) and cryo-electron microscopy (cryo-EM)<sup>29,30</sup> have

increased the speed and resolution with which DNA nanostructures can be imaged, but still require substantial equipment investments and expertise to use to their fullest extent. Super-resolution optical microscopy techniques such as DNA-PAINT<sup>31</sup> have proven very helpful for imaging structures with multiple orthogonal labels<sup>32</sup> in 3 dimensions<sup>33</sup>, but they require labeling structures with organic dyes or single-stranded extensions. In spite of these advances, it remains difficult to characterize the component composition of multidimensional DNA nanostructures in a high-throughput, label-free fashion.

In addition to imaging methods, several methods based on gel electrophoresis can be used to analyze DNA structure assembly. These methods compare the amount of material present in monomer, product, and aggregate bands, and measure structure-wide average quality or the site-specific incorporation of labeled oligonucleotides. The simplest such label is a fluorescent intercalating dye (e.g., Sybr Safe), but de-Bruijn probes can provide more quantitative estimates of the average structure quality<sup>34</sup>. In some cases, one cares more about the local assembly of particular structural features, rather than overall structure quality. In these cases, fluorescently-labeled oligonucleotides are typically employed to measure site-specific incorporation<sup>35</sup>, or FRET is used to measure the co-localization of two structure components<sup>17,36</sup>. These methods are generally simpler to employ than imaging, but they do not provide detailed information about the component composition of a structure with single component strand resolution. Thus, measuring the incorporation efficiency of all of the components of a DNA structure remains challenging.

A candidate method that could provide detailed quantitative information about the component composition of DNA nanostructures is next-generation DNA sequencing (NGS). Since each component strand in a uniquely addressable DNA nanostructure has a unique DNA sequence, it should be possible to obtain information about the composition of an entire structure with single tile or staple resolution.

NGS thus has higher multiplexing capabilities than site labeling methods, which typically are limited to labeling a few component oligonucleotides at a time. Unlike imaging methods, NGS allows for many samples to be processed in parallel using sequence barcodes, thereby increasing the throughput of the method substantially. Also, since the sequencing data are collected as an unbiased class average of many individual structures, they provide a rich picture of the statistics of self-assembly. Furthermore, the cost of NGS has dropped exponentially over the last few years, making this an increasingly attractive and affordable analysis technique. NGS has been used by biologists to measure RNA expression levels<sup>37</sup>, ribosome activity<sup>38</sup>, transcription elongation<sup>39</sup>, and protein-DNA interactions<sup>40</sup>, thus it should be possible to apply the method to study the self-assembly of DNA nanostructures in a quantitative fashion.

Here, we introduce a method for studying the assembly of DNA nanostructures that uses next-generation sequencing to quantify the relative incorporation of staples or tiles. The method works by assembling structures, segregating and isolating products and unincorporated strands using gel electrophoresis and ligating barcoded adapter sequences to the ends of denatured staples or tiles. Once the ligations are complete, the resulting libraries can be purified, amplified, and analyzed using NGS. We call our method barcode extension for analysis and reconstruction of structures, or BEARS. Using BEARS, we demonstrate the simultaneous quantification of each of the components of 2D and 3D DNA brick and DNA origami nanostructures. Thus, BEARS can be used to study the assembly of a wide range of self-assembled DNA nanostructures.

## Results

### **Developing a barcode ligation strategy**

The BEARS protocol is divided into five basic parts: structure assembly, product and monomer purification, denaturation and barcode ligation, sample pooling, and next-generation sequencing (Fig. 1a). Structure assembly, product purification, and next-generation sequencing involve standard procedures, but the barcode ligation step presents a unique challenge (Supplementary Fig. 1). Ideally, one would like to attach barcodes in an efficient, unbiased, and sequence-independent manner. To do this, we developed a barcode ligation strategy that uses single-stranded DNA ligation. First, a 5' phosphorylated adapter with a 3' dideoxycytosine modification is ligated to the 3' end of a tile or staple using T4 RNA ligase 1. The ligation product is then purified on a PAGE gel (Supplementary Fig. 2), phosphorylated, and a barcoded adapter is added to the 5' phosphate using T4 RNA ligase 1. This doubly ligated product is column-purified, amplified using PCR. The resulting samples (isolated from the product band, the monomer band, or the input oligonucleotide mix for each structure design) are pooled to enhance the throughput of the method, and then sequenced using NGS. Although BEARS requires structure purification, this is not a limitation because structures are denatured immediately after purification. Thus, we do not expect purification to alter the component composition of the structures analyzed using BEARS. For further details regarding sequencing library preparation, see the Methods section and Supplementary Figures 1 and 2.

### **Sequencing data analysis**

NGS sequencing data files are scanned to look for reads corresponding to strands present in a DNA origami design or DNA brick canvas (Supplementary Figure 1). This yields a list of read counts for each strand (for a histogram of the read counts, see Supplementary Figure 3). Read fractions are computed from the read count lists by dividing the read count for each strand by the sum of all of the read counts in a sample. Next, we determine which strands are present in the structure by taking the ratio of the read fraction in the product band to the read fraction in the entire canvas or origami design. This yields

either a unimodal or bimodal lognormal distribution, with the rightmost peak corresponding to well-incorporated strands and the left-most peak corresponding to strands that were not present in the structure (Supplementary Figure 4). If one peak is present, then all strands are well-incorporated, which is typically the case with DNA origami structures. When necessary, we apply a threshold to separate these two groups of strands. Further analysis is performed on the strands that passed the threshold, as described below.

To quantify the relative incorporation of strands into a DNA structure, one needs a metric that accounts for ligation and NGS sequence bias. There is some sequence bias inherent to NGS, as we obtain a distribution of read counts even for oligonucleotides mixed at equal stoichiometry (Supplementary Figure 3). This is consistent with observations in other NGS library preparation methods that certain library members tend to be over- or under-represented based on their sequence<sup>41</sup>. Fortunately, the sequence bias of NGS tends to be consistent, due to sequence-specific increases in ligation or amplification efficiency based on our observations (data not shown) and those in the literature<sup>41,42</sup>. Thus, by analyzing the data appropriately, one should be able to account for this bias.

To account for the sequence bias of NGS, we introduce a new metric for measuring the relative incorporation of strands into a DNA structure called structure-wide relative incorporability, or SRI. SRI is calculated by dividing the read fraction of each strand in the product by the sum of its read fractions in the product and monomer bands. As a result, the SRI varies from 0 (no incorporation) to 1 (full incorporation, no reads in the monomer), with higher SRI values indicating better incorporation of a strand relative to the other strands in a structure. SRI makes two key assumptions: that the variation in stoichiometry in the input assembly reaction is lower than the variation in incorporation into structures, and that sum of the product and monomer bands is representative of the entire assembly reaction. It is

important to emphasize that SRI measurements are relative and not absolute because they involve the ratios of read fractions between pairs of samples (i.e. the product band and the monomer band), which are effectively ratios of ratios. Thus, SRI is not a linear metric for strand incorporation, and should not be treated as such. Furthermore, it should be emphasized that the SRI is not equivalent to the bulk yield of a structure, as it is a nonlinear metric. For further details on SRI calculation, see the Methods section.

### **Validating BEARS**

BEARS is a reproducible and quantitative method for studying structure assembly. We assembled a 2D DNA brick structure on two separate occasions and prepared sequencing libraries from these replicate assemblies on separate days. We calculated SRI maps for each of these replicate assemblies using BEARS (Fig. 2a). Also, we directly compare the SRI for each brick strand in replicate structure assemblies using a scatterplot (Fig. 2b). The SRI values for each brick strand are highly correlated between replicate assemblies (correlation coefficient = 0.925). This suggests that there is relatively little variation introduced to the incorporation data due to structure assembly, gel purification, library preparation, and Illumina sequencing. It is likely that these variables would influence the monomer and product bands equally, so by taking the ratio of these samples we can control for these sources of variation.

In addition to reproducibility, we also tested the quantitative performance of Illumina sequencing by mixing pools of pre-extended oligonucleotides with varying stoichiometries (Fig. 2c). This allowed us to directly amplify and sequence the mixes without any ligation, thereby testing the effects of amplification and NGS on quantification. We counted the number of reads for each pool of oligonucleotides in each mix, and calculated the read fraction for each pool by dividing the read counts for each pool by the total number of reads in the mix. We then compared the observed read fraction for each pool with the expected read fraction based on the stoichiometry in the mix, which was set to 0%, 10%, 20%, 30%, or

40% based on the pipetted volume (Fig. 2d, see Methods section for details). We found that the expected and observed read fractions were highly correlated ( $R^2 = 0.9995$ ), with a relatively low amount of variation between mixes (error bars indicate one standard deviation). Based on the sizes of the error bars, we can detect 1.5-fold changes in stoichiometry with high reliability.

Ligation bias does not substantially alter quantification by sequencing. To test the effect of ligation bias on the quantitative performance of BEARS, we mixed pools of non-extended oligos with varying stoichiometry, and prepared libraries for sequencing using BEARS (see Methods section for details). This allowed us to test the combined effects of ligation, amplification, and NGS on quantification. As in Fig. 2d, we calculated read fractions for each pool, and compared the observed read fraction for each pool with the expected read fraction based on the stoichiometry in the mix (Fig. 2e). Although the correlation is less strong ( $R^2 = 0.9921$ ) than in the pre-extended case and the error bars are larger, we can still robustly detect 2-fold changes in oligo stoichiometry using BEARS. This indicates that ligation bias does add some variation to the read fractions measured using BEARS.

### **Using BEARS to analyze 2D DNA brick structures**

We can reconstruct structures from a 2D molecular canvas with high accuracy using BEARS. We assembled 5 arbitrarily chosen letters from a 2D molecular canvas<sup>8</sup> (the canvas consists of “pixels” that are individual DNA bricks and can be modularly combined to form a variety of shapes) and one newly-designed bear shape using a subset of the DNA bricks from the same molecular canvas, and prepared sequencing libraries from both product and monomer bands using BEARS (Fig. 3). Each panel shows a design schematic, 4 representative AFM images of fully assembled structures, thresholded images based on Gaussian fitting of read fraction ratios, and renderings of the SRI data (see Methods section for details) for that structure generated using BEARS. When we computed histograms of the number of

reads per brick strand for each structure design, we observed a bimodal lognormal distribution, with the rightmost peak corresponding to brick strands found in the structure and the leftmost peak corresponding to reads from other brick strands in the molecular canvas. We used this distribution to set thresholds for rendering the SRI data for each structure. SRI values are displayed on a color scale ranging from 0 (blue) to 1 (yellow), which are the maximum and minimum possible SRI values, respectively.

We observed good correspondence between the designs, AFM data, and the thresholded sequencing data for all of the structures (Fig. 3 a-f, first 3 rows). In a few cases, we obtained sequencing reads which did not correspond to brick strands found in a particular design. Conversely, we failed to detect fewer than 1% of brick strands in any given structure design. Reads corresponding to brick strands not found in a shape could come from contamination present in the strand mix, or introduced during sample processing. Brick strands within a shape that were not detected in either the monomer or product band by sequencing likely have poor ligation efficiency, and were more prevalent in samples with lower sequencing depth. For the bear design, there is a single missing brick strand in the hindquarters that is observed both in the AFM images and in the thresholded sequencing data (Fig. 3f). Together, these data indicate that BEARS can be used to reconstruct 2D DNA brick structures with single component resolution, which can sometimes be difficult to resolve with AFM.

SRI data from BEARS are recapitulated at the single structure level by AFM data. After analyzing the SRI data (Fig. 3, fourth row), we noticed that certain structures had some areas with poor incorporation (blue), predominantly found near the edges of structures, or in thinner features (for example, the legs of the bear in Fig. 3f). We were able to find a number of partially assembled structures which were missing these poorly incorporated areas when imaging with the AFM (Supplementary Fig. 5). However, we did

not observe a correlation between GC content or strand  $\Delta G$  and SRI (Supplementary Fig. 6). Since our method involves gel-purifying assembled structures, we conclude that the SRI data represent a class average of the structures present in the product band.

### **Using BEARS to analyze 3D DNA brick structures**

BEARS can be used to analyze 3D structures, including strands on the inside of the structures. Such analysis would be difficult to obtain using standard TEM methods. We assembled a 3D DNA brick cuboid with a tunnel along the long axis (Fig. 4a), imaged the product band using TEM, and sequenced the monomer and product bands using BEARS (Fig. 4b, c). We calculated the SRI of each brick strand as in Fig. 2e and projected the incorporation data along each of the major axes of the structure (Fig. 4b, c). The tunnel in the middle of the structure is clearly visible in the rendering from the appropriate perspective (Fig. 4c), as well as a stripe of decreased intensity from left to right along the long axis in the other two projections (Fig. 4b). This could indicate directional bias in DNA brick incorporation. Furthermore, the spatial patterns of incorporation can be visualized by displaying slices through the long axis of the structure (Fig. 4d, see Supplementary Fig. 7 for a larger image). Such data indicate that there is some spatial clustering present in the SRI data. In particular, the right edge of the structure contains two layers with poor SRI (Fig. 4d, blue slices). As with the 2D DNA brick structures, we did not observe a correlation between the strand GC content or free energy and the SRI measurements (Supplementary Fig. 8). These data highlight the power of a label-free approach, as they include many brick strands in the interior of the structure which may interfere with structure assembly when labeled.

### **Extending BEARS to DNA origami structures**

BEARS can be used to calculate the SRI of staples in DNA origami structures. We assembled a 2D DNA origami rectangle (Fig. 5a, b) and a 3D DNA origami cuboid (Fig. 5c-e), imaged the product bands using

AFM or TEM respectively, and sequenced the monomer and product bands using BEARS. Most staples had high SRI into the 2D origami rectangle, with a few exceptions (blue). For a larger image of the slices of the 3D origami cuboid, see Supplementary Fig. 9. As with the DNA brick structures, we did not observe a correlation between the strand GC content or free energy and the SRI measurements (Supplementary Fig. 10). Also, we observe that there is less spatial correlation between SRI values in origami structures compared with DNA brick structures. Thus, we conclude that BEARS can be extended to measure the SRI of staples in DNA origami structures. Given this, we expect that BEARS will be extensible to a wide variety of DNA nanostructures with uniquely addressable components.

## Discussion

Here, we have demonstrated a new method for quantifying the component composition of DNA nanostructures called BEARS. BEARS is high-throughput, label-free, and generates data that correlate with AFM images of individual structures. We believe that BEARS is complementary to other structure characterization methods, such as AFM/TEM imaging, and gel-based labeling. In particular, one can use BEARS to augment the resolution of a structure image, if one is not able to obtain component-level resolution by imaging. Alternatively, one can use BEARS to screen a new set of structure designs, and then image the designs with higher SRI values in areas of interest using AFM or TEM. In these ways, we envision that BEARS will help remove some of the existing bottlenecks in structure characterization, allowing one to design, build, and test more structures than was previously possible.

One use case for BEARS is to determine the weak points of a structure design and improve them with a new design. This is particularly important for applications in which guest molecules are attached at specific points on a structure – these should be chosen to have the highest SRI, or redesigned to

optimize the yield. At present, the incorporation data represent a class average of structures purified from a product band on a gel. This provides a useful overview of which parts of a structure assemble well and which do not, but it would also be interesting to get sequencing data at the single structure level. One limitation of our method is that one needs a clear product band in order to ensure that the reads one generates come from properly assembled structures. However, this can likely be surpassed by fractionating an entire gel lane, or by using other methods for structure purification such as PEG precipitation or glycerol gradient centrifugation<sup>43,44</sup>. These sorts of improvements to BEARS will be driven by decreases in the cost of NGS, enabling one to sequence more samples in parallel, with higher depth per sample, thereby yielding better statistics about the self-assembly of populations of structures.

In addition to NGS, DNA synthesis is also experiencing a rapid decrease in price over time. One technology which will hopefully increase this trend is chip-based DNA synthesis, which allows for the production of tens of thousands of oligonucleotides at once<sup>45</sup>. These oligo libraries can then be used to assemble large libraries of regular-sized structures, or larger structures themselves<sup>46</sup>. Larger structures can be difficult to characterize by imaging when the assembly yield is low, and it is easy to damage these structures during sample preparation and processing. Such structures are also of interest because they tend to assemble with lower yield, thus it is quite possible that their assembly pathway(s) are very suboptimal, limiting the yield. Using BEARS, one could perhaps not only get component-level resolution renderings of these large structures, but also improve the yields of the structures by getting around limitations in the assembly yield. Overall, a combination of single component resolution, high throughput, and built-in class averaging make BEARS a promising method for characterizing a wide variety of DNA nanostructures in the coming years.

## Methods

## Structure designs

The 2D DNA BRICK structures shown in Fig. 3 are derived from the R6 canvas described in Wei et. al 2012<sup>8</sup>. The letters A, B, E, R, and S were described in that work, whereas the bear structure was designed using a different subset of the R6 canvas. The 3D DNA brick structure shown in Fig. 4 has a 13 nt domain length (Ong *et al.* unpublished). The 2D origami rectangle shown in Fig. 5 is based on a twist-corrected version<sup>47</sup> of the original 2D origami rectangle<sup>5</sup>, but with a different scaffold sequence<sup>32</sup>. The 3D origami cuboid shown in Fig. 5 was designed using caDNAo. For staple or brick strand sequences and structure design schematics, see Supplementary Data.

## Structure assembly

The letters B, E, A, R and the bear shape were assembled in assembly buffer (5 mM Tris-HCl, 1 mM EDTA, adjusted to pH 8.0) supplemented with 25 mM MgCl<sub>2</sub> using the following annealing protocol: 95 °C for 1 minute, anneal from 90 °C to 60 °C at 5 minutes per degree, then from 60 °C to 25 °C at 25 minutes per degree, followed by a hold at 25 °C. For AFM imaging, the letter S was assembled in assembly buffer (5 mM Tris-HCl, 1 mM EDTA, adjusted to pH 8.0) supplemented with 25 mM MgCl<sub>2</sub> using a modified annealing protocol: 95 °C for 1 minute, anneal from 90 °C to 60 °C at 5 minutes per degree, then from 60 °C to 40 °C at 45 minutes per degree, followed by a hold at 25 °C.

For the 2D origami rectangle, scaffold (M13, purchased from Bayou Biolabs) and staples were mixed together at target concentrations of 10 nM and 100 nM, respectively, in TAE (40 mM Tris acetate, 1 mM EDTA) buffer with 12.5 mM magnesium acetate (TAE/Mg). For 2D origami folding, the mixtures were kept at 90 °C for 5 min and annealed from 90 °C to 60 °C over the course of 30 min, from 60 °C to 45 °C over the course of 90 min, and from 45 °C to 25 °C over the course of 20 min. For the 3D origami cuboid,

scaffold (p7560<sup>13</sup>) and staple strands were mixed at 10 nM and 100 nM respectively in assembly buffer (5 mM Tris-HCl, 1 mM EDTA, adjusted to pH 8.0) supplemented with 10 mM MgCl<sub>2</sub> and annealed over 3 days using the following protocol: anneal from 80 °C to 60 °C at 2 minutes per degree, then from 60 °C to 25 °C at 2 hours per degree. For the 3D DNA brick cuboid, brick strands were mixed together at 100 nM in assembly buffer (5 mM Tris-HCl, 1 mM EDTA, adjusted to pH 8.0) supplemented with 20 mM MgCl<sub>2</sub> and annealed over 3 days using the following protocol: anneal from 80 °C to 60 °C at 2 minutes per degree, then from 60 °C to 25 °C at 2 hours per degree.

### **Gel electrophoresis**

The 2D DNA brick shapes, 2D origami, and 3D origami were analyzed by electrophoresis in a native 1.5% agarose gel supplemented with 10 mM MgCl<sub>2</sub>. Electrophoresis was performed at 90 V for 2 hours in an ice-water bath. Gels were pre-stained with 1× Sybr Safe (Life Technologies). The 3D DNA brick cuboid was analyzed by electrophoresis in a native 1% agarose gel supplemented with 10 mM MgCl<sub>2</sub>.

Electrophoresis was performed at 80 V for 2 hours in an ice-water bath. Gels were pre-stained with 1× Sybr Safe (Life Technologies). Afterwards, gels were scanned with a Typhoon FLA 9000 (General Electric) using the SYBR Safe channel (excitation at 473 nm, emission ≥ 510 nm).

Gel bands were visualized using a Safe Imager 2.0 Blue-Light Transilluminator (Invitrogen) and excised from the gel using a fresh razor blade. The excised piece was then placed into a Freeze 'N Squeeze column (Bio-Rad) and crushed using a plastic pestle (USA Scientific). For the 2D DNA brick shapes, 2D origami rectangle, and the 3D origami cuboid, structures were eluted from the column by centrifugation at 400 × *g* for 3 minutes. For the 3D DNA brick cuboid, structures were eluted from the column by centrifugation at 1,200 × *g* for 3 minutes.

### **Atomic force microscopy**

Images of folded structures were obtained with a Veeco Multimode V atomic force microscope. C-type Bruker SNL-10 tips were used under tapping mode in fluid. Samples (25  $\mu$ l) were deposited on the mica surface for 1 minute. The mica surface was then rinsed 5 times with 0.5x TE (5 mM Tris, 1 mM EDTA, adjusted to pH 8.0) supplemented with 25 mM  $MgCl_2$ . For the 2D DNA brick shapes, samples were supplemented with 5 mM  $NiCl_2$  (final concentration) to aid in attachment to the mica surface before imaging. The 2D origami rectangle was imaged in 1x TE (10 mM Tris, 1 mM EDTA, adjusted to pH 8.0) supplemented with 25 mM  $MgCl_2$ .

### **Transmission Electron Microscopy**

2.5  $\mu$ l of each sample was deposited on to glow-discharged, carbon-coated EM grids for 2 minutes. The liquid was wicked off, and 2.5  $\mu$ l of stain (2% uranyl formate + 25 mM NaOH) was added. The 3D DNA brick structure was stained for 30 seconds, and the 3D origami cuboid was stained for 45 seconds. After staining, excess liquid was wicked off. All samples were imaged using a JEOL JEM-1400 TEM operating at 80 kV.

### **Oligonucleotide mixing experiments**

For the experiments described in Fig. 2d, thirty oligonucleotides of length 80 nt were divided into 5 pools of 6 oligonucleotides each. These pools, 1-5, were mixed together with systematically varying stoichiometry as follows: Mix A 1:1:1:1:1, Mix B 1:2:3:4:0, Mix C 0:1:2:3:4, Mix D 4:0:1:2:3, Mix E 3:4:0:1:2, Mix F 2:3:4:0:1. Thus, in each mix, each pool is present at 0%, 10%, 20%, 30%, or 40% of the total mix. The mixes A-F were amplified using 2 cycles of PCR and sequenced.

For the experiments described in Fig. 2e, thirty oligonucleotides of length 42 nt were divided into 5 pools of 6 oligonucleotides each. These pools, 1-5, were mixed together with systematically varying stoichiometry as follows: Mix A 1:1:1:1:1, Mix B 1:2:3:4:0, Mix C 0:1:2:3:4, Mix D 4:0:1:2:3, Mix E 3:4:0:1:2, Mix F 2:3:4:0:1. Thus, in each mix, each pool is present at 0%, 10%, 20%, 30%, or 40% of the total mix. The mixes A-F prepared for sequencing using the full BEARS protocol (see below).

Data were analyzed by calculating the average number of reads per oligo in each pool, then dividing the read fractions in Mixes B-F by the read fractions calculated from Mix A, which contains each oligo pool mixed with a 1:1:1:1:1 stoichiometry. The resulting normalized read fractions were compared with the expected read fractions based on the mix stoichiometries. The data shown in Fig. 2d and 2e are the mean normalized read fraction for each of the 5 pools at each expected read fraction, based on the normalized data from Mixes B-F (see formula below). For example, for pools present at 10%, we expect the mean read fraction to be 1.6%. Error bars were calculated by taking the standard deviation from the 5 pools with a given expected read fraction.

$$NRF_B = RF_B / RF_A$$

### Sequencing library preparation

Gel-purified monomer or product bands were concentrated using the Oligo Clean and Concentrator kit (Zymo Research), eluting with 6  $\mu$ l of milliQ H<sub>2</sub>O. Samples were denatured by heating to 95 °C for 5 minutes. The 3dC adapter sequence (see Supplementary Table 1 for details) was then ligated to the 3' ends of the staples/brick strands using T4 RNA ligase 1 (New England Biolabs). Each 10  $\mu$ l ligation reaction contained 10 units of enzyme, 25% (w/v) PEG-8000, 1 $\times$  T4 RNA ligase buffer, 5 pmol of the 3dC adapter, 2-5 pmol of brick strands/staples, and ATP at a final concentration of 1 mM. Ligation reactions were incubated overnight at room temperature, then heat-inactivated at 65 °C for 20 minutes.

After heat-inactivation, samples were analyzed using denaturing polyacrylamide gel electrophoresis (Supplementary Fig. 2a). Samples were mixed 1:1 with 2× RNA loading dye (New England Biolabs), denatured for 10 minutes at 70 °C, and loaded onto a precast 10% TBE-urea gel. Electrophoresis was performed at 65 °C for 35-55 minutes at 180 volts using 0.5× TBE as a running buffer. Gels were post-stained with 1× Sybr Gold (Invitrogen) for 30 minutes in an orbital shaker. Ligation product bands were visualized using a Safe Imager 2.0 Blue-Light Transilluminator (Invitrogen), and excised from the gel using razor blades or 1.1×6.5 mm gel cutting tips (MidSci).

Gel slices were placed into dialysis tubes (Slide-A-Lyzer MINI Dialysis Device, 2K MWCO, 0.1 mL) and 50 µl of 0.5× TBE was added to submerge the gel slices. Electroelution was performed at 90 V for 30 minutes in 0.5× TBE, followed by reversing the leads and running for ~30 seconds to prevent the DNA from being stuck to the surface of the dialysis tube. Electroeluates were then concentrated using the Oligo Clean and Concentrate kit (Zymo Research), eluting with 6 µl of milliQ H<sub>2</sub>O.

The gel-purified ligation product was phosphorylated using T4 polynucleotide kinase (New England Biolabs). Phosphorylation reactions were carried out in a 10 µl reaction volume containing 5-10 units of enzyme, 1× T4 PNK buffer, and 1 mM final concentration of ATP. Reactions were incubated for 30 minutes at 37 °C, then heat-inactivated at 65 °C for 20 minutes.

The 5bc adapter sequence (containing indexing barcodes, see Supplementary Table 1 for details) were then ligated to the 3' ends of the staples/brick strands using T4 RNA ligase 1 (New England Biolabs). Each 20 µl ligation reaction contained 2 µl of enzyme at 10 units/µl, 10 µl of 50% (w/v) PEG-8000 (final concentration: 25% (w/v)), 2 µl of 10× T4 RNA ligase buffer, 5 pmol of the 3dC (1 µl at 5 µM), 3 µl of

phosphorylated ligation 1 product, and 2  $\mu$ l of 10 mM ATP (final concentration: 1 mM). Ligation reactions were incubated overnight at room temperature, then heat-inactivated at 65 °C for 20 minutes. We then purified the samples using the Oligo Clean and Concentrate kit (Zymo Research), eluting with 6-10  $\mu$ l of milliQ H<sub>2</sub>O.

### **Library amplification and quantification**

Individual samples were amplified before pooling using Q5 polymerase (from 2 $\times$  master mix purchased from NEB) and previously validated Illumina qPCR primers at a final concentration of 300 nM. Between 2 and 10  $\mu$ l of template was used in a 50  $\mu$ l PCR reaction. Sequencing libraries were quantified using quantitative PCR. Q5 polymerase (NEB) was used according to the manufacturer's instructions. Previously validated Illumina qPCR primers<sup>42</sup> were synthesized by IDT and used at a final concentration of 300 nM. Syto13 (Molecular Probes / Life Technologies) was used as a fluorescent indicator dye according to the manufacturer's instructions. DNA standards 1-6 from the Kapa NGS library quantification kit were used to make a standard curve for absolute concentration determination (sample qPCR data are shown in Supplementary Fig. 2b).

### **Next-generation sequencing**

Libraries were sequenced using an Illumina MiSeq machine according to the manufacturer's instructions using the MiSeq V2 paired end 50 kit (Illumina Inc., San Diego, CA). In some cases, we used a modified library denaturation and loading protocol optimized for lower-concentration libraries<sup>42</sup>.

### **Data analysis**

Sequence processing was done using custom MATLAB software (Supplementary xxx). Fastq files from the MiSeq were parsed and partitioned based on the index reads. Sequence matching was done using

regular expressions to query for an exact match to the first 20 bases of a brick strand or staple, allowing between 4 and 7 random nucleotides at the 5' end of the read. These random nucleotides are part of the 5' sequencing adapter, and add sequence diversity to first few cycles of the sequencing (necessary for the machine to focus properly). Data were filtered by disregarding strands with fewer than 25 reads present in the product band, as there is not sufficient information to accurately quantify the incorporation for these strands. Read counts were normalized to the overall number of reads in a sample, resulting in "read fraction" data for both the product and monomer bands.

After calculating the read fraction for each strand in a structure, we applied a threshold to determine which strands were actually present in a structure and which were not. Specifically, we took the ratio of the read fraction for each strand in the product band and the read fraction for each strand in the canvas mix. This resulted in a bimodal lognormal distribution for most shapes, since they contain a subset of all of the strands in the canvas. By taking the ratio, we control for how well each individual strand is ligated, amplified, and sequenced. Thresholds were determined by fitting the sum of two Gaussian distributions to histograms (with 20 evenly spaced bins) of the  $\log_2$  of the read fraction ratio (product/canvas) mentioned above using MATLAB. Specifically, threshold was taken as the minimum value of the sum of the Gaussian fits, rounded up to by the width of one bin from the histogram. Strands which passed the threshold were analyzed further, as described below.

SRI values were calculated based on the ratio of the read fraction in the product band, divided by the sum of the read fractions in the monomer and product bands (see formula).

$$SRI = \frac{\frac{P_i}{P_{tot}}}{\frac{P_i}{P_{tot}} + \frac{M_i}{M_{tot}}}$$

Where  $P_i$  and  $M_i$  are the read counts for oligonucleotide  $i$  in the product and monomer bands, respectively, and  $P_{\text{tot}}$  and  $M_{\text{tot}}$  are the total read counts in the product and monomer bands. We divide  $P_i$  and  $M_i$  by  $P_{\text{tot}}$  and  $M_{\text{tot}}$  respectively because the total number of reads varies between samples; thus the read count is inherently a relative measure rather than an absolute measure. This metric has two key assumptions. The first assumption is that the variation in the starting concentrations of the oligonucleotides used for self-assembly is lower than the variation expected to be found in the product band. Otherwise, the method will tend to measure variation in the starting stoichiometry, rather than in the relative incorporability of DNA oligonucleotides into the structures themselves. The second assumption is that the even distribution of oligonucleotide stoichiometry is not influenced by aggregation. If aggregation is strand-specific, then certain oligonucleotide might get sequestered preferentially in aggregates, thereby biasing the incorporation measurements for those oligonucleotides. So long as these two assumptions hold, the SRI is a good proxy for the relative incorporability of an oligonucleotide in a DNA nanostructure. We expect the variation in starting stoichiometry to be small (~10%) based on spectrophotometric measurements of oligo concentrations (data not shown). Also, we expect that aggregation will not be strand-specific, as it is likely mediated by nonspecific interactions between oligonucleotides or structures.

### **Structure rendering**

Coordinates of each strand were parsed from the caDNAno design files using custom MATLAB software, and are colored based on the SRI of each brick strand or staple. Brick strands with fewer than 25 reads in the product band are colored gray. Half-brick strands and edge protector brick strands are also colored gray. Data are rendered using an inter-helix distance of 2 nm and 10.67 nm per helical turn, based on the square lattice model.

### Author contributions

CM and MB conceived of the project and designed the sequencing protocol. NH and LO imaged the structures. CM, NH, JG, and performed sequencing experiments. PY supervised the project. CM wrote the manuscript, and MB, NH, and PY edited the manuscript.

### Acknowledgements

We acknowledge Omar Yaghi and Thomas Schaus for useful discussions. We acknowledge Sungwook Woo for help with the 2D origami rectangle structures, and Sarah Boswell for help with MiSeq sequencing. This work was funded by the NSF grant #1162459. CM is funded by the Fannie and John Hertz Foundation, and NH by the German National Academic Foundation and German Academic Exchange Service.

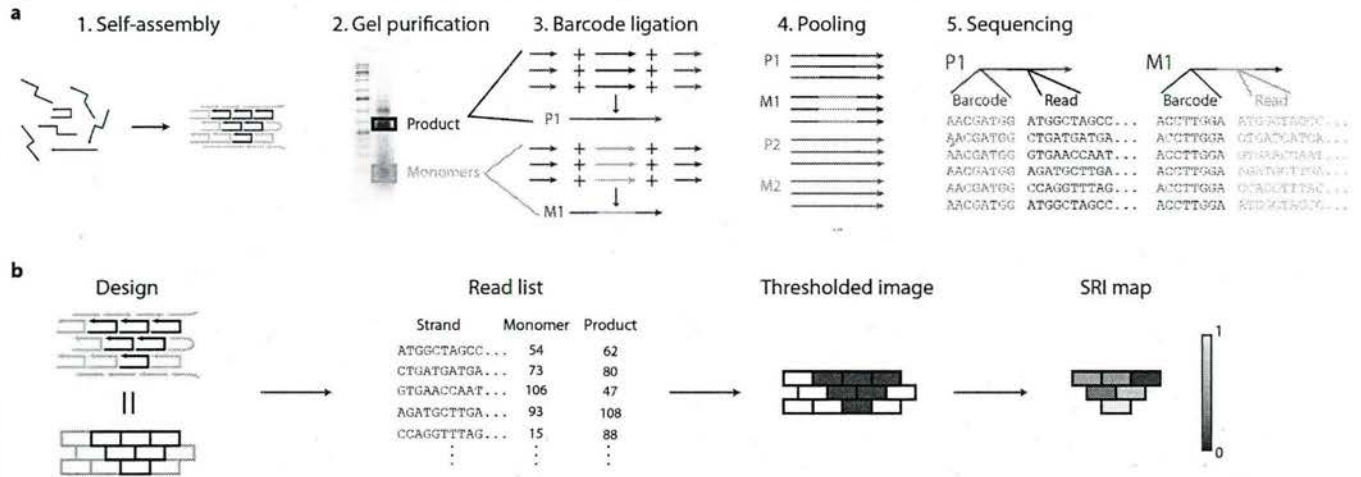
## References

1. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol* **99**, 237–47 (1982).
2. Mao, C., Sun, W. & Seeman, N. C. Assembly of Borromean rings from DNA. *Nature* **386**, 137–8 (1997).
3. Winfree, E., Liu, F., Wenzler, L. a & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–44 (1998).
4. Yan, H., Park, S. H., Finkelstein, G., Reif, J. H. & LaBean, T. H. DNA-templated self-assembly of protein arrays and highly conductive nanowires. *Science (80-. )*. **301**, 1882–4 (2003).
5. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
6. Han, D. *et al.* DNA gridiron nanostructures based on four-arm junctions. *Science* **339**, 1412–5 (2013).
7. Yin, P. *et al.* Programming DNA tube circumferences. *Science (80-. )*. **321**, 824–6 (2008).
8. Wei, B., Dai, M. & Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **485**, 623–6 (2012).
9. Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science (80-. )*. **338**, 1177–1183 (2012).
10. Benson, E. *et al.* DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523**, 441–444 (2015).
11. Andersen, E. S. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73–6 (2009).
12. Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science (80-. )*. **325**, 725–30 (2009).
13. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
14. Douglas, S. M., Chou, J. J. & Shih, W. M. DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. *Proc. Natl. Acad. Sci. USA* **104**, 6644–8 (2007).
15. Delebecque, C. J., Lindner, A. B., Silver, P. a & Aldaye, F. a. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333**, 470–4 (2011).
16. 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**, 5516–9 (2012).
17. Liu, M. *et al.* A DNA tweezer-actuated enzyme nanoreactor. *Nat. Commun.* **4**, 2127 (2013).
18. Kuzyk, A. *et al.* DNA-based self-assembly of chiral plasmonic nanostructures with tailored optical response. *Nature* **483**, 311–4 (2012).
19. Schreiber, R. *et al.* Hierarchical assembly of metal nanoparticles, quantum dots and organic dyes using DNA origami scaffolds. *Nat. Nanotechnol.* **9**, 74–78 (2013).

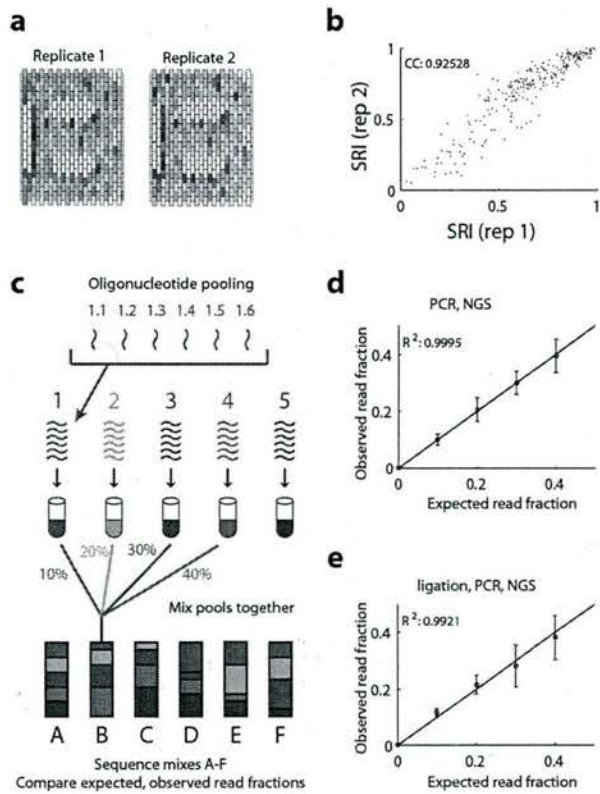
20. Shen, X. *et al.* Rolling up gold nanoparticle-dressed dna origami into three-dimensional plasmonic chiral nanostructures. *J. Am. Chem. Soc.* **134**, 146–149 (2012).
21. Lee, H. *et al.* Molecularly self-assembled nucleic acid nanoparticles for targeted in vivo siRNA delivery. *Nat. Nanotechnol.* **7**, 389–93 (2012).
22. Liu, X. *et al.* A DNA nanostructure platform for directed assembly of synthetic vaccines. *Nano Lett.* **12**, 4254–9 (2012).
23. Douglas, S. M., Bachelet, I. & Church, G. M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science (80-. )*. **335**, 831–4 (2012).
24. Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–3 (2011).
25. Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* **37**, 5001–5006 (2009).
26. Castro, C. E. *et al.* A primer to scaffolded DNA origami. *Nat. Methods* **8**, 221–229 (2011).
27. Sobczak, J.-P. J., Martin, T. G., Gerling, T. & Dietz, H. Rapid Folding of DNA into Nanoscale Shapes at Constant Temperature. *Science (80-. )*. **338**, 1458–61 (2012).
28. Myhrvold, C., Dai, M., Silver, P. a & Yin, P. Isothermal self-assembly of complex DNA structures under diverse and biocompatible conditions. *Nano Lett.* **13**, 4242–8 (2013).
29. Bai, X.-C., Martin, T. G., Scheres, S. H. W. & Dietz, H. Cryo-EM structure of a 3D DNA-origami object. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 20012–7 (2012).
30. Tian, Y. *et al.* Prescribed nanoparticle cluster architectures and low-dimensional arrays built using octahedral DNA origami frames. *Nat. Nanotechnol.* **10**, 637–644 (2015).
31. Jungmann, R. *et al.* Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami. *Nano Lett.* **10**, 4756–61 (2010).
32. Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11**, 313–8 (2014).
33. Inuma, R. *et al.* Polyhedra self-assembled from DNA tripods and characterized with 3D DNA-PAINT. *Science* **344**, 65–9 (2014).
34. Wagenbauer, K. F., Wachauf, C. H. & Dietz, H. Quantifying quality in DNA self-assembly. *Nat. Commun.* **5**, 3691 (2014).
35. Sadowski, J. P., Calvert, C. R., Zhang, D. Y., Pierce, N. A. & Yin, P. Developmental self-assembly of a DNA tetrahedron. *ACS Nano* **8**, 3251–3259 (2014).
36. You, M. *et al.* Photon-Regulated DNA-Enzymatic Nanostructures by Molecular Assembly. *ACS Nano* **5**, 10090–10095 (2011).
37. Nagalakshmi, U. *et al.* The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science (80-. )*. **320**, 1344–1349 (2008).
38. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–23 (2009).

39. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).
40. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–502 (2007).
41. Aird, D. *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**, R18 (2011).
42. Quail, M. A. *et al.* A large genome center's improvements to the Illumina sequencing system. *Nat. Methods* **5**, 1005–10 (2008).
43. Stahl, E., Martin, T. G., Praetorius, F. & Dietz, H. Facile and scalable preparation of pure and dense DNA origami solutions. *Angew. Chemie - Int. Ed.* **53**, 12735–12740 (2014).
44. Lin, C., Perrault, S. D., Kwak, M., Graf, F. & Shih, W. M. Purification of DNA-origami nanostructures by rate-zonal centrifugation. *Nucleic Acids Res.* **41**, (2013).
45. Kosuri, S. *et al.* Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol.* **28**, 1295–9 (2010).
46. Marchi, A. N., Saaem, I., Vogen, B. N., Brown, S. & Labean, T. H. Toward larger DNA origami. *Nano Lett.* **14**, 5740–5747 (2014).
47. Woo, S. & Rothmund, P. W. K. Programmable molecular recognition based on the geometry of DNA nanostructures. *Nat. Chem.* **3**, 620–627 (2011).

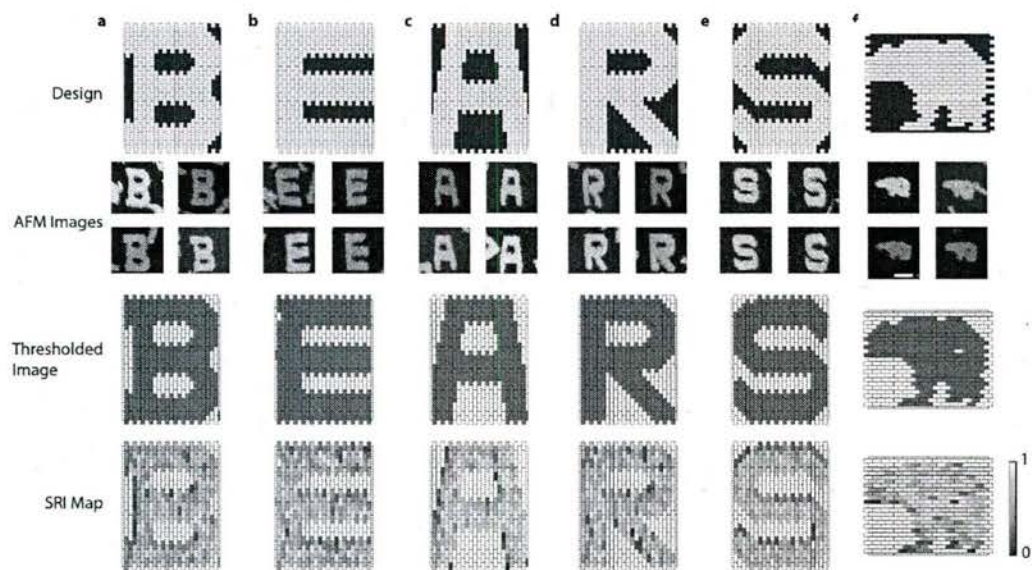
### BEARS: Barcode Extension for Analysis and Reconstruction of Structures



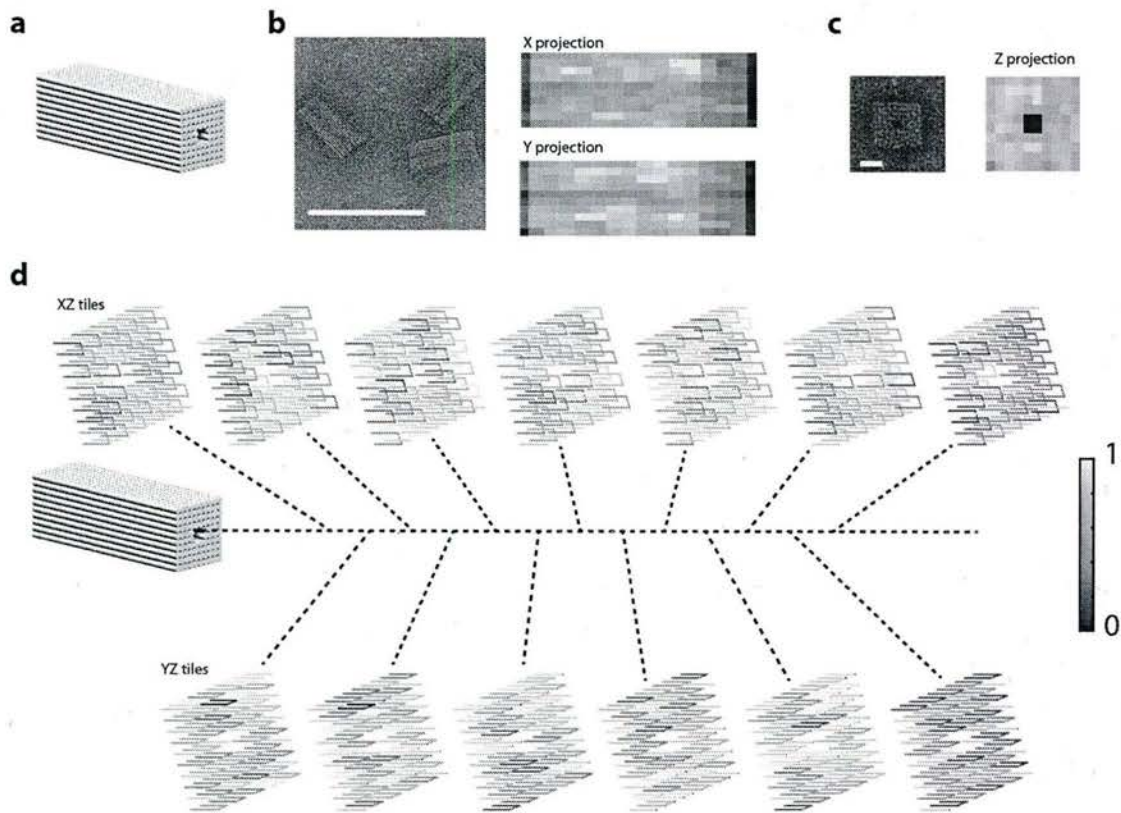
**Figure 1.** BEARS uses sequencing to assay DNA nanostructure assembly. We show schematics of each of the 5 major parts of BEARS: structure assembly, purification, barcode ligation, pooling, and next-generation sequencing (a). Structures are assembled, and well-assembled products and unincorporated monomers are isolated on an agarose gel. Barcodes and sequencing adapters are ligated onto each sample in separate reactions, and then samples are pooled and sequenced using NGS. We also show an overview of the data analysis pipeline used to generate a quantitative, spatial map of strand incorporation (b). We start with a molecular canvas and a structure design containing a subset of the DNA brick strands within the canvas. After self-assembly and sequencing, we have a list of read counts for each brick strand in the product and monomer bands. We threshold the sequencing data to determine which brick strands are part of the assembled structure and which are not (see Methods section for details). Finally, we compute the SRI for each strand that passed the threshold (see Methods for details).



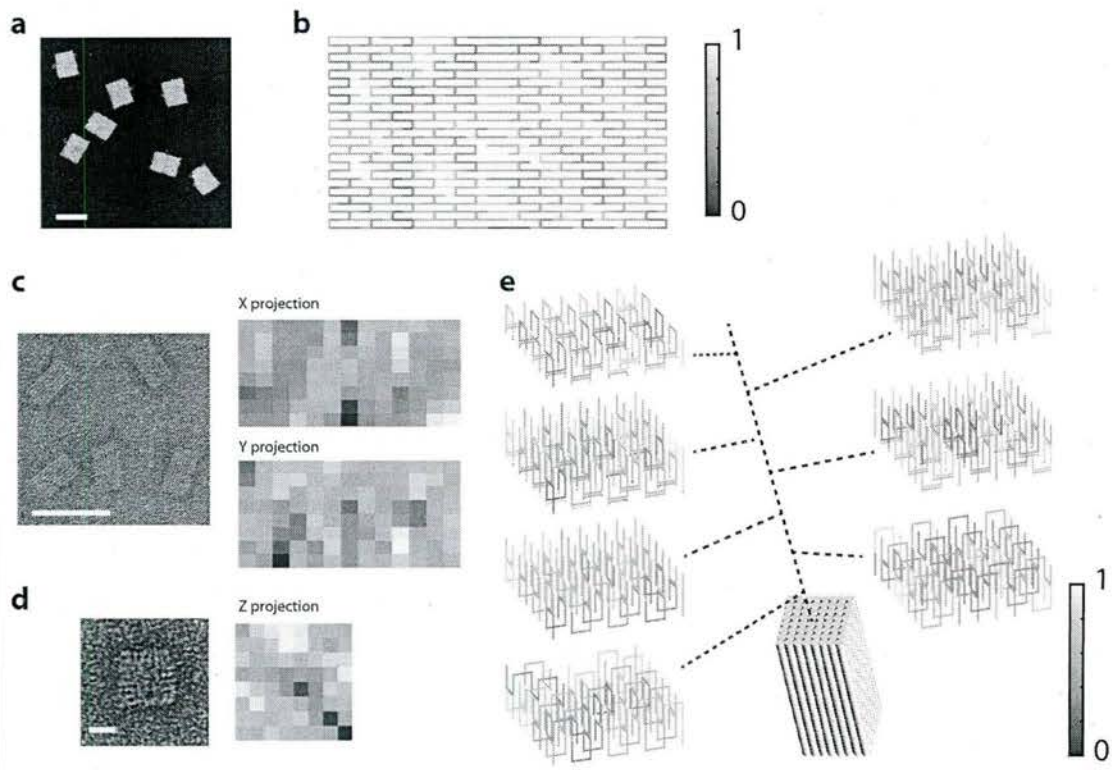
**Figure 2.** Sequencing is reproducible and quantitative. We assembled and sequenced a 2D DNA brick structure (the letter B) on two separate occasions using BEARS, and calculated the SRI for each replicate experiment (a). A scatterplot of the SRI values from each replicate is shown (b). The correlation coefficient between the two replicates is indicated in the upper left corner of the graph. We performed oligo mixing experiments in which 5 pools containing 6 unique oligonucleotides species each were mixed together with varying stoichiometries (c). We compared the observed and measured read fractions of these oligonucleotide pools after PCR and NGS using pre-extended oligonucleotides (d) and after ligation, PCR, and NGS (e). Error bars indicate one standard deviation based on measuring each of the 5 pools with a given expected read fraction.



**Figure 3.** BEARS recapitulates the shape and SRI of tiles into 2D DNA brick structures. We tested 6 structures: the letter B (a), the letter E (b), the letter A (c), the letter R (d), the letter S (e), and a bear shape (f). In each panel, we show design schematics (top row), 4 representative AFM images of fully-assembled structures (second row), thresholded images generated from SRI data (third row) and SRI data from BEARS (bottom row). Each box in the schematics and incorporation renderings indicates one tile. In the AFM images, the scale bars are 50 nm long.



**Figure 4.** Extending BEARS to 3D DNA brick structures. We self-assembled a 3D DNA brick cuboid structure containing a hole along the long axis (a). We show TEM images of the structure and SRI data from BEARS are projected along each of the 3 major dimensions using MATLAB (b, c). Scale bar: 100 nm in (b), 10 nm in (c). It is not possible to distinguish the X-projections and Y-projections using the TEM. We render incorporation data from BEARS in 3 dimensions for each layer of the structure using MATLAB (d). The location and order of each layer is indicated using the dotted lines. Spacing between adjacent helices is set to 2 nm, and each layer is approximately 8.1 nm thick.



**Figure 5.** Extending BEARS to 2D and 3D DNA origami structures. We show AFM images and BEARS incorporation data for a 2D origami rectangle (a, b). Scale bar in (a): 100 nm. We show TEM images and BEARS incorporation projections for a 3D origami cuboid (c, d). Scale bar: 50 nm in (c), 10 nm in (d). It is not possible to distinguish the X-projections and Y-projections using the TEM. We render SRI data from BEARS in 3 dimensions for slices through the 3D origami cuboid using MATLAB (e). The location and order of each slice is indicated using the dotted lines. Spacing between adjacent helices is set to 2 nm, and each slice is 5 nm thick.

# Eliminating Kinetic Barriers through Sequence Design Allows Rapid Assembly of DNA Origami at Room Temperature

Nikhil Gopalkrishnan<sup>\*,†</sup>, Thomas Schaus<sup>†</sup> and Peng Yin<sup>†</sup>

<sup>\*</sup>*nikhil.gopalkrishnan@wyss.harvard.edu*

<sup>†</sup>*Wyss Institute for Biologically Inspired Engineering, Harvard University*

DNA Origami [3] is a robust and programmable technique to make DNA nanostructures. A key use of DNA origami structures has been as nanoscale breadboards that organize components like small proteins, nanoparticles, small molecules and other DNA strands. Typical DNA origami assembly requires that the scaffold strand be annealed with staple strands in the appropriate buffer. The high temperatures ( $> 80C$ ) used in this process are not always compatible with temperature-sensitive components or contexts. In response, there have been attempts to isothermally fold DNA. Studies have shown that DNA origami can assemble isothermally if held at temperatures around  $60C$  [4, 5], or at room temperature by using denaturing agents [1, 2]. However, rapid assembly of DNA origami at room temperature without the use of denaturing agents, which can be toxic or have other unintended consequences, has not been achieved.

We assembled a rectangular DNA origami from a 1000 base long DeBruijn derived scaffold sequence and a corresponding staple set. The origami assembles isothermally in as little as five minutes at temperatures ranging from  $25C$  to  $50C$ .

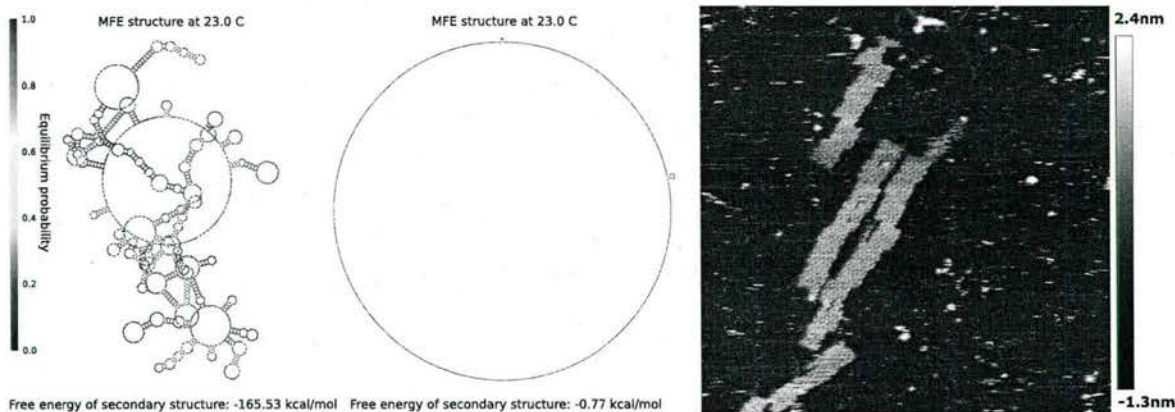


Figure 1: Left: A random 1000 base four letter scaffold sequence has significant secondary structure. Center: A three letter scaffold sequence has almost no secondary structure in the same conditions. Analyzed with NUPACK. Right: DNA origami assembled from a 1000 base long DeBruijn derived scaffold sequence and a corresponding staple set. The scaffold and staple strands were mixed together and incubated at room temperature for five minutes in 1X TE buffer supplemented with  $10mMMg^{++}$ . Image size  $1\mu m \times 1\mu m$ .

Traditional DNA origami designs are faced with prohibitive kinetic barriers to assembly at room temperature. Heating, or using denaturing agents, helps overcome these barriers. Here, we take an alternate approach. We use computational sequence design to eliminate prohibitive kinetic barriers and thus assemble DNA origami at room temperature in as little as five minutes.

We conjectured that the key contributions to kinetic barriers are: i) secondary structure of the scaffold, ii) secondary structure of staple strands and iii) unintended staple-staple interactions. We circumvent these issues by the use of a three letter code. The scaffold sequence is chosen from the three letter alphabet  $\{A, T, C\}$  and the staple sequences from the complementary alphabet  $\{A, T, G\}$ . Mfold and NUPACK analysis of the resulting sequences shows dramatic reduction in predicted secondary structure (Fig. 1) and unintended staple-staple interactions through this strategy. A potential drawback of a three letter code is reduced sequence diversity. A four letter code affords  $4^8 \approx 65000$  distinct 8-mers while a three letter code affords only  $3^8 \approx 6500$ . Thus, a poorly chosen three letter scaffold sequence could have poor sequence diversity that results in unintended staple-scaffold interactions, resulting in kinetic barriers to assembly. We attempt to minimize such bad choices by using DeBruijn sequences to maximize sequence diversity. A DeBruijn sequence  $S$  of order  $n$  over the three letter alphabet  $\{A, T, C\}$  is a sequence where every possible subsequence of length  $n$  appears exactly once. DeBruijn sequences can be generated computationally in time linear in the length of the sequence. We adapt DeBruijn sequences to filter out degenerate subsequences and ensure that  $C$  content is evenly distributed across the sequence. We call the resulting sequence DeBruijn derived and choose it as the scaffold sequence.

## References

- [1] Ralf Jungmann, Tim Liedl, Thomas L. Sobey, William Shih, and Friedrich C. Simmel. Isothermal assembly of dna origami structures using denaturing agents. *Journal of the American Chemical Society*, 130(31):10062–10063, 2008. PMID: 18613687.
- [2] Andreas Kopyelski, Anne Schneider, Andrea Csaki, and Wolfgang Fritzsche. Isothermal dna origami folding: avoiding denaturing conditions for one-pot, hybrid-component annealing. *Nanoscale*, 7:2102–2106, 2015.
- [3] Paul Rothemund. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature*, 440:297–302, 2006.
- [4] Jean-Philippe J. Sobczak, Thomas G. Martin, Thomas Gerling, and Hendrik Dietz. Rapid folding of dna into nanoscale shapes at constant temperature. *Science*, 338(6113):1458–1461, 2012.
- [5] Jie Song, Zhao Zhang, Shuai Zhang, Lei Liu, Qiang Li, Erqing Xie, Kurt Vesterager Gothelf, Flemming Besenbacher, and Mingdong Dong. Isothermal hybridization kinetics of dna assembly of two-dimensional dna origami. *Small*, 9(17):2954–2959, 2013.

# Quantitative super-resolution imaging with qPAINT

Ralf Jungmann<sup>1,2,5,6</sup>, Maier S Avendaño<sup>1,2,6</sup>, Mingjie Dai<sup>1,3</sup>, Johannes B Woehrstein<sup>1,5</sup>, Sarit S Agasti<sup>1,5</sup>, Zachary Feiger<sup>4</sup>, Avital Rodal<sup>4</sup> & Peng Yin<sup>1,2</sup>

Counting molecules in complexes is challenging, even with super-resolution microscopy. Here, we use the programmable and specific binding of dye-labeled DNA probes to count integer numbers of targets. This method, called quantitative points accumulation in nanoscale topography (qPAINT), works independently of dye photophysics for robust counting with high precision and accuracy over a wide dynamic range. qPAINT was benchmarked on DNA nanostructures and demonstrated for cellular applications by quantifying proteins *in situ* and the number of single-molecule FISH probes bound to an mRNA target.

Optical super-resolution microscopy is revolutionizing the way we study biology. It allows researchers to achieve a spatial resolution below the diffraction limit of light<sup>1</sup>, providing insights that were previously impossible<sup>2</sup>. Super-resolution studies often focus on the visualization of synthetic or cellular structures with sub-diffraction spatial resolution. However, data sets obtained by stochastic switching and readout methods<sup>3–5</sup> contain a wealth of information that can be explored for quantitative studies<sup>6–13</sup> beyond ‘just’ binning molecule localizations for spatial visualization. These counting techniques typically rely on complex modeling of blinking properties for target-bound fluorescent dyes. However, with current methods it is difficult to accurately and precisely quantify the number of target-bound fluorophores, especially for a large number of fluorescent proteins or dyes in dense clusters. First, the dyes typically have complex, environmentally sensitive photophysics that are difficult to model. Different switching properties for dissimilar dyes further complicate multiplexed quantitative imaging<sup>14</sup>. Furthermore, irregular sample illumination or varying excitation and activation intensities can lead to variation in switching kinetics and thus inaccurate quantification<sup>13</sup>. Additionally, target-bound dyes can be bleached prematurely, before sufficient localizations are collected to allow for accurate and precise quantification.

We introduce a simple and robust quantitative super-resolution method called quantitative PAINT (or qPAINT) based on DNA-PAINT<sup>15–19</sup>. Instead of relying on stochastic switching of target-bound dyes<sup>3,4</sup>, DNA-PAINT achieves apparent blinking of targets via transient binding of free-floating dye-labeled ‘imager’ strands to complementary target-bound ‘docking’ strands. By analyzing the predictable binding kinetics between the imager and docking strands, qPAINT counts the number of targets without spatially resolving them. As compared to existing quantification methods, qPAINT has two unique features: it explicitly decouples blinking from dye photophysics and it is immune to photobleaching (as dye-labeled imager strands are continuously replenished from the solution).

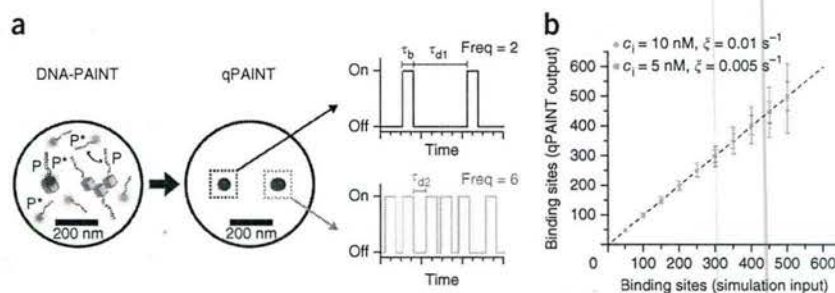
Thanks to these features, qPAINT represents a conceptual framework that can simultaneously achieve high accuracy, precision, a wide dynamic range, robustness, and multiplexing capability for quantifying the number of labeled targets. More specifically, by explicitly avoiding analyzing the hard-to-predict and illumination-dependent photophysical kinetics of dyes, qPAINT achieves high quantification accuracy. Immunity to photobleaching permits arbitrarily long imaging time and the collection of a large number of blinking events, leading to high counting precision. The easily adjustable influx rate of the imager strands makes such high accuracy and precision achievable over a wide dynamic range. Consistent, predictable, and easily calibratable (when required) binding kinetics allows qPAINT to work robustly under diverse conditions and with different dyes. Finally, because it decouples the apparent blinking from the photophysical properties of dyes, qPAINT is easily multiplexable (spectrally and sequentially<sup>17</sup>).

**Figure 1** illustrates the principle of qPAINT with the example of protein quantification in resolution-limited spots in a fixed cell. The region in **Figure 1a** consists of one protein spaced about 200 nm away from a small cluster formed by three proteins (which are spaced ~5 nm apart). Current super-resolution techniques fall short of resolving individual proteins in the small cluster. Thus, simple spatial counting would underestimate the total number of proteins. However, by analyzing the predictable and programmable binding kinetics of imager strands in DNA-PAINT rather than spatially resolving individual targets, it is possible to quantify integer numbers of molecules in these resolution-limited areas. Single-molecule DNA hybridization and dissociation can be described using a simple kinetic model with a second-order association rate  $k_{on}$  and a first-order dissociation rate  $k_{off}$ . These kinetic constants determine the fluorescence on- and off-times ( $\tau_b$  for bright-time and  $\tau_d$  for dark-time, respectively).  $\tau_b$  is linked to  $k_{off}$  via  $\tau_b = k_{off}^{-1}$  and  $\tau_d$  is linked to the influx rate of imager strands  $\xi = k_{on} \times c_i$  by  $\tau_d = (k_{on} \times c_i)^{-1} = \xi^{-1}$ , where

<sup>1</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts, USA. <sup>2</sup>Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA. <sup>3</sup>Harvard Program in Biophysics, Harvard University, Boston, Massachusetts, USA. <sup>4</sup>Rosenstiel Basic Medical Sciences Research Center, Department of Biology, Brandeis University, Waltham, Massachusetts, USA. <sup>5</sup>Present addresses: Max Planck Institute of Biochemistry and Ludwig Maximilian University of Munich, Munich, Germany (R.J. and J.B.W.); Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Jakkur, Bangalore, India (S.S.A.). <sup>6</sup>These authors contributed equally to this work. Correspondence should be addressed to P.Y. (py@hms.harvard.edu).

## BRIEF COMMUNICATIONS

**Figure 1** | qPAINT principle. (a) In DNA-PAINT, fluorescently labeled ‘imager’ strands ( $P^*$ ) transiently bind from solution to complementary ‘docking’ strands ( $P$ ) attached to a target. Intensity vs. time traces show characteristic fluorescence on- and off-times ( $\tau_b$  and  $\tau_d$ , respectively). qPAINT uses the predictable blinking kinetics to deduce molecule numbers. (b) The number of binding sites can be calculated given a known probe influx rate  $\xi = k_{on} \times c_i$ . Stochastic simulations of DNA-PAINT binding events show a linear relationship between simulated and ‘measured’ numbers of binding sites (mean  $\pm$  s.d.). The counting precision for a given number of sites is dependent on the probe influx rate  $\xi$  (green,  $0.01 \text{ s}^{-1}$ ; orange,  $0.005 \text{ s}^{-1}$ ).



$c_i$  is the imager strand concentration. If a single protein molecule labeled with a single docking strand ‘blinks’ with a frequency of 2 in a certain time interval, then three molecules (containing three docking sites) will blink with three times the frequency, given a constant influx rate  $\xi$  (Fig. 1a). To practically quantify the number of binding sites from the intensity versus time traces, we first determine the mean dark-time  $\tau_d$  from the cumulative distribution function (Supplementary Fig. 1) in an area of interest, and then calculate the number of binding sites as  $(k_{on} \times c_i \times \tau_d)^{-1} = (\xi \times \tau_d)^{-1}$ .

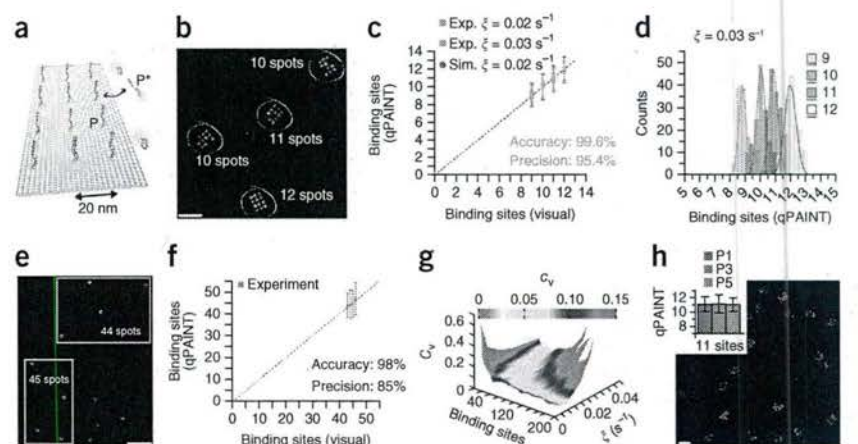
We first performed qPAINT *in silico* by stochastically simulating DNA-PAINT data and plotting the results obtained by qPAINT versus the ‘true’ molecule number used as input (Fig. 1b). The linear relationship between the results and the true molecule number over a wide range of binding sites shows that counting with qPAINT is feasible (see Supplementary Fig. 2 for discussion regarding binding frequency readout). qPAINT’s counting precision ( $1 - c_v$  with  $c_v = \sigma/\mu$ : coefficient of variation with  $\sigma/\mu$  s.d. and mean, respectively) can be increased by optimizing the probe influx rate  $\xi$  or extending image acquisition (Supplementary Fig. 3).

Next, we experimentally benchmarked qPAINT *in vitro* and compared it to our *in silico* results. We adopted a DNA origami<sup>20</sup>-based imaging quality benchmark platform and an accompanying drift correction and high-resolution image analysis

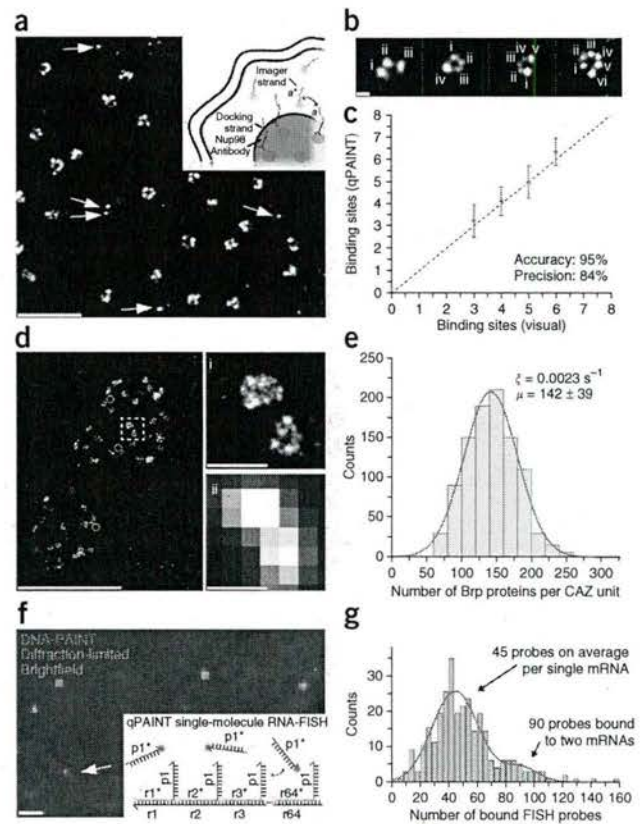
method<sup>21</sup>. The origami was designed to display 12 binding sites that can be visually separated using an advanced image analysis method that can achieve ultra-high-resolution imaging to discretely visualize each target in a densely packed cluster<sup>21</sup> (Fig. 2a, design details in Supplementary Fig. 4). The visually counted (‘ground truth’) number of sites can be compared to the qPAINT analysis for the same structure, omitting any a priori spatial information (Fig. 2b). High-resolution DNA-PAINT imaging revealed that not all origami showed 12 binding sites (Fig. 2b; see also Supplementary Fig. 5), likely resulting from missing docking strands<sup>15</sup> (see Supplementary Fig. 6 for incorporation efficiency measurement). Comparing qPAINT *in vitro* and *in silico* (Fig. 2c) versus the ground truth showed good agreement (90% precision, 97% accuracy). Adjusting the imager influx rate to  $\xi = 0.03 \text{ s}^{-1}$  and extending image acquisition time further increased the precision and accuracy to 95.4% and 99.6%, respectively (Fig. 2c and Supplementary Fig. 7). This then allowed us to distinguish between integer numbers of sites (i.e. between 9, 10, 11, and 12, Fig. 2d).

While counting a few molecules is possible using stepwise photobleaching approaches<sup>22</sup> or photon statistics<sup>23</sup>, these approaches are less effective for counting higher fluorophore densities<sup>24</sup>. For this reason, we sought to demonstrate similar qPAINT performance for a higher target density. Grouping four DNA

**Figure 2** | qPAINT *in vitro* benchmarking. (a) DNA origami structure with 12 designed docking sites. (b) DNA-PAINT image of the structures. (c) Visual counting ( $x$ -axis), *in silico* simulation (gray), and *in vitro* experimental qPAINT data (orange and green) are in good agreement (97% accuracy and 90% precision for  $\xi = 0.02 \text{ s}^{-1}$  and 25 min imaging time; 99.6% accuracy and 95.4% precision for  $\xi = 0.03 \text{ s}^{-1}$  and 166 min imaging time; error bars, 1 s.d.). (d) Distributions plotted from data in c (green data points) demonstrate qPAINT’s ability to distinguish between integer numbers of binding sites (i.e., 9 vs. 10 vs. 11 vs. 12; Tukey *post hoc* test:  $F_{3,605} = 1,032.52$ ,  $*P < 0.01$ , Supplementary Fig. 7). (e) Dynamic range. Same DNA-PAINT data set as in c (orange data points) reanalyzed by grouping four DNA origami structures together. (f) Comparison between visual counting ( $x$ -axis) and *in vitro* (orange) qPAINT analysis; error bars, 1 s.d. (g) *In silico* analysis of the counting error (coefficient of variation,  $c_v$ ) dependency on the number of binding sites and imager strand influx rate. Tuning  $\xi$  (experimentally adjustable over a wide range) can reach optimal conditions with low counting errors ( $<10\%$ ) for virtually any number of bindings sites. (h) Multiplexed qPAINT. Three distinct DNA origami structures (similar to those in c) with orthogonal docking strand sequences (red P1, green P3 and blue P5; error bars, 1 s.d.) were imaged sequentially using Exchange-PAINT. Inset shows qPAINT analysis on 11 binding sites structures. Scale bars: 100 nm (b,h), 500 nm (e).



**Figure 3 | qPAINT *in situ*.** (a–c) qPAINT *in situ* benchmarking with Nup98. (a) DNA-PAINT image of Nup98 proteins in NPCs. Inset: labeling and imaging schematic for NPC proteins. Single targets (marked with arrows) are used for influx rate calibration. (b) NPC structures displaying three, four, five, and six distinct Nup98 protein clusters were respectively grouped for qPAINT analysis. (c) Comparison between visual counting (*x*-axis) and qPAINT data (orange); error bars, 1 s.d. (d,e) Brp qPAINT experiments. (d) DNA-PAINT image of *Drosophila* NMJs obtained by using secondary DNA-antibody conjugates and primary monoclonal antibodies against Brp<sup>Nc82</sup>. Zoomed-in view (i, DNA-PAINT; ii, diffraction limited) of the highlighted area in d showing two separate CAZ units (assemblies of Brp proteins into multiprotein clusters<sup>12</sup>). Green circles indicate single targets. (e) qPAINT quantification ( $n = 981$ ) indicating that the average number of Brp molecules per CAZ unit is  $142 \pm 39$  (mean  $\pm$  s.d.). (f,g) smFISH qPAINT experiments. (f) SUZ12 mRNAs molecules are tagged using single-stranded oligonucleotides with binding sequences unique to a part of the target mRNA (r1\*–r64\*), a fixed Cy3B label, and a single-stranded DNA-PAINT docking strand (p1). (g) qPAINT quantification ( $n = 301$ ) yields  $\sim 45$  probes bound to a single mRNA molecule ( $\sim 90$  probes for two mRNAs), revealing  $\sim 70\%$  hybridization efficiency of the FISH probes to the mRNA target. Scale bars: 500 nm (a, insets i and ii in d), 50 nm (b), 1  $\mu\text{m}$  (d,f).



origami from the previous data set into one region of interest for qPAINT analysis allowed us to assay qPAINT's performance for counting 48 sites (Fig. 2e; note that spatial information was still disregarded). Again, qPAINT versus ground truth analysis showed good agreement (Fig. 2f; 98% accuracy, 85% precision). Additionally, we performed qPAINT on DNA origami with 12, 48, and 150 binding sites, respectively, yielding similar performance (see Supplementary Figs. 4 and 8 for design and results).

Next, we performed an *in silico* study to determine the counting error for a given set of binding sites and imager strand influx rate  $\xi$  (Fig. 2g). The results show that  $\xi$  can be tuned to select an optimal 'working' point depending on the expected number of binding sites, thus achieving high counting precision over a wide dynamic range. Specifically, one should be able to obtain counting errors below 10% independent of the number of binding sites. Finally, we demonstrated the ability of qPAINT to count multiple target species in the same sample with similar performance by performing Exchange-PAINT<sup>17</sup> on three distinct DNA origami structures in the same sample (Fig. 2h) using orthogonal imager strand sequences.

Moving from a clean *in vitro* to a more complex *in situ* environment, we next evaluated qPAINT on DNA origami structures bound to cell surface proteins or microinjected in the nuclear and cytoplasmic regions of fixed cells (Supplementary Fig. 9). In both cases, qPAINT performed favorably, achieving accuracy and precision similar to those for the *in vitro* DNA origami results. Comparing nuclear and cytoplasmic counting results yielded similar numbers, demonstrating the consistent binding kinetics of qPAINT in different locations in fixed cells and suggesting its applicability, for example, for counting proteins or DNA or RNA molecules *in situ*.

We next benchmarked qPAINT's performance *in situ* by quantifying the number of individual units that form the symmetrical ring structure of the nuclear pore complexes (NPCs) by targeting the nucleoporin Nup98 (anchored mainly to the inner ring of the NPC<sup>25,26</sup>) in U2OS cells using monoclonal primary antibodies directly conjugated to docking strands (Fig. 3a). Due to variation of Nup98 protein units present in each NPC and potentially imperfect antibody labeling (Fig. 3a), not all complexes show the same number of individual proteins units (Fig. 3b and Supplementary Fig. 10). We first used single isolated Nup98 protein clusters as

calibration for the probe influx rate. The association rate was determined by using the mean value from the Gaussian fitting of the dark-time distribution from the individual protein localizations (Fig. 3a). We subsequently grouped NPC structures according to their number of visible Nup98 protein units. We then plotted for each group the distribution obtained by qPAINT versus the visually determined ground truth and we obtained 95% accuracy and 84% precision (Fig. 3c).

We then applied qPAINT to quantify the number of single Bruchpilot proteins (Brp), which are structural and functional components of the cytomatrix at the synaptic active zone (CAZ) in the *Drosophila* neuromuscular junction (NMJ)<sup>27</sup>. Here, we used monoclonal primary antibodies (Brp<sup>Nc82</sup>)<sup>12,27</sup> and DNA-conjugated secondary antibodies for labeling. Similarly to what was done in a previous study<sup>12</sup>, we defined a CAZ unit as an interconnected region of Brp molecules with an elliptical shape (Fig. 3d, i, and Supplementary Fig. 11). Subsequent qPAINT yielded  $142 \pm 39$  (mean  $\pm$  s.d.) Brp molecules per CAZ, consistent with previously reported numbers using quantitative dSTORM<sup>12</sup> (Fig. 3e). Note that more than one secondary antibody can be bound per primary antibody. This, together with the variability of antibodies bound per subunit, will lead to a larger 'labeling' variability. However, unlike previous work involving elaborate calibration using antibody titration<sup>12</sup>, qPAINT simply uses single isolated targets to calibrate the influx rate as described above.

Finally, we applied qPAINT to quantify the number of probes bound to an mRNA target (SUZ12) in single-molecule fluorescence *in situ* hybridization (smFISH) experiments (Fig. 3f; see Supplementary Fig. 12 for a larger view). We used a probe set of 64 smFISH probes<sup>28</sup> consisting of a 5'-Cy3B label and a docking

strand on the 3' end (inset Fig. 3f). After cell fixation, labeling<sup>29</sup> (see Online Methods for details), and imaging, colocalization between Cy3B and DNA-PAINT was observed (Fig. 3f). Subsequent qPAINT analysis yielded—for the first time, to our knowledge—the number of *in situ*-bound smFISH probes per mRNA molecule. The bimodal distribution showed a first peak at ~45 and a second peak at ~90, consistent with one and two mRNA molecules in a resolution-limited area, respectively (Fig. 3g). Note that counting less than the designed 64 smFISH strands is expected, as it is likely that not all FISH probes were bound at each mRNA molecule. This experiment thus indicates that ~70% of all designed smFISH probes were bound to their targets, with a variance of ~36%.

As qPAINT decouples blinking kinetics from dye photophysics and is immune to photobleaching (see Supplementary Fig. 13), it avoids typical 'undercounting' errors due to already photobleached or 'inactive' dye labels, as well as 'overcounting' errors due to blinking artifacts that are unaccounted for<sup>10,30</sup>. qPAINT works robustly under diverse experimental conditions with consistent high accuracy and precision (see Supplementary Fig. 14 for day-to-day reproducibility). Although qPAINT performs robustly, it is still reliant on stoichiometric labeling of protein targets. Imperfect labeling could potentially lead to undercounting. Similarly to the approach taken with the previous quantitative approach, dSTORM<sup>12</sup>, we opted to use standard immunostaining to label endogenous protein targets with antibodies and we obtained comparable quantification results, but with a simpler *in situ* calibration using isolated single targets. For future quantification of proteins in compact clusters, alternative labeling techniques (for example, nanobodies<sup>31</sup>, aptamers<sup>32</sup>, modified amino acids<sup>33</sup>, or small-molecule binders) need to be developed. In our present work, we demonstrated qPAINT using DNA-PAINT, but the general concept may be generalized to other PAINT methods (for example, recently developed IRIS probes<sup>34</sup>).

## METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

## ACKNOWLEDGMENTS

We thank M.T. Strauss and J. Lara for help with DNA origami design and J. Werbin (Department of Systems Biology, Harvard Medical School) for the donation of CHO cells and fruitful discussions. We also thank A. Raj for providing the smFISH sequences, F. Schueder for initial Brp imaging experiments, and H. Soundararajan and J. Paulson for fruitful discussions. This work is supported by a National Institutes of Health (NIH) Director's New Innovator Award (1DP20D007292), an NIH Transformative Research Award (1R01EB018659), an NIH grant (5R21HD072481), an Office of Naval Research (ONR) Young Investigator Program Award (N000141110914), ONR grants (N000141010827 and N000141310593), a National Science Foundation (NSF) Faculty Early Career Development Award (CCF1054898), an NSF grant (CCF1162459) and a Wyss Institute for Biologically Engineering Faculty Startup Fund to P.Y., and a Pew Scholar Award to A.R. The BRP antibodies were obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH. R.J. acknowledges support from the Alexander von Humboldt Foundation through a Feodor-Lynen Fellowship. M.S.A. and M.D. acknowledge support from HHMI International Student Research Fellowships.

## AUTHOR CONTRIBUTIONS

R.J. and M.S.A. conceived of the study, designed and performed the experiments, analyzed the data, developed software, and wrote the manuscript. M.D. developed software and wrote the manuscript. J.B.W. helped with *in vitro* experimental design and wrote the manuscript. S.S.A. helped with DNA-dye conjugation and developed the antibody labeling protocol. Z.F. and A.R. helped with Brp experiments. P.Y. conceived of and supervised the study, interpreted the data and wrote the manuscript. All authors reviewed and approved the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Hell, S.W. *Nat. Methods* **6**, 24–32 (2009).
- Xu, K., Zhong, G. & Zhuang, X. *Science* **339**, 452–456 (2013).
- Betzig, E. *et al. Science* **313**, 1642–1645 (2006).
- Rust, M.J., Bates, M. & Zhuang, X. *Nat. Methods* **3**, 793–795 (2006).
- Sharonov, A. & Hochstrasser, R.M. *Proc. Natl. Acad. Sci. USA* **103**, 18911–18916 (2006).
- Lee, S.H., Shin, J.Y., Lee, A. & Bustamante, C. *Proc. Natl. Acad. Sci. USA* **109**, 17436–17441 (2012).
- Renz, M., Daniels, B.R., Vamosi, G., Arias, I.M. & Lippincott-Schwartz, J. *Proc. Natl. Acad. Sci. USA* **109**, E2989–E2997 (2012).
- Endesfelder, U. *et al. Biophys. J.* **105**, 172–181 (2013).
- Puchner, E.M., Walter, J.M., Kasper, R., Huang, B. & Lim, W.A. *Proc. Natl. Acad. Sci. USA* **110**, 16015–16020 (2013).
- Durisc, N., Laparra-Cuervo, L., Sandoval-Alvarez, A., Borbely, J.S. & Lakadamyali, M. *Nat. Methods* **11**, 156–162 (2014).
- Zhao, Z.W. *et al. Proc. Natl. Acad. Sci. USA* **111**, 681–686 (2014).
- Ehmann, N. *et al. Nat. Comm.* **5**, 4650 (2014).
- Nieuwenhuizen, R.P. *et al. PLoS ONE* **10**, e0127989 (2015).
- Dempsey, G.T., Vaughan, J.C., Chen, K.H., Bates, M. & Zhuang, X. *Nat. Methods* **8**, 1027–1036 (2011).
- Jungmann, R. *et al. Nano Lett.* **10**, 4756–4761 (2010).
- Lin, C. *et al. Nat. Chem.* **4**, 832–839 (2012).
- Jungmann, R. *et al. Nat. Methods* **11**, 313–318 (2014).
- Iinuma, R. *et al. Science* **344**, 65–69 (2014).
- Schlichthaerle, T., Strauss, M.T., Schueder, F., Woehrstein, J.B. & Jungmann, R. *Curr. Opin. Biotechnol.* **39**, 41–47 (2016).
- Rothmund, P.W. *Nature* **440**, 297–302 (2006).
- Dai, M., Jungmann, R. & Yin, P. *Nat. Nanotechnol.* (in the press).
- Ulbrich, M.H. & Isacoff, E.Y. *Nat. Methods* **4**, 319–321 (2007).
- Ta, H., Kiel, A., Wahl, M. & Hertel, D.P. *Phys. Chem. Chem. Phys.* **12**, 10295–10300 (2010).
- Coffman, V.C. & Wu, J.Q. *Trends Biochem. Sci.* **37**, 499–506 (2012).
- Hoelz, A., Debler, E.W. & Blobel, G. *Annu. Rev. Biochem.* **80**, 613–643 (2011).
- Strambio-De-Castillia, C., Niepel, M. & Rout, M.P. *Nat. Rev. Mol. Cell Biol.* **11**, 490–501 (2010).
- Fouquet, W. *et al. J. Cell Biol.* **186**, 129–145 (2009).
- Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A. & Tyagi, S. *Nat. Methods* **5**, 877–879 (2008).
- Shaffer, S.M., Wu, M.T., Levesque, M.J. & Raj, A. *PLoS ONE* **8**, e75120 (2013).
- Endesfelder, U. *et al. Molecules* **16**, 3106–3118 (2011).
- Ries, J., Kaplan, C., Platonova, E., Eghlidi, H. & Ewers, H. *Nat. Methods* **9**, 582–584 (2012).
- Opazo, F. *et al. Nat. Methods* **9**, 938–939 (2012).
- Milles, S. *et al. J. Am. Chem. Soc.* **134**, 5187–5195 (2012).
- Kiuchi, T., Higuchi, M., Takamura, A., Maruoka, M. & Watanabe, N. *Nat. Methods* **12**, 743–746 (2015).

## ONLINE METHODS

**Materials.** Non-modified and amino-modified DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Fluorescently-labeled DNA oligonucleotides were purchased from Biosynthesis (Lewisville, TX). Biotinylated monoclonal antibody against EGF receptor was purchased from Cell Signaling (Cat. No. 6627, Danvers, MA). Streptavidin was purchased from Invitrogen (S-888, Carlsbad, CA). Bovine serum albumin (BSA), BSA-Biotin and Triton X-100 was obtained from Sigma Aldrich (A8549, St. Louis, MO). Whole cell blue stain was obtained from Thermo Scientific (8403501, Rockford, IL). Glass slides and coverslips were purchased from VWR (Radnor, PA). Lab-Tek II chambered coverglass was purchased from Thermo Fisher Scientific (Billerica, MA). M13mp18 scaffold was obtained from New England BioLabs (Ipswich, MA). Freeze 'N Squeeze columns were ordered from Bio-Rad (Hercules, CA). Paraformaldehyde and glutaraldehyde were obtained from Electron Microscopy Sciences (Hatfield, PA). Protocatechuic acid (PCA), Protocatechuate-3,4-dioxygenase (PCD), and Trolox (TX) were purchased from Sigma-Aldrich (37580-25G-F (PCA), P8279-25UN (PCD), 238813-5G (TX)) (St. Louis, MO).

Four buffers were used for sample preparation and imaging: buffer A (10 mM Tris-HCl, 100 mM NaCl, 0.05% Tween-20, pH 7.5), buffer B (5 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM EDTA, 0.05% Tween-20, pH 8), buffer B<sup>+</sup> (5 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM EDTA, 0.05% Tween-20, pH 8, supplemented with 1 mM PCA, 1 mM PCD, and 1 mM TX), and buffer C (1× PBS, 500 mM NaCl, pH 8).

**Super-resolution optical setup.** Fluorescence imaging was carried out on an inverted Nikon Eclipse Ti microscope (Nikon Instruments, Melville, NY) with the Perfect Focus System, applying an objective-type TIRF configuration using a Nikon TIRF illuminator with an oil-immersion objective (CFI Apo TIRF 100×, NA 1.49, Oil), corresponding to a final pixel size of 160 nm. Three lasers were used for excitation: 488 nm (200 mW nominal, Coherent Sapphire, Santa Clara, CA), 561 nm (200 mW nominal, Coherent Sapphire) and 647 nm (300 mW nominal, MBP Communications, Canada). The laser beam was passed through cleanup filters (ZT488/10, ZET561/10, and ZET640/20, Chroma Technology, Bellows Falls, VT) and coupled into the microscope objective using a multi-band beam splitter (ZT488rdc/ZT561rdc/ZT640rdc, Chroma Technology). Fluorescence light was spectrally filtered with emission filters (ET525/50m, ET600/50m, and ET700/75m, Chroma Technology) and imaged on an EMCCD camera (iXon X3 DU-897, Andor Technologies, North Ireland).

**Confocal setup.** Confocal imaging was carried out on a Zeiss Axio Observer with LSM 710 scanning laser confocal system (Zeiss, Thornwood, NY) using the following excitation and emission lines: Whole cell blue stain: 25 mW Argon laser (458/488/514 nm) and 490–560 nm emission filter. Cy3: 20 mW DPSS laser (561 nm) and 563–593 nm emission filter.

**Data analysis.** Super-resolution DNA-PAINT images were reconstructed using custom-programmed software written in either LabVIEW<sup>15,17</sup> or MATLAB<sup>21</sup> for spot-finding and 2D-Gaussian fitting<sup>17</sup>. Drift correction and super-resolution imaging analysis for the 20 nm grid origami sample was performed using custom

MATLAB software to visually separate the binding sites<sup>21</sup>. Subsequent qPAINT analysis was performed using a custom-written software implemented in LabVIEW. Both reconstruction and analysis programs are available upon request. After single-molecule reconstruction, the qPAINT analysis software uses a region of interest as input (either automatically or interactively selected) for the binding kinetics analysis. The software determines the mean dark time  $\tau_{d^*}$  by first creating a cumulative distribution function (cdf) from all dark times  $t$  and subsequently fitting this distribution to a single exponential function<sup>15</sup>:

$$P(t) = 1 - \exp(-t/\tau_{d^*}),$$

with  $P$  representing the binding probability for a DNA-PAINT imaging probe at time  $t$  after a previous unbinding event. Finally—after calibrating the imager probe influx rate  $\xi = k_{\text{on}} \times c_i$  from a known calibration sample (i.e. previously established from experiments under similar conditions or *in situ* DNA origami calibration standards)—the number of binding sites in the selected region of interest is obtained by

$$\text{binding sites} = \frac{1}{\tau_{d^*} \cdot \xi}$$

**qPAINT data simulation.** *In silico* qPAINT experiments were performed using COPASI<sup>35</sup>. In brief, a bimolecular chemical reaction model—modeling DNA hybridization and dissociation reactions—with two species (docking and imager strands) and two rate constants (association rate  $k_{\text{on}}$  and dissociation rate  $k_{\text{off}}$ ) was used. Using a direct stochastic solver, we performed time-lapsed simulations that yielded binarized single-molecule intensity versus time traces. These traces were then subjected to the same qPAINT analysis software that was used to analyze the experimental data (see above). Simulation parameters were selected in accordance to the respective experimental conditions.

Simulation conditions: for the simulation in **Figure 1b**, two probe influx rates (0.005 s<sup>-1</sup> and 0.01 s<sup>-1</sup>) were used, corresponding to 5 nM and 10 nM imager strand concentration at a constant association rate  $k_{\text{on}}$  of 10<sup>6</sup> M<sup>-1</sup>s<sup>-1</sup>. Simulations were run for 15,000 “frames” at a sampling interval (or “integration” time) of 0.1 s. 200 simulations per data point were performed (mean and s.d. are plotted in **Fig. 1b**). More detailed simulation conditions for simulations accompanying experimental data are given in the following paragraphs.

**DNA origami self-assembly.** DNA origami structures displaying 12, 44, 48 and 150 DNA-PAINT docking sites were self-assembled in a one-pot reaction with 50  $\mu$ l total volume containing 10 nM scaffold strand M13mp18, 100 nM folding staples, 100 nM biotinylated staples, and 1  $\mu$ M DNA-PAINT docking staple strands in folding buffer (1× TE Buffer with 12.5 mM MgCl<sub>2</sub>). The solution was annealed using a thermal ramp cooling from 90 °C to 20 °C over the course of 1.5 h. After self-assembly, structures were purified by agarose gel electrophoresis (2% agarose, 0.5× TBE, 10 mM MgCl<sub>2</sub>, 0.5× SybrSafe) at 4.5 V/cm for 1.5 h. Gel bands were cut, crushed and filled into a Freeze 'N Squeeze column and spun for 5 min at 800 × g at 4 °C. After this, structures were ready for microscopy sample preparation and image acquisition.

**DNA sequences.** 20 nm grid structures can be found in **Supplementary Table 1**. Staple sequences for 12 docking sites origami can be found in **Supplementary Table 2**. Staple sequences for 48 docking sites origami can be found in **Supplementary Table 3**. Staple sequences for 150 docking sites origami can be found in **Supplementary Table 4**. Staple sequences for 20 nm grid structures with fixed Cy3 dyes can be found in **Supplementary Table 5**. Staple sequences for 44 docking sites origami with fixed Cy3 dyes can be found in **Supplementary Table 6**. M13mp18 scaffold sequence for DNA origami structures can be found in **Supplementary Table 7**. Sequences for FISH-PAINT probes can be found in **Supplementary Table 8**. DNA-PAINT docking and imager sequences and biotin docking sequence can be found in **Supplementary Table 9**.

**Sample preparation, imaging, and analysis of DNA origami structures.** *Sample preparation.* For sample preparation, a piece of coverslip (No. 1.5,  $18 \times 18 \text{ mm}^2$ ,  $\sim 0.17 \text{ mm}$  thick) and a glass slide ( $3 \times 1 \text{ inch}^2$ , 1 mm thick) were sandwiched together by two strips of double-sided tape to form a flow chamber with inner volume of  $\sim 20 \mu\text{l}$ . First,  $20 \mu\text{l}$  of biotin-labeled bovine albumin (1 mg/ml, dissolved in buffer A) was flown into the chamber and incubated for 2 min. The chamber was then washed using  $40 \mu\text{l}$  of buffer A.  $20 \mu\text{l}$  of streptavidin (0.5 mg/ml, dissolved in buffer A) was then flown through the chamber and allowed to bind for 2 min. After washing with  $40 \mu\text{l}$  of buffer A and subsequently with  $40 \mu\text{l}$  of buffer B,  $20 \mu\text{l}$  of biotin labeled DNA structures ( $\sim 300 \text{ pM}$  concentration) in buffer B were finally flown into the chamber and incubated for 5 min. The chamber was washed using  $40 \mu\text{l}$  of buffer B<sup>+</sup> for *in vitro* imaging of 20 nm grid structures and buffer B for all other *in vitro* DNA origami experiments.

*Imaging conditions.* The imaging buffer contained 10 nM (Fig. 2b,e) or 15 nM (Fig. 2c,d) Cy3B-labeled imager strands in buffer B<sup>+</sup>, or 20 nM Atto 655-labeled imager strands in buffer B (Supplementary Fig. 8). Imaging chambers were sealed with epoxy before imaging. Image acquisition was carried out with a CCD readout bandwidth of 3 MHz at 14 bit and 5.1 pre-amp gain. No EM gain was used. Imaging was performed using TIR illumination with an excitation intensity of  $\sim 5 \text{ mW}$  using the 561 nm laser line (Fig. 2b–e) or  $\sim 40 \text{ mW}$  using the 647 nm laser line (Supplementary Fig. 8). 15,000 frames at 10 Hz frame rate were acquired (Fig. 2b,e,h and Supplementary Fig. 8). 50,000 frames at 5 Hz frame rate were acquired (Fig. 2c green data points).

*Data analysis.* Super-resolution images were reconstructed in either custom LabVIEW or MATLAB software as described above. qPAINT analysis on DNA origami structures was performed interactively by selecting a region of interest that was large enough to fully comprise a single DNA origami (compare gray “ROIs” (regions of interest) in Fig. 2b). First, 100 random DNA origami structures displaying the same number of binding sites (i.e. 10 spots) were used to calibrate the probe influx rate  $\xi$ . For the imaging conditions used, a probe influx rate of  $0.02 \text{ s}^{-1}$  was determined (Fig. 2c,f, orange data points, and Fig. 2h) and  $0.03 \text{ s}^{-1}$  (Fig. 2c, green data points). These influx rates were then used to perform qPAINT analysis on all DNA origami structures in the sample. qPAINT analysis for the sample in **Supplementary Figure 8** was performed similarly. Here, the probe influx rate was determined by assuming that the 12 binding site origami structures carry on average 10.5 binding sites (determined by visual

counting of similar structures in Fig. 2b). Here, a probe influx rate of  $0.0201 \text{ s}^{-1}$  was determined for the imaging conditions used in this experiment.

*Simulation conditions.* qPAINT simulations for **Figure 2c** were performed using 9, 10, 11, or 12 binding sites as model input in combination with association and dissociation rates obtained from the experimental data. qPAINT simulations used in **Supplementary Figure 8** were performed in a similar fashion, here, however, using a normal distribution of binding sites with means determined from the experimental data. s.d. of the normal distribution were scaled based on the simulated mean value taking the 12 binding site data from **Figure 2b** as input. qPAINT simulations for **Figure 2g** were performed similarly to simulations in **Figure 1b**. Here, binding sites ranging from 2 to 200 binding sites and probe influx rates ranging from  $0.001 \text{ s}^{-1}$  to  $0.05 \text{ s}^{-1}$  were used to perform 200 stochastic simulations per value pair. After analysis, the coefficient of variation ( $c_v = \sigma/\mu$ ), reflecting the counting error, is plotted against the number of binding sites and probe influx rate. As before, 15,000 frames at an “integration time” of 0.1 s are simulated.

**Sample preparation, acquisition, and analysis of artificial receptor clusters on cell surfaces.** *Sample preparation.* CHO cells were transiently transfected with the construct encoding a human wild-type EGFR fused with the fluorescent protein mEos (pQCEGFRmEos; a kind gift from Jeff Werbin, Harvard Medical School) using Lipofectamine (Invitrogen) according to the manufacturer’s instructions. Briefly, for each transfection reaction,  $10 \mu\text{L}$  Lipofectamine was incubated with  $250 \mu\text{L}$  Opti-MEM for 10 min. Next,  $4 \mu\text{g}$  of EGFR-mEos vector plasmid was added and incubated for 30 min. This mixture was added to cells and incubated overnight. Transfections were carried out when CHO cells were at 90% confluence. After 24 h, approximately 30% confluence cells per well were seeded into Lab-Tek II chambers for 24 h before fixation. EGF receptors were immunostained using the following procedure: washing in PBS; fixation in a mixture of 3% paraformaldehyde and 0.1% glutaraldehyde in PBS for 10 min; three washes with PBS; reduction with  $\sim 1 \text{ mg/ml}$   $\text{NaBH}_4$  for 7 min; three washes with PBS; permeabilization with 0.25% (v/v) Triton X-100 in PBS for 10 min; three washes with PBS; blocking with 3% (w/v) bovine serum albumin for 30 min and staining overnight with the biotinylated EGFR monoclonal antibody (diluted to  $10 \mu\text{g/ml}$  in 5% BSA); three washes with PBS; incubation with  $20 \mu\text{g/ml}$  streptavidin (dissolved in 5% BSA) for 30 min; three washes with PBS; incubation with  $\sim 0.4 \text{ nM}$  of biotinylated Cy3-labeled DNA origami structures (dissolved in 5% BSA) for 1 h; three washes with PBS. Finally, imager strands were added in buffer C for image acquisition.

*Imaging conditions.* For surface-bound origami in **Supplementary Figure 9**, 12 nM Atto 655-labeled imager strands in buffer C were used. The CCD readout bandwidth was set to 5 MHz at 16 bit and 5.1 pre-amp gain with 255 EM gain was used. Imaging was performed using highly inclined (HILO) illumination<sup>36</sup> with an excitation intensity of  $\sim 23 \text{ mW}$  using the 647 nm laser line.

*Data analysis.* Super-resolution images were reconstructed as described above. Here, ROI selection for qPAINT was performed in a semi-automated fashion. In brief, the co-localization between diffraction-limited and DNA-PAINT images was used to select the structures (ROIs) bound to the membrane of the cells.

Then, diffraction-limited and super-resolved images were aligned by computing the cross-correlation between them and discarding non-co-localizing molecules. After selection, qPAINT quantification was performed as described above. As before, the probe influx rate was determined by assuming that the 12 binding site origami structures carry on average 10.5 binding sites (determined by visual counting of similar structures in Fig. 2b). Here, a probe influx rate of  $0.02163 \text{ s}^{-1}$  was determined, resulting in an association rate of  $k_{\text{on}} = 1.78 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$  under these imaging conditions.

**Sample preparation, acquisition, and analysis of DNA origami microinjected into fixed cells.** *Sample preparation.* DNA origami structures with 44 binding sites for DNA-PAINT and fixed Cy3-labeled strands were injected into fixed HeLa cells using a Femtojet (Eppendorf, NY). Multiple injections per cell were performed into nuclear and cytoplasmic regions with the origami sample at  $\sim 1 \text{ nM}$  concentration. The incorporation of the structures inside the cells was confirmed by confocal microscopy (Supplementary Video 1) and by deconvolution of a wide-field image stack using Huygens Professional image processing software (Supplementary Video 2).

*Imaging conditions.* For microinjected DNA origami in Supplementary Figure 9, 5 nM 655-labeled imager strands in buffer C were used. The CCD readout bandwidth was set to 3 MHz at 14 bit and 5.1 pre-amp gain with no EM gain was used. Imaging was performed using HILO illumination with an excitation intensity of  $\sim 50 \text{ mW}$  using the 647 nm laser line.

*Data analysis.* Super-resolution DNA-PAINT images were reconstructed as described above. For qPAINT analysis, DNA origami structures in the cytoplasm were separated from structures in the nucleus by using the DAPI signal for segmentation (Supplementary Fig. 9). After structure selection, quantification was done as described above in order to determine hybridization kinetics for DNA-PAINT probes in nuclear and cytoplasmic regions. For calculating the number of binding sites, the same association rate determined in experiments on cell surfaces ( $k_{\text{on}} = 1.78 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ ) was used, resulting in a probe influx rate of  $0.0089 \text{ s}^{-1}$ . We used the same apparent association rate, as imaging and buffer conditions are the same as in the cell surface experiments.

**Sample preparation, image acquisition, and analysis of qPAINT applied to NPCs.** *Sample preparation.* The nucleoporin Nup98 was labeled using a custom DNA-conjugated monoclonal antibody. In brief, first a 5'-thiolated DNA-PAINT docking strand (p1 sequence) was reduced with DTT and purified using NAP5 columns from GE Healthcare (Pittsburgh, PA). In a second step, the monoclonal antibody (#2598, Cell Signaling) was cross-linked with a 50 $\times$  excess of a maleimide-PEG-succinimidyl ester crosslinker (746223, Sigma-Aldrich). Excess of linker that did not react with antibody was removed using Zeba columns (89883, Life Technologies). Third, the antibody-linker was reacted with the reduced 5'-thiolated DNA-PAINT docking strand at 4C overnight. Finally, after conjugation, the oligo-labeled antibodies were purified by using 100 kDa Amicon ultra filter (UFC510096, Emdmillipore) from unreacted DNA. The sample was then characterized by mass spectrometry, determining that on average each antibody is coupled with two docking strands.

U2OS cells were cultured in DMEM supplemented with 10% FBS. Immunostaining was performed using a standard protocol<sup>37</sup>. Cells were grown overnight on 8-well Lab-Tek Chambered Coverglass. The cells were then rinsed with 2.4% PFA for 20 s in PBS and extracted with 0.4% Triton X-100 in PBS for 3 min. Next, the cells were fixed for 30 min with 2.4% PFA in PBS, quenched for 5 min with 50 mM NH<sub>4</sub>Cl, then extensively washed with PBS (3 $\times$ , 10 min each), and blocked with 5% BSA in PBS (blocking buffer) for 1 h. Subsequently, the cells were incubated overnight at 4 °C with the DNA-conjugated monoclonal antibody freshly diluted in the blocking buffer and then incubated with DAPI plus drift markers for 5 min. Then cells were washed once with buffer C before imaging and, finally, imager strands were added in buffer C for image acquisition.

*Imaging conditions.* For Figure 3a, 5 nM Cy3B-labeled imager strands in buffer C were used. The CCD readout bandwidth was set to 3 MHz at 14 bit and 5.1 pre-amp gain with no EM gain. Imaging was performed using HILO illumination with an excitation intensity of  $\sim 8 \text{ mW}$  using the 561 nm laser line. 15,000 frames at 5 Hz frame rate were acquired.

*Data analysis.* Super-resolution DNA-PAINT images were reconstructed as described above. For qPAINT analysis, NPC structures in the nucleus were masked from the signal in the cytoplasm using the DAPI signal. Single Nup98 proteins or complete structures were masked using particle filter and automatic structure selection. Quantification was done as described above in order to determine hybridization kinetics for DNA-PAINT probes. The association rate was determined by using the mean value from the Gaussian fitting of the dark time distribution from the individual protein locations (see arrows in Fig. 3a) to  $k_{\text{on}} = 1.75 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ , resulting in a probe influx rate of  $0.00875 \text{ s}^{-1}$ . This value was then used to perform all qPAINT calculations.

Distances between each individual Nup98 protein in an NPC were determined by first calculating the coordinates of the center for each structure using circular Hough transform. Then we projected the histograms for the cluster of localization from the center forming a linearized intensity profile with a given number of peaks (3–7 spots, see Supplementary Fig. 10). Distances between spots were calculated by fitting the histograms with minimal number of Gaussians, and calculating the intervals between the peaks. This routine was implemented using custom-written MATLAB software.

**Sample preparation, image acquisition, and analysis of qPAINT for quantifying Brp proteins in CAZ-units.** *Sample preparation.* The Brp protein was labeled using a monoclonal antibody Brp<sup>Nc82</sup> (nc82, DSHB). For DNA-PAINT, we used custom secondary anti-mouse goat-DNA-labeled antibodies. The secondary antibody was labeled using the same protocol used for the conjugation of the Nup98 antibody.

*Drosophila* crawling third instar larvae fillets were prepared by dissection in ice-cold HL-3.1 and fixed for 10 min using 4% paraformaldehyde in HL-3.1. The fillets were then permeabilized with 3 washes of 15 min each in PBT (PBS with 0.05% Triton X-100) followed by a 30 min block in PBT containing 1% normal goat serum and 2% BSA. Preparations were incubated with primary antibodies at 4 °C overnight. After one short and three  $\sim 20$  min washing steps, the fillets were incubated with secondary DNA conjugated antibodies for 2 h followed by another

three washing steps. Then fillets were washed once with buffer C before imaging and, finally, imager strands were added in buffer C for image acquisition.

**Imaging conditions.** For **Figure 3d**, 10 nM Cy3B-labeled imager strands in buffer C were used. The CCD readout bandwidth was set to 3 MHz at 14 bit and 5.1 pre-amp gain with no EM gain. Imaging was performed using HILO illumination with an excitation intensity of ~8 mW using the 561 nm laser line. 15,000 frames at 10 Hz frame rate were acquired.

**Data analysis.** Super-resolution DNA-PAINT images were reconstructed as described above. For qPAINT analysis, DNA-PAINT signal from synaptic boutons on neuromuscular junctions (NMJ) were masked using anti horseradish peroxidase (HRP-A488). Single Brp molecules or complete CAZ-units were masked using particle filter and automatic structure selection. Quantification was done as described above in order to determine hybridization kinetics for DNA-PAINT probes. The association rate was determined by using the mean value from the Gaussian fitting of the dark time distribution from the individual Brp molecule locations (see **Fig. 3d** green circles) resulting in a probe influx rate of  $0.0023 \text{ s}^{-1}$ . This value was then used to perform all qPAINT calculations on individual CAZ-units. This routine was implemented using custom-written MATLAB software.

**Sample preparation, acquisition, and analysis of smRNA-FISH in fixed cells.** **Sample preparation.** SUZ12 mRNAs molecules were tagged in fix HeLa cells using single-stranded oligonucleotides in a FISH-like hybridization scheme. mRNAs were labeled with a unique set of 64 DNA single strand oligonucleotides (**Supplementary Table 8**). Each probe consists of a unique binding sequence to the mRNA (20 nucleotides length), Cy3B coupled to the 5'-end. The 3'-end carries a single-stranded extension for DNA-PAINT (**Fig. 3f**). For coupling, the oligonucleotides were ordered with 5'-amino modifications from IDT (Coralville, IA). Cy3B was ordered as a succinimidyl ester derivative from GE Healthcare (Pittsburgh, PA). After coupling, the dye-labeled DNA strands were purified by using high pressure liquid chromatography<sup>28</sup>. To perform RNA-FISH, we followed a standard protocol<sup>29</sup>. Cells were fixed with pre-chilled methanol ( $-20 \text{ }^{\circ}\text{C}$ ) for 30 min. Following fixation, cells were hybridized with the probes at 16 mM each in hybridization buffer consisting of 10% formamide,  $2\times$  SSC, and 10% dextran sulfate (w/v). Samples were then hybridized overnight in a humidified chamber at  $37 \text{ }^{\circ}\text{C}$ . Following hybridization, samples were washed twice with wash buffer ( $2\times$  SSC with 10% formamide) for 30 min at  $37 \text{ }^{\circ}\text{C}$ , and then incubated with DAPI plus drift markers for 5 min. Then cells were washed once with buffer C before imaging and, finally, imager strands were added in buffer C for image acquisition.

**Imaging conditions.** For **Figure 3f**, 5 nM Atto 655-labeled imager strands in buffer C were used. The CCD readout bandwidth was set to 3MHz at 14 bit and 5.1 pre-amp gain with no EM gain was used. Imaging was performed using HILO illumination with an excitation intensity of ~50 mW using the 647 nm laser line.

**Image analysis.** Super-resolution DNA-PAINT images were reconstructed as described above. Here, Bright field and DAPI images were used to identify individual cells. Diffraction-limited Cy3B and DNA-PAINT images were used to detect RNA transcripts. For image processing, a binary brightfield image was merged with a binary DAPI image. The resulting mask was used to segment the cells. Diffraction-limited spots were detected using a script that operates as follows: First, a median filter followed by a Laplacian filter was applied to each image. Then, a threshold was selected to detect individual spots. This procedure enabled us to identify the total number of diffraction-limited spots within each cell. For quantification of diffraction-limited and super-resolved images, they were aligned by computing the cross-correlation between them. After selection, qPAINT quantification was performed as described above for each individual spot in order to determine the number of probes bound to each individual mRNA. For calculating the number of binding sites, the same association rate determined previously ( $k_{\text{on}} = 1.78 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ ) was used, resulting in a probe influx rate of  $0.0089 \text{ s}^{-1}$ . We used the same apparent association rate as imaging and buffer conditions are the same as in the experiments in **Supplementary Figure 9**.

**Drift correction.** *In vitro* imaging. Drift correction was performed with the custom-written MATLAB software<sup>21</sup>. The positions of all DNA origami structures were tracked throughout the duration of each movie and averaged for use as the drift correction trace. In the 20 nm grid super-resolution image (**Fig. 2b**), an advanced drift correction algorithm<sup>21</sup> was performed to visually separate the individual grid points.

*In situ* imaging. For cellular imaging, 100 nm gold nanoparticles (Sigma-Aldrich; 10 nM in buffer C, added before imaging) were used as drift markers. The gold nanoparticles adsorb non-specifically to the glass bottom of the imaging chambers. Drift correction is performed in a similar fashion as for the *in vitro* imaging (see above). The apparent "movement" of all gold nanoparticles in a field of view is tracked throughout the movie. The obtained trajectories are then averaged and used for global drift correction of the final super-resolution image.

35. Hoops, S. *et al. Bioinformatics* **22**, 3067–3074 (2006).

36. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. *Nat. Methods* **5**, 159–161 (2008).

37. Szymborska, A. *et al. Science* **341**, 655–658 (2013).

# Optical imaging of individual biomolecules in densely packed clusters

Mingjie Dai<sup>1,2</sup>, Ralf Jungmann<sup>1,3†</sup> and Peng Yin<sup>1,3\*</sup>

**Recent advances in fluorescence super-resolution microscopy have allowed subcellular features and synthetic nanostructures down to 10–20 nm in size to be imaged. However, the direct optical observation of individual molecular targets (~5 nm) in a densely packed biomolecular cluster remains a challenge. Here, we show that such discrete molecular imaging is possible using DNA-PAINT (points accumulation for imaging in nanoscale topography)—a super-resolution fluorescence microscopy technique that exploits programmable transient oligonucleotide hybridization—on synthetic DNA nanostructures. We examined the effects of a high photon count, high blinking statistics and an appropriate blinking duty cycle on imaging quality, and developed a software-based drift correction method that achieves <1 nm residual drift (root mean squared) over hours. This allowed us to image a densely packed triangular lattice pattern with ~5 nm point-to-point distance and to analyse the DNA origami structural offset with ångström-level precision (2 Å) from single-molecule studies. By combining the approach with multiplexed exchange-PAINT imaging, we further demonstrated an optical nanodisplay with 5 × 5 nm pixel size and three distinct colours with <1 nm cross-channel registration accuracy.**

Biological and synthetic biomolecular systems exhibit complex structures at the nanoscale. Understanding the spatial arrangement of their individual components is critical for unravelling the molecular mechanism that underlies complex molecular behaviour (Fig. 1a). Super-resolution fluorescence techniques have bypassed the traditional diffraction limit and demonstrated imaging resolution down to 10–20 nm (refs 1–13). In particular, single-molecule localization microscopy (SMLM) builds up super-resolution images from single-emitter localizations and typically achieves photon-limited localization precision (down to ~1 nm) for single-emitter blinking events<sup>13–15</sup>. Previous single-molecule and SMLM studies separately demonstrated single-target visualization in isolation or in sparse arrangements<sup>16,17</sup>, and high localization precision compatible with molecular-scale resolution<sup>18–24</sup>. However, the discrete visualization and precise localization of each individual molecular target (~5 nm) in a densely packed biomolecular cluster, which we refer to as ‘discrete molecular imaging’ (DMI), remains difficult.

Several factors limit the performance of current super-resolution techniques, such as the finite fluorophore photon budget, unsatisfactory fluorophore imaging efficiency, or limited control over target blinking kinetics<sup>9,22,25–30</sup>. These restrictions respectively translate to a limited photon count per localization, a limited number of blinking events per target and a high fraction of false localizations, which ultimately restrict the final imaging resolution, signal-to-noise ratio and the visualization of individual targets within dense clusters. Moreover nanometre-level accuracy stage noise<sup>8</sup> and drift compensation is critical for high imaging resolution and quality.

We addressed the above challenge and demonstrated DMI using the DNA-PAINT super-resolution method<sup>10,23,31–33</sup> (Supplementary Fig. 1). DNA-PAINT, a variation of PAINT<sup>9</sup>, exploits the transient binding of fluorophore-labelled imager strands to target-bound docking strands to achieve the necessary blinking for super-resolution reconstruction<sup>10</sup> (Fig. 1b; Supplementary Fig. 2). The continuous replenishment of imager strands renders DNA-PAINT immune to photobleaching, allowing high localization precision

by extracting a large number of photons per single-molecule localization and a high target separability by collecting a large number of blinking events from each target. Additionally, due to independent and programmable control of blinking on- and off-rates, DNA-PAINT permits low imaging background in dense clusters from appropriately adjusted blinking duty cycle based on the target density (Fig. 1c).

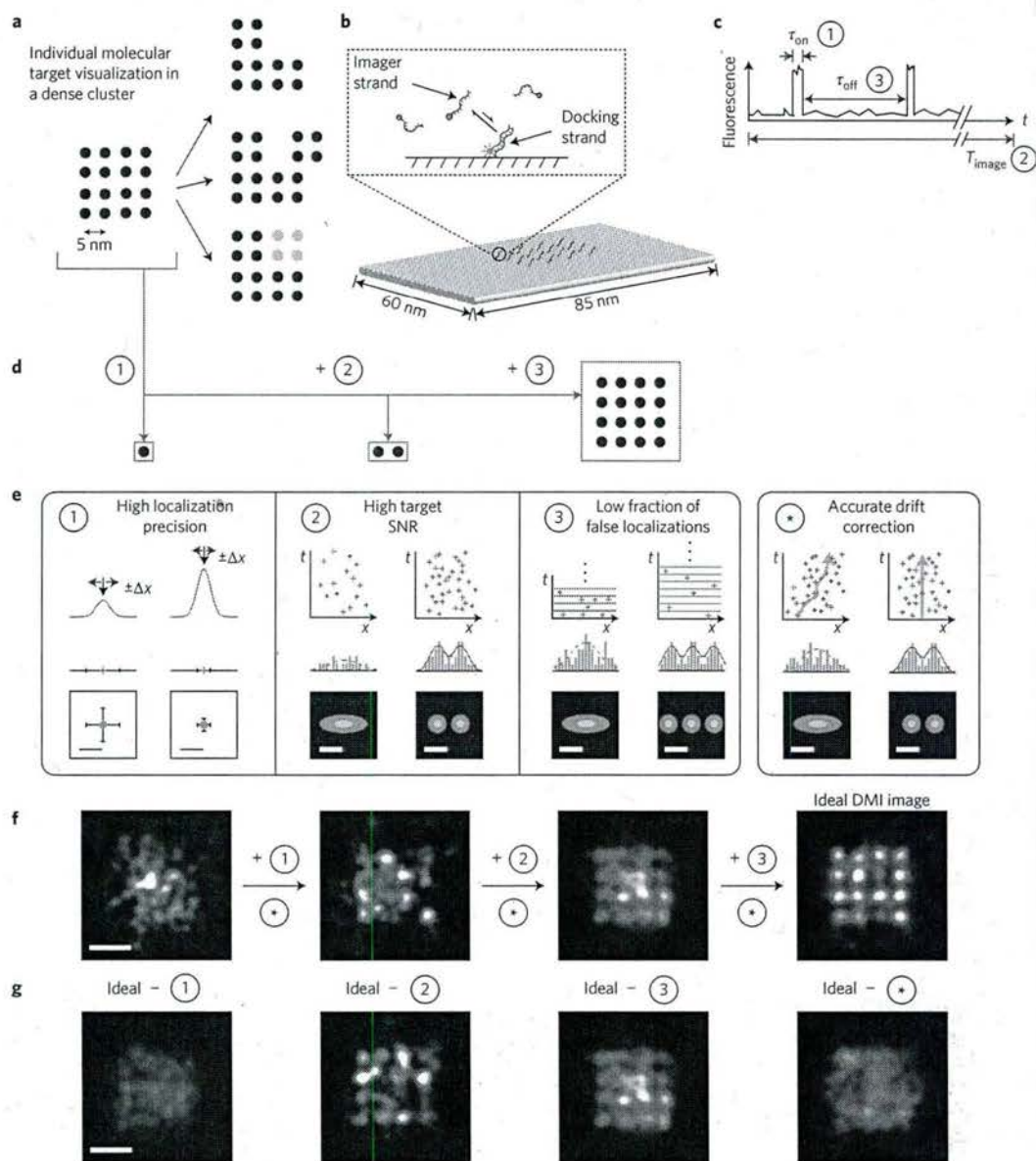
We developed a framework for achieving DMI in the context of localization microscopy, including the technical requirements and quality assay methods (Supplementary Fig. 1). Specifically, we introduced an image-based assay for measuring the localization precision and the maximally achievable resolution, a target signal-to-noise ratio assay for measuring single-target separation and a method for estimating the fraction of false double-blinking localizations. We systematically studied their effects, and demonstrated stringent control for each of them with DNA-PAINT. In particular, we achieved a high localization precision (<1 nm single-molecule fitting precision, from up to 50,000 photons per single-molecule localization), a high target separability (from ~80 blinking events per target), a low imaging background using appropriately tuned blinking duty cycles and high-accuracy (<1 nm root mean squared (r.m.s.) over hours of imaging) microscope stage drift correction with a novel method based on synthetic nanostructure drift markers with designed geometric patterns. Finally, we used DMI to visualize individual targets in a compactly labelled molecular grid of targets (with a point-to-point spacing of ~5 nm) and demonstrated multiplexed DMI on a three-colour nanodisplay board with ~5 nm pixels.

## Technical requirements for DMI

The technical requirements for DMI depend on the spatial distribution of the targets, particularly the closest spacing between targets and their local density<sup>26,34</sup> (within a diffraction-limited region, see Supplementary Note 7.1). Even with only two targets, a high localization precision that allows a full width at half maximum (FWHM) resolution equal to or smaller than the

<sup>1</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, USA. <sup>2</sup>Biophysics Program, Harvard University, Boston, Massachusetts 02115, USA. <sup>3</sup>Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.

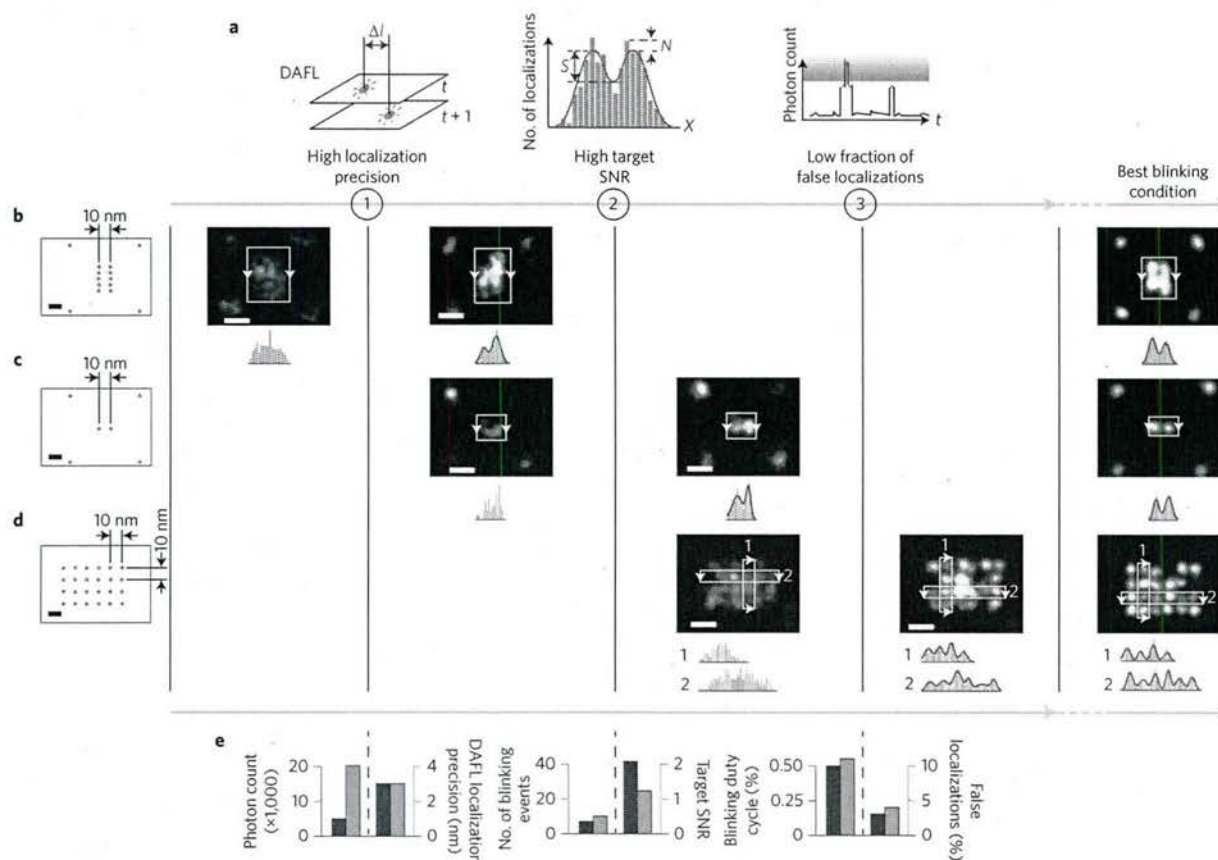
<sup>†</sup>Present address: Max Planck Institute of Biochemistry and LMU, Munich 82152, Germany. \*e-mail: py@hms.harvard.edu



**Figure 1 | Principle and requirements of DMI.** **a**, The concept of super-resolution DMI, illustrated with a point array representation (blue dots represent individual molecular targets, yellow dots represent chemical modifications). Left: A regular 16-component biomolecular complex. Right: Its various structural and chemical variations. **b**, Illustration of the DNA-PAINT principle: transient binding between a docking strand and dye-conjugated imager strands (top) on a synthetic DNA nanostructure, where each cylinder represents a DNA double helix (bottom). **c**, Schematic DNA-PAINT blinking time trace of a single imaging target. The three blinking characteristics measure the blinking on-time  $\tau_{\text{on}}$  (1), the total imaging time  $T_{\text{image}}$  (2) and the blinking off-time  $\tau_{\text{off}}$  (3), and can be tuned to meet the three blinking requirements in **e**. **d**, Schematics of different substructures from the complex in **a**: a single target (left), a pair of targets in close proximity (middle) and a dense lattice (right). All need different blinking requirements in **e** to be clearly visualized. **e**, Technical requirements for achieving discrete molecular imaging. Each panel outlines one technical requirement and schematically depicts the effect on imaging quality before (left column) and after (right column) the requirement is satisfied. For requirement (1), the intensity profile in 1D (top), fitted Gaussian centres in 1D (middle) and 2D (bottom) are shown. For requirements (2), (3) and (\*) the localization time trace in 1D (top), localization histogram in 1D (middle) and 2D (bottom) are shown. The orange lines and crosses indicate localizations. The orange bars depict localization histograms. The solid red lines and dotted grey lines indicate successful and failed Gaussian fittings on localization histograms, respectively. For (3), grey crosses indicate true localizations eclipsed by false double-blinking localizations. The same numbering for technical requirements (1)–(3) is also used in Fig. 2, Table 1 and Supplementary Fig. 3. **f, g**, Simulations of the imaging effects of the technical requirements for the complex in **a** under increasingly improved imaging conditions without stage drift (**f**) or under non-ideal imaging conditions with one of the four requirements unsatisfied (**g**). See Methods and Supplementary Methods 2 for simulation details, and Supplementary Note 7 for discussions. Scale bars, 5 nm (**e**); 10 nm (**f, g**).

spacing between them is necessary but not sufficient for their clear separation<sup>23,35</sup>. Targets arranged in dense clusters impose even more stringent requirements. Although general requirements and quality

guidelines for super-resolution imaging have been discussed<sup>26,27,34,36</sup>, a systematic formulation of the technical requirements for DMI and its quality control methods is still lacking.



**Figure 2 | Systematic characterization of the blinking requirements and optimization of the DNA-PAINT imaging quality.** **a**, Methods for the systematic characterization of the three blinking requirements depicted in Fig. 1. (1) DAFL measures the distance ( $\Delta l$ ) between pairs of spatially close localizations originating from adjacent camera frames. (2) Target SNR measures the separability of the peaks in the localization histogram in the super-resolved image.  $S$ , signal;  $N$ , noise. The red curve indicates the two-peak Gaussian fit. (3) The photon count cut-off in the blinking trace measures the fraction of false localizations. The blue shaded area indicates identified false localizations. The orange markers, bars and curves indicate localizations, histograms and time traces, respectively. **b–d**, Designed DNA origami standards with 10 nm spacing under different blinking conditions. The left column shows the design schematics of the DNA origami standards; green dots indicate DNA-PAINT docking strands; four corners in **b,c** are used as alignment markers. The right columns show DNA-PAINT images under increasingly better blinking conditions (one condition per column). The histograms below the images show the projection profiles from the areas indicated by white boxes along the directions of arrows. **e**, Quantitative characterization and pairwise comparisons of the imaging conditions used in **b–d** before and after meeting each extra requirement, assayed with the methods in **a**. For each comparison, the left y axis (blue) shows the control parameter and right y axis (green) shows the experimental measurement. For more details see Supplementary Figs 4–6 on origami designs, Supplementary Figs 7–14 for super-resolution images, Methods and Supplementary Methods 3 and 5 for DNA-PAINT imaging conditions and analysis methods. Scale bars, 10 nm in schematics and 20 nm in super-resolution images.

We examined the technical requirements for DMI with an example square lattice pattern (Fig. 1d). It is increasingly more difficult to discretely identify and precisely position a single isolated target (Fig. 1d, left), a pair of targets in close proximity (middle) and finally a dense lattice of targets (right), thus requiring increasingly more stringent imaging conditions (Fig. 1e). These four conditions are described as follows. Requirement (1): A high localization precision. This can be obtained by collecting a high photon count per single-molecule localization, and allows the precise localization of an isolated target. Requirement (2): A high target signal-to-noise ratio (target SNR) in the super-resolved image. This can be achieved by collecting a large number of blinking events per molecular target, and allows for a clear separation between two nearby targets. Requirement (3): A low fraction of false localizations from double-blinking events. This can be achieved by using a low blinking on-off duty cycle, which minimizes falsely localized background noise and hence allows the discrete visualization of each target within a densely packed complex. Lastly, requirement (\*): An accurate compensation

mechanism for microscope stage drift. This is important for the accurate identification and localization of any molecular structures of interest, especially over extended imaging times. Computer-simulated super-resolution images reveal increasingly better image qualities when more of these requirements are satisfied (Fig. 1f) and that each of the four requirements is indispensable for DMI (Fig. 1g). See online methods, Supplementary Method 2 and Supplementary Note 7.2 for details.

We termed requirements (1)–(3) collectively the three blinking requirements, as they can all be met by appropriate single-molecule blinking properties (see Fig. 1c). As DNA-PAINT allows flexible tuning of target blinking kinetics, it therefore provides a promising route for implementing DMI.

### Systematic characterization and quality control for DMI

To quantitatively characterize the effects of the above DMI requirements on imaging quality, we proposed a set of assay methods (Fig. 2a): (1) an image-based assay of the localization precision

and the maximal achievable resolution by comparing the positions of super-localized centres from neighbouring frames (termed the distance between adjacent-frame localizations, or DAFL); (2) a target SNR assay based on the analysis of the distribution of super-localized centres, which directly measures the separability of neighbouring targets in super-resolved images; and (3) a localization time trace-based assay for estimating the false localization ratio. These assay methods provide a general, sample-agnostic method for the stringent quality control of general super-resolution microscopy studies as well as DMI (see Methods, Supplementary Fig. 3 and Supplementary Methods 3 for details).

Using these assay methods, we first simulated super-resolution movies with varying photon counts, numbers of blinking events and blinking duty cycles, and measured the resultant single-molecule localization precision, the target SNR and the fraction of false localizations from double-blinking events (see Methods, Supplementary Fig. 3 and Supplementary Methods 2 and 3 for details). We observed a high localization precision ( $<1$  nm, supporting  $<2$  nm FWHM resolution) with a high photon count ( $>30,000$ ), a consistent target separation with a high target SNR ( $>2$  under our definition) and low background noise under a low fraction of false localizations (down to  $<5\%$ ), allowing DMI imaging.

Applying the same assay methods, the three blinking requirements (localization precision, target SNR and false localization ratio) for DMI were each experimentally verified by subjecting synthetic DNA origami nanostructure<sup>37,38</sup> standards with three designed target patterns to different DNA-PAINT imaging conditions (Fig. 2b–d: the leftmost columns are the designed pattern schematics; the five right columns are DNA-PAINT images, see Methods and Supplementary Methods 4 for details). Synthetic DNA nanostructures provide a programmable and geometrically precise molecular patterning platform for single-molecule and super-resolution studies. These structures were self-assembled from a long single-stranded DNA scaffold and a collection of short staple strands<sup>37</sup>. By extending a selected subset of these staple strands, DNA-PAINT docking strands could be arranged into user-prescribed, geometrically precise nanopatterns and used as super-resolution imaging standards or auxiliary markers (see the next section, Supplementary Figs 4–6 and Supplementary Note 8 for details).

In Fig. 2b, two 10-nm spaced lines (each consisting of 5 points, spaced 5 nm apart to satisfy the Nyquist criterion) only turned from unresolvable speckles (left image) to separable lines (right image) with an increased photon count per single-molecule localization and hence a higher localization precision. However, the high localization precision alone under this imaging condition failed to resolve two points spaced by 10 nm (Fig. 2c, left image) due to the reduced number of targets (from 5 to 1 on each side). The two points only became resolvable (Fig. 2c, right image) with a larger number of blinking events per target and hence an increased target SNR. However, this imaging condition (with a high localization precision and target SNR) still failed to resolve a 24-target 10-nm grid (Fig. 2d, left image) due to the increased target density. The grid points only became individually resolvable (Fig. 2d, right image) with a decreased blinking duty cycle and hence a lower false localization ratio. Quantitative pairwise comparisons of these imaging conditions and the imaging quality assay results before and after meeting each of the three blinking requirements are shown in Fig. 2e (see Methods, Supplementary Figs 7–12 and Supplementary Methods 5.1 and 6.1 for details). Finally, we imaged these structures under the best imaging conditions and obtained clear images of the designed patterns (Fig. 2b–d, rightmost column, see Supplementary Figs 13 and 14 for details).

#### Subnanometre accuracy software-based drift correction

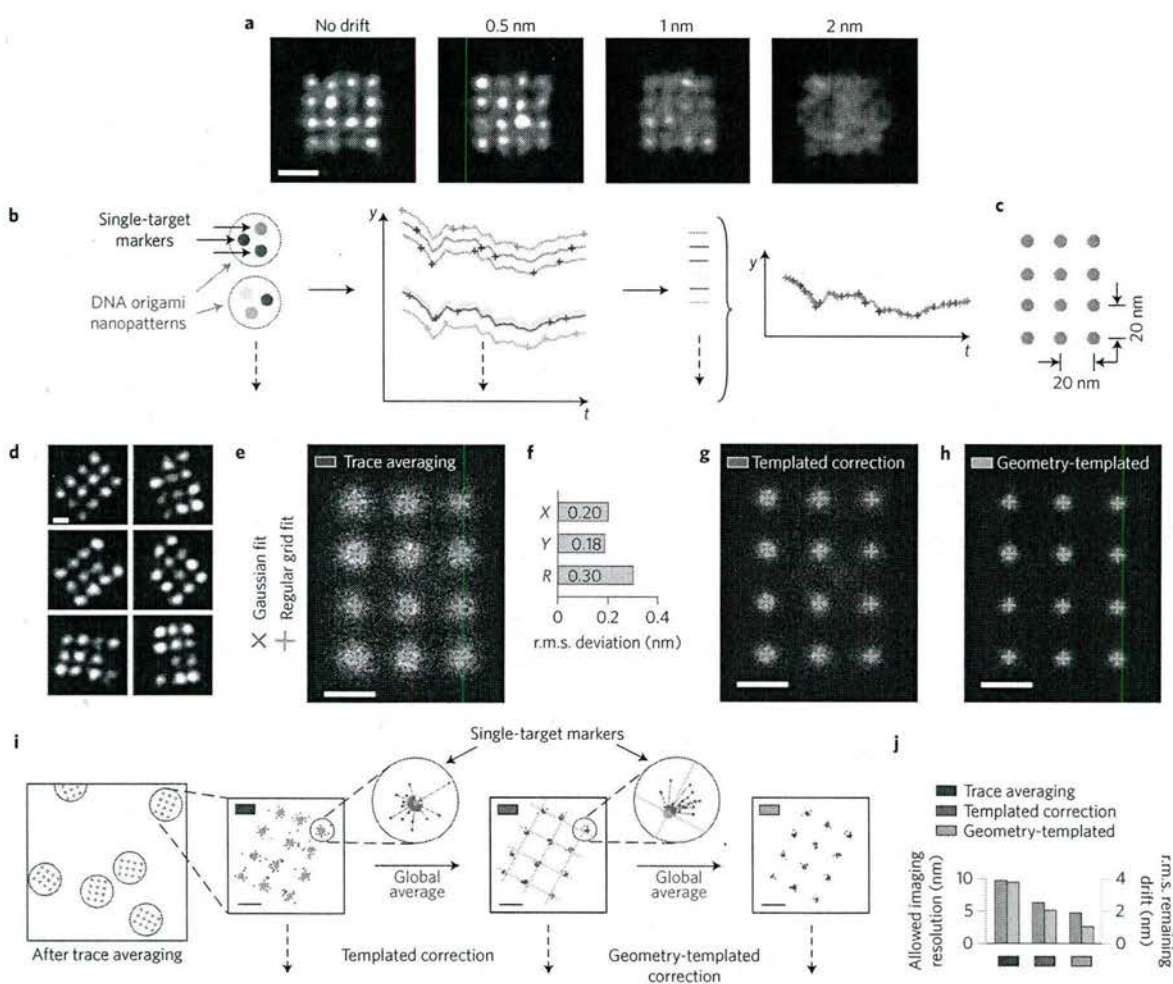
DMI also imposes stringent requirements on microscope stage drift compensation. For example, the clear separation of targets with

5 nm target-to-target spacing requires an accurate drift correction of  $<1$  nm (r.m.s. drift) due to the compounded effects of stage drift and finite localization precision (assuming 1 nm localization precision, Fig. 3a; see Methods and Supplementary Methods 2 for details). Previous solutions either used an active feedback system to reach  $<1$  nm residual drift, which requires a complicated hardware set-up and is technically involved to implement<sup>21</sup>, or used software-based post-processing methods with embedded nanoparticle fiducial markers, which are typically limited by imperfect tracking accuracy<sup>5,7,22</sup> and do not support DMI resolution ( $\sim 5$  nm).

We demonstrate a novel, synthetic nanostructure fiducial marker-based stage drift correction and noise compensation method—termed templated drift correction—which achieves high-accuracy drift and noise cancellation ( $<1$  nm r.m.s. residual drift) over both short and long timescales (from subsecond to  $\sim 5$  h), without using specialized hardware. The key principle is to use super-resolved single-molecule targets (such as a single DNA docking strand) on pre-designed geometrically precise nanopatterns (such as a DNA nanostructure) as drift markers (Fig. 3b). Unlike conventional fiducial markers, these single-target drift markers possess unique advantages as they can be stably anchored on the surface, localized with high precision and without bleaching, flexibly rotatable around their anchor points (thus avoiding the fixed-dipole effect) and free of offset from the fluorescence centre to the drift marker centre position, such as in micrometre-sized beads and previous DNA origami drift markers<sup>10,31</sup>. We arranged these single-target markers in pre-designed, well-separated nanopatterns (the templates) to help pack a number of single-target markers within a diffraction-limited area and to allow their identification and separation during software processing steps (Fig. 3b). Moreover, the precise geometry of a nanopattern correlates all of the single-target drift markers on this pattern and effectively increases their on-fraction (the fraction of time that the marker is bright), producing a more accurate drift correction. We term this method geometry-templated drift correction.

We designed a square lattice pattern with 20 nm spacing as our nanopattern template to implement this strategy (Fig. 3c): a grid consists of 12 targets (docking strands), each of which will be treated as a single-target drift marker. After performing DNA-PAINT imaging and applying a round of simple trace-averaging drift correction, all 12 targets are clearly separated from each other in the reconstructed image, allowing them to serve as single-target drift markers (Fig. 3d, see also Supplementary Fig. 15). Note that the missing grid points were probably due to defects in the origami self-assembly or DNA synthesis (see Supplementary Note 9.1 for discussion), rather than imaging incompleteness. Single-particle analysis confirmed the completeness of the nanopattern and showed a high degree of regularity of the grid geometry as designed (Fig. 3e, see Methods, Supplementary Fig. 16 and Supplementary Method 6.2 for details). Specifically, we performed two-dimensional (2D) Gaussian fitting on each grid point followed by a regular grid fit to these 12 Gaussian-fitted centres, and observed that the average deviation between the Gaussian-fitted and regular grid-fitted centres was well below 1 nm ( $<0.30$  nm r.m.s., Fig. 3f; see Methods, Supplementary Methods 6.2 and Supplementary Note 9.2 for details).

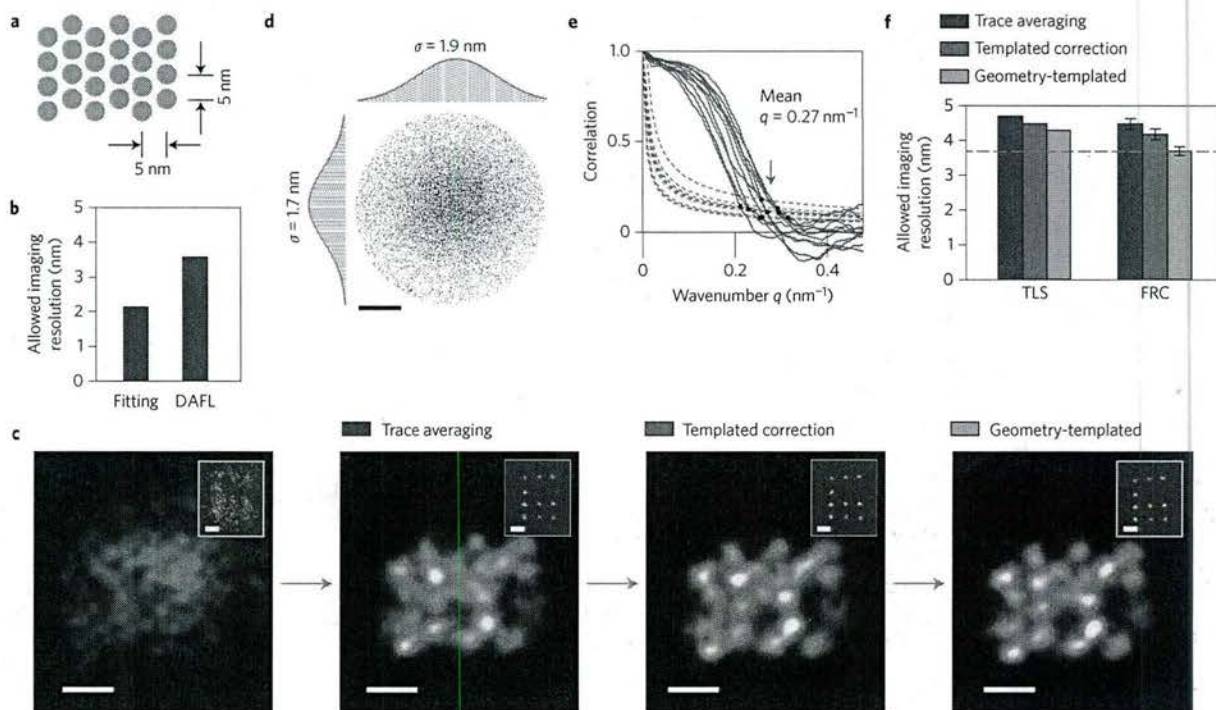
We then tested the performance of templated drift correction on the grid structures using the analysis workflow in Fig. 3i (left arrow, see Methods and Supplementary Methods 5.2 for details). Briefly, we first identified a pool of separable single-target drift markers, extracted their blinking time traces and determined the expected centre position for each target. Then for every frame we collected all of the localizations originating from these drift markers and calculated a global, photon-weighted average of 'offset vectors' from each localization to its expected centre position and used it as the drift correction vector. Single-particle analysis after templated drift correction showed a sharper image of the grid (Fig. 3g).



**Figure 3 | Principle and performance of DNA nanostructure templated drift correction.** **a**, The effect of drift on imaging quality is simulated for the biomolecular complex in Fig. 1a with 1 nm localization precision and different levels of stage drift. The structure can only be clearly visualized with 1 nm (r.m.s.) of drift or less. **b**, The principle of the templated drift correction method with pre-designed nanostructure patterns. Illustrating a three-target marker example, the schematics show a nanopattern design with single-target markers (left), the localization time traces from individual single-target markers (middle) and averaged drift correction trace after combining the traces from many markers (right). The targets and traces are colour-matched. Dashed arrows indicate many more structures used for drift correction. **c**, Design schematic of a  $3 \times 4$  square grid with 20 nm point-to-point spacing on a DNA origami nanostructure. Each green dot indicates a docking strand. **d**, Representative DNA-PAINT super-resolution images of the 20 nm grid structure in **c** imaged with 300 ms frame time, 30,000 total frames and 3 nM imager strands. Missing grid points were probably due to synthesis or incorporation defects (see Supplementary Note 9.1 for more discussion). **e**, Single-particle averages of 20 nm grid images ( $n = 700$ ) after trace averaging. The overlaid crosses indicate Gaussian fitted centres and regular grid-fitted centres using the red crosses as fitting targets. **f**, Root-mean-square (r.m.s.) deviation between the Gaussian fitted and regular grid-fitted centres in **e** in X, Y projections and 2D distance. **g, h**, Single-particle averages of the 20 nm grid images ( $n = 700$ ) after templated (**g**) and geometry-templated (**h**) drift correction. The overlaid crosses indicate Gaussian fitted and regular grid-fitted centres as in **e**. The same colour code for different stages of drift correction in **e, g** and **h** are also used in **i, j** and Fig. 4. **i**, Procedure for templated and geometry-templated drift correction with 20 nm grid structures as templates. The schematics show a large field-of-view image with many drift markers after simple trace averaging (left). Each grey circle indicates a 20 nm drift marker. Three zoomed-in (squares) schematics show a super-resolved 20 nm grid marker after simple trace averaging (left), after templated drift correction (middle) and after geometry-templated correction (right). Dashed arrows indicate many more structures used for drift correction. Further zoomed-in schematics (circles) show one single-target marker and the calculation of the offset vectors. In the zoomed-in schematics (both squares and circles) grey dots indicate localizations, green dots and lines indicate Gaussian-fitted centres and regular grid-fitted lattices as guides for templated and geometry-templated drift correction calculations, and the red lines represent the calculated offset vectors. **j**, Comparison of the allowable imaging resolution (measured in FWHM, orange) and estimated remaining drift (turquoise) at different stages of drift correction. For more details, see Methods and Supplementary Methods 2 and 5 for the simulation and analysis methods, and Supplementary Fig. 15 for the super-resolution images. Scale bars, 10 nm (**a**); 20 nm (**d, e, g, h** and zoomed-in images in **i**).

Furthermore, the superior regularity of these 20 nm grid structures allowed us to perform another round of geometry-templated drift correction, again using these structures as drift markers (Fig. 3i, second arrow; see Methods and Supplementary Methods 5.2 for details), and producing an even sharper single-particle averaged

image (Fig. 3h). We also compared the imaging resolution before and after templated and geometry-templated drift corrections, respectively, and estimated  $<1$  nm (r.m.s.) residual drift after templated and geometry-templated corrections (Fig. 3j; see Methods and Supplementary Methods 6.2 for details).



**Figure 4 | Systematic quality analysis of 5 nm grid super-resolution image.** **a**, Design schematics of a  $4 \times 6$  triangular grid structure with  $\sim 5$  nm point-to-point spacing on a DNA origami nanostructure. Each green dot indicates a docking strand. **b**, Allowable imaging resolution assayed by two methods before drift correction: single-molecule fitting uncertainty and DAFL, both measured in FWHM. **c**, Comparison of the DNA-PAINT images of a 5 nm grid structure and a 20 nm grid drift marker (blue, inset) at different stages of drift correction. **d**, Target localization spread (TLS), where the point cloud shows the overlapped localizations from individually separable targets aligned by the centre of mass. Histograms are shown for the horizontal (left) and vertical (top) projections. The red curves indicate Gaussian fits. **e**, Fourier ring correlation (FRC). The correlation curves (blue solid lines) and noise-based cutoff (red dotted lines) are shown for 10 representative images; black dots indicate crossing points. **f**, Comparison of the measured imaging resolution at different stages of drift correction assayed by TLS and FRC. The red dashed line indicates the localization precision-limited best allowable resolution (as determined by DAFL). The DNA-PAINT imaging conditions used for this experiment were 400 ms frame time, 40,000 total frames and 1 nM imager strand concentration. See Methods, Supplementary Figs 17–21 and Supplementary Methods 3 and 5 for more details on the assay methods and results. Scale bars, 10 nm in the images in **c**, 20 nm in the insets in **c**; and 2 nm in **d**.

Note that during templated and geometry-templated drift correction processes the global average of all of the offset vectors calculated from many origami grids across the entire field of view (rather than those from a single origami grid) was used for drift correction. In addition, although here we treat the 20 nm grid as both drift markers and imaging samples, in general applications (such as the 5 nm grid image in the next section) no previous knowledge of the imaging sample is required—only that of the origami grid markers is used.

### 5 nm grid DMI and analysis

To demonstrate the imaging capability of DMI, we designed a triangular grid structure with  $\sim 5$  nm point-to-point spacing (Fig. 4a). This is the densest clustering pattern possible on our origami breadboard<sup>37</sup> and it also mimics the monomer spacing and arrangement in a microtubule segment (with a  $\sim 5 \times 4$  nm monomer size)<sup>39</sup>.

Compared with the 10 nm grid images (Fig. 2b–d), the higher target density in this sample imposes more stringent imaging conditions—namely, an even higher photon count, a larger number of blinking events per target and a lower blinking duty cycle. We used a short (7 nt) docking strand to accommodate the high target density and to avoid potential spatial cross-talk between neighbouring targets, then carefully adjusted the imaging conditions to satisfy all three blinking requirements (Table 1) with a high single-molecule localization precision (1.6 nm by DAFL,  $<1.0$  nm from single-molecule fitting, 1.5 nm by theoretical estimate<sup>15</sup>,

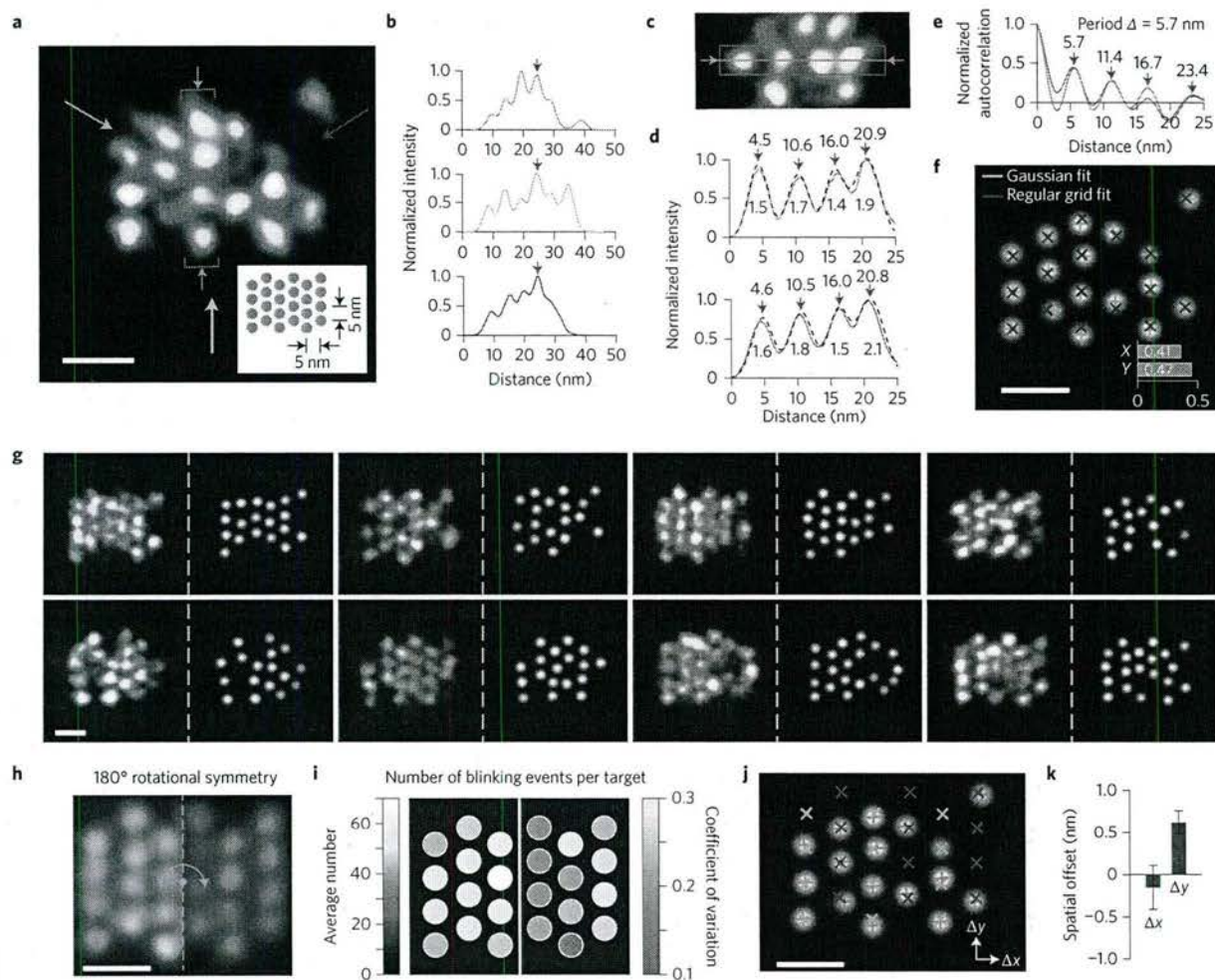
**Table 1 | Critical imaging quality parameters for the three blinking requirements.**

	Parameter	Value
①	Photon count ( $\times 1,000$ )	$50 \pm 14$
	Localization precision (nm)	1.6 (0.9)
②	Blinking events per target	$77 \pm 15$
	Target signal-to-noise ratio	$2.3 \pm 1.0$
③	Blinking duty cycle per target (%)	0.3
	Fraction of false localizations (%)	8

The localization precision value in brackets was measured by the single-molecule fitting uncertainty.

Fig. 4b). See Supplementary Figs 17–21 for details on the imaging quality characterization.

We used the 20 nm grid structures as drift markers. After applying each step of the drift correction procedure (Fig. 3i), both the 20 nm grid drift markers (Fig. 4c, insets) and 5 nm grid samples (Fig. 4c) became sharper and more regular. We employed two methods for assaying the integral imaging quality: target localization spread (or TLS, carried out by overlaying the localization clouds from well-separated targets and measuring the spread of the overlaid cloud with 2D Gaussian fitting; see Fig. 4d) and Fourier ring correlation (FRC, carried out by computing the correlation between the 2D Fourier transform spectra of independent half images, Fig. 4e). A comparison of the maximal allowable resolution

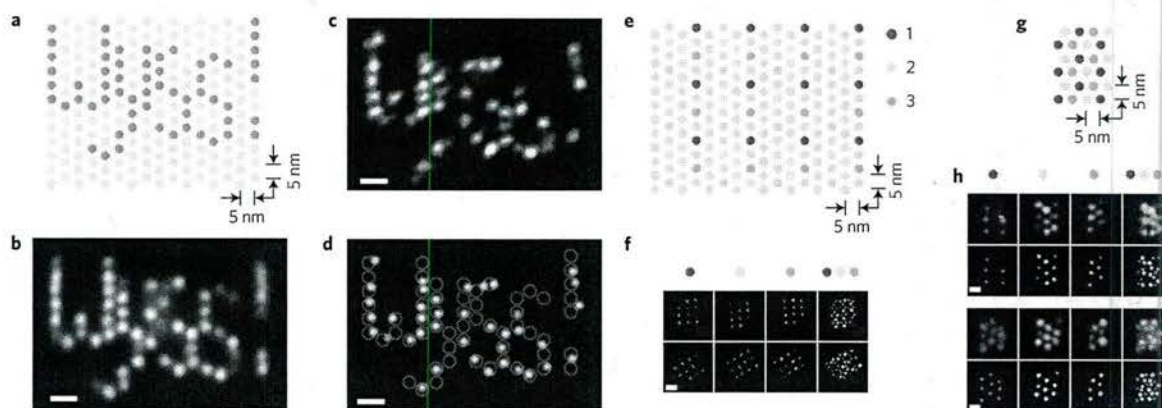


**Figure 5 | Discrete molecular imaging of 5 nm grid structure.** **a**, Representative DMI image of a 5 nm triangular grid structure obtained with DNA-PAINT. Inset: Design schematics, where each green dot indicates a docking strand. Arrows indicate the projection directions and areas of study for **b–e**. **b**, Intensity projection profiles from the image in **a** along the directions indicated by the colour-matched arrows. The profiles are aligned by the central peaks indicated by red arrows. **c**, Cropped-out image from **a** showing the central region (grey rectangle) and central pixel line (the magenta line and arrows) used for the analysis in **d** and **e**. **d**, Intensity profile along the central line (magenta) and projection from the central region (grey), as indicated by the colour-matched regions in **c**. The four-peak Gaussian fit is shown by the dashed black lines. Numbers indicate the fitted centre positions (top) and the standard deviation values (bottom) for each peak, with an average of 1.7 nm, supporting a 4.0 nm FWHM resolution. **e**, Auto-correlation analysis from the colour-matched profiles in **d**, showing a consistent periodicity of 5.7 nm. **f**, Automatic multitarget fit of the 5 nm grid image in **a**. The crosses indicate Gaussian-fitted centres (green) and regular grid-fitted centres using the green crosses as targets (blue). Inset: r.m.s. deviation between the green and blue crosses (<0.5 nm in 1D and <0.7 nm in 2D). **g**, Representative images of the 5 nm grid structures, showing structural regularity and heterogeneity. For each structure the left panel shows the super-resolution rendered image and the right panel shows the automatic fitted image. **h**, Single-particle class average of the 5 nm grid ( $n = 25$ ). The green dashed line and arrow indicate the axis and operation of symmetry of the structure. **i**, Uniformity of the blinking kinetics as represented on a 5 nm degenerate grid. The colour maps show averages (left) and coefficients of variation (right) of the numbers of blinking events for each distinguishable target. **j**, Automatic multitarget fit (grey) and two-component grid fit of the 5 nm image in **a**, allowing an offset between two groups of targets with opposite staple strand orientations, coloured in green and blue respectively. **k**, Offsets between the two groups of staples in **j** measured from single-molecule images. The error bars indicate standard deviation ( $n = 10$ ). It is important to note that no previous knowledge of the sample structure (the 5 nm grid) was used to produce the above results. The DNA-PAINT imaging conditions used for this experiment were 400 ms frame time, 40,000 total frames and 1 nM imager strand concentration. See Supplementary Figs 17, 22, 23 and Supplementary Methods 5 and 6 for super-resolution images and analysis details, and Supplementary Note 9 for discussions. Scale bars, 10 nm in all panels.

is shown in Fig. 4f—4.3 nm (by TLS) and 3.7 nm (by FRC) after geometry-templated drift correction, which allowed us to estimate the residual drift to be <1 nm (r.m.s.). See Methods for details on image quality characterization.

Plotting the projection profiles along the three symmetry axes of the triangular grid structure further confirmed the DMI imaging quality and the structural regularity of the nanostructures on a

representative single-molecule image of a 5 nm grid (Fig. 5a,b). Variations in the peak intensities resulted from missing grid points, which we note again were probably due to nanostructure synthesis defects (see Supplementary Note 9.1 for more details). A projection histogram of the four targets in a central column and the auto-correlation analysis showed an averaged 1.7 nm target standard deviation (equivalent to 4.0 nm resolution in FWHM; Fig. 5c,d) and



**Figure 6 | Discrete molecular imaging with complex patterns and multiplexed visualization. a–d**, DMI of a five-character pattern ('Wyss!') on a DNA origami nanodisplay board with 5 nm pixel size. **a**, Design schematics. Each dot indicates a staple strand. Green dots were extended with DNA-PAINT docking strands. **b**, Single-particle class average of the Wyss! pattern ( $n = 85$ ). **c**, A representative single-molecule image of the Wyss! pattern under DMI. **d**, Overlay of the design schematics on top of the automatically fitted single-molecule image in **c**. **e–h**, Three-colour multiplexed DMI, each colour indicates a separate imaging channel with an orthogonal DNA-PAINT sequence. **e**, Design schematic of a three-colour dual-purpose drift and alignment marker. **f**, Cross-channel alignment. Three single-channel images (left three columns) and one composite image (right column) are shown for two example alignment markers. **g**, Design schematics of a three-colour 5 nm grid structure. **h**, Representative multiplexed DMI image of three-colour 5 nm grid pattern in **g**. DNA-PAINT super-resolution images (top row) and automatically fitted images (bottom row) are shown for all three single-colour channels (left three columns) and the combined image (right column) for two representative 5 nm grid structures. The DNA-PAINT imaging conditions used in these experiments are as follows. 'Wyss!' letter pattern image: 500 ms frame time, 100,000 total frames and 0.4 nM imager strand concentration. Multicolour pattern image: 400 ms frame time, 2–3 nM imager strand concentration and 20,000 total frames for each colour channel. See Supplementary Figs 24 and 25 for more super-resolution images, Methods and Supplementary Methods 5 and 6 for the image analysis methods. Scale bars, 10 nm (**b,d,h**); 20 nm (**f**).

5.7 nm intertarget spacing, consistent with the design (Fig. 5e). We next performed automatic target detection and 2D Gaussian fitting to each target, followed by regular grid fitting to the Gaussian-fitted centres (Fig. 5f, see Methods and Supplementary Methods 6.2 for details). The average deviation between the Gaussian-fitted and grid-fitted centres was measured to be below 1 nm (with a single-target localization precision of 0.18 nm; Fig. 5f).

Each imaging session produced ~50–100 single-molecule images of well-resolved 5 nm grids (Fig. 5g). We next performed single-particle averaging on them and studied the single-molecule imaging heterogeneity among them on the degenerate (half) grid (due to the 180° rotational symmetry; see Fig. 5h, Supplementary Fig. 16 and Supplementary Note 9.2). Specifically, we computed the coefficient of variation in the number of blinking events across different targets in the degenerate grid (0.07) and across different single-molecule images at the same target (0.20). These results demonstrated a relatively uniform imaging efficiency (Fig. 5i, see Supplementary Fig. 22 and Supplementary Methods 6.2 for details), which further supports the hypothesis that the missing grid points probably resulted from strand synthesis or incorporation defects rather than non-uniform imaging efficiency (see Supplementary Note 9.1 for details).

We analysed the origami staple wiring pattern on single-molecule DMI images with ångström-level precision (<2 Å, Fig. 5j,k). By performing an automatic two-component grid fitting algorithm on single-molecule images, we determined the structural offset between two groups of staples with opposing orientations ( $0.6 \pm 0.1$  nm), which is consistent with the prediction from structural analysis (0.6 nm, see Supplementary Fig. 23 for details).

### DMI in complex and multiplexed samples

We also demonstrated DMI in a complex arrangement by constructing and visualizing a custom-designed letter pattern ('Wyss!') on the  $60 \times 85$  nm origami nanodisplay breadboard with 5 nm display pixel size (Fig. 6a and Supplementary Fig. 24). The single-particle

class average showed an average resolution of 4.6 nm in FWHM, allowing the visualization of individual targets (Fig. 6b). A representative single-molecule super-resolution image (Fig. 6c) displays individually distinguishable targets arranged in the designed pattern (Fig. 6d).

Multiplexed DMI further requires accurate registration between multiple imaging channels. For example, to achieve 5 nm imaging resolution, <1 nm registration accuracy is required. Traditional multitarget imaging and colocalization studies on molecular scale rely on registration between multiple spectral channels and either require a specialized and complicated hardware set-up and calibration process, and hence could be practically difficult to extend beyond two-colour registration or to the entire field of view, or achieve a suboptimal registration accuracy that does not allow molecular resolution imaging<sup>21,40</sup>.

We combined DMI with Exchange-PAINT<sup>10</sup> and demonstrated highly accurate (<1 nm) three-colour registration, in addition to highly accurate drift correction (<1 nm r.m.s.) within each channel. The method uses only one fluorophore and one optical path, thus greatly simplifies cross-channel registration and naturally extends to the entire field of view. To implement this method, we first designed a three-colour nanogrid dual-purpose drift and alignment marker, where each colour comprises a regular square grid with 20 nm lattice spacing, similar to the 20 nm grid used above (Fig. 6e). We designed three orthogonal imager sequences with optimized binding on-time, and labelled the substructure in each colour with a unique sequence. We performed multiplexed DMI imaging through serial buffer exchange (Exchange-PAINT), and then performed drift correction using these dual-purpose markers for each channel individually, followed by cross-channel alignment between each pair of different colours (Fig. 6f, see Methods, Supplementary Fig. 25 and Supplementary Methods 6.2 for details). The high-precision visualization of individual targets (down to the ångström level, as shown above) allows highly accurate alignment across all channels (<1 nm), enabling DMI imaging (Fig. 6f).

We then imaged a three-colour mixture structure of the 5 nm grid with multiplexed DMI with an average DAFL localization precision of 2.0 nm and a TLS resolution of 5.9 nm in FWHM (Fig. 6g,h, Supplementary Fig. 25). Automatic target fitting to single-molecule structures revealed the regular grid pattern as designed.

### Conclusion

DMI differs from current super-resolution fluorescence microscopy techniques in a similar manner to digital versus analogue signal processing. Whereas current super-resolution demonstrations typically focus on depicting the continuous morphology of macromolecular structures and the biomolecule spatial distributions (analogue)<sup>22,26,41–43</sup> with the resolution limited by the sampling density (the Nyquist sampling criterion), DMI imaging aims at discrete visualization of each individual molecular component within these structures. Combined with the high multiplexing power of Exchange-PAINT, DMI could potentially enable the determination of the position and identity of each molecular component in a complex biological or synthetic nanoscale system, and thus offers a complementary method to electron microscopy and crystallography with single-molecule sensitivity. Furthermore, the ångström-level precision (<2 Å) structural study of DNA nanostructures suggests that DMI could be applicable to studying sub-molecular and even potentially atomic-level features (for example, the macromolecular composition and component positions, geometric distortions and conformational changes).

We see two challenges in the further development of DMI. The first comes from the physical trade-off between spatial and temporal resolution<sup>13,26</sup>. Specifically, achieving higher spatial resolutions requires longer blinking on-times, a larger number of blinking events and a lower blinking on-off duty cycle—all of which necessarily increase the total image acquisition time. Potential ways to shorten the imaging time include engineering brighter fluorophores to shorten the blinking on-time required to collect enough photons. The second challenge is imposed by the imperfect labelling of biomolecular targets. Although our method is not limited by the Nyquist sampling criterion (the number of probes labelled per unit space), it is still limited by the absolute labelling efficiency (the average number of probes labelled per molecular target). Conventional immunostaining methods with IgG antibodies both introduce a large offset from the target epitope to the chemically conjugated probe (~10 nm) and often have limited labelling efficiencies (due to the large size and sometimes the low affinity of the antibodies). Several approaches could potentially address these issues: genetically labelled tags (for example SNAP-tags, unnatural amino acids) can provide a smaller probe size and higher labelling efficiency; aptamers, small-molecule labels, single-chain antibody fragments and camelid single-chain antibodies (nanobodies) could also provide effective alternatives for smaller affinity probes.

DMI allows the direct visualization of each individual component in densely packed biomolecular environment. In the future, we expect the development of DMI to eventually allow quantitative molecular features to be studied in diverse biological systems such as the molecular composition and architecture of diverse cellular systems (for example, cell membrane receptor clusters, neuronal synapses), the molecular states of individual protein components within macromolecular context (for example, the binding and rotational states of ring ATPase complexes, patterns of histone modifications) and the 3D architecture of chromosomes with high spatial and genomic resolution.

### Methods

Methods and any associated references are available in the online version of the paper.

Received 2 December 2015; accepted 9 May 2016; published 4 July 2016

### References

- Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).
- Klar, T. A. & Hell, S. W. Subdiffraction resolution in far-field fluorescence microscopy. *Opt. Lett.* **24**, 954–956 (1999).
- Gustafsson, M. G. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* **198**, 82–87 (2000).
- Gustafsson, M. G. Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. *Proc. Natl Acad. Sci. USA* **102**, 13081–13086 (2005).
- Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
- Hess, S. T., Girirajan, T. P. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258–4272 (2006).
- Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**, 793–795 (2006).
- Heilemann, M. *et al.* Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. *Angew. Chem. Int. Ed.* **47**, 6172–6176 (2008).
- Sharonov, A. & Hochstrasser, R. M. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proc. Natl Acad. Sci. USA* **103**, 18911–18916 (2006).
- Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11**, 313–318 (2014).
- Hell, S. W. Microscopy and its focal switch. *Nat. Methods* **6**, 24–32 (2009).
- Huang, B., Bates, M. & Zhuang, X. Super-resolution fluorescence microscopy. *Ann. Rev. Biochem.* **78**, 993–1016 (2009).
- Sahl, S. J. & Moerner, W. E. Super-resolution fluorescence imaging with single molecules. *Curr. Opin. Struct. Biol.* **23**, 778–787 (2013).
- Moerner, W. E. & Kador, L. Optical detection and spectroscopy of single molecules in a solid. *Phys. Rev. Lett.* **62**, 2535–2538 (1989).
- Russell, E. T., Daniel, R. L. & Watt, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* **82**, 2775–2783 (2002).
- Löschberger, A. *et al.* Super-resolution imaging visualizes the eightfold symmetry of gp210 proteins around the nuclear pore complex and resolves the central channel with nanometer resolution. *J. Cell Sci.* **125**, 570–575 (2012).
- Szymborska, A. *et al.* Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* **341**, 655–658 (2013).
- Yildiz, A. *et al.* Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science* **300**, 2061–2065 (2003).
- Rittweger, E., Han, K., Irvine, S. E., Eggeling, C. & Hell, S. W. STED microscopy reveals crystal colour centres with nanometric resolution. *Nature Photon.* **3**, 144–147 (2009).
- Shtengel, G. *et al.* Interferometric fluorescent super-resolution microscopy resolves 3D cellular ultrastructure. *Proc. Natl Acad. Sci. USA* **106**, 3125–3130 (2009).
- Pertsinidis, A., Zhang, Y. & Chu, S. Subnanometre single-molecule localization, registration and distance measurements. *Nature* **466**, 647–651 (2010).
- Vaughan, J. C., Jia, S. & Zhuang, X. Ultrabright photoactivatable fluorophores created by reductive caging. *Nat. Methods* **9**, 1181–1184 (2012).
- Raab, M., Schmiech, J. J., Jusuk, I., Forthmann, C. & Tinnefeld, P. Fluorescence microscopy with 6 nm resolution on DNA origami. *Chemphyschem* **15**, 2431–2435 (2014).
- Kaplan, C. *et al.* Absolute arrangement of subunits in cytoskeletal septin filaments in cells measured by fluorescence microscopy. *Nano Lett.* **15**, 3859–3864 (2015).
- Fernández-Suárez, M., Chen, T. S. & Ting, A. Y. Protein-protein interaction detection in vitro and in cells by proximity biotinylation. *J. Am. Chem. Soc.* **130**, 9251–9253 (2008).
- Shroff, H., Galbraith, C. G., Galbraith, J. A. & Betzig, E. Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. *Nat. Methods* **5**, 417–423 (2008).
- Dempsey, G. T., Vaughan, J. C., Chen, K. H., Bates, M. & Zhuang, X. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat. Methods* **8**, 1027–1036 (2011).
- Xu, K., Babcock, H. P. & Zhuang, X. Dual-objective STORM reveals three-dimensional filament organization in the actin cytoskeleton. *Nat. Methods* **9**, 185–188 (2012).
- Giannone, G. *et al.* Dynamic superresolution imaging of endogenous proteins on living cells at ultra-high density. *Biophys. J.* **99**, 1303–1310 (2010).
- Schoen, L., Ries, J., Klotzsch, E., Ewers, H. & Vogel, V. Binding-activated localization microscopy of DNA structures. *Nano Lett.* **11**, 4008–4011 (2011).

31. 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).
32. Johnson-Buck, A. *et al.* Super-resolution fingerprinting detects chemical reactions and idiosyncrasies of single DNA pegboards. *Nano Lett.* **13**, 728–733 (2013).
33. Inuma, R. *et al.* Polyhedra self-assembled from DNA tripods and characterized with 3D DNA-PAINT. *Science* **344**, 65–69 (2014).
34. Nieuwenhuizen, R. P. *et al.* Measuring image resolution in optical nanoscopy. *Nature Methods* **10**, 557–562 (2013).
35. Cordes, T. *et al.* Resolving single-molecule assembled patterns with superresolution blink-microscopy. *Nano Lett.* **10**, 645–651 (2010).
36. Thompson, M. A., Lew, M. D. & Moerner, W. E. Extending microscopic resolution with single-molecule imaging and active control. *Annu. Rev. Biophys.* **41**, 321–342 (2012).
37. Rothmund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
38. Schmied, J. *et al.* Fluorescence and super-resolution standards based on DNA origami. *Nature Methods* **9**, 1133–1134 (2012).
39. Alushin, G. M. *et al.* High-resolution microtubule structures reveal the structural transitions in  $\alpha$ -Tubulin upon GTP hydrolysis. *Cell* **157**, 1117–1129 (2014).
40. Pertsinidis, A. *et al.* Ultrahigh-resolution imaging reveals formation of neuronal SNARE/Munc18 complexes *in situ*. *Proc. Natl Acad. Sci. USA* **110**, E2812–E2820 (2013).
41. Chmyrov, A. *et al.* Nanoscopy with more than 100,000 'doughnuts'. *Nat. Methods* **10**, 737–740 (2013).
42. Xu, K., Zhong, G. & Zhuang, X. Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons. *Science* **339**, 452–456 (2013).
43. Dong, L. *et al.* Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics. *Science* **349**, aab3500 (2015).

### Acknowledgements

We thank M. Avendaño, J. Wöhrstein, F. Schüder, G. Church, W. Shih, D. Zhang, S. Agasti, E. Winfree and P. Rothmund for helpful discussions, and C. Myhrvold, N. Liu, W. Sun, X. Chen, F. Vigneault and H. Tang for helpful comments on the manuscript. This work is supported by a National Institutes of Health (NIH) Director's New Innovator Award (1DP2OD007292), an NIH Transformative Research Award (1R01EB018659), an NIH grant (5R21HD072481), an Office of Naval Research (ONR) Young Investigator Program Award (N000141110914), ONR grants (N000141010827 and N000141310593), a National Science Foundation (NSF) Faculty Early Career Development Award (CCF1054898), an NSF grant (CCF1162459) and a Wyss Institute for Biologically Engineering Faculty Startup Fund to P.Y. M.D. acknowledges support from the HHMI International Predoctoral Fellowship. R.J. acknowledges support from the Alexander von Humboldt-Foundation through a Feodor-Lynen Fellowship.

### Author contributions

M.D. conceived of and designed the study, designed and performed the experiments, developed the software, analysed the data and wrote the manuscript. R.J. conceived of and designed the study, interpreted the data and critiqued the manuscript. P.Y. conceived of, designed and supervised the study, interpreted the data and wrote the manuscript. All authors reviewed and approved the manuscript.

### Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.Y.

### Competing financial interests

The authors have filed a patent application. P.Y. and R.J. are co-founders of Ultivue, Inc., a start-up company with interests in commercializing the reported technology.

## Methods

**Materials and buffers.** Unmodified and biotin-labelled DNA oligonucleotides were purchased from Integrated DNA Technologies. Fluorescently modified DNA oligonucleotides were purchased from Biosynthesis. Streptavidin was purchased from Invitrogen (catalog number S-888). Biotinylated bovine serum albumin (BSA-biotin, catalog number A8549), Protocatechuic acid (PCA, catalog number 37580) and 6-Hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid (Trolox, catalog number 238813) were purchased from Sigma-Aldrich. Glass slides and coverslips were purchased from VWR. The M13mp18 scaffold was purchased from New England Biolabs. Freeze 'N Squeeze columns were purchased from Bio-Rad.

The following buffers were used for sample preparation and imaging: DNA origami folding buffer (12.5 mM MgCl<sub>2</sub>, 1× TE buffer), buffer A (10 mM Tris-HCl, 100 mM NaCl, 0.1% (v/v) Tween 20, pH 8.0), buffer B (10 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM EDTA, 0.1% (v/v) Tween 20, pH 8.0) and buffer TP (1× buffer B, 10 nM PCD, 2.5 mM PCA, 1 mM Trolox).

**Fluorescence microscopy set-up.** Fluorescence imaging was carried out on an inverted Nikon Eclipse Ti microscope (Nikon Instruments) with the Perfect Focus System, applying an objective-type TIRF configuration using a Nikon TIRF illuminator with an oil-immersion objective (CFI Apo TIRF 100×, numerical aperture (NA) 1.49). Laser excitation with a 561 nm laser (200 mW, Coherent Sapphire) was passed through a clean-up filter (ZET561/10, Chroma Technology) and coupled into the microscope using a beam splitter (ZT488rdc/ZT561rdc/ZT640rdc, Chroma Technology). Fluorescence light was spectrally filtered with emission filter (ET60050m, Chroma Technology). Super-resolution movies were recorded with either an electron multiplying charge-coupled device (EMCCD, used without EM gain option) camera (iXon X3 DU-897, Andor Technologies) or a scientific complementary metal-oxide-semiconductor device (sCMOS) camera (Zyla 4.2, Andor Technologies).

**Simulation of the microscopy data set.** Simulation of the microscopy data sets was performed with custom-written MATLAB software for Figs 1, 3 and Supplementary Fig. 3, with realistic parameters determined from fluorescence microscopy experiments, including image and pixel sizes, camera conversion factor and noise level, fluorophore photon emission rate and imaging background noise. Stochastic and independent blinking kinetics was simulated for all images apart from the first two blinking requirement tests in Supplementary Fig. 3. Intensity distributions from single-molecule blinking events were generated with finite pixel approximation of Gaussian profiles and corrupted with Poisson noise and Gaussian background and readout noise. See Supplementary Methods 2 for more details.

**Imaging quality characterization for three blinking requirements.** For blinking requirement (1) the photon count was calculated by converting camera counts to photons using camera manufacture-provided conversion factor. The localization precision was characterized by two methods. DAFL was calculated for pairs of localizations that originated from the same blinking events but were separated into two adjacent frames; the distribution of all of the distances between the pairs was fitted to the theoretical distribution function and the localization precision was determined from the fit. The Gaussian fitting uncertainty reports the Cramer–Rao lower bound (CRLB) for 2D Gaussian fitting for each localization<sup>44</sup>.

For blinking requirement (2) the number of blinking events was calculated for each imaging target from the single-molecule blinking time trace by counting the number of on–off switchings within the time trace. The target SNR was calculated for each pair of neighbouring targets by either automatically or manually selecting a region of interest that enclosed both targets and computing the localization distribution along the axis connecting them; two-peak Gaussian fitting was performed and the peak-to-valley distance and residual noise were used as signal and noise.

For blinking requirement (3) the blinking duty cycle was calculated for each structure from the structure's blinking time trace by calculating the characteristic on-time and off-time respectively. These were obtained by fitting the cumulative distribution of all on- and off-times both of and between blinking events to the expected distribution functions. False localizations were determined from abnormally high photon counts by a photon count threshold ( $2\sigma$  above the mean photon count). The effective localizations for the simulations were determined by a distance cutoff between the localized position and the simulated (true) positions with a  $3\sigma$  threshold.

For each 10 nm comparison structure under each imaging condition the corresponding technical requirement was measured using the methods described above. In addition, a projection histogram from the marked region in the image was generated and fitted to a multi-peak Gaussian distribution. See Supplementary Fig. 3 and Supplementary Methods 3 for more details regarding these methods.

**DNA origami design and self-assembly.** All DNA origami nanostructures were designed with the caDNA software<sup>45</sup>, and were based on a twist-corrected variant of the rectangular structure<sup>37</sup> (see Supplementary Tables 1–5 for the sequence details). The DNA origami structures used as imaging standards with specific dimensions (20 nm square grid, 10 nm comparison standards and 5 nm triangular

grid) were designed on the basis of length measurements from AFM (Supplementary Fig. 4). Eight staple strands were biotin-modified for surface fixation. Drift marker structures used in the 20 nm comparison pattern experiments were folded with DNA-PAINT extension on all possible staple strands. Staple strands used as imaging targets were extended with DNA-PAINT docking sequences (7–10 nt in length) with one or two thymine base(s) spacers. See Supplementary Figs 4–6 for more details.

The DNA origami 20 nm square grid structures were self-assembled in a one-pot annealing reaction with 50  $\mu$ l total volume containing 10 nM scaffold strand (m13mp18), 100 nM unmodified staple strands, 120 nM biotin-modified strands and 1  $\mu$ M strands with DNA-PAINT extensions in DNA origami folding buffer. Drift markers for 20 nm grid image were self-assembled with 400 nM of all of the staple strands with DNA-PAINT extensions. The 10 nm comparison patterns, 5 nm grid and 'Wyss!' pattern structures were self-assembled with 500 nM biotin-modified staple strands and 1  $\mu$ M staple strands with DNA-PAINT extensions. The three-colour 5 nm grid structure was self-assembled with 120 nM biotin-modified staple strands and 1  $\mu$ M staple strands with DNA-PAINT extensions. For the 20 nm square grid and 10 nm comparison pattern structures the solution was annealed with a thermal ramp cooling from 90 °C to 25 °C over the course of 75 min. For the 5 nm grid and 'Wyss!' pattern structures, the solution was annealed with a thermal ramp cooling from 90 °C to 20 °C over the course of 3 h and for the three-colour 20 nm grid and 5 nm grid structures the solution was annealed with a thermal ramp cooling from 90 °C to 20 °C over the course of 72 h.

The self-assembled DNA origami structures were characterized and purified (except for the 20 nm square grid structures) by agarose gel electrophoresis (2% agarose, 0.5× TBE, 10 mM MgCl<sub>2</sub>, 0.5× SybrSafe pre-stain) at 4.5 V cm<sup>-1</sup> for 1.5 h. For purification, the gel bands were cut, crushed and filled into a Freeze 'N Squeeze column and spun for 10 min at 800g at 4 °C.

**DNA-PAINT sample preparation and imaging.** DNA-PAINT sample preparation was performed in custom-constructed flow chambers between a piece of coverslip and a glass slide or on a commercial flow chamber slide (ibidi). Sample structures were fixed on the surface via a biotin–streptavidin–biotin bridge by serially flowing BSA-biotin (1.0 mg ml<sup>-1</sup>), streptavidin (0.5 mg ml<sup>-1</sup>) and biotin-labelled samples. The sample concentration was calibrated for different structure and imager combinations to make sure that similar numbers of blinking events are detected per camera frame. The flow chamber was filled with an imaging buffer (an appropriate concentration of dye-labelled imager strands in buffer TP) and incubated for 10 min before imaging. For imaging with the custom-constructed flow chambers the flow chamber was sealed with epoxy glue before imaging. See Supplementary Methods 4 for the flow chamber protocol details and Supplementary Table 6 for the imager sequences.

Exchange-PAINT imaging for the three-colour samples was performed based on protocol adapted from our previous work<sup>10</sup>. In brief, the DNA origami sample was prepared in a flow chamber and a buffer exchange was performed by serially flowing in 400  $\mu$ l of buffer B and then 200  $\mu$ l of the next imaging buffer into the imaging chamber.

**DNA-PAINT super-resolution imaging.** DNA-PAINT super-resolution movies for the 10 nm standard patterns were captured with 5 Hz camera frame rate (200 ms per frame) for all images. The laser intensity was varied from 0.3–1.0 kW cm<sup>-2</sup> before and after meeting requirement (1). The imaging length was varied from 2,500 to 12,500 frames before and after meeting requirement (2), and to 40,000 frames after meeting requirement (3). The imager concentration was varied from 20 nM to 5 nM before and after meeting requirement (3). DNA-PAINT movies for the 'best condition' 10 nm standard patterns in Fig. 2 was captured with 4 Hz camera frame rate (250 ms per frame) for (b) and (c), 2.5 Hz frame rate (400 ms) for (d); 30,000 total imaging frames for (b) and (c), 50,000 frames for (d); 15 nM imager concentration for (b) and (c) and 5 nM for (d). The laser intensity was 1.0 kW cm<sup>-2</sup> for all images. DNA-PAINT movies for the 20 nm grid images were captured with 3.3 Hz camera frame rate (300 ms per frame), with 1.0 kW cm<sup>-2</sup> laser intensity, and 3 nM of dye-labelled imager strand for 30,000 frames. DNA-PAINT movies for the 5 nm grid images were captured with 2.5 Hz camera frame rate (400 ms per frame), with 1.0 kW cm<sup>-2</sup> laser intensity and 1 nM of dye-labelled imager strand for 40,000 frames. The DNA-PAINT movie for the 'Wyss!' pattern was captured with 2 Hz frame rate (500 ms) and 0.4 nM imager strand concentration for 100,000 frames. The DNA-PAINT movie for the three-colour 5 nm grid was captured with 2.5 Hz frame rate (400 ms) and 2–3 nM imager strand concentrations with 20,000 frames per colour.

Extra drift markers were supplemented for some images. For the 20 nm grid structure, extra drift markers with DNA-PAINT docking extensions on all strands were used. For the 10 nm two-target structure, 5 nm grid and 'Wyss!' pattern images, extra drift markers of the 20 nm grid structures were used. For the three-colour 5 nm grid images, further dual-purpose drift and alignment markers of the three-colour 20 nm grid structures were used.

**Super-resolution data processing and image analysis.** DNA-PAINT super-resolution movies were processed with custom-written MATLAB software. In general, images were processed and analysed in three steps: (i) spot detection and

localization, (ii) drift correction, (iii) super-resolution rendering and quality analysis. Spot detection and localization was performed with efficient and accurate Gaussian fitting algorithms as reported in ref. 44. For the 10 nm comparison pattern images, drift correction was performed with simple trace averaging only. For the 20 nm grid, 5 nm grid, Wyss! pattern and three-colour 5 nm grid images, drift correction was performed with simple trace averaging followed by templated and geometry-templated drift correction methods with the 20 nm grid markers, as described in the sections below. For the three-colour 5 nm grid images, alignment across different channels was performed following drift correction procedures with the same three-colour 20 nm grid dual-purpose markers described in sections below. The imaging quality was characterized separately for each of the three DMI requirements and for the final super-rendered image by a variety of methods, as described below. Final super-resolution images were rendered with Gaussian blurring with the standard deviation set to the estimated localization precision. A simplified version of the software processing suite can be obtained at <http://molecular.systems/software> or <http://www.dmi-imaging.net>. See Supplementary Methods 5 and 6 for more details.

**DNA nanostructure-based drift correction.** Drift compensation based on the DNA origami marker structures was performed using the following steps. First we selected (either automatically or manually) a pool of isolated structures as drift markers (either the 20 nm grid markers, all-modified drift markers or the samples themselves) and took a simple trace average of their blinking time traces as the drift correction trace. This was the only drift correction method used for the 10 nm comparison patterns images. For the 20 nm grid, 5 nm grid, Wyss! pattern and the three-colour 5 nm grid samples, templated drift correction and geometry-templated drift correction methods were then performed with the aid of 20 nm origami grid markers in a frame-by-frame manner. For each frame, the algorithm identified all of the localizations and assigned an 'offset vector' to each, to be used for averaging. A photon-weighted global average of all of the calculated offset vectors was used for drift correction of the current frame. In the templated drift correction method the offset vector was determined from the localization position in relation to the target site position, which was determined by a local 2D Gaussian fitting of all of the localizations originated from that target. In geometry-templated drift correction method the offset vector was determined from the localization position relative to the regular 20 nm grid-fitted target position instead of the Gaussian-fitted target position. The remaining drift in the corrected images was estimated by comparing the maximal allowed imaging resolution (measured by the DAFL localization precision) and the measured imaging resolution (either by Gaussian fitting of the single-particle averaged image or by the TLS method, see below) using the quadratic sum principle. It is important to note that in producing the 5 nm grid, the 'Wyss!' pattern and the three-colour 5 nm grid images, no previous information about the sample structures was used for the templated and geometry-templated drift correction procedures. See Supplementary Methods 5.2 for more details.

**Imaging quality characterization for super-rendered images.** The imaging resolution for super-rendered images was characterized using a few methods. Target localization spread (TLS) was calculated by overlaying the localization cloud for all of the separable targets on top of each other, aligned by the centre of mass position, and measuring the standard deviation of the overlaid cloud of localizations; the FWHM

value was reported as the resolution estimate. Fourier ring correlation (FRC) was calculated by splitting the image into two independent half images by cutting the super-resolution movie stack into sections of ~100 frames and arbitrarily assigning half of the sections to each half image. The spatial frequency correlation between their respective 2D fast Fourier transform (FFT) spectra was calculated and the crossover point with the noise-based  $2\sigma$  threshold curve was reported as the supported imaging resolution. See Supplementary Methods 6.1 for more details.

**Single-particle class averaging analysis.** The single-particle averaging analysis was carried out with the EMAN2 software package (version 2.0). Images of the individual particles were automatically selected and super-rendered with a pixel size set to less than the localization precision in a custom MATLAB program and processed with a reference free class averaging functionality in EMAN2 (e2refine2d), allowing only rotational and translational transformations during alignment. A number of particles were used for the averaging ( $N = 700$  for the 20 nm square lattice,  $N = 25$  for the 5 nm grid standard and  $N = 85$  for the Wyss! pattern) and from each session the most representative class average image was selected as the final result. See Supplementary Methods 6.2 for more details.

**Automatic fitting, regular grid fitting and cross-channel alignment.** Automatic fitting was performed on the 20 nm grid (single-particle class average), the 5 nm grid (non-averaged single-molecule), the 'Wyss!' pattern (non-averaged single-molecule) and the three-colour 5 nm grid (non-averaged single-molecule) images. Spot detection was performed after a Gaussian filter was used to suppress the background variation. Automatic 2D Gaussian fitting was performed for each detected centre with a fixed standard deviation determined by the overall image resolution. The fitted image was rendered with the fitted positions and intensity values.

Regular grid fitting analysis was carried out for the 20 nm grid (single-particle class average) and the 5 nm grid (both non-averaged super-resolution images and single-particle class average) with an automatic algorithm based on the individually fitted centres (above). The grid geometry (a square lattice for 20 nm grid and a triangular lattice for 5 nm grid) and number of grid points were input manually and an initial estimate of the grid boundaries was also manually set to reduce the fitting time. The best fitted grid was determined by minimizing the r.m.s. deviation of the fitted points.

Cross-channel alignment with three-colour 20 nm grid alignment markers was performed by applying automatic fitting and regular grid fitting to all three substructures of the alignment marker and recording the offset between their fitted positions. The cross-channel registration offset was then calculated by comparing the recorded offset with the pre-designed offset across different channels. A number of (>10) high-quality alignment markers were identified and their offset computed in this way; the average from all of them was used as the final registration offset. See Supplementary Methods 6.2 for more details.

## References

- Smith, C. S., Joseph, N., Rieger, B. & Lidke, K. A. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nature Methods* **7**, 373–375 (2010).
- Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucl. Acids Res.* **37**, 5001–5006 (2009).



**Cite this article:** Grun C, Werfel J, Zhang DY, Yin P. 2015 DyNAMiC Workbench: an integrated development environment for dynamic DNA nanotechnology. *J. R. Soc. Interface* **12**: 20150580.  
<http://dx.doi.org/10.1098/rsif.2015.0580>

Received: 30 June 2015  
 Accepted: 21 August 2015

**Subject Areas:**

nanotechnology, computational biology, synthetic biology

**Keywords:**

DNA nanotechnology, molecular programming, self-assembly, software, sequence design, integrated development environment

**Authors for correspondence:**

Casey Grun  
 e-mail: [casey.grun@wyss.harvard.edu](mailto:casey.grun@wyss.harvard.edu)  
 Peng Yin  
 e-mail: [py@hms.harvard.edu](mailto:py@hms.harvard.edu)

<sup>†</sup>Present Address: Department of Bioengineering, Rice University.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2015.0580> or via <http://rsif.royalsocietypublishing.org>.

# DyNAMiC Workbench: an integrated development environment for dynamic DNA nanotechnology

Casey Grun<sup>1</sup>, Justin Werfel<sup>1</sup>, David Yu Zhang<sup>1,†</sup> and Peng Yin<sup>1,2</sup>

<sup>1</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA

<sup>2</sup>Department of Systems Biology, Harvard Medical School, Boston, MA, USA

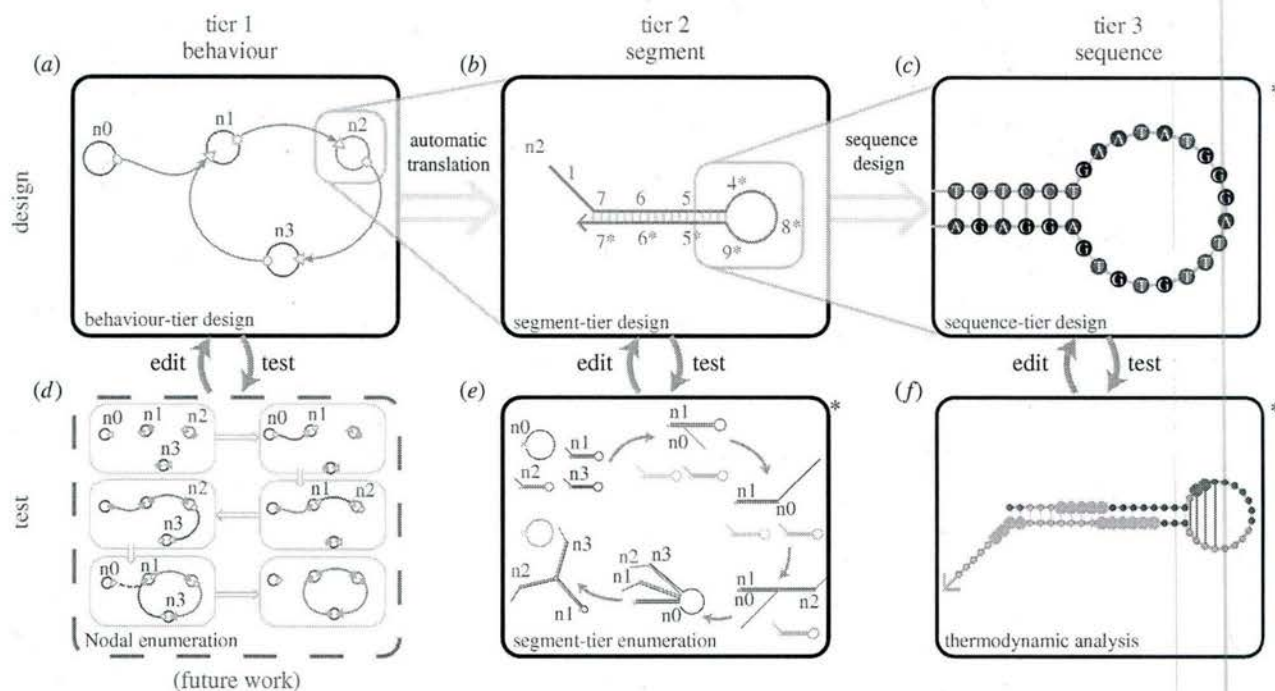
CG, 0000-0002-4563-6358

Dynamic DNA nanotechnology provides a promising avenue for implementing sophisticated assembly processes, mechanical behaviours, sensing and computation at the nanoscale. However, design of these systems is complex and error-prone, because the need to control the kinetic pathway of a system greatly increases the number of design constraints and possible failure modes for the system. Previous tools have automated some parts of the design workflow, but an integrated solution is lacking. Here, we present software implementing a three ‘tier’ design process: a high-level visual programming language is used to describe systems, a molecular compiler builds a DNA implementation and nucleotide sequences are generated and optimized. Additionally, our software includes tools for analysing and ‘debugging’ the designs *in silico*, and for importing/exporting designs to other commonly used software systems. The software we present is built on many existing pieces of software, but is integrated into a single package—accessible using a Web-based interface at <http://molecular-systems.net/workbench>. We hope that the deep integration between tools and the flexibility of this design process will lead to better experimental results, fewer experimental design iterations and the development of more complex DNA nanosystems.

## 1. Introduction

DNA has been demonstrated to be a robust and versatile substrate for engineering static nanostructures [1,2] and dynamic nanodevices [3,4]. The specificity of Watson–Crick base pairing [5,6], combined with recent improvements in thermodynamic predictive models [7,8] and rapidly decreasing costs for commercial oligonucleotide synthesis [9], have resulted in an explosion of research in DNA nanotechnology [10], in which hybridization interactions (rather than enzymatic processes) are primarily used to implement the desired molecular behaviour. Recently, the field has progressed beyond primarily demonstrations of static equilibrium structure formation to the design of dynamic systems with kinetically controlled, non-equilibrium dynamics [4]—including molecular machines [11,12], motors [13,14], walkers [15–17], amplifiers [16,18,19], self-assembly processes [16,20], logic circuits [21,22] and other sophisticated computational devices [23,24].

Design of dynamic nucleic acid systems—both at the high level of abstraction and at the low level of sequence design—requires consideration of a different set of design parameters and metrics from the design of static DNA nanostructures. Static structures are designed to adopt a single, minimum free energy (MFE) structure; dynamic systems additionally require design of a kinetic pathway—a series of reactions. Disruption of any of these intended reactions or prevalence of unintended ‘side’ reactions can disturb the intended behaviour. For instance, poor kinetics of individual reactions can lead to slow performance for a molecular calculation; unanticipated ‘leak’ reactions can cause unexpected product formation; and side reactions that produce undesired products can result in low yields of the intended product.



**Figure 1.** Tiers of abstraction and the overview of the design workflow. The design process occurs at three ‘tiers’ of decreasing abstraction: at the level of (1) behavioural description, (2) segment-level complementarity design and (3) nucleotide sequences. The design workflow involves two forms of tools: (a–c) describe tools for *creating* dynamic nucleic acid systems, while panels (d–f) describe tools for *analysing and testing* these systems *in silico*. Box (d) is future work, not yet incorporated by DyNAMiC Workbench; asterisks denote tools that use previously described software. Images in all panels except (d,f) were generated automatically by DyNAMiC Workbench. (a) Users begin by describing an abstract behaviour, using a formalism such as the Nodal language of Yin *et al.* [16] (future extensions may incorporate alternative behavioural designers). A built-in compiler automatically generates a segment-level representation (when possible) of a system of nucleic acid strands which implements the desired behaviour, through a systematic labelling of segments corresponding to the nodes and connections specified (see §3.1). (b) A segment-tier design (specifying the identity and complementarity relationships between all segments in the system) can be either generated from the behavioural designer or assembled directly by the user (see §3.2). (c) From a segment-level design, nucleotide sequences can be generated with one of a variety of sequence design packages. DyNAMiC Workbench allows users to interact with and modify systems at all three levels (see §3.3). (d,e) Reaction enumerators identify all possible reaction paths, highlighting possible undesired interactions. Enumeration, in principle, can be performed at the behavioural tier among behavioural species, or at the segment tier among molecular species. The results of both enumerations should be comparable, but the segment tier enumeration may reveal unintended side-reactions or kinetic traps not prescribed by the behavioural enumeration (see §3.4.1). (f) Base-tier sequences can be analysed using thermodynamic methods to identify unintended secondary structure in monomers [25], as well as some unintended interactions between species (see §3.4.2).

These problems are amplified in more complex systems, since the number of possible unintended interactions grows at least quadratically with the number of species. Recent demonstrations of nucleic acid logic circuits now exceed over 100 different molecular species [23,24], motivating the need for novel software to assist the design of complex systems.

To describe the design process for a dynamic DNA system, we propose a system of three different ‘tiers’ of abstraction (figure 1). The first tier describes desired system behaviour, using a set of high-level abstractions that represent molecular species and their interactions (figure 1a). The second tier gives a set of idealized DNA strands and interactions (in terms of prescribed regions of Watson–Crick complementarities) that implement the first-tier behavioural species (figure 1b). The lowest tier gives specific sequences of nucleotides to implement the prescribed complementarities among the second-tier species (figure 1c). In principle, one may begin the design process at any of the tiers and software should automatically translate to lower tiers; similarly, computer-aided verification at any tier should be possible. In practice, the first two tiers are traditionally designed by hand, and the final ‘sequence design’ step uses one of a variety of computational sequence optimization [26–30] and verification [25,31–33] packages; that is, only sequence-tier design and analysis is automated.

Recent software packages have made impressive advances in automating and integrating various parts of this design

process. Visual DSD [34,35] has been used to automate the ‘segment-level’ (tier 2) design and analysis of systems containing hundreds of distinct species [25,34]. However, Visual DSD currently does not support the design and enumeration of branched junction structures (used in many demonstrations of self-assembly [16] and molecular computation [21]), nor does it support enumeration of certain reaction types, such as four-way branch migration or branch migration with remote toeholds [36,37]. Qian & Winfree [22,23] have presented a compiler for their ‘seesaw gate’ systems, which has been used in demonstrations of sophisticated molecular computation; however, this abstraction is built around a single structural motif (the seesaw gate) and is therefore of limited use for applications in self-assembly that require a wider variety of structural features (e.g. hairpins, branched junctions, etc.). The NUPACK software package [8,25,30] integrates thermodynamic design and evaluation of nucleotide sequences; however, it does not currently provide a programming language for tier 1 (behavioural) design, nor does it allow for analysis or evaluation at the segment tier. Therefore, no current package integrates a full-featured DNA programming language with sequence designers, as well as analysis and verification tools at the segment and sequence tiers. See table 1 for a detailed comparison.

Here, we present DyNAMiC Workbench (the **D**ynamic **N**ucleic **A**cid **M**echanism **C**ompiler), which provides a tightly

**Table 1.** Comparison of currently available software packages for dynamic DNA nanotechnology. ✓, feature is implemented by this package; ×, feature is not implemented by this package; and ✓\*, feature is provided by an external package with which this package directly interfaces.

		DyNAMiC Workbench <sup>a</sup>	DSD [34] <sup>a</sup>	Seesaw Compiler [22] <sup>a</sup>	NUPACK [25] <sup>a</sup>	Mfold [38] <sup>a</sup>	Vienna RNAfold [32] <sup>a</sup>	Multistrand [39] <sup>a</sup>
implementation	behavioural	✓	✓	✓	×	×	×	×
	segment	✓	✓	✓	✓	×	×	×
	sequence	✓*	✓	✓	✓	×	✓	×
reaction enumeration	behavioural	×	×	×	×	×	×	×
	segment	✓*	✓	×	×	×	×	×
	sequence	×	×	×	×	×	×	✓
kinetic simulation	behavioural	×	×	×	×	×	×	×
	segment	×	✓	✓*	×	×	×	×
	sequence	×	×	×	×	×	×	✓
thermodynamic simulation	structure	✓*	×	×	✓	✓	✓	×
	prediction							
	partition function	✓*	×	×	✓	✓	✓	×

<sup>a</sup>Package is available as an online Web service at the time of publication.

integrated graphical user interface for all three tiers of design, as well as automated enumeration and analysis of potential interactions at the segment and sequence tiers. Users may begin by designing an abstract behaviour for a dynamic system and are assisted in translating that behaviour first to segments, then to nucleotide sequences. In this way, the design process is *hierarchical*, progressing through several layers of abstraction towards the ultimate implementation. However, the user may easily enter or exit the software at any stage of the design process, perform analysis *in silico* and make changes to the design based on the results of those tests. In this way, the design process can be *iterative*.

Furthermore, DyNAMiC Workbench supports integration with a number of commonly used software packages for each tier, allowing the user to 'mix-and-match' different aspects of preferred software. Finally, the system also includes a pluggable framework for expansion and inclusion of new tools and interfaces—for instance, a kinetic simulation package could be easily added. The software is deployed as a free Web service, available at <http://molecular-systems.net/workbench>, providing a cross-platform graphical interface without requiring installation of new software; a downloadable version is also available (at the same URL) to support local installations if preferred. We believe that the inherent flexibility in this design process, combined with the deep integration between the various tools, will help eventually enable fully automated design of dynamic systems. Expanded and improved *in silico* design and analysis will allow better experimental results with fewer experimental design iterations.

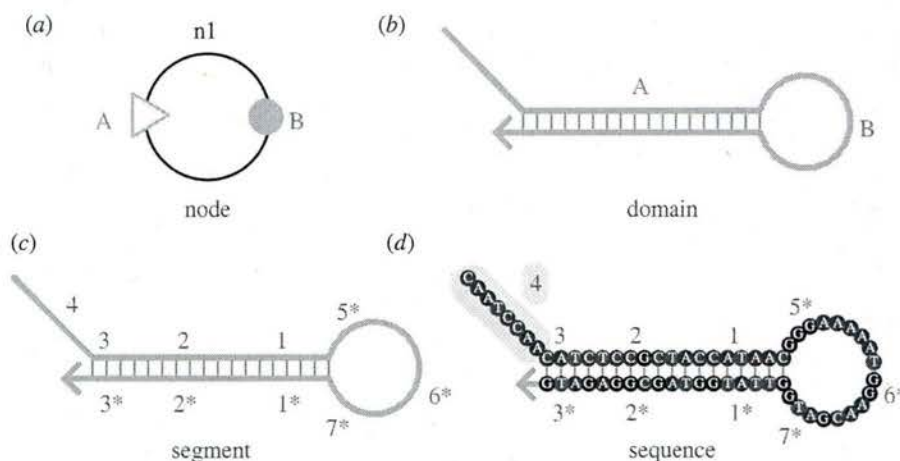
## 2. Designing a system with DyNAMiC Workbench

To illustrate the various features of DyNAMiC Workbench and the tiers of a typical design process for a dynamic nucleic acid system, we first present a sketch of the process for designing a triggered, catalytic three-arm branched junction, which has been experimentally demonstrated in [16]. We also

provide four additional design examples in the electronic supplementary material.

We adapt the 'Nodal' formalism [16] to describe the design procedure, following the three-tiered process outlined above. The terminology and abstractions of the Nodal formalism are summarized in figure 2. This formalism maps abstract behavioural units ('node types', figure 3*a*) to concrete molecular implementations ('molecule types', figure 3*b*)—DNA strands or complexes of strands. Once a node type has been defined, many 'Nodal species'—instances of the node type—may be created. The process that we wish to demonstrate—triggered catalytic three-arm junction formation—is shown in figure 3*c*; we will implement this process by instantiating node types and composing Nodal species (figure 3*d*). The Nodal language allows a complicated behavioural process to be described by composing small, reusable modules, much like functions in a programming language (tier 1). DyNAMiC Workbench can then use the underlying node type definitions to automatically generate a complete segment-level (tier 2) molecular implementation (figure 3*e*). Ultimately, sequences are designed for each of these molecular species (tier 3), and many DNA molecules are produced for each species by commercial oligonucleotide synthesis (figure 3*f*).

The connectivity of our target structure (the three-arm junction)—along with our prescription that the assembly process be *stepwise*, *triggered* and *catalytic* (see below)—suggests a particular architecture of nodes. DyNAMiC Workbench currently provides a pool of approximately 20 pre-defined node types, based on previous work [16,20,40] (electronic supplementary material, figure S4); many of these node types have tested DNA implementations. For this example, we choose hairpins for the monomers and a linear strand without secondary structure for the initiator (node types *m0* and *m1*). 'Ports' on the nodes represent 'domains' of the underlying molecular implementation (see figure 2 for terminology); the pattern of desired interactions between domains of the molecular species is encoded in the connections between ports on the nodes. Each domain in turn is



**Figure 2.** Abstractions and their definitions. (a) A strand or set of strands with a defined behaviour may be represented as a ‘node’, as described by Yin *et al.* [16]. In a node, domains are drawn as coloured ‘ports’. A triangle represents an ‘input port’, which can trigger opening of a circular ‘output port’. Output ports may in turn bind to downstream input ports. (b) Nucleic acid strands may be drawn as lines; behaviourally relevant portions of the molecule are called ‘domains’ and represented by capital letters; we have highlighted an input domain A in orange and an output domain B in blue—corresponding to the ports on the node pictured in panel (a). (c) ‘Segments’ represent contiguous regions of several nucleotides that act as discrete units of complementarity and are labelled by numbers or lowercase letters. For instance, segment 4 (highlighted in yellow in panel (d)) has the sequence ‘CAATCAA’; each domain can comprise multiple segments. (d) Bubbles represent individual nucleotides; hashes indicate base pairing, and an arrow indicates the 3’ end of a strand.

subdivided into several ‘segments’—continuous regions of bases that are designed to act as discrete units. Figure 3g shows the intended execution of the Nodal reaction graph presented in figure 3d.

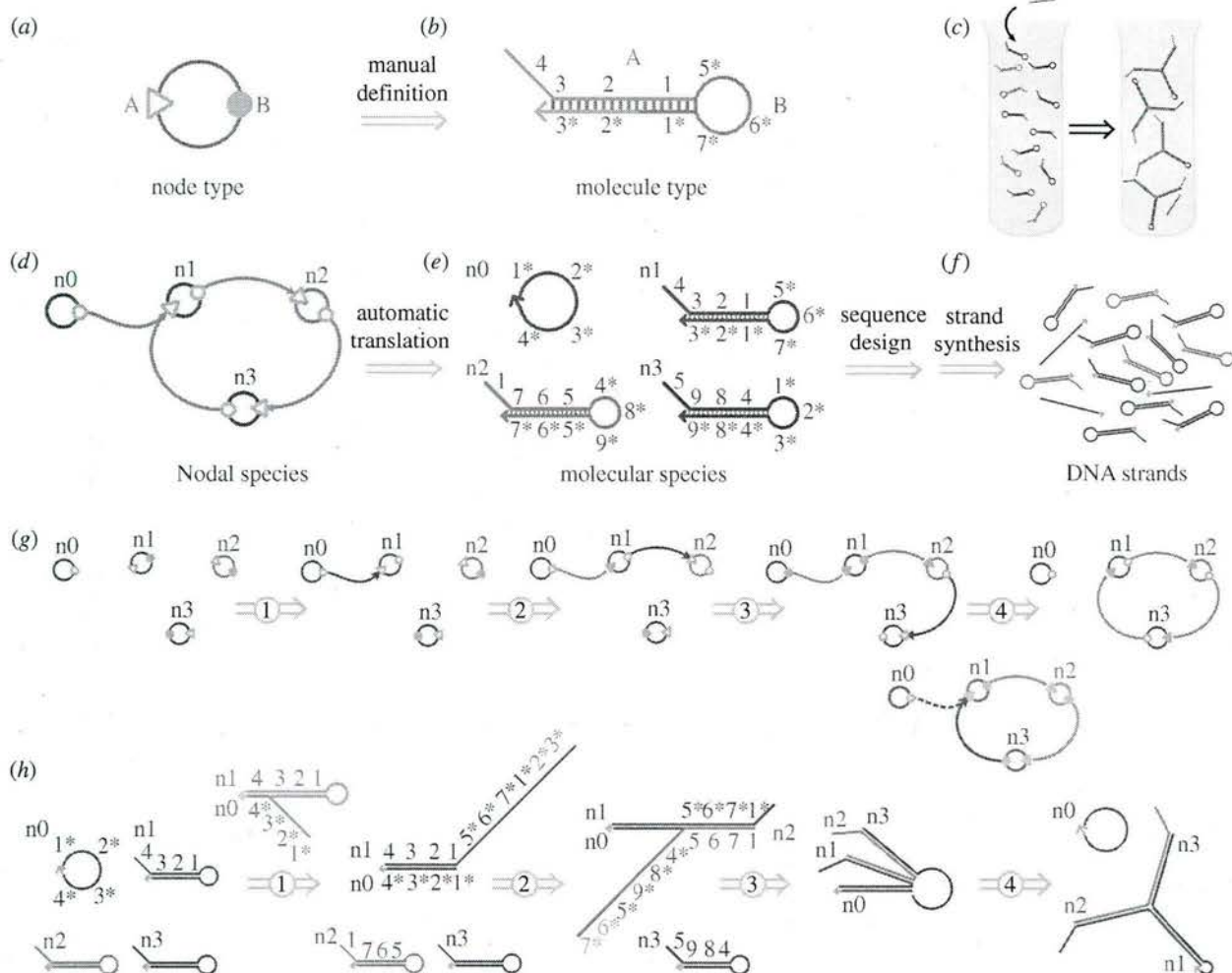
The target structure in this example, a three-arm junction (figure 3c), is designed to be formed in a kinetically guided pathway out of three monomer species and an initiator species (figure 3e). In the absence of the initiator, the monomers are to be metastable; this is indicated by drawing their output ports in a closed state (filled circles). Since the monomers are metastable, the initiator is required for target structure formation (‘triggered’). The Nodal language allows for two types of reactions: ‘assembly’ reactions occur between an open input and an open output port, while ‘disassembly’ reactions occur between an open output and a closed input. In the intended reaction pathway (figure 3g), the initiator  $n_0$  reacts with node  $n_1$  via an assembly reaction; this binding causes the output port on  $n_1$  (circle) to be switched from closed to open, reconfiguring  $n_1$  and allowing it to react with node  $n_2$  via another assembly reaction. The sequential opening of ports allows the assembly process for a single species of the target to proceed via several steps in a prescribed order (‘stepwise’). Once reacted with  $n_2$ , a reconfigured  $n_3$  can displace the initiator  $n_0$  through a disassembly reaction; this reaction frees the initiator to react with another copy of node  $n_1$ , implementing multiple turnovers (‘catalytic’).

This Nodal reaction pathway can be compiled to the more detailed reaction mechanism—shown at the segment level in figure 3h (compare each step to the Nodal mechanism in figure 3g). The initiator  $n_0$  first binds to hairpin  $n_1$  via the interaction between the ‘toehold’ segments 4 and 4\* (where  $x^*$  is the Watson–Crick complement of segment  $x$ ). After  $n_1$  and  $n_0$  are thus co-localized, the 3, 2 and 1 segments on  $n_0$  can displace their analogues—hybridizing to the 3\*, 2\* and 1\* segments—via a process known as ‘toehold-mediated branch migration’ [41–44]. At the conclusion of branch migration, monomer  $n_1$  is reconfigured so that the toehold segment 1\* is no longer sequestered in the duplex and thus becomes available for a similar downstream reaction with hairpin  $n_2$ . This process continues similarly for the remaining species.

The Nodal design can be easily constructed by the user via a drag-and-drop graphical interface. After completing the system design using the tier 1 Nodal formalism, the user invokes the DyNAMiC Workbench compiler, which automatically generates a valid set of tier 2 segment labels in order to implement the behaviour. The software can also identify potential interactions between molecules due to complementary segment interactions (‘reaction enumeration’)—including spurious reactions that may necessitate a redesign (§3.4.1). The reaction enumerator in DyNAMiC Workbench was used to generate the mechanism shown in figure 3h. The segment-tier (tier 2) descriptions capture all relevant structures and interactions of the system, assuming that all non-complementary segments have effectively orthogonal sequences (so as to not significantly interact with one another).

In reality, some degree of spurious interaction between segments is unavoidable, since real sequences are not perfectly orthogonal; careful sequence design is needed in order for the system to implement the behaviour as intended—both ensuring desired interactions are favourable and minimizing undesired side reactions that may occur between segments—both within or across different molecules [27]. For this step, DyNAMiC Workbench converts the segments generated in the previous tier into poly-N sequences of the proper lengths. The poly-N sequences are then mutated into a set of non-interacting sequences (a tier 3 design) using one of several existing software tools: Domain Design (DD) [29], NUPACK [25,30] or Multisubjective [45] (§3.3).

Finally, DyNAMiC Workbench interfaces with tools [25,32,38] to perform thermodynamic calculations as heuristic tests to assess the quality of the final sequences. For instance, we may compute the MFE structures of each of the strands or complexes in our ensemble to verify that they adopt the intended secondary structures in their monomeric form. Additionally, DyNAMiC Workbench can be instructed to look for strong pairwise interactions between species which are intended to be non-interacting. Once the user is satisfied that the sequences will implement the intended complementarity scheme, DyNAMiC Workbench can export the sequences of all relevant DNA strands; these sequences can



**Figure 3.** Design example. The Nodal formalism [16] maps (a) simple behavioural units (nodes) to (b) molecular architectures. Commonly used molecular motifs, or 'molecule types', may be expressed as node types by describing the primary and secondary structure of the molecule type and then assigning functional roles (e.g. input or output) to the molecule's domains. (c) Our target process is the conversion of a set of metastable hairpins into a three-arm junction, with a single-stranded initiator serving as a trigger. (d) Behavioural description of our catalytic three-arm branched junction, described by composing behavioural nodes. Once a node type has been defined, multiple Nodal species with the given node type may be instantiated and connected together. Each species of a given type will have the same basic structure, but will correspond to a distinct molecular species with a unique sequence identity. Connections between Nodal species (arrows) represent desired behavioural interactions between their domains. (e) These behavioural interactions also thus imply Watson–Crick complementarity relationships between the domains. DyNAMiC Workbench can use the node type definition to automatically map a set of Nodal species to a set of molecular species whose sequence complementarity relationships implement the intended behavioural interactions. This panel shows the segment-level representation of the hairpin monomer and initiator species which will make up our system. (f) Sequences may be designed to implement the molecular species, and DNA molecules may be produced by commercial oligonucleotide synthesis. (g) Intended 'execution' of the Nodal complementarity graph. (h) Segment-level enumeration of possible reactions between four starting complexes. Short, single-stranded regions at the hairpin termini serve as nucleation sites—'toeholds' to prime branch migration reactions. The opening of the hairpin by branch migration exposes new toeholds, implementing the cascade.

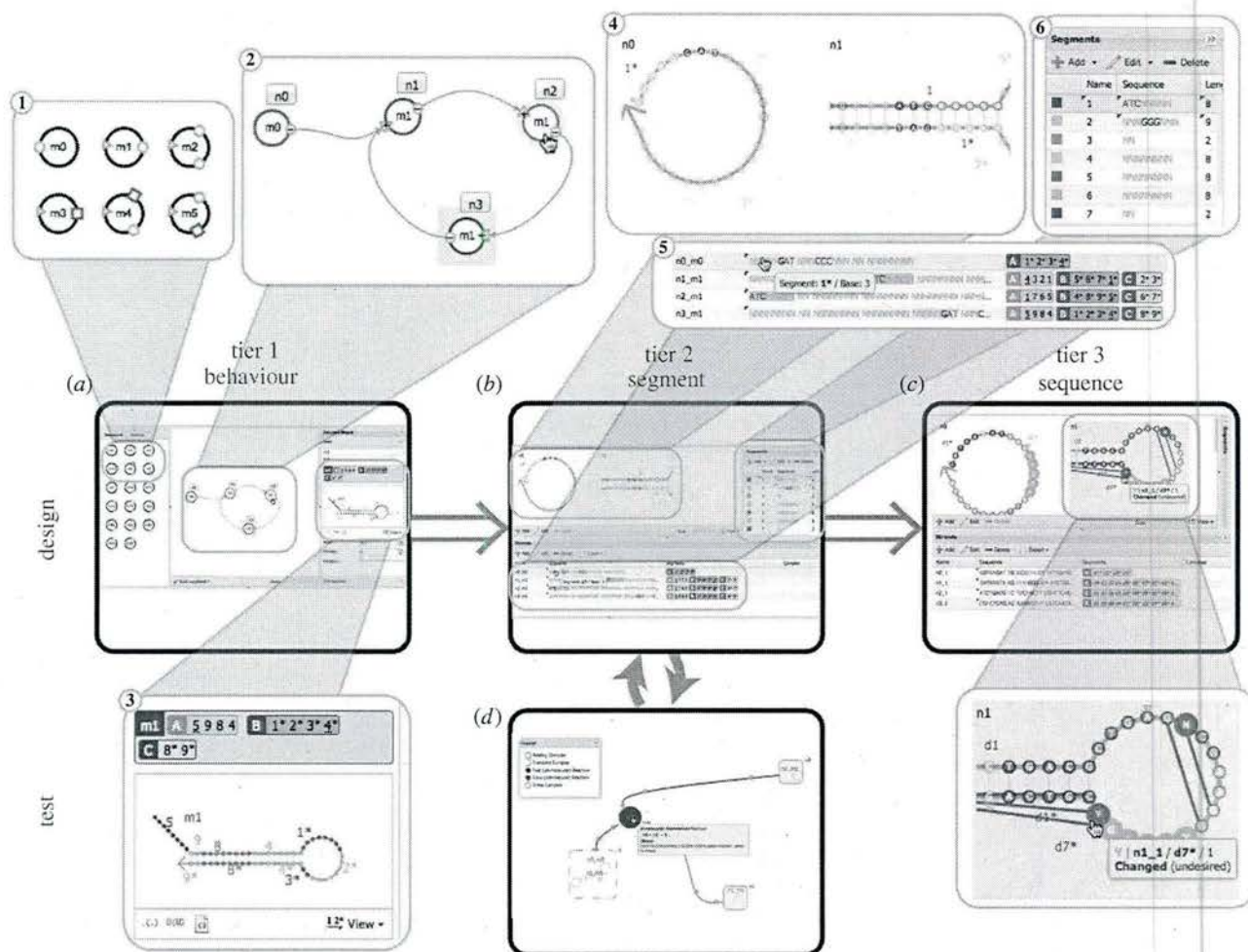
then be directly provided to a commercial supplier of oligonucleotides for synthesis.

### 3. Methods and implementation

We have developed and unified graphical tools for designing and analysing systems at the three tiers of abstraction. Figure 1 describes the various tools we have implemented and/or integrated, and several tools that could be implemented in the future, as well as the intended workflow for a user developing a system with DyNAMiC Workbench. This section describes the input to, computation performed by, and output from each of these tools in detail, as well as discussing important features of the tools and their interfaces. Figure 4 shows screenshots of various tools.

#### 3.1. Tier 1: behavioural/Nodal design

As described above, the Nodal formalism maps abstract behavioural units (node types) to known molecular implementations. The input to the Nodal compiler is a set of instances (Nodal species) of these node types, along with a prescribed connectivity between the ports of Nodal species. Input may be constructed by a 'drag-and-drop' graphical interface (figure 4a and electronic supplementary material, figure S1), or by describing the system using a text-based language called DyNAML (the *Dynamic Nucleic Acid Markup Language*) [46]. The hairpin monomers and the single-stranded initiator described above (§2) are two examples of molecule types. In the example of the previous section, there were three distinct 'species' of the hairpin monomer molecule type. Like the three distinct hairpin monomers, each corresponding Nodal species will have a unique identity, but will



**Figure 4.** DyNAMiC Workbench screenshots. (High-resolution image—view PDF for details.) (a) Nodal (tier 1) design interface. Systems are composed by dragging-and-dropping nodes from the palette on the left 1, then connecting nodes in the centre panel 2. The right panel shows a preview of the molecular implementation 3. See electronic supplementary material, figure S1, for details. (b) Segment-level (tier 2) design interface. Centre panel shows secondary structure view of each complex in the system 4. Lower panel lists sequences and composition of strands in the system 5. Right panel shows name and sequence of each distinct segment in the system 6. See electronic supplementary material, figure S2, for details. (c) Multisubjective sequence (tier 3) design interface. Similar to the tier 2 design interface, different panels show complexes, strands and segments. The secondary structure view also highlights unintended interactions and shows bases flagged for modification by the analysis. See electronic supplementary material, figure S3, for details and images of other sequence design interfaces. (d) Reaction enumerator interface. Rectangular nodes represent complexes, joined by circular nodes representing reactions between intermediates. See figure 5 for details.

share the same basic shape with the other species of the same type. In the same way, generated molecular species of the same type will have different sequences but similar dimensions and secondary structures.

The Nodal compiler uses the definition of each node type (the mapping to a known molecular implementation), as well as the network of connections between Nodal species, to produce a segment-level representation of the system—such that connections between nodes are implemented as appropriate Watson–Crick complementarity relationships between segments. To do this, the Nodal compiler first creates an instance of each molecular species (one for each Nodal species in the system), then labels the segments in the molecular species such that they satisfy the prescribed connections between nodes. The compiler begins with an initiator node, then traverses the network of connected nodes in a breadth-first search until all segments are labelled. The compiler will return an error if two complementarity statements are in conflict and cannot be fulfilled (e.g. if one connection requires a segment be labelled 7 and another connection requires that it be labelled 7\*), or if a complementarity statement exists between domains of incompatible shape (e.g. if a domain A,

comprising an 8-nt segment, is to be complementary to a domain B which contains an 8-nt segment and a 2-nt segment).

The output from the Nodal compiler is a segment-level (tier 2) representation of the system, encoded in the DyNAMiC Intermediate Language (DIL) [46]. This representation can be further edited by the user or converted to a sequence-level (tier 3) representation.

The Nodal language and compiler are highly general—rather than enforcing specific invariants on the form or structure of the underlying molecular primitives, or supporting only specific, pre-defined molecular motifs, our compiler allows the inclusion of systems with arbitrary nucleic acid secondary structures and behaviours. This flexibility is essential to supporting the previously discussed hairpin motifs, as well as more complex multi-stranded motifs (such as branched junctions)—both of which are important for structure formation applications. The language also allows arbitrary new motifs to be developed. DyNAMiC Workbench contains a collection of built-in node types with DNA implementations based on published literature (electronic supplementary material, figure S4), but also supports definition of new node types. A new node type can be defined by the following procedure:

the segment-wise primary and secondary structure of the underlying molecule type are described, the molecule's segments are grouped into domains and the domains are assigned functional roles such as 'input' and 'output'. New node types can also be defined from portions of existing systems. The ability to create 'composite node types' allows designed systems to be extended and re-used, by easily abstracting portions of an existing system into reusable components.

The software environment is built to be intuitive and easily usable, but also to provide detailed information to advanced users if necessary. A graphical interface (electronic supplementary material, figure S1a) reduces the barrier to entry for those unfamiliar with computer programming. Users construct Nodal species by dragging and dropping from a palette of pre-defined and custom node types. New node types with new molecular implementations can be defined by using a graphical interface (electronic supplementary material, figure S1b) or by selecting components of a Nodal program and exposing relevant domains on the nodes (electronic supplementary material, figure S1). For example, to define a reusable three-arm junction node type, the user could first define a three-arm junction system by composing nodes, then 'wrap' that system to define a new motif, exposing one input port. New node types can also be defined using DyNAML. The interface provides real-time feedback, generating interactive previews of the compiled molecular species and highlighting errors (as described above).

The Nodal language, compiler and design interface are described in detail in [46].

### 3.2. Tier 2: segment design

A tier 2 design comprises a description of each of the strands in a system, a division of the strands into segments, a statement of which segments should be complementary (or equal) to one another and an intended initial secondary structure for each of the strands. Essentially, a tier 2 design captures all of the desired inter-molecular and intra-molecular interactions between species in the system without assigning specific nucleotide sequences to segments. If the user has described their initial behavioural design using the Nodal formalism, DyNAMiC Workbench can automatically generate a tier 2 design as discussed above. The user may also choose to begin the design process at the segment tier and use DyNAMiC Workbench primarily for sequence design (§3.3); in this case, the user can input a tier 2 design directly using a graphical interface (figure 4b), or using a number of standard text-based input formats.

DyNAMiC Workbench uses the DIL as a common intermediate format for the tier 2 design; a graphical interface allows the user to modify the design at this tier and to transfer the design to various other tools (electronic supplementary material, figure S2). The user may change the number of nucleotides in existing segments, or impose specific constraints on the segments to be generated in tier 3. The user may inspect and modify the primary and secondary structures of the complexes in the system. The user may send the tier 2 design to *in silico* analysis tools, such as a segment-level reaction enumerator (§3.4.1). Similarly, the tier 2 design may be transferred to a sequence designer (tier 3) to generate specific nucleotide sequences that implement the design; this process is described next.

### 3.3. Tier 3: sequence design

The input to each of the 'sequence designers' is a segment-level (tier 2) representation of the system, along with any constraints or restrictions specified by the user (for instance, the user may want to hold the sequence of some particular segment fixed in order to incorporate a restriction enzyme cutting site).

Sequence design is the process of generating sequences of nucleotides to implement a particular set of complementarity and orthogonality relationships between nucleic acid strands. Several sequence design methodologies exist [27]. One approach uses calculations of the partition function [47–50] for an ensemble based on a detailed thermodynamic model of nucleic acid secondary structure [51]. This method attempts to maximize the probability that the ensemble adopts the intended secondary structure(s), while minimizing the chance that unintended interactions occur [27]. Recent work has extended this paradigm to entire test tubes of complexes [52]. It should be noted that these thermodynamic designers do not explicitly consider kinetic behaviour and that sequences with desirable thermodynamics may have undesirable kinetic properties [29]. One objective of future work is to explicitly incorporate kinetic models in sequence design.

Earlier methods focused on minimizing sub-sequence repeats or maintaining a minimum edit distance between non-complementary sequences in an ensemble—an approach known as 'sequence symmetry minimization' [26,53]. Evolutions of this approach [28,29] combine sequence symmetry minimization with some insights from thermodynamic models, as well as various heuristic methods, to produce sequences that meet additional criteria desirable for dynamic systems. These recent heuristic approaches may have favourable computational complexity compared to the thermodynamic approaches (scaling quadratically with the number of segments or strands, rather than cubically with the number of bases), but the actual computation time depends on the algorithm and the implementation. Finally, recent software combines these two approaches by passing limited subsets of a system between various designers—while independently analysing sequences according to a variety of criteria (e.g. attempting to explicitly eliminate secondary structure in key 'toehold' regions) [45].

Rather than selecting a single design methodology, our software incorporates and provides interfaces to various software tools that implement these different approaches (figure 4c and electronic supplementary material, figure S3). First, we have extended the core algorithms of Zhang's DD package, which performs heuristic, 'segment-based sequence design' [29], to add flexible stopping conditions and increase sequence diversity. Segment-based sequence design uses an approach similar to sequence symmetry minimization to design individual segments, which are then 'threaded' together to form full strands. In DD, a scoring function/hill-climbing algorithm is used to combine this technique with heuristic metrics of features believed to be important for dynamic systems. We have also developed a graphical interface to DD (electronic supplementary material, figure S3a). This interface allows the user to graphically add, remove and edit segments, as well as to thread these segments into strands. The individual segments and the full design can be visualized in real time as DD automatically tunes segments.

Additionally, we provide a direct interface to the NUPACK multi-objective thermodynamic sequence design [25] Web

service (electronic supplementary material, figure S3b). Several tools in our package generate input scripts for the NUPACK designer, which can then be readily submitted to the NUPACK Web server.

Finally, we have developed an interface to the 'Multisubjective' sequence design package [45] (electronic supplementary material, figure S3c). Multisubjective uses several sequence design packages (the NUPACK multi-objective designer and DD) to eliminate unintended secondary structure from key regions of strands in dynamic systems. This is an iterative process—sequence design is initially performed with another sequence designer (DD or NUPACK), then sequences are evaluated by Multisubjective. Multisubjective identifies unintended secondary structure, long regions of repeated nucleotides, etc. and proposes mutations to the designed sequences. The sequences, with mutations, are then re-submitted to the primary sequence designer and the process is repeated. Our interface allows the user to inspect the analysis performed by Multisubjective and edit the suggested modifications before redesigning.

We note briefly that, though our focus is on DNA, the design considerations are essentially identical for RNA (though thermodynamic designers will use different sets of parameters [51,54,55]). All three packages discussed also support the design of RNA sequences. Design and evaluation of mixed DNA/RNA sequences is outside the scope of current work.

The output of each of these sequence designers is a set of nucleotide sequences for each of the complete strands in the system. These sequences can be exported to a text file or a spreadsheet, which can be used to order sequences from a commercial oligonucleotide synthesis provider.

## 3.4. Analysis tools

### 3.4.1. Enumerator

Reaction enumeration is the process of predicting the network of possible reactions between a given set of starting complexes. For dynamic nucleic acid systems, a reaction enumeration at the level of segments (rather than individual nucleotides) can provide valuable insights into the behaviour of the system (by culling much of the detail of a sequence-level enumeration and focusing on larger interactions). Results of a segment-level enumeration could also be used to perform stochastic or analytical mass-action kinetic simulations of the reaction networks in order to predict physical behaviours of systems *in vitro*. Finally, segment-level reaction enumeration could be used for formal verification of system behaviour.

We have built an interface for performing segment-level reaction enumeration [56] and visualizing the results (figures 4d and 5). Our interface displays a graph of starting, intermediate and resting/end-state complexes, connected by the reactions between these complexes. The interface automatically colours fast versus slow reactions, transient/short-lived versus resting/long-lived complexes and initial complexes versus enumerated complexes. The graph can be readily traversed by zooming and panning, and the interface highlights relevant substrates or reactions proximal to a complex of interest (for instance, all reactions producing or consuming a given complex can be highlighted). The graph can be re-arranged by dragging and dropping to aid visualization. Each complex or reaction can also be isolated and visualized as an interactive secondary structure diagram.

The reaction enumerator is described in greater detail in ref. [56], as well as in an upcoming publication.

### 3.4.2. Sequence-level analysis

Once system-level behaviour has been designed and analysed, and nucleotide sequences have been generated, sequences can be analysed to determine experimental suitability. A variety of analyses can be performed. For small self-assembly systems, the MFE structure and base pair probabilities of the entire ensemble can be determined using a full physical model which computes the partition function for the system [49,50]. For larger dynamic systems (beyond tens of species or hundreds of bases), this exhaustive computation becomes impractical. However, the MFE structure and base pair probabilities of individual strands or small subsets of strands may still be examined to discover unintended secondary structure within strands or spurious interactions between strands.

Many packages exist which can perform these computations. Our software integrates directly with several publicly available Web services for sequence-level analysis. The NUPACK Web server can perform full partition function and pair probability calculation [25]. The Mfold and DINAmelt Web servers provide several types of analysis [31,38]. The TBI Vienna RNAfold Web server allows computation of the MFE structure and partition function [32]. These software tools provide a wide range of options and allow the use of various available thermodynamic parameter sets for both DNA and RNA. DyNAMiC Workbench allows users to easily submit any sequence displayed within the software to these Web servers for analysis through a simple, unified interface.

## 3.5. Utilities

### 3.5.1. Structure visualization

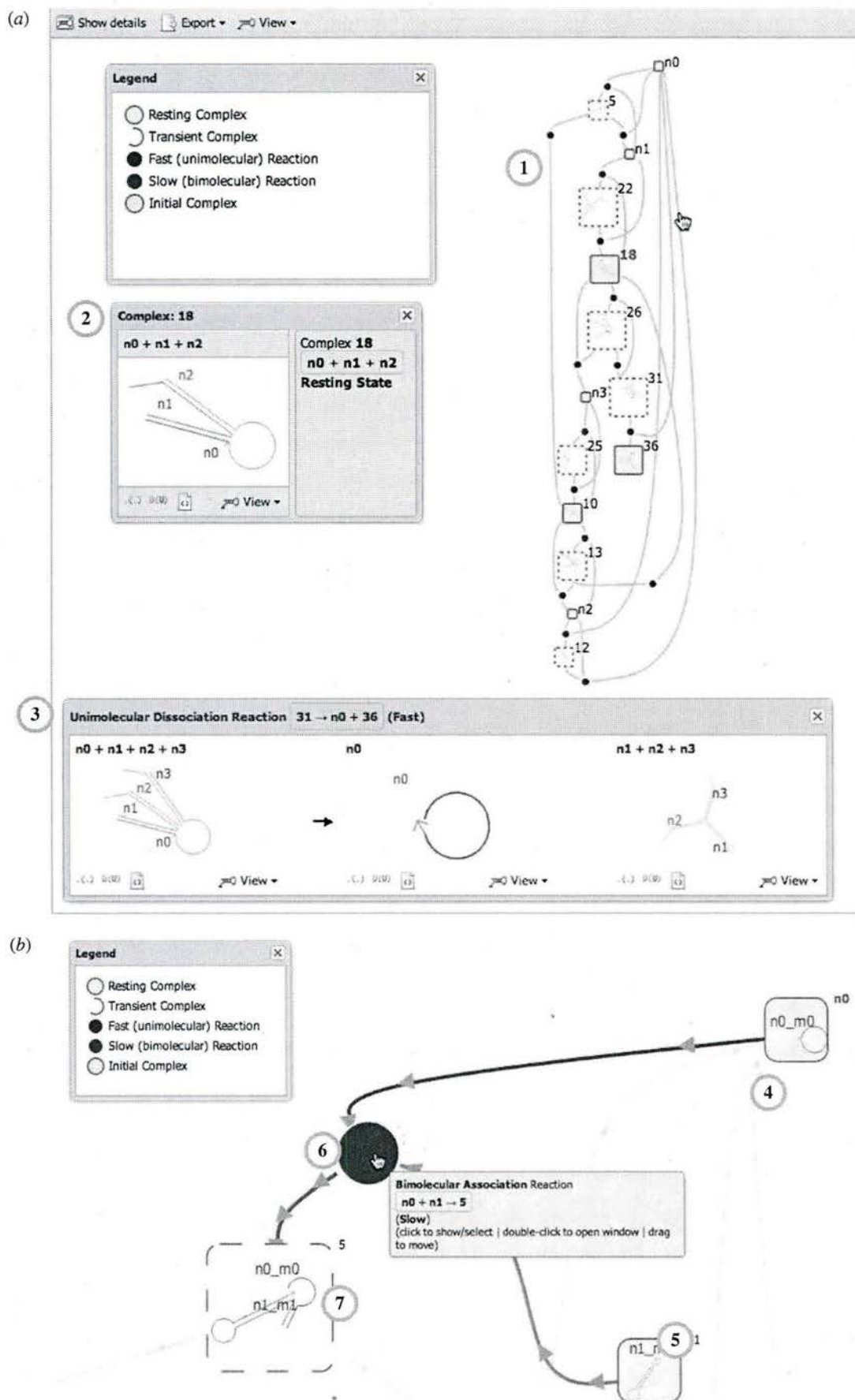
DyNAMiC Workbench includes a flexible system for visualizing arbitrary unspseudoknotted nucleic acid secondary structures, which was used to generate all images of secondary structure that appear in this paper. The basic visualization combines a traditional tree-based 'planar graph' [57,58] layout with a linear representation (allowing branched structures to be more easily visualized) and an interactive colouring scheme. This visualization is used throughout the software to display secondary structures, and can also be accessed as a separate utility.

### 3.5.2. Sequence manipulation

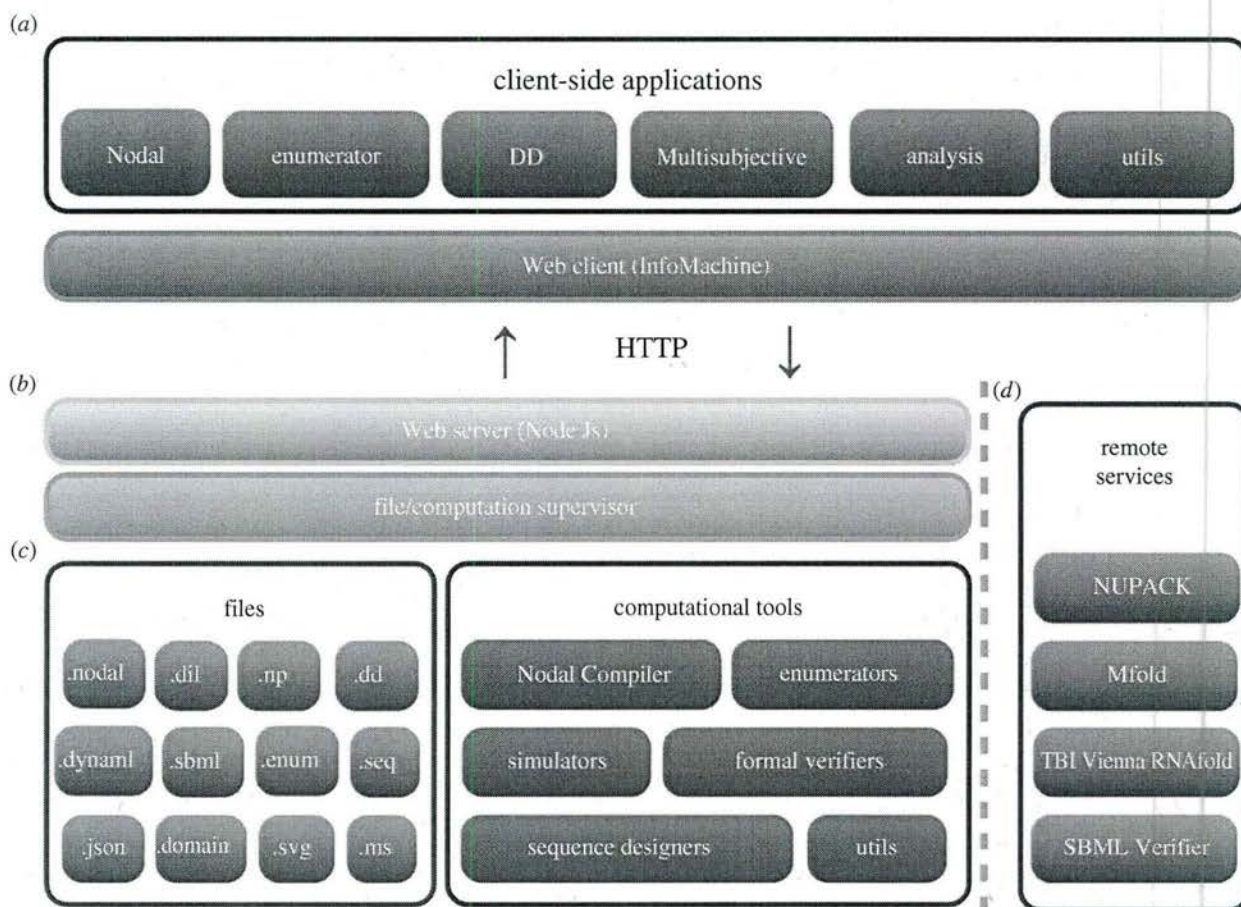
Finally, a suite of utilities is provided in DyNAMiC Workbench to perform various common transformations on nucleic acid sequence strings. These range from Watson–Crick complementation to Levenshtein distance calculation [59] to threading segments into strands. The user can also quickly perform sequence-level thermodynamic analysis (through the interfaces described above). Finally, sequences can be exported to/imported from various standard formats (FASTA, CSV, etc.) These utilities are available in context anywhere the user may interact with sequences.

## 4. Architecture

The software is deployed as a Web service—software tools are installed on a server and managed by a supervisor layer, which is tightly connected to a Web server (figure 6). The Web server exposes a rich client-side interface. The supervisor manages



**Figure 5.** Graphical interface to a reaction enumerator [56]. (a) The enumerator calculates a network of possible reactions **1** between the starting complexes, as well as the possible intermediate complexes formed. The user can pan and zoom throughout the network to view intermediate complexes. Double-clicking a complex **2** or reaction **3** allows its structure to be inspected. The reaction network shown here is the execution of the three-arm junction system—a detailed version of the schematic in figure 3*h*. (b) The user can hover over components of the graph to highlight connections to neighbouring components—in this case, arrows showing the reactants **4**, **5** of a bimolecular association reaction (red dot, **6**) are highlighted in blue and purple, while the arrow showing the product **7** is highlighted in red. Additionally, the graph layout can be re-arranged by dragging-and-dropping complexes, reactions and arrows. These interactive features allow larger reaction graphs to be explored and interpreted.



**Figure 6.** Software architecture. (a) The architecture supports a rich set of high-level, client-side applications. These applications are presented graphically to the user. (b) A Web server exposes this interface and handles requests from the client-side applications, dispatching those requests to the computational tools and the file system. (c) A wide variety of file formats are supported, and many conversions between related formats can be performed automatically. These files are then consumed by computational tools, which may run on the same machine as the Web server/supervisor, or be located remotely (e.g. run on a cluster). (d) The Web interface interacts directly with tools that are provided independently as Web services. For instance, the Web interface is capable of submitting requests directly to various sequence analysis servers [25,31,32,38].

user files, converting between different formats and launching computational tools. Examples of server-side computational tools include the Nodal compiler and the segment-level enumerator. The core Web server and supervisor layers are written in JavaScript and are executed by NodeJS—an evented, asynchronous runtime and Web server based on Google’s V8 JavaScript engine.

A benefit of the client–server architecture is that server-side tools can be written in any language, compiled or deployed for a single architecture (that of the server), and made available to clients running on many platforms through the Web interface. Various server tools are currently written in JavaScript, Python, C and C++.

The client-side interface comprises a file manager and many applications that provide graphical interfaces to the underlying server tools. Examples of client-side applications include the Nodal designer, DD and the enumerator interface. The client-side interface is written in JavaScript and uses a variety of open-source user interface toolkits. User registration allows designs and other user files to be saved on the server across multiple sessions.

## 5. Discussion

We have presented software that facilitates a design process that is hierarchical and iterative. A user designs an abstract behaviour

which can be compiled into segments and then sequences. The ability to easily enter the design process, to perform analysis, and to make edits to the design at any tier of abstraction (behavioural, segment or sequence) makes DyNAMiC Workbench an effective platform for iterative design. The goal of this flexible design process is to allow for fewer experimental design iterations, by moving *in silico* much of the testing and ‘debugging’ that would have been done by costly and time-consuming laboratory experiments. We have shown an example system designed using the software, and we present four additional designs in the electronic supplementary material—a larger, six-arm junction (electronic supplementary material, §S2.1); a bounded dendrimeric structure (electronic supplementary material, §S2.2); an exponentially-amplifying catalytic circuit (electronic supplementary material, §S2.3); and a self-assembling three-dimensional tetrahedron (electronic supplementary material, §S2.4). With these designs, we discuss in greater detail the practical challenges of implementing large systems using the software.

We have chosen to incorporate several existing tools for each stage of the design process. For instance, when designing sequences, users can easily choose one of the several tools discussed above, or can combine the results of several tools. The availability and interoperability between multiple tools, even for the same task, provides two benefits. First, different tools may be better suited to perform a given task under different

circumstances; for instance, DD may be more useful for design of large kinetic systems, while a thermodynamic designer such as NUPACK may provide a more rigorous design for smaller systems. Second, easy rapid switching between different tools will allow for better comparison of tools' performance and improved characterization of tools' strengths and weaknesses for different tasks.

Our software is deployed as a publicly available Web service. This design decision was motivated by several observations. First, installation of software (especially software which requires the user to build an executable from source code, or to use the command line) can be a significant barrier to adoption. Second, use within the field of a variety of file formats can require tedious manual juggling when working in traditional command-line or desktop interfaces. Third, many software tools in the field require significant computational resources; providing these tools as a Web service allows users to use or at least experiment with these tools without making the significant upfront investment in computational resources. The Web-based interface allows the user to access a graphical interface to the service, using any device with a modern Web browser. This makes the service truly cross-platform and also obviates the need for the user to install any software or to maintain an updated version of the software—updates to both the interface and the tools themselves can be centrally managed. For users who manage their own research computing facilities, the software is also available as a downloadable package.

The software has been designed with extensibility in mind; while we have worked to integrate many existing tools, this is only a starting point. Many excellent packages exist which have not yet been integrated. To this end, the architecture allows new server-side tools to be added and new client-side applications to be easily developed. We intend this software and the associated service to serve as a 'clearinghouse' for dynamic DNA/RNA nanotechnology software development and deployment. We have taken several steps towards this goal: first, the entire source code for the project is to be released under the GNU General Public License. We hope this will encourage users to modify and contribute to the software, in addition to encouraging usage of the public Web server. Second, we have developed and provided extensive documentation of the application programming interface (API) for the software, to facilitate development of third party extensions; this API documentation is linked to from within DyNAMiC Workbench's built-in user manual, and can be accessed from the main 'Help' page within the software.

As summarized in table 1, there are many possible features that are yet to be implemented. Reaction enumeration could be

performed at the behavioural tier, so that the intended reaction pathway can be compared directly to the expected reaction pathway at the segment tier. The behavioural- and segment-tier reaction enumerations could also be used to perform kinetic simulations and determine the time-evolution of concentrations of intermediate complexes. Electronic supplementary material, figure S5, demonstrates some of these possible extensions.

In future work, we intend to improve capabilities for *in silico* analysis, as well as provide higher level design tools for self-assembling systems. Specifically, we intend to incorporate various simulation tools based on the segment-level enumerator—for instance, ODE-based or stochastic kinetic simulators—as well as separate, base-level kinetic simulation tools [39]. It would be interesting to develop tools for higher level design of self-assembly processes; for instance, a two- or three-dimensional 'molecular canvas', which would eventually allow users to quickly and easily translate two-dimensional images and three-dimensional structures to molecular implementations. Expanding individual tools to better handle more diverse structural motifs (for instance, pseudoknots, which are currently prohibited by the compiler and enumerator) and reaction types is another area for future work. Finally, in the Nodal compiler included with DyNAMiC Workbench, there is a one-to-one mapping between pre-defined node types and corresponding molecular implementations. However, this need not be the case for every behavioural designer. For instance, a future behavioural designer could operate on abstract chemical reaction networks, adopting one of several translation schemes [60,61] to convert the behaviour into a molecular structure. We anticipate many such future designers are possible.

Finally, an ongoing goal will also be to seek out and integrate valuable existing tools; we hope this will be a collaborative effort embraced by the entire community.

Authors' contributions. C.G. wrote the software, with contributions from J.W. and D.Y.Z. All authors contributed to the preparation of the manuscript.

Funding. Office of Naval Research (ONR) Young Investigator Program Award N000141110914, ONR grants N000141010827, N000141310593 and N000141410610, NIH Director's New Innovator Award 1DP2OD007292, NSF Faculty Early Career Development Award CCF1054898, NSF Expedition in Computing Award CCF1317291, NSF grants CCF1162459 and Wyss Institute funding to P.Y.

Competing interests. We declare no competing interests.

Acknowledgements. The authors thank John Sadowski and Erik Winfree for allowing integration and incorporation of their software; we also thank them, as well as Niles Pierce, Radhika Nagpal, William Shih, and Jocelyn Kishi for comments on the manuscript and testing of the software.

## References

- Seeman NC. 2010 Nanomaterials based on DNA. *Annu. Rev. Biochem.* **79**, 65–87. (doi:10.1146/annurev-biochem-060308-102244)
- Pinheiro AV, Han D, Shih WM, Yan H. 2011 Challenges and opportunities for structural DNA nanotechnology. *Nat. Nanotech.* **6**, 763–772. (doi:10.1038/nnano.2011.187)
- Krishnan Y, Simmel FC. 2011 Nucleic acid based molecular devices. *Angew. Chem. Int. Ed.* **50**, 3124–3156. (doi:10.1002/anie.200907223)
- Zhang DY, Seelig G. 2011 Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Chem.* **3**, 103–113. (doi:10.1038/nchem.957)
- SantaLucia J. 1998 A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA* **95**, 1460–1465. (doi:10.1073/pnas.95.4.1460)
- SantaLucia J, Hicks D. 2004 The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415–440. (doi:10.1146/annurev.biophys.32.110601.141800)
- Mathews DH, Turner DH. 2006 Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.* **16**, 270–278. (doi:10.1016/j.sbi.2006.05.010)
- Dirks RM, Bois JS, Schaeffer JM, Winfree E, Pierce NA. 2007 Thermodynamic analysis of interacting

- nucleic acid strands. *SIAM Rev.* **49**, 65–88. (doi:10.1137/060651100)
9. Carlson R. 2009 The changing economics of DNA synthesis. *Nat. Biotechnol.* **27**, 1091–1094. (doi:10.1038/nbt1209-1091)
  10. Service RF. 2011 DNA nanotechnology grows up. *Science* **332**, 1140–1143. (doi:10.1126/science.332.6034.1140)
  11. Yurke B, Turberfield AJ, Mills AP, Simmel FC, Neumann JL. 2000 A DNA-fuelled molecular machine made of DNA. *Nature* **406**, 605–608. (doi:10.1038/35020524)
  12. Turberfield AJ, Mitchell JC, Yurke B, Mills AP, Blakey MI, Simmel FC. 2003 DNA fuel for free-running nanomachines. *Phys. Rev. Lett.* **90**, 118102. (doi:10.1103/PhysRevLett.90.118102)
  13. Ding B, Seeman NC. 2006 Operation of a DNA robot arm inserted into a 2D DNA crystalline substrate. *Science* **314**, 1583–1585. (doi:10.1126/science.1131372)
  14. Venkataraman S, Dirks RM, Rothmund PWK, Winfree E, Pierce NA. 2007 An autonomous polymerization motor powered by DNA hybridization. *Nat. Nanotech.* **2**, 490–494. (doi:10.1038/nnano.2007.225)
  15. Shin J-S, Pierce NA. 2004 A synthetic DNA walker for molecular transport. *J. Am. Chem. Soc.* **126**, 10 834–10 835. (doi:10.1021/ja047543j)
  16. Yin P, Choi HMT, Calvert CR, Pierce NA. 2008 Programming biomolecular self-assembly pathways. *Nature* **451**, 318–322. (doi:10.1038/nature06451)
  17. Omabegho T, Sha R, Seeman NC. 2009 A bipedal DNA Brownian motor with coordinated legs. *Science* **324**, 67–71. (doi:10.1126/science.1170336)
  18. Dirks RM, Pierce NA. 2004 Triggered amplification by hybridization chain reaction. *Proc. Natl Acad. Sci. USA* **101**, 15 275–15 278. (doi:10.1073/pnas.0407024101)
  19. Zhang DY, Turberfield AJ, Yurke B, Winfree E. 2007 Engineering entropy-driven reactions and networks catalyzed by DNA. *Science* **318**, 1121–1125. (doi:10.1126/science.1148532)
  20. Sadowski JP, Calvert CR, Zhang DY, Pierce NA, Yin P. 2014 Developmental self-assembly of a DNA tetrahedron. *ACS Nano* **8**, 3251–3259. (doi:10.1021/nn4038223)
  21. Seelig G, Soloveichik D, Zhang DY, Winfree E. 2006 Enzyme-free nucleic acid logic circuits. *Science* **314**, 1585–1588. (doi:10.1126/science.1132493)
  22. Qian L, Winfree E. 2011 A simple DNA gate motif for synthesizing large-scale circuits. *J. R. Soc. Interface* **8**, 1281–1297. (doi:10.1098/rsif.2010.0729)
  23. Qian L, Winfree E. 2011 Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **332**, 1196–1201. (doi:10.1126/science.1200520)
  24. Qian L, Winfree E, Bruck J. 2011 Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368–372. (doi:10.1038/nature10262)
  25. Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, Dirks RM, Pierce NA. 2011 NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173. (doi:10.1002/jcc.21596)
  26. Seeman NC. 1990 De novo design of sequences for nucleic acid structural engineering. *J. Biomol. Struct. Dyn.* **8**, 573–581. (doi:10.1080/07391102.1990.10507829)
  27. Dirks RM, Lin M, Winfree E, Pierce NA. 2004 Paradigms for computational nucleic acid design. *Nucleic Acids Res.* **32**, 1392–1403. (doi:10.1093/nar/gkh291)
  28. Tulpan D, Andronescu M, Chang SB, Shortreed MR, Condon A, Hoos HH, Smith LM. 2005 Thermodynamically based DNA strand design. *Nucleic Acids Res.* **33**, 4951–4964. (doi:10.1093/nar/gki773)
  29. Zhang DY. 2011 Towards domain-based sequence design for DNA strand displacement reactions. In *DNA Computing and Molecular Programming: 16th Int. Conf., DNA 16, Hong Kong, China, 14–17 June* (eds Y Sakakibara, Y Mi). *Lecture Notes in Comput. Sci.* **6518**, 162–175. (doi:10.1007/978-3-642-18305-8\_15)
  30. Zadeh JN, Wolfe BR, Pierce NA. 2011 Nucleic acid sequence design via efficient ensemble defect optimization. *J. Comput. Chem.* **32**, 439–452. (doi:10.1002/jcc.21633)
  31. Markham NR, Zuker M. 2005 DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.* **33**, W577–W581. (doi:10.1093/nar/gki591)
  32. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. 2008 The Vienna RNA websuite. *Nucleic Acids Res.* **36**, W70–W74. (doi:10.1093/nar/gkn188)
  33. Lorenz R, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011 Vienna RNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26. (doi:10.1186/1748-7188-6-26)
  34. Lakin MR, Youssef S, Polo F, Emmott S, Phillips A. 2011 Visual DSD: a design and analysis tool for DNA strand displacement systems. *Bioinformatics* **27**, 3211–3213. (doi:10.1093/bioinformatics/btr543)
  35. Lakin MR, Youssef S, Cardelli L, Phillips A. 2012 Abstractions for DNA circuit design. *J. R. Soc. Interface* **9**, 470–486. (doi:10.1098/rsif.2011.0343)
  36. Dabby NL. 2013 The kinetics of toehold-mediated four-way branch migration. In *Synthetic molecular machines for active self-assembly: prototype algorithms, designs, and experimental study*. PhD thesis, California Institute of Technology Pasadena, CA.
  37. Genot AJ, Zhang DY, Bath J, Turberfield AJ. 2011 Remote toehold: a mechanism for flexible control of DNA hybridization kinetics. *J. Am. Chem. Soc.* **133**, 2177–2182. (doi:10.1021/ja1073239)
  38. Zuker M. 2003 Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415. (doi:10.1093/nar/gkg595)
  39. Schaeffer JM, Thachuk C, Winfree E. 2015 Stochastic simulation of the kinetics of multiple interacting nucleic acid strands. In *DNA Computing and Molecular Programming: 21st Int. Conf., DNA 21, Boston and Cambridge, MA, 17–21 August* (eds A Phillips, P Yin). *Lecture Notes in Comput. Sci.* **9211**, 194–211. (doi:10.1007/978-3-319-21999-8\_13)
  40. Zhang DY. 2011 Cooperative hybridization of oligonucleotides. *J. Am. Chem. Soc.* **133**, 1077–1086. (doi:10.1021/ja109089q)
  41. Radding CM, Beattie KL, Holloman WK, Wiegand RC. 1977 Uptake of homologous single-stranded fragments by superhelical DNA. *J. Mol. Biol.* **116**, 825–839. (doi:10.1016/0022-2836(77)90273-X)
  42. Green C, Tibbetts C. 1981 Reassociation rate limited displacement of DNA strands by branch migration. *Nucleic Acids Res.* **9**, 1905–1918. (doi:10.1093/nar/9.8.1905)
  43. Biswas I, Yamamoto A, Hsieh P. 1998 Branch migration through DNA sequence heterology. *J. Mol. Biol.* **279**, 795–806. (doi:10.1006/jmbi.1998.1769)
  44. Panyutin IG, Hsieh P. 1994 The kinetics of spontaneous DNA branch migration. *Proc. Natl Acad. Sci. USA* **91**, 2021–2025. (doi:10.1073/pnas.91.6.2021)
  45. Sadowski JP. In preparation. Multisubjective: better nucleic acid design through fast removal of undesired secondary structure.
  46. Grun C. 2014 Automated design of dynamic nucleic acid systems. BA thesis, School of Engineering and Applied Sciences, Harvard University, Boston, MA.
  47. McCaskill JS. 1990 The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**, 1105–1119. (doi:10.1002/bip.360290621)
  48. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. 1994 Fast folding and comparison of RNA secondary structures. *Mon. hefte Chem. Chem. Mon.* **125**, 167–188. (doi:10.1007/BF00818163)
  49. Dirks RM, Pierce NA. 2003 A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J. Comput. Chem.* **24**, 1664–1677. (doi:10.1002/jcc.10296)
  50. Dirks RM, Pierce NA. 2004 An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. *J. Comput. Chem.* **25**, 1295–1304. (doi:10.1002/jcc.20057)
  51. Mathews DH, Sabina J, Zuker M, Turner DH. 1999 Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911–940. (doi:10.1006/jmbi.1999.2700)
  52. Wolfe BR, Pierce NA. 2014 Sequence design for a test tube of interacting nucleic acid strands. *ACS Synth. Biol.* (doi:10.1021/sb5002196)
  53. Seiffert J, Huhle A. 2008 A full-automatic sequence design algorithm for branched DNA structures. *J. Biomol. Struct. Dyn.* **25**, 453–466. (doi:10.1080/07391102.2008.10507193)
  54. SantaLucia J, Allawi HT, Seneviratne PA. 1996 Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **35**, 3555–3562. (doi:10.1021/bi951907q)
  55. Tinoco I, Uhlenbeck OC, Levine MD. 1971 Estimation of secondary structure in ribonucleic

- acids. *Nature* **230**, 362–367. (doi:10.1038/230362a0)
56. Grun C, Sarma K, Wolfe BR, Shin SW, Winfree E. 2014 A domain-level DNA strand displacement reaction enumerator allowing arbitrary non-pseudoknotted secondary structures. In *Verification of Engineered Molecular Devices and Programs (VEMDP)*, 17 July, Vienna, Austria (eds M Kwiatkowska, A Phillips, C Thachuk), pp. 1–29. See <http://arxiv.org/abs/1505.03738>.
57. Shapiro BA, Maizel J, Lipkin LE, Currey K, Whitney C. 1984 Generating non-overlapping displays of nucleic acid secondary structure. *Nucleic Acids Res.* **12**, 75–88. (doi:10.1093/nar/12.1Part1.75)
58. Brucoleri RE, Heinrich G. 1988 An improved algorithm for nucleic acid secondary structure display. *Comput. Appl. Biosci.* **4**, 167–173. (doi:10.1093/bioinformatics/4.1.167)
59. Levenshtein VI. 1966 Binary codes capable of correcting deletions, insertions and reversals. *Sov. Phys. Dokl.* **10**, 707–710.
60. Soloveichik D, Seelig G, Winfree E. 2010 DNA as a universal substrate for chemical kinetics. *Proc. Natl Acad. Sci. USA* **107**, 5393–5398. (doi:10.1073/pnas.0909380107)
61. Cardelli L. 2013 Two-domain DNA strand displacement. *Math. Struct. Comput. Sci.* **23**, 247–271. (doi:10.1017/S0960129512000102)

## Resource

# Paper-Based Synthetic Gene Networks

Keith Pardee,<sup>1,2</sup> Alexander A. Green,<sup>1,2</sup> Tom Ferrante,<sup>1</sup> D. Ewen Cameron,<sup>2,3</sup> Ajay DaleyKeyser,<sup>1</sup> Peng Yin,<sup>1</sup> and James J. Collins<sup>1,2,3,\*</sup>

<sup>1</sup>Wyss Institute for Biological Inspired Engineering, Harvard University, Boston, MA 02115, USA

<sup>2</sup>Department of Biomedical Engineering and Center of Synthetic Biology, Boston University, Boston, MA 02215, USA

<sup>3</sup>Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

\*Correspondence: [jcollins@bu.edu](mailto:jcollins@bu.edu)

<http://dx.doi.org/10.1016/j.cell.2014.10.004>

## SUMMARY

Synthetic gene networks have wide-ranging uses in reprogramming and rewiring organisms. To date, there has not been a way to harness the vast potential of these networks beyond the constraints of a laboratory or in vivo environment. Here, we present an in vitro paper-based platform that provides an alternate, versatile venue for synthetic biologists to operate and a much-needed medium for the safe deployment of engineered gene circuits beyond the lab. Commercially available cell-free systems are freeze dried onto paper, enabling the inexpensive, sterile, and abiotic distribution of synthetic-biology-based technologies for the clinic, global health, industry, research, and education. For field use, we create circuits with colorimetric outputs for detection by eye and fabricate a low-cost, electronic optical interface. We demonstrate this technology with small-molecule and RNA actuation of genetic switches, rapid prototyping of complex gene circuits, and programmable in vitro diagnostics, including glucose sensors and strain-specific Ebola virus sensors.

## INTRODUCTION

The field of synthetic biology aims to re-engineer the molecular components of the cell to harness the power of biology. In doing so, synthetic biologists have created whole-cell biosensors (Kobayashi et al., 2004; Kotula et al., 2014), synthetic probiotics, new sources of drugs (Fossati et al., 2014), green energy, (Torella et al., 2013) and chemistry (Zhang et al., 2012; Martin et al., 2003). At the heart of many of these cellular technologies are synthetic gene networks (Elowitz and Leibler, 2000; Gardner et al., 2000), which are decision-based circuits often composed of a sensor element followed by a transducer that regulates a measurable output (Figure 1A). With sequence-specific sensing of nucleic acids and small-molecule recognition, these engineered logic and output elements hold great potential for broad biotechnology and medical applications. However, the application of cell-based synthetic gene networks outside of the laboratory has been restricted by concerns of biosafety and the practicality of the cellular host.

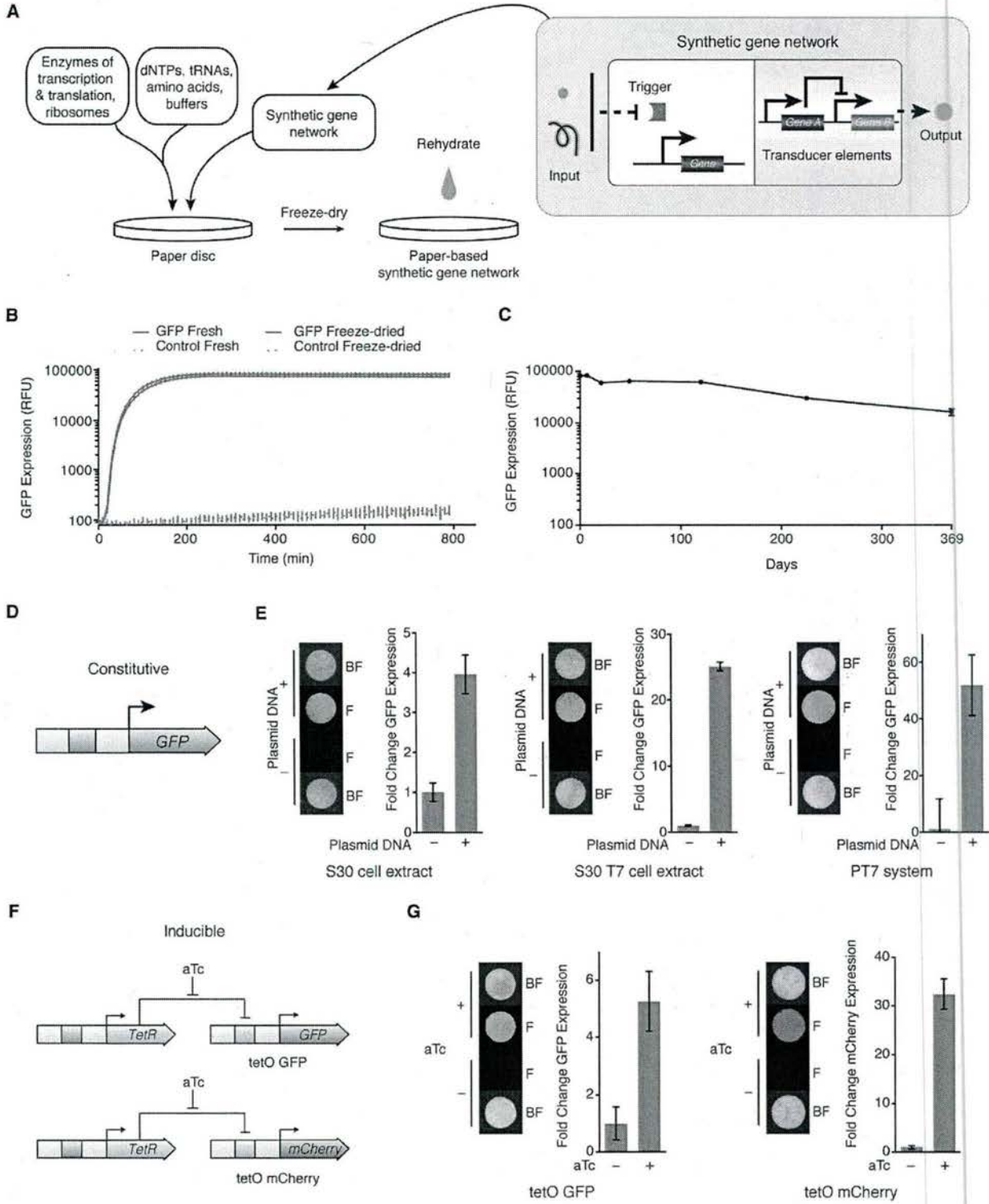
Earlier studies in the area of in vitro synthetic biology and cell-free systems have made important contributions to our understanding of fundamental molecular biology and biochemistry and, more recently, in the study of molecular switch dynamics and complex gene circuits (Hong et al., 2014; Karzbrun et al., 2014; Sun et al., 2014; Takahashi et al., 2014). These efforts, however, have focused on solution-phase reactions using fresh from frozen cell-free systems and often in liposomes with the goal of assembling artificial cells (Kuruma et al., 2009; Kobori et al., 2013). These solution-phase reactions are not stable or practical for handling outside of the lab and therefore miss the opportunity to leverage the abiotic and sterile nature of these systems.

Here, we report a method for embedding cell-free synthetic gene networks onto paper and other materials for in vitro applications in the clinic, global health, industry, research, and education. This is achieved by freeze drying cell-free systems into paper and other porous substrates to create materials with the fundamental transcription and translation properties of a cell but that are sterile and abiotic (Figure 1A). Stable at room temperature, these embedded materials are readily stored, distributed, and can be activated by simply adding water. We begin by demonstrating the successful operation of synthetic networks in this alternative medium for biomolecular reactions and then move on to show rapid screening of gene constructs and application of the paper-based platform for programmable in vitro diagnostics, including strain-specific Ebola sensors.

## RESULTS

### Development of Paper-Based Technology

To begin, we tested to see whether the enzyme activity required for transcription and translation could be reconstituted from freeze-dried cell-free expression systems, which normally require storage at  $-80^{\circ}\text{C}$ . The commercially available PT7 expression system, which is assembled from ribosomes and 35 purified bacterial proteins (Shimizu et al., 2001; Shimizu and Ueda, 2010), was freeze dried to form a pellet, reconstituted in water, and then used to transcribe and translate a constitutive GFP expression cassette. Remarkably, the freeze-dried and reconstituted PT7 solution-phase reactions showed activity comparable to the fresh from frozen PT7 reactions (Figure 1B). This approach was then extended to systems with greater molecular complexity based on whole-cell extracts. Using expression constructs dependent on either *E. coli* RNA polymerase (RNAP; S30) or T7



(legend on next page)

RNAP (S30 T7), constitutive expression of GFP was supported in small, solution-phase reactions started from freeze-dried pellets (Figures S1A and S1B available online). Importantly, these freeze-dried pellets are stable over time, with PT7 transcription and translation activity remaining after a year of room temperature storage (Figure 1C).

Although freeze-dried preparations offer great potential for the storage of poised synthetic biology tools, the handling of solution-phase reactions is awkward for applications outside of the lab. To address this, we embedded the cell-free systems and synthetic gene networks onto paper and other porous materials (Figure 1A). Previously used in measuring pH and more recently in chemistry-based diagnostics (Martinez et al., 2007), we speculated that paper could also serve as high-capillary-action matrix to hold small-volume molecular and biochemical reactions. To begin, we tested whether simple cell-free expression could be supported while embedded in the cellulose matrix of paper.

Small 2 mm filter paper discs were freeze dried with bacterial cell-free systems containing either *E. coli* (S30) or T7 RNAPs (S30 T7, PT7) and a corresponding *GFP* expression plasmid (Figure 1D). For all experiments, unless otherwise specified, the gene construct or network was freeze dried as a DNA copy into paper discs, along with the appropriate cell-free system (Figure S1G). After being freeze dried and rehydrated, these paper-based reactions yielded consistent GFP fluorescence under the regulation of either RNAP. Fluorescence imaging of the paper discs confirm GFP expression, and reactions were monitored by placing the paper discs at the bottom of black, clear-bottom 384-well plates for incubation in a 37°C plate reader (Figure 1E).

### Synthetic Gene Networks

Next, we tested inducible expression systems to determine whether more complex synthetic gene networks could be supported by the freeze-dried, paper-based reactions (Figure 1A). We started with tetO-regulated expression. Regulation of this system is mediated by the Tet repressor (TetR) that binds to the tetO promoter, preventing transcription (Figure 1F; Lutz and Bujard, 1997). Such regulation performed in vitro required that constitutive *TetR* expression also be encoded into the synthetic gene network. Expression is then induced by addition of the tetracycline analog anhydrotetracycline (aTc), which disrupts TetR binding to the promoter, allowing transcription of the regulated gene.

Freeze-dried discs were prepared with a cell-free system containing S30 *E. coli* cell extract, pretranslated TetR, and the

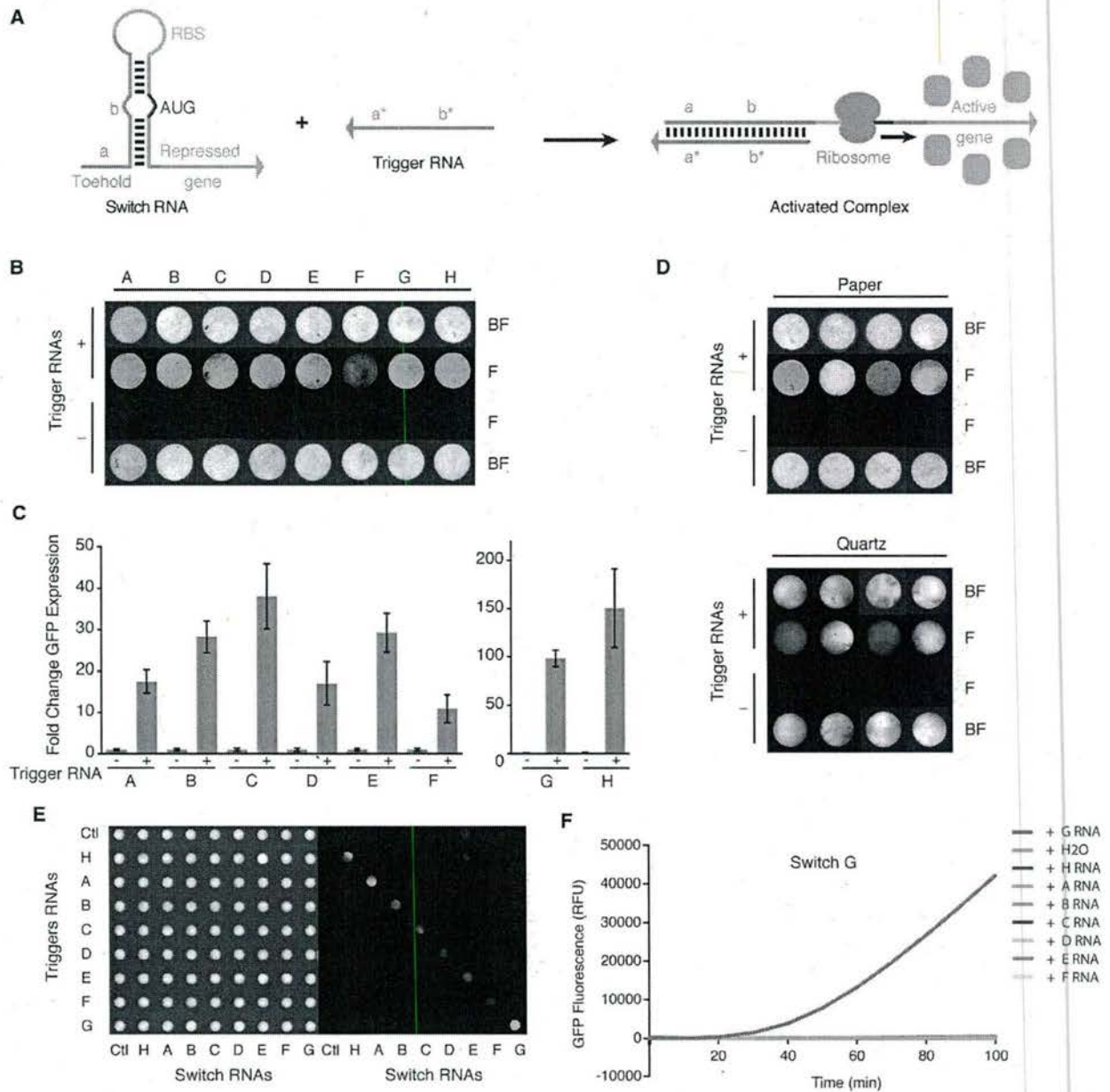
network elements encoding the constitutive expression of *TetR* and tetO-regulated *GFP* or *mCherry*. Varying only the presence of aTc, discs were rehydrated and incubated at 37°C. The aTc-induced discs yielded expression of GFP and mCherry at 5- and 32-fold, respectively (Figures 1G, S1C, and S1D). To optimize performance, we found that regulatory control from the tetO promoter could be tightened if cell extracts were supplemented, prior to freeze drying, with TetR to control promoter leakage prior to the expression of the encoded *TetR* (Figures S1E and S1F). This successful TetR augmentation of cell-free systems prior to freeze drying suggests that protein doping may be a useful way to tailor and enhance the regulatory and enzymatic environment of cell-free applications.

An important trend for synthetic biology is the use of RNA-actuated gene circuits (Callura et al., 2012). Unlike circuits based on small-molecule-based regulation, which are limited by the number of available unique recognition domains, RNA-based switches can be rationally programmed to recognize sequence-specific triggers and accordingly have the potential to offer essentially unlimited design and signaling space (Isaacs et al., 2004; Callura et al., 2010; Lucks et al., 2011; Mutalik et al., 2012). Here, we test a new generation of riboregulators in an in vitro demonstration of toehold switches (Green et al., 2014 [this issue of *Cell*]). These robust biomolecular switches provide tight translational regulation over transcripts and exhibit excellent orthogonality. Moreover, as we will demonstrate, by controlling the expression of alternative RNA polymerases (e.g., T3 RNAP), toehold switches can also trigger the synthesis of additional switch RNAs, enabling cascades and more sophisticated synthetic circuits.

The rational programmability of toehold switches comes from their design (Figure 2A). Riboregulators are composed of two cognate RNAs: a transducer RNA that encodes the output signal of the system (e.g., a GFP mRNA) and a trigger RNA that modulates the output signal. Conventional riboregulators have historically repressed translation by sequestering the ribosomal binding site (RBS) of the transducer RNA within a hairpin. This hairpin is unwound upon binding of a cognate trigger RNA, exposing the RBS and enabling translation of the downstream protein. However, this design restricts the potential trigger RNAs to those that contain RBS sequences. Toehold switches have removed this constraint by moving the RBS to a loop region of the hairpin, leaving the trigger RNA binding site free to adopt virtually any sequence. The transducer or “switch” RNA of toehold switches also contains a single-stranded domain known as a toehold at

### Figure 1. Paper-Based Synthetic Gene Networks

- (A) The enzymes of transcription and translation are combined with engineered gene circuits and then embedded and freeze dried onto paper to create stable and portable synthetic gene networks outside of the cell. These networks are genetically encoded tools with trigger, regulatory transducer, and output elements.
- (B) GFP expression in solution phase from fresh and freeze-dried PT7 cell-free reactions.
- (C) Freeze-dried pellets of the PT7 cell-free expression system are stable for over a year at room temperature, yielding GFP expression in solution phase when rehydrated.
- (D) A schematic of the constitutive *GFP* expression constructs used on paper.
- (E) Images of paper discs following a 2 hr incubation and maximum fold-change measurements of constitutive paper-based GFP expression from freeze-dried S30, S30 T7, and PT7 cell-free systems during the first 90 min of incubation.
- (F) Schematic of tetO regulation for *GFP* and *mCherry*.
- (G) Images of paper discs following a 2 hr incubation and maximum fold-change measurements of GFP and mCherry from the tetO-regulated promoter during the first 2 hr of incubation ± aTc inducer (11 μM) from freeze-dried S30 reactions. BF indicates bright-field images, and F indicates fluorescence images. Error bars represent SD.



**Figure 2. Freeze-Dried RNA-Actuated Gene Circuits on Paper**

(A) Schematic of the RNA-based toehold switch regulating translation of GFP.

(B) Images of paper-based GFP expression from eight toehold switches (A–H) in the PT7 cell-free system,  $\pm$  complementary trigger RNAs following a 2 hr incubation.

(C) Maximum fold-change measurement of GFP expression from paper-based toehold switches A–H during the first 90 min of incubation.

(D) RNA-actuated expression of GFP, Venus, mCherry, and Cerulean fluorescent proteins from toehold switch H on paper and quartz microfiber discs following a 2 hr incubation.

(E) Bright field and fluorescence images of an orthogonality screen between toehold switches and trigger RNAs using paper-based reactions arrayed in a 384-well plate. Images collected after the overnight data collection.

(F) Quantification of fluorescence over time from paper discs containing switch G (bottom row of the array). All data were generated from freeze-dried, cell-free reactions embedded into paper with their respective gene circuits. Trigger RNA concentration used for toehold switch activation was  $5 \mu\text{M}$ . BF indicates bright-field images, and F indicates fluorescence images.

Error bars represent SD.

its 5' end. This toehold domain, first developed in in vitro molecular programming studies (Yurke et al., 2000), provides the initial reaction site for binding between the trigger and switch RNAs and greatly improves the ON/OFF ratio of the switches.

We began by demonstrating the regulation of GFP expression under the control of eight different toehold switches on paper. Experiments were performed by freeze drying the recombinant PT7 expression system onto paper discs, along with linear DNA encoding specific switch RNAs. The paper discs were rehydrated with or without the complementary RNA trigger 24 hr after drying and then incubated at 37°C in humid chamber for 2 hr. As expected, discs that received the trigger RNA exhibited strong GFP expression, whereas control discs exhibited little, if any, expression (Figure 2B). Toehold switch and trigger RNAs can also be directly applied to the paper at the appropriate times or synthesized from DNA on the paper. By freeze drying the paper with pretranscribed toehold switch RNA, we were able to detect positive reactions 6 to 8 min earlier than having the RNAs transcribed in place (data not shown). However, due to the greater stability of DNA, we chose to freeze dry all paper reactions with DNA copies of the respective toehold switches.

Parallel experiments were done to measure reaction kinetics. We found that, within 90 min of 37°C incubation, the maximum ON/OFF ratios ranged between 10- and 140-fold (Figure 2C), with signal detection in as little as 20 min (Figure S2A). Maximum expression rates occurred between 60 and 120 min, and the reaction output generally reached a plateau by 200 min. Taking advantage of the nuclease-free nature of the PT7 recombinant system, we titrated RNA directly into the reactions and found a linear response to RNA concentration between low nanomolar and low micromolar concentrations (Figure S2B). We continued to use purified trigger RNAs so that we could control the concentration and chose 5 μM trigger RNA to maximize response in demonstration reactions. It is also of note that switch behavior in freeze-dried lysates replicates that of toehold switches in vivo in *E. coli* (Figure S2C; Green et al., 2014). Importantly, the clear target-dependent induction of the toehold switches in the RNA-rich environment of *E. coli* helps to demonstrate sequence specificity of these RNA components.

Whole-cell extracts and cellulose fibers exhibit an inherent autofluorescence around the emission spectrum of GFP (Schmidt, 2010), so we also constructed gene circuits to contain output fluorescent proteins with alternate emission spectra for both paper and quartz microfiber discs (Figures 2D and S2D). With these distinct outputs, a single paper disc can also be used to host multiple toehold switches, each capable of individual activation and measurement (Figure S2E). We also successfully tested toehold switches in S30 T7 *E. coli* cell extracts, a lower cost alternative, using circularized DNA plasmid as a template for both toehold switch and trigger RNAs (data not shown).

By arraying the paper discs into a 384-well plate, multiplexed reactions can be easily quantified and measured over time. We tested the orthogonality of the toehold switches by combining each of the eight toehold hairpins against each of the trigger RNAs. As seen in Figure 2E, activation of the toehold switches was essentially limited to the diagonal, with only one weak off-target interaction between switch E and trigger H (Figures 2E, 2F, S2G, and S2H). Despite not being explicitly designed to be

orthogonal to one another, significant orthogonality was also reflected in the specific activation demonstrated above in the in vivo milieu of bacteria cells (Figure S2C). This arrayed approach also underscores another important advantage of our in vitro paper-based system, namely, the rapidity with which it can be used to rationally design and test synthetic constructs. To collect similar data in vivo would require 81 separate transformations, overnight cultures on plates, and subsequent culturing in multiwell plates and could take several days. Using in vitro paper-based reactions, this experiment took under 90 min to set up, and, upon incubation, signals from activated switches were detected in as little as 20 min (Figure S2H).

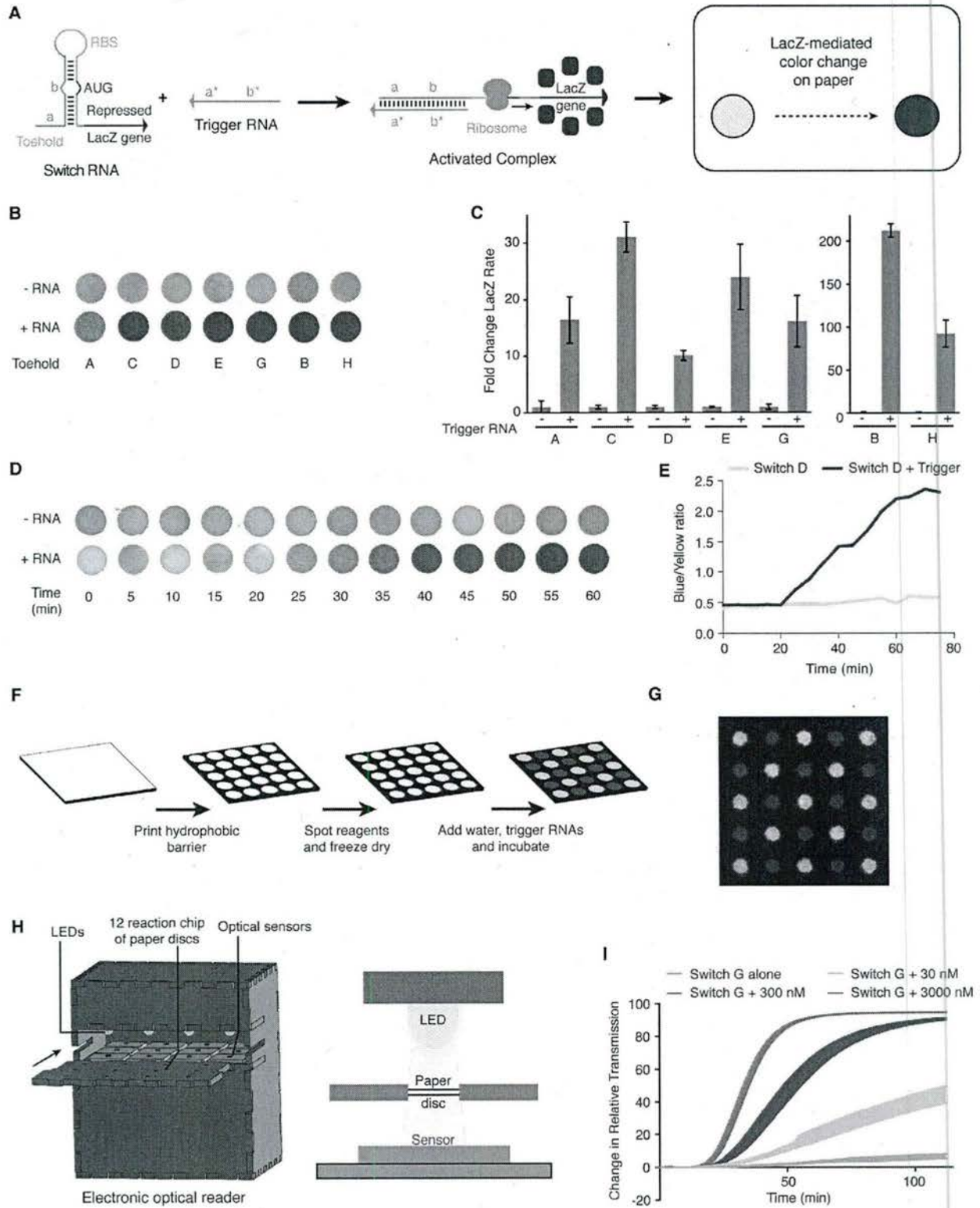
### Enabling Technologies

In keeping with our goal of making synthetic biology-based tools widely available outside of the laboratory, we designed our paper-based system to generate a colorimetric output visible to the naked eye. To do so, we replaced GFP with the enzyme β-galactosidase (LacZ) to create a synthetic gene network that generates a dramatic enzyme-mediated color change in response to conditional inputs (Figure 3A). LacZ cleaves the yellow substrate, chlorophenol red-β-D-galactopyranoside, embedded into the freeze-dried paper discs, to produce a purple chlorophenol red product that is visible to the naked eye and can be measured on standard plate readers by monitoring the absorbance at 570 nm.

As with fluorescent outputs from the GFP toehold switches, each triggered LacZ switch produces a unique switch-dependent response, with maximum expression rates occurring between 50 and 120 min (Figures 3B, 3C, and S3A). Some toehold switches rapidly reached their maximum output, whereas others produced a relatively linear output throughout overnight experiments. To maximize detection speed and sensitivity, we used the enzymatic reaction rate determined from the slope of absorbance at 570 nm over time to calculate fold change for the family of LacZ toehold switches (Figure 3C). As with the fluorescent reactions, paper-based reactions with colorimetric outputs can be arrayed for large-scale, quantitative experiments (Figure S3B).

Using switch D to demonstrate color development, detection of the reaction can be seen by the naked eye to begin at 25 min and saturates in ~1 hr (Figure 3D). Quantification of these reactions can also easily be done by measuring the intensity of the color channels red, green, and blue from images produced by most cameras, including those widely available in cellphones. By tracking the generation of signal in the blue channel, relative to yellow signal (red + green channels), the progress of the activated freeze-dried toehold switch reaction can be assessed and quantified (Figure 3E). Similar results were observed for the whole family of toehold switches (Figure 3B).

We have also incorporated an alternative colorimetric reporter enzyme, chitinase, into synthetic gene networks for systems based on extracts from cells that contain a LacZ background (Figure S3C). As demonstrated with our tetO-GFP and tetO-mCherry switches, the presence of aTc inducer results in the expression of chitinase, which cleaves a colorless substrate (4-Nitrophenyl N,N'-diacetyl-β-D-chitobioside) to yield a yellow p-nitrophenol product. The colorimetric output is visible to the naked eye and can be quantified using a standard plate reader



(legend on next page)

(410 nm; Figures S3D–S3F). Color development with this system was linear from the onset of detectable output at about 60 min, until the end of the experiment at 800 min, with a maximum induction of 38-fold at 420 min.

An important advantage of paper-based distribution of synthetic gene networks is their potential for low cost and relative ease to manufacture. As a proof-of-principle demonstration, we created printed arrays of stable synthetic gene networks using a standard computer printer and chromatography paper (Figure 3F). A commercially available wax-based ink serves as a hydrophobic barrier separating respective reaction spaces to create custom arrays and layouts (Carrilho et al., 2009). As a demonstration of the technique, we printed  $5 \times 5$  arrays to host the freeze-dried LacZ-expressing toehold switch E and then rehydrated the array with and without trigger RNA to create a checkerboard pattern (Figure 3G).

For other practical applications, we built a low-cost electronic optical reader to enable quantification and ultimately automation of the paper-based reactions. Such a device could also impart these molecular devices with the ease of use and convenience of a home glucose monitor. To measure light transmission through these reactions, we placed LacZ-based toehold switches on paper discs between LED light sources (570 nm) and electronic sensors (Figure 3H). In the event of a positive reaction, light transmission is progressively blocked by the production of the purple LacZ cleavage product. LEDs and sensors were coordinated through multiplexers connected to an Arduino that controlled the read pattern and rate parameters (Figure S3G). Electronic hardware was housed using computer-designed parts laser cut from acrylic to create a device for under \$100 USD (Figures 3H and S3H). Freeze-dried paper discs were placed into holes on a chip, rehydrated, incubated in the device at 37°C, and monitored in real time through an attached laptop. As a proof of concept, we tested the LacZ toehold switch G and observed consistent and significant reads from different concentrations of RNA trigger (Figure 3I).

### An In Vitro Diagnostics Platform

With a device in hand, we next turned our focus to developing toehold switch sensors capable of detecting full-length active mRNA targets, a key feature for a diagnostics platform. Our first goal was the detection of GFP and mCherry mRNAs. Using an

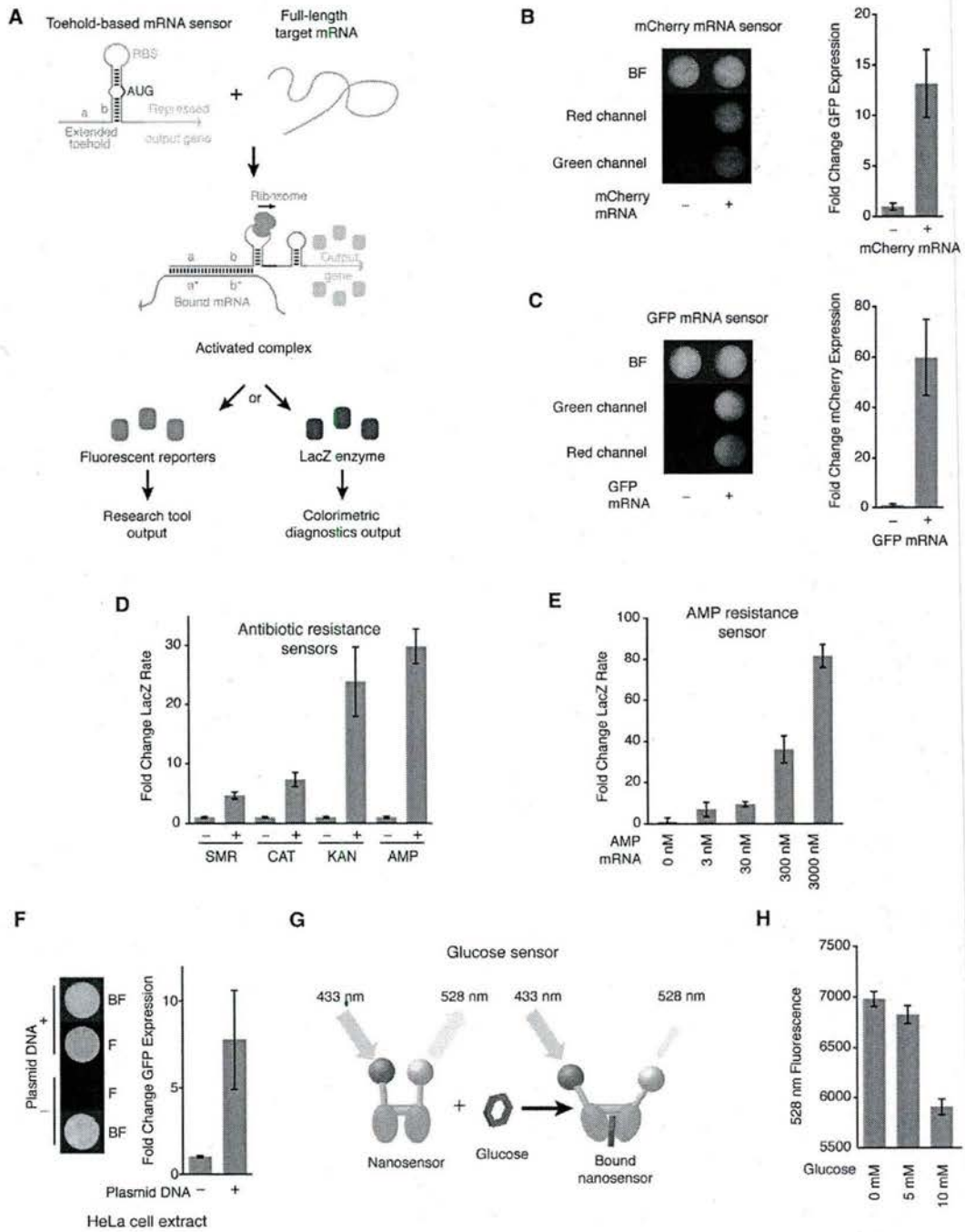
algorithm that predicts RNA secondary structure, mRNA sensors were built to target sequences likely to be accessible for binding (Green et al., 2014) and, when tested, yielded fluorescent induction in the presence of GFP (60-fold) and mCherry (13-fold) mRNAs, respectively (Figures 4B and 4C). As a demonstration of the potential for paper-based synthetic gene networks as an in vitro diagnostics platform, we next developed colorimetric mRNA sensors for antibiotic resistance genes (Figure 4A). In freeze-dried, paper-based reactions, mRNA sensors for spectinomycin (5-fold), chloramphenicol (7.5-fold), kanamycin (24-fold), and ampicillin (30-fold) resistance genes yielded significant LacZ induction in the presence of their respective mRNAs at 3,000 nM (Figures 4D and S4A). mRNAs used in these proof-of-concept experiments were provided in their purified form. However, given the fidelity of the toehold switches in the RNA-rich environment of the cell (Figure S2C; Green et al., 2014), we are confident that in vitro paper-based toehold sensors will exhibit similar specificity in the presence of mixed RNA samples.

Switching from the plate reader, we then tested the ampicillin resistance mRNA sensor using the electronic optical reader and found significant detection of target transcripts as low as 3 nM (Figures 4E and S4B). Moreover, the signal from our in-house device was about three times greater than the plate reader at equivalent trigger RNA concentrations (Figures 4D, 4E, and S4A).

In addition to the purpose-built synthetic gene networks presented above, this in vitro paper-based approach also holds tremendous potential for hosting other gene circuits, including those already in the literature as synthetic biology tools. Moreover, although the work above focused on bacterial components, the platform should in principle be readily extendible to mammalian-based systems. As confirmation, we tested a paper-based system using freeze-dried HeLa cell extracts and found strong expression of GFP in the presence of a constitutive expression plasmid (Figure 4F). To demonstrate the use of a pre-existing molecular tool, we chose to build a portable, in vitro glucose sensor on paper using a FRET-based nanosensor originally designed for measuring glucose in mammalian cells (Figure 4G; Takanaga and Frommer, 2010). The DNA construct for the nanosensor was freeze dried, along with HeLa cell extracts onto paper discs, and then rehydrated in the presence or absence of glucose. Upon rehydration, we observed the glucose-related shift previously reported in mammalian cells in

### Figure 3. Colorimetric Output from Paper-Based Synthetic Gene Networks

- (A) A schematic of the modified, LacZ-expressing toehold switches used to generate colorimetric outputs.
- (B) Images of the paper-based, colorimetric output from toehold switches  $\pm$  complementary RNA triggers following a 2 hr incubation.
- (C) Maximum fold-change measurements from LacZ toehold switches during the first 2 hr of incubation. Fold induction based on the rate of color change from LacZ toehold switches.
- (D) The paper-based development of color from LacZ toehold switch D over 60 min.
- (E) Color intensities from (D) were converted to a ratio of blue over yellow (red + green channels) channels and graphed over time.
- (F) Schematic describing the process of arraying synthetic gene networks on paper using printed arrays.
- (G) A 25-reaction printed array (14  $\times$  14 mm) of toehold switch E, containing positive reactions (purple) and negative control reactions (yellow) distributed in a checkerboard pattern following a 2 hr incubation.
- (H) Schematic of the low-cost, electronic optical reader developed to read colorimetric output from paper-based synthetic gene networks. Paper reactions held in a chip between an LED light source (570 nm) and electronic sensor.
- (I) Time-course data from the electronic optical reader of toehold switch G in the presence of 0, 30, 300 and 3,000 nM trigger RNA. Data were collected every 10 s with SD indicated by line thickness, and all data were generated from freeze-dried, cell-free reactions embedded into paper with their respective gene circuits. Trigger RNA concentrations used for toehold switch activation were 5  $\mu$ M or as specified. Error bars represent SD.



**Figure 4. Application of Paper-Based Synthetic Gene Networks**

(A) Schematic of the paper-based mRNA sensors based on toehold switches.

(B) Images and fold-change measurements of a paper-based mCherry mRNA sensor in the presence and absence of full-length target mRNA, following a 2 hr incubation. GFP is produced in response to detection of mCherry mRNA.

(C) Images and fold-change measurements of a paper-based GFP mRNA sensor in the presence and absence of full-length target mRNA. mCherry is produced in response to detection of GFP mRNA.

(D) Maximum fold change during the first 90 min of the LacZ-mediated color output rate from sensors for mRNAs encoding resistance to spectinomycin, chloramphenicol, ampicillin, and kanamycin antibiotics using a plate reader.

(legend continued on next page)

our portable paper-based system within a physiologically relevant glucose concentration range (Figures 4H and S4C). Highlighting the stability of this FRET-based sensor, the SD of reads remains low even when compiled from data collected over a 100 min period. Importantly, de novo translation of the nanosensor seems to be critical to function because freeze-dried preparations of pretranslated protein did not exhibit the characteristic shift in fluorescence.

### Rapid Sensor Prototyping

The low cost of manufacturing (4–65¢/sensor) is a key feature of the paper-based platform for diagnostics. However, perhaps even more important for the adoption and potential impact of this technology is the time and cost of developing new sensors. To test how rapidly new sensors could be developed, we chose to build mRNA sensors for the viral pathogen Ebola. Our goal was to construct and test 24 sensors that could distinguish between the Sudan and Zaire strains of the virus in under a day. Using our algorithm, sensors were designed to target mRNA from 12 regions (A–L) of the mRNA encoding the Ebola nucleoprotein, which differs in length by only three nucleotides between the Sudan and Zaire strains (Figure 5A). Construction began with PCR amplification of synthetic 135-nt DNA oligos bearing the toehold switch sensor cassettes, followed by ligation of the modules to our LacZ reporter chassis. The ligated product was then amplified by PCR. Followed by only column purification, sensors were tested on paper discs containing the freeze-dried PT7 cell-free system for testing in the presence and absence of 36 nt trigger RNAs. Here, we chose to add the sensor element during the rehydration phase of the experiment to demonstrate the potential for rapid prototyping and to highlight that the user does not need to know what sensors will be tested ahead of time, as might be the case with an emerging pathogen. Due to potential challenges with access to full-length material, we chose to use short 36 bp trigger RNAs rather than full-length mRNA to test the sensors. Although accessibility due to RNA secondary structure is important to keep in mind, work from our previous mRNA sensors suggests that short RNA triggers provide a representative indication of sensor function.

The colorimetric dynamics of the 240 reactions were captured using a plate reader and could be tracked by eye (Figures 5B and 5C). Each of the 24 sensors was triggered in the presence of their target (3,000 nM), with maximum induction during the first 90 min ranging 4- to 77-fold. Remarkably, beginning with the arrival of the DNA oligos, this phase of the screen was completed in less than 12 hr. We then selected four matching sets (D, E, G, and H) of the Sudan and Zaire sensors to test specificity and found a high degree of strain-specific discrimination (Figure 5D), as

well as sensitivity down to a concentration of 30 nM trigger RNA for both strains (Figure 5E).

### Rapid Prototyping and Testing of Complex Synthetic Gene Networks

As a demonstration of the capacity of the paper-based system for the prototyping and testing of more complex synthetic gene circuits, we assembled a converging transcription cascade that can convert the transcription activity of the native *E. coli* RNAP in a S30 cell extract into T3 and/or T7 RNAP activity (Figure 6A). To this point, we have validated the paper-based system using more than 45 toehold switches, each with unique combinations of inputs and measurable outputs. Here, we assembled toehold switches to create a series of molecular reactions that lead to the de novo production of transcriptional tools. In the presence of DNA encoding the toehold switch and trigger for the T3 module, T3 RNAP is transcribed and translated, which activates T3-mediated transcription and translation of GFP from an otherwise passive component (Figure 6B). Alternatively, in the presence of DNA encoding the toehold switch and trigger for the T7 module, T7 RNAP is transcribed and translated, leading to the expression of GFP from a T7-dependent promoter (Figure 6C). If both the T3 and T7 toehold switch modules are combined, *E. coli* RNAP generates both T3 and T7 RNAPs. This converging transcription activity leads to their respective trigger and toehold switch RNAs, producing a third route to GFP expression (Figure 6D). This work shows that the platform can sustain multiple rounds of transcription and translation and be utilized to construct, test, and debug complex circuits in a modular fashion.

### DISCUSSION

We have developed a method for embedding synthetic gene networks into paper and other porous materials, creating a much-needed path for moving synthetic biology out of the lab and into the field. Here, we have constructed a number of purpose-built synthetic gene networks responsive to synthetic RNA, mRNA, and small molecules. This includes paper-based toehold switches, which generate networks with greater than 20-fold (Figures 2C and 3C) and as high as 350-fold induction (Figure S2F). As demonstrated with the glucose sensor, this paper-based approach also promises to convert existing constructs designed for basic research and biotechnology into portable and readily accessible molecular tools (Figures 4G and 4H).

The in vitro nature of these reactions affords many benefits. For instance, without having to contend with maintaining the plasmid over generations of cell division, complex genetically

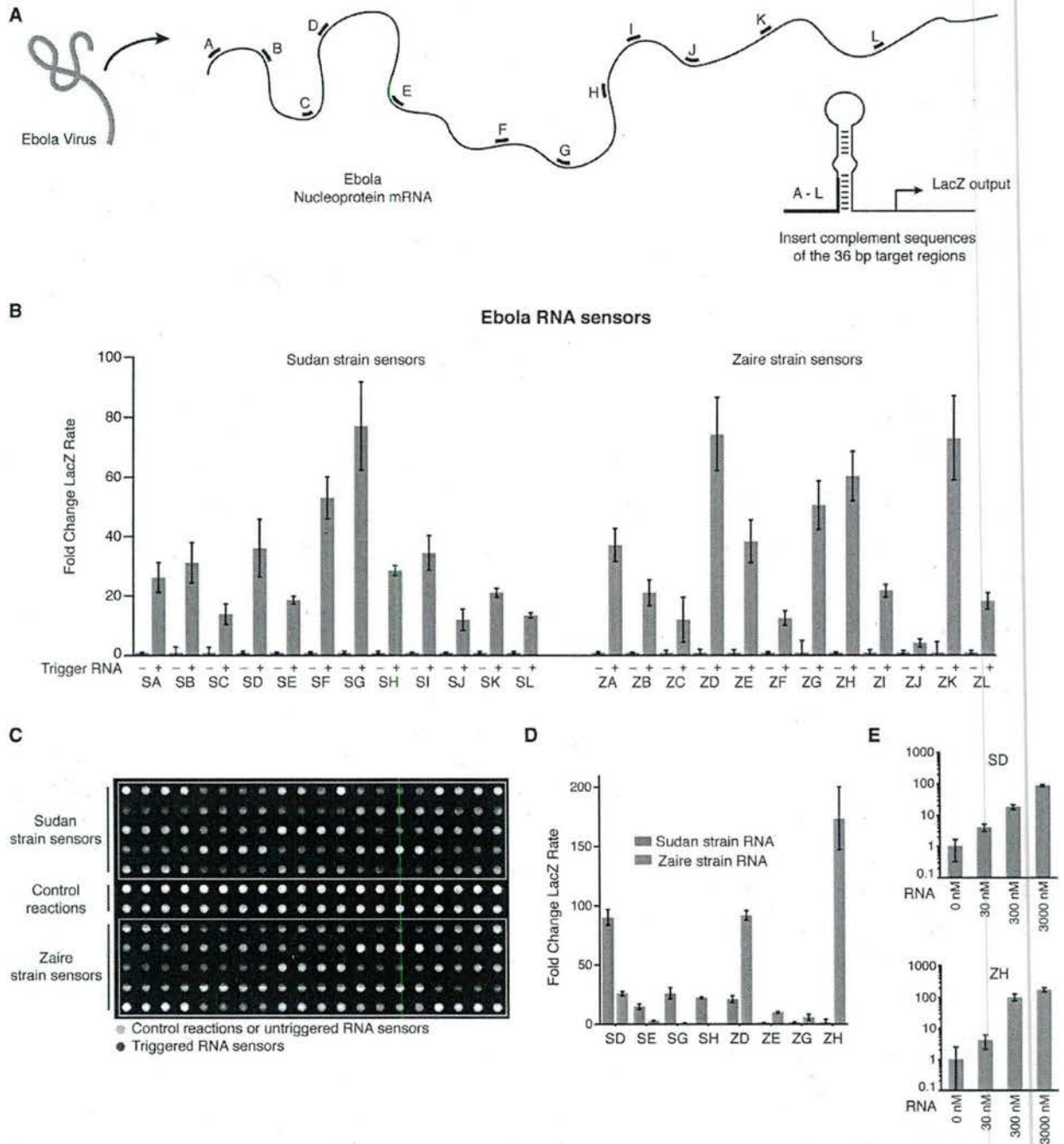
(E) Maximum fold change during the first 90 min of the color output rate from the ampicillin resistance sensor using the in-house electronic optical reader over a titration of mRNA concentrations.

(F) Images of paper discs following a 2 hr incubation and fold-change measurement of constitutive paper-based GFP expression from a freeze-dried HeLa cell extract.

(G) Schematic of the FRET-based mechanism used in the glucose nanosensor.

(H) Using an average of the data collected every 10 min between 390 min and 490 min, the 528 nm fluorescence is reduced in response to glucose binding to the FRET-based glucose nanosensor expressed on paper. All data were generated from freeze-dried, cell-free reactions embedded into paper with their respective gene circuits.

Error bars represent SD.



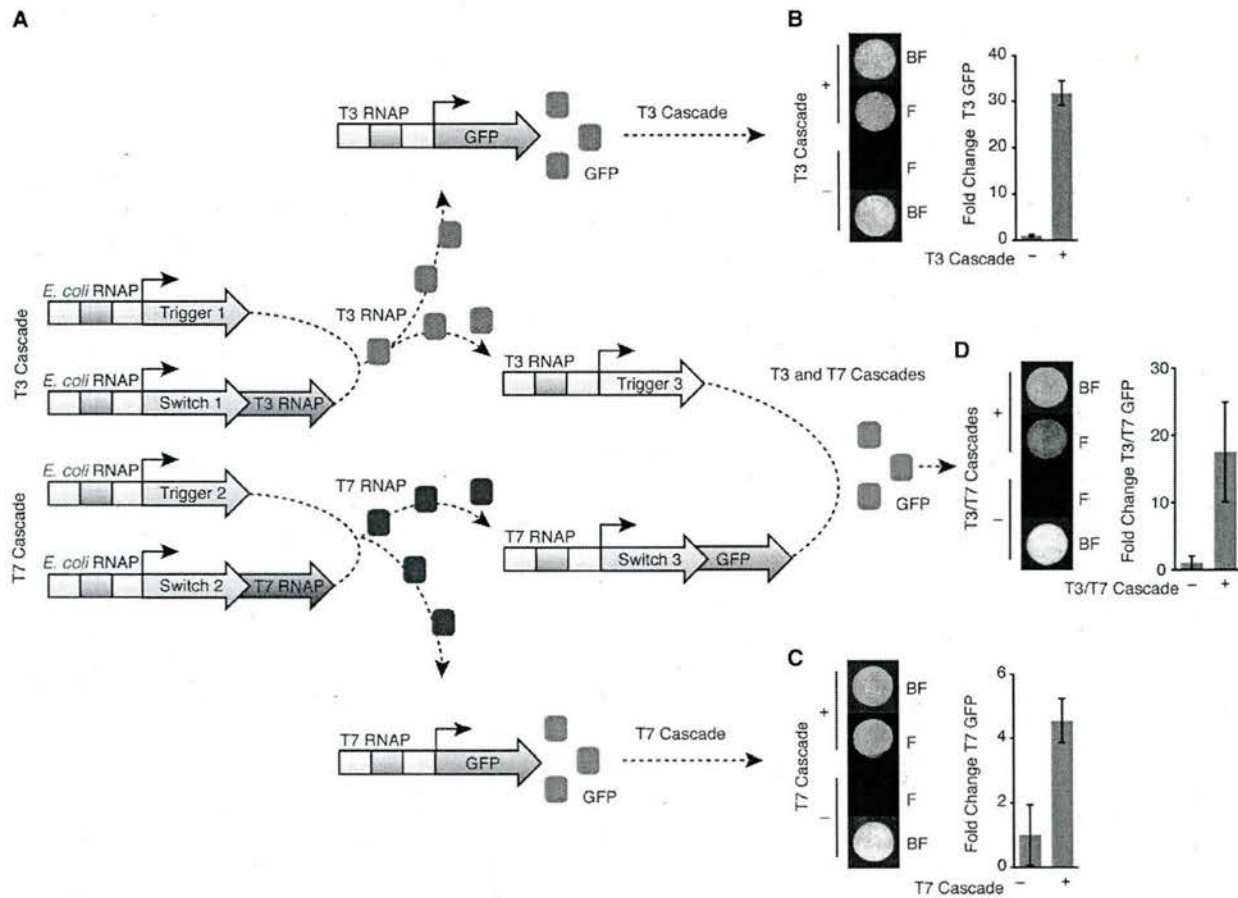
**Figure 5. Rapid Prototyping of Paper-Based RNA Sensors for Sequences from Sudan and Zaire Strains of the Ebola Virus**

(A) Schematic of the generation of Ebola RNA sensors. Sensors with the same letter were targeted to identical windows in the Ebola nucleoprotein mRNAs of their respective strains.

(B) Twenty-four toehold switch-based RNA sensors were constructed and tested in a 12 hr period. Based on the RNA segment windows (A–L), maximum fold change during the first 90 min at 37°C is reported for both the Sudan and Zaire strains of the virus. Fold change rate is determined from the slope of absorbance at 570 nm over time (– control, + 3000 nM RNA trigger).

(C) Composite image of the 240 paper-based reactions used to test the 24 sensors after data collection overnight. Control and untriggered toehold sensors remain yellow, and activated toehold sensors have turned purple.

(legend continued on next page)



**Figure 6. Paper-Based Converging Transcriptional Cascade**

(A) Schematic of the genetically encoded components that convert transcription from *E. coli* RNAP into transcription from T3 RNAP and/or T7 RNAP. Expression of these new RNAPs drives the transcription of previously dormant GFP constructs containing T3 or T7 promoters, and alternatively, they can drive expression of a toehold switch and trigger pair under the regulation of T7 and T3, respectively.

(B) Images and fold-change measurements of paper-based T3 GFP expression with and without the T3 cascade module.

(C) Images and fold-change measurements of paper-based T7 GFP expression with and without the T7 cascade module.

(D) Images and fold-change measurements of paper-based T3/T7-dependent GFP expression with and without the T3 and T7 cascade modules. All data were generated from freeze-dried, cell-free reactions embedded into paper with their respective GFP expression constructs; cascade modules were added as DNA components at rehydration. Data were collected after an overnight incubation. Error bars represent SD.

encoded networks can be assembled without the need for coordinating selective antibiotic pressures. Moreover, paper-based reactions can serve as a mock cell, allowing for linear PCR products to be screened directly for function without sequencing, assembly into circularized plasmids, or even gel purification of fragments. We took advantage of these features in our paper disc arrays to measure toehold orthogonality, as well as Ebola sensor prototyping and complex circuit testing (Figures 2E and

2F, 5, 6, S2G, S2H, and S3B), which allowed us to rapidly screen for gene circuit performance.

The paper-based cell extracts also offer a unique opportunity for molecular biologists to work with difficult-to-culture organisms. By generating cell extracts from engineered cell lines, pathogens, or other specialized organisms (e.g., extremophiles and symbionts), these reactions could serve as a proxy for biology that is otherwise inaccessible to the broad research

(D) Sequence specificity tested for four Sudan and four Zaire sensors from the original set of 24. Each of the four sensors targeting Sudan sequences were treated with 3,000 nM of off-target RNA sequence from the complementing Zaire RNA sequence and vice versa.

(E) Fold change of the color output rate of sensors SD and ZH over a titration of RNA concentrations. All data were generated from freeze-dried, cell-free reactions embedded into paper. DNA containing the sensor and RNA triggers was added during rehydration.

Error bars represent SD.

community. Similarly, the incorporation of human cell extracts into our paper-based scheme is an exciting feature that leads to the prospect of diagnostic and synthetic biology applications based on the thousands of human and mammalian transcription factors (Hughes, 2011), such as nuclear receptors, with complex and nuanced regulatory features, including small-molecule-responsive regulation (Pardee et al., 2011).

Perhaps most exciting is the potential to bring the rational design approach of synthetic biology to in vitro diagnostics and sensing. Although competing antibody-based technologies such as rapid diagnostic tests (RDTs) currently have greater sensitivity, we expect advances in molecular techniques and analyte concentration methods to solve these challenges. Examples of existing technologies that could be implemented as part of the platform include isothermal nucleic acid sequence-based amplification of target RNA inputs (Yan et al., 2014) or concentration of pathogen through generic opsonin-mediated capture (Cooper et al., 2014; Kang et al., 2014). Moreover, rather than competing with technically demanding and expensive lab-based techniques like ELISA, PCR, and mass spectrometry, we envision paper-based systems creating a new, low-cost generation of sensors that could be embedded ubiquitously into daily life (including clothing). Importantly, we were able to run toehold switch reactions in a variety of porous materials, including cloth, lab membranes, and porous alumina (data not shown).

Unlike antibody-based diagnostics, toehold switch mRNA sensors offer a sequence-based method of detection that means research and clinical tools can be designed rationally, lowering development costs and allowing for significantly shorter design-to-production cycles. Here, we presented toehold-based mRNA sensors, including 24 Ebola sensors that were constructed in less than 12 hr. The DNA input cost for the Ebola series was \$21 USD/sensor. These features compare highly favorably to custom commercial antibody production, where development time is typically 2 to 6 months and development costs range from \$4,000 to \$30,000 USD. Further, sensors can be designed from sequence information alone, which is ideal for emerging pathogens (Figure 5), and the acquisition of other relevant clinical information, such as the presence of drug-resistance genes (Figures 4D and 4E) or other indicators of pathogenesis, such as biofilm-specific RNAs (Dötsch et al., 2012). Coupled to an electronic reader, these paper-based systems also offer the potential for quantitative diagnostics (Figures 3I and 4E), a much-needed feature that is largely unavailable with RDTs (Cordray and Richards-Kortum, 2012). Moreover, the added sensitivity (2.5–3.5 ×) of our purpose-built, electronic reader could be considered a hardware enhancement of the designed gene circuits (Figures 4E and S4A).

Paper-based synthetic gene networks are also potentially less expensive to manufacture than most of the standard of care options currently available. At the moment, the cost of a 1 μl paper-based sensor would be between 35¢ and 65¢ using commercial cell-free expression systems. However, these systems can be readily produced in house, reducing the cost to as little as 2¢ to 4¢ per sensor (Noireaux, 2013). This compares to \$0.45 to \$1.40 for a single RDT reaction and \$1.50 to \$4.00 (reagents only) for PCR (Cordray and Richards-Kortum, 2012). Transcription- and translation-based detection is also competitive with re-

gards to time to detection. As an example, detection of mRNA from the ampicillin resistance gene was recorded as early as 25 min for high concentrations of mRNA and about 40 min for our 3 nM treatment (Figures 4E and S4B). This compares favorably to RDTs, which can detect a single antigen in ~20 min or PCR, which can take at least 1 hr and is largely confined to laboratory settings (Cordray and Richards-Kortum, 2012).

Our construction and testing of a converging transcription cascade on freeze-dried paper discs (Figure 6) further underscores the potential of this approach to host complex network reactions for both laboratory and field applications. Although perhaps not at first obvious, for each of the three transcription cascade reactions to produce GFP (Figures 6B–6D), the system had to essentially cycle through two cycles of the central dogma. The first cycle converts the DNA of the cascade module(s) to RNA and then protein (RNAP), which sets off the second cycle that converts the DNA of dormant reporter to RNA and then protein (GFP). Thus, despite their small size, the paper discs have the transcription and translation capacity required for complex tasks. For synthetic biology applications, such a transcription cascade could be used to build more sophisticated networks with layers of outputs that remain inactive unless triggered by a circuit producing the correct RNAP. Further, as the orthogonality (Figures 2E, S2G, S2H, and S3B) and Ebola screens showed (Figure 5), the paper-based system has significant potential to increase the pace at which genetically encoded tools can be built and tested. With this in mind, we foresee this technology being extended to prototyping of engineered metabolic pathways and other complex gene circuits as a way to rapidly vet combinatorial designs before moving to cellular hosts. Thus, ready-to-use paper-based systems could not only make tools currently only available in the laboratory readily fieldable but also improve the development of new tools and the accessibility of these molecular tools to educational programs for the next generation of practitioners.

Taken together and considering the projected cost, reaction time, ease of use, and no requirement for laboratory infrastructure, we envision paper-based synthetic gene networks significantly expanding the role of synthetic biology in the clinic, global health, industry, research, and education.

## EXPERIMENTAL PROCEDURES

### TetR-Based Inducible Synthetic Gene Circuits

The TetR-based GFP, mCherry, and chitinase expression circuits were constructed by inserting *GFP*, *mCherry*, or *chitinase* downstream of the TetR-repressed pLtetO promoter in pZE11 (Lutz and Bujard, 1997). To provide constitutive TetR expression, *TetR* was cloned downstream of the constitutive *placIQ* promoter, and this expression cassette was inserted into pZE11-gfp and pZE11-mcherry using XhoI and AatII.

### Toehold Switch Construction

Toehold switch plasmids were constructed using conventional molecular biology techniques. Synthetic DNA templates (Integrated DNA Technologies) were amplified using PCR, inserted into plasmids using Gibson assembly (Gibson et al., 2009) with 30 bp overlap regions, and then successfully constructed plasmids identified using DNA sequencing. All plasmids were derived from pET system parent plasmids (EMD Millipore) and constitutively express *lacI* and antibiotic resistance genes. Additional descriptions of the toehold switch plasmids and their sequences are provided in Green et al. (2014).

### Preparation of Matrix Materials

Although our initial paper-based reactions were successful, we suspected that nonspecific interactions between the components of the cell-free system and the cellulose matrix of paper could be impeding the activity of the reactions. Cellulose holds a pH-dependent charge (Budd, and Herrington, 1989), and thus it seemed likely that components from the complex biomolecular systems could adsorb to the high surface area of cellulose fibers, reducing reaction efficiency. Accordingly, we treated the filter paper with bovine serum albumin (BSA) and other commonly used blocking reagents. We found an overall improvement in fluorescent output from treated paper, with 5% BSA yielding the greatest increase (Figures S5A and S5B). For quartz microfiber, an alternate substrate, the performance enhancement of blocking with the surfactant Tween-20 was even more pronounced (Figures S5C and S5D).

### Cell-Free Reactions

Cell-free reactions were assembled (4°C) as generally described in the instructions of the respective manufacturers and then immediately incubated at 37°C. Briefly, for S30 and S30T7 reactions (Promega, L1020 and L1110), cell extracts and premix containing amino acids were combined at a ratio of 0.33 and 0.51, respectively. The volume was then brought up with RNase inhibitor (Roche; 0.005), plasmid DNA constructs comprising the gene circuits, and nuclease-free ddH<sub>2</sub>O. For tetO-based gene circuits, reactions were supplemented with prerun cell-free reactions expressing constitutive TetR (0.05) to reduce promoter leakage. For the PT7 cell-free system (NEB, E6800L), solutions A (0.4) and B (0.3) were combined together, with the remaining volume composed of RNase inhibitor (Roche; 0.005), linear DNA constructs, and nuclease-free ddH<sub>2</sub>O. HeLa cell cell-free systems (Thermo Scientific, 88881) were assembled by combining HeLa cell lysate (0.5), accessory proteins (0.1), and reaction mix (0.2) with RNase inhibitor (Roche; 0.005), plasmid DNA constructs comprising the gene circuits, and nuclease-free ddH<sub>2</sub>O. Long-term storage of freeze-dried pellets was done protected from light, under nitrogen gas and in the presence silica gel desiccation packages. For colorimetric reactions, the enzymatic substrates chlorophenol red-β-D-galactopyranoside and 4-Nitrophenyl N,N'-diacetyl-β-D-chitobioside were supplied to the reactions prior to freeze drying at final concentrations of 0.6 mg/ml and 3 mg/ml, respectively.

The concentration for the gene circuits were as follows: Figures 1B and 1C, T7\_GFP plasmid DNA 5 ng/μl (pBR939b\_T7\_GFP; Anderson et al., 2007); Figure 1E, PN25-GFP plasmid DNA 30 ng/μl and T7\_GFP plasmid DNA 30 ng/μl; Figure 1G, TetO-GFP and TetO-mCherry plasmid DNA 30 ng/μl. Toehold switch and sensor sequences are available as a supplemental information file (Data S2); Figure 2, linear DNA 33 nM, RNA triggers 5 μM; Figure 3, linear DNA 33 nM, RNA triggers 5 μM or as specified; Figures 4B and 4C, mCherry and GFP mRNA sensors linear DNA 33 nM, respective mRNAs 2.5 μM; Figures 4D and 4E, antibiotic resistance gene mRNA sensors for spectinomycin, chloramphenicol, kanamycin, and ampicillin linear DNA 33 nM, respective mRNAs 3 μM or as specified; Figure 4F, pT7CFE1\_GFP plasmid DNA (Thermo Scientific) 30 ng/μl, pcDNA3.1 FLIPglu-30uDelta13V plasmid DNA (Takanaga and Frommer, 2010; Addgene:18015) 30 ng/μl; Figure 5, PCR product was not gel purified, so a 5x linear DNA concentration of 150 ng/μl was used to ensure adequate sensor product, respective trigger RNAs 3 μM or as specified; Figure 6, T3 and T7 cascade modules plasmid DNA 30 ng/μl (*E. coli* RNAP\_trigger\_1, *E. coli* RNAP\_switch\_1\_T3RNAP, *E. coli* RNAP\_trigger\_2, *E. coli* RNAP\_switch\_2\_T7RNAP), T3\_GFP and T7\_GFP plasmid DNA 40 ng/μl, T3 RNAP\_trigger\_3, and T7 RNAP\_switch\_3\_GFP plasmid DNA 40 ng/μl.

### Preparation of Reactions and Incubation

Assembled cell-free reactions were applied (1.8 μl/disc) to 2 mm paper discs (Whatman, 1442-042), which were then flash frozen in liquid nitrogen and freeze dried overnight. Paper discs were cut using a 2 mm biopsy punch. Similarly, quartz microfiber (Spectrum, 884-66171) was cut and treated with cell-free reactions (3 μl/disc) prior to freeze drying. Freeze-dried solution phase reactions (7 μl) were similarly flash frozen and placed on the lyophilizer to create pellets. After 24 hr, paper reactions for Figures 1, 2, 3, and 4 were rehydrated with either nuclease-free ddH<sub>2</sub>O or inducer at the concentrations specified. For these reactions, DNA for gene circuits was added to paper before freeze drying. For Figure 5 reactions, DNA containing the sensor and RNA trig-

gers were both added during rehydration. For Figure 6 reactions, DNA for the GFP reporters were added prior to freeze drying (T3\_GFP, T7\_GFP, or T3/T7\_GFP). Rehydration of the paper discs was then done in the presence or absence of DNA constructs for the T3 and/or T7 cascade modules. Cell-free reactions without synthetic gene networks were also included to provide the background signal for subtraction from control and treatment reactions. Rehydrated reactions were incubated at 37°C using either a plate reader (BioTek NEO HTS) or our in-house electronic optical reader placed inside a tissue culture incubator. For the plate reader, paper discs were placed into black, clear-bottom 384 well plates, and for the in-house reader, paper discs were placed into 2 mm holes in the removable acrylic chip (Figures 3H and S3H). Cell-based reactions presented in Figure S2C were performed in *E. coli* as described by Green et al. (2014). Briefly, log-phase cultures containing a toehold switch construct and either an on-target or off-target IPTG-inducible RNA trigger construct were diluted 1:100. Cells were grown at 37°C overnight in the presence or absence of 1 mM IPTG, and GFP expression was measured using a plate reader.

### Microscopy and Image Processing

Images of paper discs were collected on a Zeiss Axio Zoom V16 Macroscope (magnification 7x) with an AxioCam MRm in a humidified glass chamber or through the bottom of a clear-bottom 384-well plate. Collected images were then stitched together using Zeiss Zen software into large composite images for further processing in ImageJ. Experiments were arranged so that images of matching control and treatment paper-based reactions were collected together such that parameters could be adjusted for all samples simultaneously. Once optimized, images of individual paper discs were cropped and arrayed into figures. For GFP expression in the S30 cell-free system, which exhibits a high level of autofluorescence, we used a Nuance camera to collect multispectral images between 500 and 620 nm. Perkin Elmer Nuance 3.0.2 software was then used to unmix the spectral signature of the GFP from that of the cell extract. A similar approach was used to create an absorbance signature (420 to 720 nm) for imaged paper discs with p-nitrophenol, the chitinase cleavage product of 4-Nitrophenyl N,N'-diacetyl-β-D-chitobioside. Images collected with the Nuance camera were scaled using bilinear transformation. For a few composite images, areas around the paper discs were masked to remove extraneous signal.

### Printed Arrays

Patterns for printed arrays were generated using Adobe Illustrator and then printed onto chromatography paper (Whatman, 3001-861) using a Xerox 8570 printer (Carrilho et al., 2009). Once printed, the wax was reflowed using a hot plate (120°C) so that the wax was present throughout the entire thickness of the paper, creating hydrophobic barriers to contain each reaction. The printing was performed at the Center for Nanoscale Systems at Harvard University, a member of the National Nanotechnology Infrastructure Network (NNIN), which is supported by the National Science Foundation under NSF award ECS-0335765.

### Electronic Optical Reader

The portable device consists of four layers housed within a laser-cut acrylic box (Figure 3H). The top layer holds 12 LEDs (Digi-Key, 754-1262-ND), which have a very narrow viewing angle and an emission of 570 nm to match the absorbance maximum of the product of the LacZ reaction. The LEDs were placed in close proximity to the chip in the middle layer, which holds 12 paper disks within 2 mm apertures. The apertures prevented transmission of stray light and were coaxial with the LEDs in the top layer and the array of 12 TSL2561 sensors (Adafruit, 439) in the third layer below. The bottom layer contains the Arduino Micro and associated electronics such as the multiplexers (Sparkfun, BOB-09056), breadboard, resistors, and connectors. To prevent crosstalk between reads, reactions were read in series by sequentially activating each LED and sensor pair. The read frequency and pattern of the reader can be easily adjusted by modifying and uploading alternative sketches to the Arduino Micro. Both the raw data and the data processed using the per disc formula:  $100 \cdot (100 \cdot (\text{Current}/\text{Max}))$  were calculated for each time point and sent to a laptop. A diagram of the circuit and an overview of the laser cut parts can be found in the supplemental figures (Figures S3G and S3H) and laser

cutting patterns and Arduino sketch files are provided as supplemental information files (Data S1).

#### Calculation of Fold Change

Fluorescence and absorbance data were first smoothed to reduce measurement noise using a moving three-point average of the time point and the data both preceding and following that read. Background signal was subtracted from each well using the average of control blank reactions that contained the relevant cell-free system alone. The minimum value of each well was then adjusted to zero. For fluorescence data, fold change calculations were done by dividing the wells at each time point by the average signal from the corresponding uninduced control wells. For absorbance measurements, fold change values reflect the difference in the rate of color change between induced and uninduced wells. This was done by calculating the rate of change using slope; therefore, at each 10 min time point, the rate reported was calculated as follows:  $S_n = (T_{n+1} - T_n)/10$ , where  $T$  is the normalized data at a time point ( $T_n$ ) and the time point 10 min later ( $T_{n+1}$ ), and  $S_n$  is the slope reported for  $T_n$ . Fold change was then calculated as with the fluorescence data. Experiments were run in either triplicate or quadruplicate. Due to high autofluorescence of the S30 reactions, a few experiments exhibited high variability and were accordingly subjected to the Modified Thompson Tau Test to identify outliers to be discarded (Anbarasi et al., 2011; Byrd et al., 2014). Briefly, the Tau test was performed by calculating the difference (delta) between a measured value (replicate) and the mean of the group. Using the Modified Thompson Tau value for quadruplicate data (1.4250), the test was performed by multiplying the SD of the group by this Tau value. If the resulting number was smaller than the delta calculated for a measured value, that replicate was considered an outlier and was not included in the analysis.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and two data files and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.004>.

#### ACKNOWLEDGMENTS

K.P. is a Canadian Institutes of Health Research-funded postdoctoral fellow at the Wyss Institute at Harvard University. This work was supported by the Wyss Institute; and NIH Director's New Innovator Award (1DP2OD007292), an ONR Young Investigator Program Award (N000141110914), and an NSF Expedition in Computing Award (CCF1317291) through P.Y. and the Howard Hughes Medical Institute, the Office of Naval Research MURI program, and the Defense Threat Reduction Agency grant HDTRA1-14-1-0006 through J.J.C.

Received: September 10, 2014

Revised: September 29, 2014

Accepted: October 3, 2014

Published: October 23, 2014

#### REFERENCES

Anbarasi, M.S., Ghaayathri, S., Kamaleswari, R., and Abirami, I. (2011). Outlier Detection for Multidimensional Medical Data. *Int. J. Comp. Sci. Info. Tech.* 2, 512–516.

Anderson, J.C., Voigt, C.A., and Arkin, A.P. (2007). Environmental signal integration by a modular AND gate. *Mol. Syst. Biol.* 3, 133.

Budd, J., and Herrington, T.M. (1989). Surface charge and surface area of cellulose fibres. *Colloids Surf.* 36, 273–288.

Byrd, K.B., O'Connell, J.L., Tommaso, S.D., and Kelly, M. (2014). Evaluation of sensor types and environmental controls on mapping biomass of coastal marsh emergent vegetation. *Remote Sens. Environ.* 149, 166–180.

Callura, J.M., Dwyer, D.J., Isaacs, F.J., Cantor, C.R., and Collins, J.J. (2010). Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proc. Natl. Acad. Sci. USA* 107, 15898–15903.

Callura, J.M., Cantor, C.R., and Collins, J.J. (2012). Genetic switchboard for synthetic biology applications. *Proc. Natl. Acad. Sci. USA* 109, 5850–5855.

Carrilho, E., Martinez, A.W., and Whitesides, G.M. (2009). Understanding wax printing: a simple micropatterning process for paper-based microfluidics. *Anal. Chem.* 81, 7091–7095.

Cooper, R.M., Leslie, D.C., Domansky, K., Jain, A., Yung, C., Cho, M., Workman, S., Super, M., and Ingber, D.E. (2014). A microdevice for rapid optical detection of magnetically captured rare blood pathogens. *Lab Chip* 14, 182–188.

Cordray, M.S., and Richards-Kortum, R.R. (2012). Emerging nucleic acid-based tests for point-of-care detection of malaria. *Am. J. Trop. Med. Hyg.* 87, 223–230.

Dötsch, A., Eckweiler, D., Schniederjans, M., Zimmermann, A., Jensen, V., Scharfe, M., Geffers, R., and Häussler, S. (2012). The *Pseudomonas aeruginosa* transcriptome in planktonic cultures and static biofilms using RNA sequencing. *PLoS ONE* 7, e31092.

Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338.

Fossati, E., Ekins, A., Narcross, L., Zhu, Y., Falguyret, J.P., Beaudoin, G.A., Facchini, P.J., and Martin, V.J. (2014). Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydroanguinarine in *Saccharomyces cerevisiae*. *Nat Commun.* 5, 3283.

Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345.

Green, A.A., Silver, P.A., Collins, J.J., and Yin, P. (2014). Toehold switches: de novo-designed regulators of gene expression. *Cell* 159, Published online October 23, 2014. <http://dx.doi.org/10.1016/j.cell.2014.10.002>.

Hong, S.H., Ntai, I., Haimovich, A.D., Kelleher, N.L., Isaacs, F.J., and Jewett, M.C. (2014). Cell-free protein synthesis from a release factor 1 deficient *Escherichia coli* activates efficient and multiple site-specific nonstandard amino acid incorporation. *ACS Synth. Biol.* 3, 398–409.

Hughes, T., ed. (2011). Handbook of transcription factors. *Subcellular Biochemistry, First Edition, Volume 52* (New York: Springer).

Isaacs, F.J., Dwyer, D.J., Ding, C.M., Pervouchine, D.D., Cantor, C.R., and Collins, J.J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.* 22, 841–847.

Kang, J.H., Super, M., Yung, C.W., Cooper, R.M., Domansky, K., Graveline, A.R., Mammoto, T., Berthet, J.B., Tobin, H., Cartwright, M.J., et al. (2014). An extracorporeal blood-cleansing device for sepsis therapy. *Nat. Med.* 20, 1211–1216. <http://dx.doi.org/10.1038/nm.3640>.

Karzbrun, E., Tayar, A.M., Noireaux, V., and Bar-Ziv, R.H. (2014). Synthetic biology. Programmable on-chip DNA compartments as artificial cells. *Science* 345, 829–832.

Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T.S., Cantor, C.R., and Collins, J.J. (2004). Programmable cells: interfacing natural and engineered gene networks. *Proc. Natl. Acad. Sci. USA* 101, 8414–8419.

Kobori, S., Ichihashi, N., Kazuta, Y., and Yomo, T. (2013). A controllable gene expression system in liposomes that includes a positive feedback loop. *Mol. Biosyst.* 9, 1282–1285.

Kotula, J.W., Kerns, S.J., Shaket, L.A., Siraj, L., Collins, J.J., Way, J.C., and Silver, P.A. (2014). Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl. Acad. Sci. USA* 111, 4838–4843.

Kuruma, Y., Stano, P., Ueda, T., and Luisi, P.L. (2009). A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells. *Biochim. Biophys. Acta* 1788, 567–574.

Lucks, J.B., Qi, L., Mutalik, V.K., Wang, D., and Arkin, A.P. (2011). Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proc. Natl. Acad. Sci. USA* 108, 8617–8622.

- Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* *25*, 1203–1210.
- Martin, V.J., Pitera, D.J., Withers, S.T., Newman, J.D., and Keasling, J.D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* *21*, 796–802.
- Martinez, A.W., Phillips, S.T., Butte, M.J., and Whitesides, G.M. (2007). Patterned paper as a platform for inexpensive, low-volume, portable bioassays. *Angew. Chem. Int. Ed. Engl.* *46*, 1318–1320.
- Mutalik, V.K., Qi, L., Guimaraes, J.C., Lucks, J.B., and Arkin, A.P. (2012). Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol.* *8*, 447–454.
- Noireaux, V. (2013). [http://www.openwetware.org/wiki/Biomolecular\\_Breadboards:Protocols:cost\\_estimate](http://www.openwetware.org/wiki/Biomolecular_Breadboards:Protocols:cost_estimate)
- Pardee, K., Necakov, A., and Krause, H. (2011). Nuclear receptors: small molecule sensors that coordinate growth, metabolism and reproduction. In *A Handbook of Transcription Factors. Subcellular Biochemistry, Volume 52*, First Edition, Hughes, T.R., ed. (New York: Springer), pp. 123–153.
- Schmidt, J.A. (2010). Electronic spectroscopy of lignans. In *Lignin and Lignans: Advances in Chemistry*, C. Heitner, D. Dimmel, and J. Schmidt, eds. (Boca Raton, FL: CRC Press), pp. 49–102.
- Shimizu, Y., and Ueda, T. (2010). PURE technology. *Methods Mol. Biol.* *607*, 11–21.
- Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001). Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* *19*, 751–755.
- Sun, Z.Z., Yeung, E., Hayes, C.A., Noireaux, V., and Murray, R.M. (2014). Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synth. Biol.* *3*, 387–397.
- Takahashi, M.K., Chappell, J., Hayes, C.A., Sun, Z.Z., Kim, J., Singhal, V., Spring, K.J., Al-Khabouri, S., Fall, C.P., Noireaux, V., Murray, R.M., and Lucks, J.B. (2014). Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription-translation (TX-TL) systems. *ACS Synth. Biol. March 28*, A–M. <http://dx.doi.org/10.1021/sb400206c>.
- Takanaga, H., and Frommer, W.B. (2010). Facilitative plasma membrane transporters function during ER transit. *FASEB J.* *24*, 2849–2858.
- Torella, J.P., Ford, T.J., Kim, S.N., Chen, A.M., Way, J.C., and Silver, P.A. (2013). Tailored fatty acid synthesis via dynamic control of fatty acid elongation. *Proc. Natl. Acad. Sci. USA* *110*, 11290–11295.
- Yan, L., Zhou, J., Zheng, Y., Gamson, A.S., Roembke, B.T., Nakayama, S., and Sintim, H.O. (2014). Isothermal amplified detection of DNA and RNA. *Mol. Biosyst.* *10*, 970–1003.
- Yurke, B., Turberfield, A.J., Mills, A.P., Jr., Simmel, F.C., and Neumann, J.L. (2000). A DNA-fuelled molecular machine made of DNA. *Nature* *406*, 605–608.
- Zhang, F., Carothers, J.M., and Keasling, J.D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* *30*, 354–359.

## Resource

# Toehold Switches: De-Novo-Designed Regulators of Gene Expression

Alexander A. Green,<sup>1</sup> Pamela A. Silver,<sup>1,2</sup> James J. Collins,<sup>1,3</sup> and Peng Yin<sup>1,2,\*</sup>

<sup>1</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA

<sup>2</sup>Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

<sup>3</sup>Howard Hughes Medical Institute, Department of Biomedical Engineering and Center of Synthetic Biology, Boston University, Boston, MA 02215, USA

\*Correspondence: [py@hms.harvard.edu](mailto:py@hms.harvard.edu)

<http://dx.doi.org/10.1016/j.cell.2014.10.002>

## SUMMARY

Efforts to construct synthetic networks in living cells have been hindered by the limited number of regulatory components that provide wide dynamic range and low crosstalk. Here, we report a class of de-novo-designed prokaryotic riboregulators called toehold switches that activate gene expression in response to cognate RNAs with arbitrary sequences. Toehold switches provide a high level of orthogonality and can be forward engineered to provide average dynamic range above 400. We show that switches can be integrated into the genome to regulate endogenous genes and use them as sensors that respond to endogenous RNAs. We exploit the orthogonality of toehold switches to regulate 12 genes independently and to construct a genetic circuit that evaluates 4-input AND logic. Toehold switches, with their wide dynamic range, orthogonality, and programmability, represent a versatile and powerful platform for regulation of translation, offering diverse applications in molecular biology, synthetic biology, and biotechnology.

## INTRODUCTION

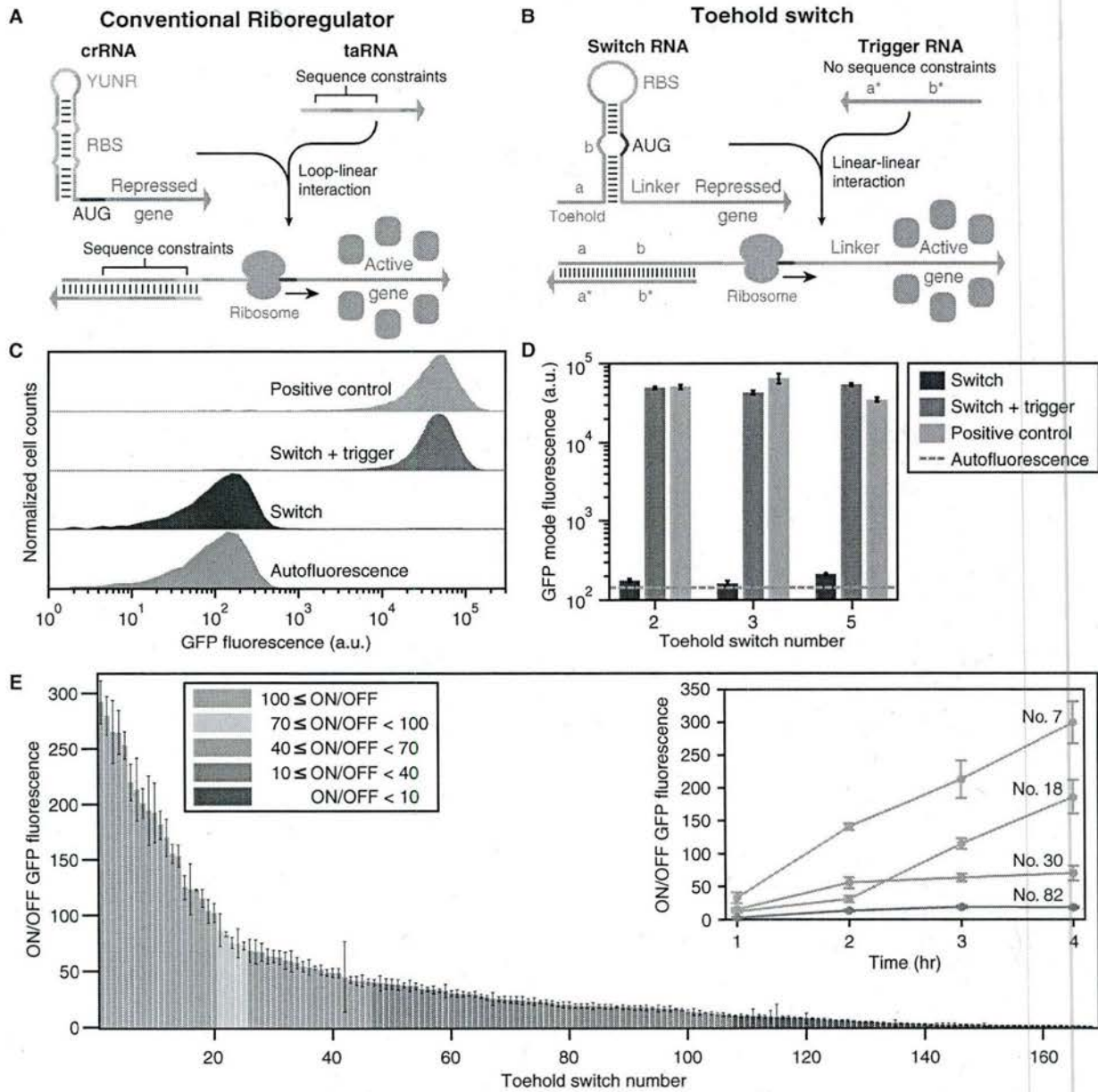
Synthetic biology seeks to apply engineering design principles to predict and control the behavior of living systems. Synthetic gene networks have been used to construct a wide range of biological devices, including molecular counters (Friedland et al., 2009), oscillators (Elowitz and Leibler, 2000), toggle switches (Gardner et al., 2000), logic gates (Ausländer et al., 2012; Moon et al., 2012; Siuti et al., 2013; Win and Smolke, 2008), cell classifiers (Xie et al., 2011), and analog signal processors (Daniel et al., 2013). Despite these developments, an underlying problem in synthetic biology remains the limited number of composable, high-performance parts for constructing genetic circuits and difficulties that arise when integrating multiple components into a large, complex synthetic network (Purnick and Weiss, 2009). Unlike electronic circuit elements, which can be electrically and spatially insulated from each other, biological components can

interact with one another in the complex cellular environment and suffer from unwanted crosstalk between components. Limitations imposed by the number and orthogonality of available biological circuit components hinder the construction of more complex circuits that can operate robustly in living cells. Consequently, new classes of regulatory components that offer wide dynamic range, low system crosstalk, and design flexibility represent a much-needed, enabling step toward fully realizing the potential of synthetic biology in areas such as biotechnology and medicine (Khalil and Collins, 2010).

RNA-based regulatory elements offer a potential solution to this component bottleneck. Biological parts constructed from RNA take advantage of predictable Watson-Crick base pairing to control cellular behavior and can harness sophisticated software tools for predicting RNA-RNA interactions. Nature has already developed a wide assortment of RNA-based parts operating at the transcriptional and posttranscriptional level (Brantl and Wagner, 2000; Gulyaev et al., 1997; Winkler et al., 2002). Starting from these natural systems, researchers have developed a number of engineered RNA regulatory elements, including riboregulators that control translation and transcription in response to cognate RNAs (Callura et al., 2010, 2012; Isaacs et al., 2004; Lucks et al., 2011; Mutalik et al., 2012; Rodrigo et al., 2012), and mRNA transducers that convert signals from small-molecule or protein ligands to protein outputs (Bayer and Smolke, 2005; Culler et al., 2010; Qi et al., 2012; Win and Smolke, 2008).

Engineered riboregulators consist of cognate pairs of RNAs: a transducer strand that regulates translation or transcription and a *trans*-acting RNA that binds to the transducer to modulate its biological activity. Riboregulator designs can be classified according to the initial RNA-RNA interaction that drives hybridization between the transducer and *trans*-acting RNAs. Reactions initiated between loop sequences in both RNAs are termed loop-loop interactions, whereas those that occur between a loop sequence and an unstructured RNA are termed loop-linear (Takahashi and Lucks, 2013).

A common limitation for riboregulators has been their dynamic range (Liu et al., 2012). Previous prokaryotic translational riboregulators have typically modulated biological signals by up to a maximum of ~55-fold for activators (Callura et al., 2012) and up to ~10-fold for repressors (Mutalik et al., 2012). In contrast, protein-based transcriptional regulators have demonstrated dynamic ranges over an order of magnitude higher, with



**Figure 1. Toehold Switch Design and In Vivo Characterization**

(A and B) Design schematics of conventional riboregulators (A) and toehold switches (B). Variable sequences are shown in gray, whereas conserved or constrained sequences are represented by different colors.

(A) Conventional riboregulator systems repress translation by base pairing directly to the RBS region. RNA-RNA interactions are initiated via a loop-linear or loop-loop interaction at the YUNR loop in an RNA hairpin.

(B) Toehold switches repress translation through base pairs programmed before and after the start codon (AUG), leaving the RBS and start codon regions completely unpaired. RNA-RNA interactions are initiated via linear-linear interaction domains called toeholds. The toehold domain a binds to a complementary a\* domain on the trigger RNA. Domains a and b are 12 and 18 nts, respectively.

(C) Flow cytometry GFP fluorescence histograms for toehold switch number 2 compared to *E. coli* autofluorescence and a positive control. Autofluorescence level measured from induced cells not bearing a GFP-expressing plasmid.

(D) GFP mode fluorescence levels measured for switches in their ON and OFF states in comparison to positive control constructs and autofluorescence. Error bars are the SD from at least three biological replicates.

(legend continued on next page)

widely-used inducible promoters regulating protein expression over 350-fold (Lutz and Bujard, 1997) and sigma factor-promoter pairs providing up to 480-fold modulation (Rhodius et al., 2013). Despite the inherent programmability of RNA-based systems, efforts at constructing large sets of orthogonal riboregulators have been limited to libraries of at most seven parts with crosstalk levels of ~20% (Takahashi and Lucks, 2013). Typical RNA-based regulators employ interaction domains consisting of ~30 nts, which corresponds to a sequence space of over  $10^{18}$  potential regulatory elements. Thus, the sheer diversity of possible RNA-based parts suggests that previous devices have not come close to realizing the potential of highly orthogonal regulation.

Much of this discrepancy arises from the significant sequence constraints imposed on riboregulators engineered thus far (Figure 1A). Like natural riboregulators, engineered riboregulators of translation have invariably used base pairing to the ribosome binding site (RBS) to prevent ribosome binding, thereby preventing translation (Callura et al., 2012; Isaacs et al., 2004; Mutalik et al., 2012; Rodrigo et al., 2012). Because repression is caused by RBS binding, trigger RNAs that activate translation are engineered to contain an RBS sequence to displace the repressing sequence, which in turn reduces the potential sequence space for the riboregulator.

Previous riboregulators have also relied on U-turn loop structures to drive loop-loop and loop-linear interactions between RNAs (Figure 1A) (Callura et al., 2012; Isaacs et al., 2004; Lucks et al., 2011; Takahashi and Lucks, 2013). U-turn loops are common RNA structural motifs formed by tertiary interactions that have been identified in ribozymes, ribosomal RNAs, and transfer RNA anticodon loops (Gutell et al., 2000). Although recent work has begun to show that loops with canonical U-turn sequences are not essential for riboregulators (Mutalik et al., 2012; Rodrigo et al., 2012), the engineered systems reported to date have continued their reliance on the loop-mediated RNA interactions from natural systems. Although these loop interactions have been selected by evolution in nature, alternative approaches employing linear-linear RNA interactions are amenable to rational engineering and exhibit more favorable reaction kinetics and thermodynamics, factors that could be exploited to increase riboregulator dynamic range.

To address these fundamental limitations, we designed a class of de-novo-designed riboregulators that enable posttranscriptional activation of protein translation through mechanisms employed in artificial systems rather than natural ones. Based on their interaction mechanism and near-digital signal processing behavior, we term these riboregulator systems toehold switches. Unlike conventional riboregulators, our synthetic riboregulators take advantage of toehold-mediated linear-linear interactions developed *in vitro* (Dirks and Pierce, 2004; Yin et al., 2008; Yurke et al., 2000) to initiate RNA-RNA strand displacement interactions. Furthermore, they rely on sequestration of the region

around the start codon to repress protein translation, eschewing any base pairing to the RBS or start codon itself to regulate translation. As a result, toehold switches can be designed to activate protein translation in response to a trigger RNA with an arbitrary sequence, enabling substantial improvements in component orthogonality. The absence of binding to the RBS and use of thermodynamically favorable linear-linear interactions also enables facile tuning of translational efficiency via RBS engineering. Consequently, these systems routinely enable modulation of protein expression over two orders of magnitude.

Here, we demonstrate the utility of toehold switches by validating more than 100 functional systems in *E. coli*. We exploit the expanded RNA sequence space afforded by the new riboregulator class to construct libraries of components with unprecedented orthogonality, including a set of 26 systems that exhibit less than 12% crosstalk. Heuristic design principles are used to generate forward-engineered switches that exhibit average ON/OFF ratios exceeding 400, a dynamic range typically reserved for protein-based transcriptional regulators. We also use toehold switches to sense endogenous RNAs *in vivo* and to regulate endogenous gene expression by integrating them into the genome. For potential applications in synthetic biology, we demonstrate that toehold switches can be used to regulate a dozen components in the cell at the same time and to construct a genetic circuit that computes a 4-input AND expression. The versatility, dynamic range, orthogonality, and programmability of toehold switches provide them with the potential to become a powerful platform for sensing and programming the internal states of living cells. We anticipate that these devices will be enabling tools for constructing increasingly complex genetic circuits in synthetic biology.

## RESULTS

### Design of a Biological Device from First Principles

A riboregulator that activates gene expression must switch from a secondary structure that prevents translation to a configuration that promotes translation upon binding of a cognate *trans*-acting RNA. We developed toehold switches based on previous reports describing the influence of secondary structure on mRNA translation and knowledge of RNA thermodynamics and kinetics.

Although the Shine-Dalgarno sequence is an important factor in determining the efficiency of translation from a given mRNA, studies have found that secondary structure in regions nearby the start codon also plays a critical role (Kudla et al., 2009). Furthermore, genome-wide analyses have revealed strong biases toward low secondary structures around the start codon of mRNAs from a panel of hundreds of bacterial genomes (Bentley et al., 2013). Thus, we ensured that toehold switches sequestered the region around the start codon to repress translation, rather than binding to either the RBS or the start codon.

(E) ON/OFF GFP fluorescence levels obtained 3 hr after induction for 168 first-generation toehold switches. Inset: ON/OFF GFP fluorescence measured for toehold switches of varying performance levels at different time points following induction. Relative errors for the switch ON/OFF ratios were obtained by adding the relative errors of the switch ON and OFF state fluorescence measurements in quadrature. Relative errors for ON and OFF states are from the SD of at least three biological replicates.

See also Figure S1 and Table S1.

Instead of using loop regions to initiate interactions, we recognized the design advantages afforded by linear-linear nucleic acid interaction strategies developed *in vitro* and focused in particular on toehold-based strand displacement reactions (Yurke et al., 2000). In these systems, interactions between reaction species are kinetically controlled through hairpins or multistranded complexes that feature exposed single-stranded domains called toeholds. These domains serve as reaction initiation sites for input nucleic acid species and do not require U-turn structural motifs to increase their accessibility. Such toehold mechanisms have been extensively used to engineer complex dynamic systems utilizing reconfigurable DNA hairpins in test tubes (Dirks and Pierce, 2004; Yin et al., 2008) and RNA hairpins in fixed tissues (Choi et al., 2010) and for complex *in vitro* information processing systems (Qian and Winfree, 2011; Qian et al., 2011; Yin et al., 2008).

Toehold switch systems are composed of two RNA strands referred to as the switch and trigger (Figure 1B). The switch RNA contains the coding sequence of the gene being regulated. Upstream of this coding sequence is a hairpin-based processing module containing both a strong RBS and a start codon that is followed by a common 21 nt linker sequence coding for low-molecular-weight amino acids added to the N terminus of the gene of interest. A single-stranded toehold sequence at the 5' end of the hairpin module provides the initial binding site for the trigger RNA strand. This trigger molecule contains an extended single-stranded region that completes a branch migration process with the hairpin to expose the RBS and start codon, thereby initiating translation of the gene of interest.

The hairpin processing unit functions as a repressor of translation in the absence of the trigger strand. Unlike previous riboregulators, the RBS sequence is left completely unpaired within the 11 nt loop of the hairpin. Instead, the bases immediately before and after the start codon are sequestered within RNA duplexes that are 6 bp and 9 bp long, respectively. The start codon itself is left unpaired in the switches we tested, leaving a 3 nt bulge near the midpoint of the 18 nt hairpin stem. Because the repressing domain b (Figure 1B) does not possess complementary bases to the start codon, the cognate trigger strand in turn does not need to contain corresponding start codon bases, thereby increasing the number of potential trigger sequences. The sequence in the hairpin added after the start codon was also screened for the presence of stop codons, as they would prematurely terminate translation of the gene of interest when the riboregulator was activated. We employed a 12 nt toehold domain at the 5' end of the hairpin to initiate its interaction with the cognate trigger strand. The trigger RNA contains a 30 nt single-stranded RNA sequence that is complementary to the toehold and stem of the switch RNA.

#### **In Silico Design of Toehold Switches**

We used the NUPACK nucleic acid sequence design package (Zadeh et al., 2011) to generate libraries of de-novo-designed translational activators satisfying the desired device parameters (see section S6 of the Extended Experimental Procedures and Figure S1C available online). We initially designed a set of 24 toehold switches (see Table S1 for RNA sequences) to gauge *in vivo* performance.

We then designed an extended library of toehold switches containing orthogonal elements selected for low crosstalk with the rest of the library. The orthogonal library was generated by assembling a set of 646 unique toehold switch designs and simulating all 417,316 pairwise interactions in the set to evaluate crosstalk interactions (see section S6.4 of the Extended Experimental Procedures and Figure S1D). A Monte Carlo algorithm was used to identify a subset of 144 switches that exhibited the lowest level of overall component crosstalk (see Table S1 for RNA sequences). The resulting library of orthogonal regulators provided a large set of components to independently regulate translation *in vivo* and comprised devices well separated from one another in the sequence space to provide valuable information regarding sequence-dependent effects.

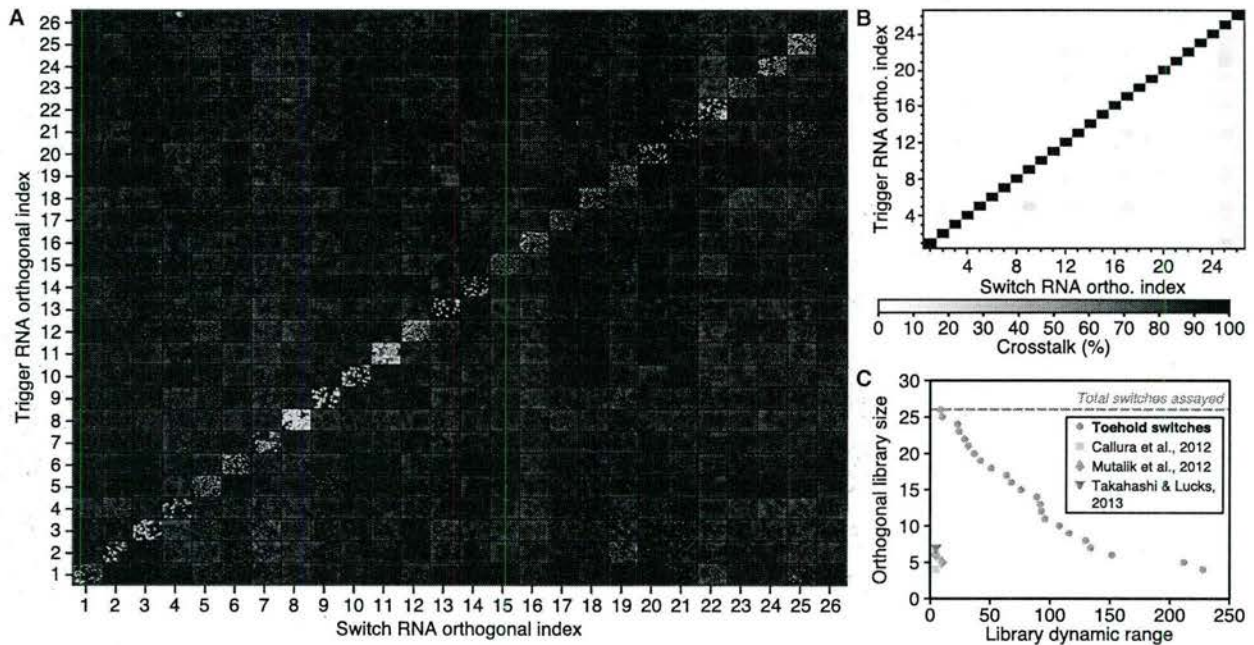
#### **In Vivo Component Validation**

The toehold switches were tested in *E. coli* BL21 Star DE3 with the switch and trigger RNAs expressed from separate medium and high copy plasmids, respectively. Although BL21 Star DE3 is an RNase-deficient strain, we found that toehold switches also performed well in strains with wild-type RNase levels (see section S10 of the Extended Experimental Procedures, Figure S2C, and Table S3). Expression of both strands was induced using IPTG, which triggered production of both RNA species through T7 RNA polymerase. GFP was used to characterize switch output performance via flow cytometry.

Representative flow cytometry histograms of GFP output from toehold switch number 2 are shown in Figure 1C. ON state fluorescence measured from cells expressing both the switch and its cognate trigger was near the fluorescence intensity of control constructs with unexpressed GFP expression. OFF state fluorescence from cells expressing the switch and a noncognate trigger revealed a slight increase in GFP levels over *E. coli* autofluorescence. The mode fluorescence value from these histograms was used to calculate the ON/OFF ratios of each toehold switch design (see section S3 of the Extended Experimental Procedures and Figure S1B). Cell autofluorescence was not subtracted from either the ON or OFF state fluorescence for determination of ON/OFF ratios.

Figure 1D presents the mode GFP fluorescence obtained from three high-performance toehold switches in their ON and OFF states. The OFF state fluorescence of the three switches was near the background autofluorescence levels measured for induced cells not expressing GFP. The ON state fluorescence for the toehold switches was comparable to the fluorescence obtained from positive controls that recapitulated the secondary structure of the toehold switch in its activated state (see section S7 of the Extended Experimental Procedures, Figure S1E, and Table S1). The activated toehold switches all showed higher GFP output than a standard GFP expression construct featuring a completely single-stranded region upstream of the RBS and gene, likely as a result of the increased mRNA stability afforded by the secondary structures of the switch RNA and the trigger/switch RNA complex.

Of the 168 first-generation systems tested, 20 exhibit ON/OFF ratios exceeding 100, and nearly two thirds display at least an ON/OFF level greater than 10 (Figure 1E). In comparison, we also characterized the widely used engineered riboregulators



**Figure 2. Assessment of Toehold Switch Orthogonality**

(A) GFP fluorescence from colonies of *E. coli* expressing 676 pairwise combinations of switch and trigger RNAs. GFP-expressing colonies are visible as green points along the diagonal in cells containing cognate switch and trigger strands. Off-diagonal, noncognate components have low fluorescence. (B) Crosstalk measured by flow cytometry for all trigger-switch combinations. Crosstalk was determined by taking GFP output measured for a given trigger-switch combination and dividing it by the GFP output measured for the switch with its cognate trigger. (C) Comparison of overall library dynamic range and orthogonal library size for the toehold switches and previous riboregulators. The overall library dynamic range corresponds to the minimum ON/OFF ratio to expect in a network employing this library of switches. See also Table S2.

(Isaacs et al., 2004) crRNA 10 and 12 in identical conditions. These earlier riboregulators exhibited lower dynamic range with ON/OFF values of  $11 \pm 2$  and  $13 \pm 4$  for crRNA systems 10 and 12, respectively (see section S8 of the Extended Experimental Procedures and Figure S1F). Time-course measurements revealed system activation within 1 hr of induction and ON/OFF ratios increased over time with continued production of GFP (Figure 1E, inset). The switches were also successful in regulating four different output proteins (see section S10.1 of the Extended Experimental Procedures and Figure S1H).

#### Evaluation of Toehold Switch Orthogonality

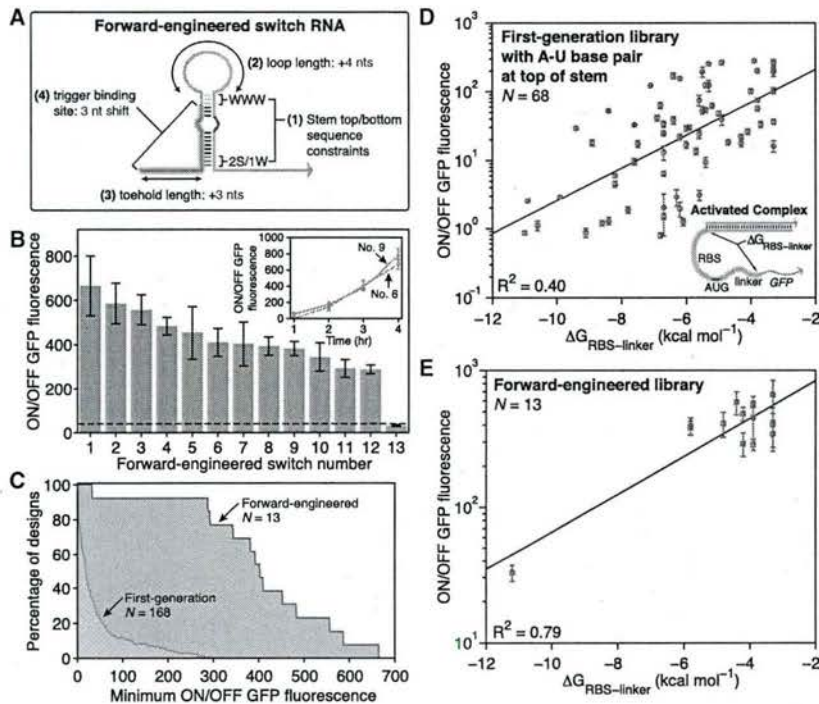
To measure the orthogonality of the devices, we performed additional in silico screening to isolate a subset of 26 that displayed extremely low levels of predicted crosstalk (see section S6.4 of the Extended Experimental Procedures) and assayed in *E. coli* all 676 pairwise combinations of switch and trigger plasmids. Figure 2A displays images of GFP fluorescence from colonies of *E. coli* induced on LB plates (see section S5 of the Extended Experimental Procedures). Clearly visible is the strong emission from cognate switch and trigger pairs along the diagonal of the grid with the final switch at index 26 displaying lower fluorescence as a result of its low ON/OFF ratio. Low fluorescence levels are observed for the off-diagonal elements featuring noncognate trigger/switch RNA pairs.

We used flow cytometry to quantify GFP output from all trigger-switch interactions (see section S4 of the Extended Experimental Procedures). Crosstalk was calculated by dividing the GFP fluorescence obtained from a noncognate trigger and a given switch RNA by the fluorescence of the switch in its triggered state (Figure 2B). The full set of 26 switches tested displayed less than 12% crosstalk. Because the number of switches in an orthogonal set is defined by its threshold crosstalk level, we identified orthogonal subsets for a range of different crosstalk thresholds (see section S11 of the Extended Experimental Procedures and Table S2). For instance, a subset of 18 of the toehold switches exhibited less than 2% subset-wide crosstalk.

A relevant metric for assessing orthogonal library performance is the reciprocal of the threshold crosstalk level. For example, the set of 18 switches with less than 2% crosstalk has a reciprocal crosstalk threshold of 50. This metric corresponds to the minimum output dynamic range to expect in a network employing this library of 18 switches. Figure 2C plots this library dynamic range metric against the maximum orthogonal subset size for the toehold switches, as well as a number of previous riboregulator systems.

#### Forward Engineering of New High-Performance Devices

In-silico-designed riboregulators forward engineered to exhibit high performance in vivo have the potential to significantly



**Figure 3. Forward Engineering and Thermodynamic Analysis of Toehold Switches**

(A) Schematic of the design modifications made for the forward-engineered switches compared to the first-generation switches. (B) ON/OFF GFP fluorescence ratios obtained for the set of 13 forward-engineered toehold switches after 3 hr of induction. Dashed black line marks the mean ON/OFF fluorescence measured for the 168 first-generation toehold switches. Inset: time course measurements for forward-engineered switches number 6 and number 9. Relative errors for the switch ON/OFF ratios were obtained by adding the relative errors of the switch ON and OFF state fluorescence measurements in quadrature. Relative errors for ON and OFF states are from the SD of at least three biological replicates. (C) Percentage of first-generation and forward-engineered library components that had ON/OFF ratios that exceeded the value defined on the x axis. (D) Correlation between  $\Delta G_{\text{RBS-linker}}$  and ON/OFF ratio measured for the 68 first-generation toehold switches that had an A-U base pair at the top of the hairpin stem. Inset: Schematic showing the RNA sequence range in the trigger-switch complex used to define  $\Delta G_{\text{RBS-linker}}$ . Relative errors for the switch ON/OFF ratios were obtained by adding the relative errors of the switch ON and OFF state fluorescence measurements in quadrature. Relative errors for ON and OFF states are from the SD of at least three biological replicates.

(E) Strong correlation ( $R^2 = 0.79$ ) between  $\Delta G_{\text{RBS-linker}}$  and ON/OFF ratio measured for the complete set of forward-engineered switches. Relative errors for the switch ON/OFF ratios were obtained by adding the relative errors of the switch ON and OFF fluorescence measurements in quadrature. Relative errors for ON and OFF states are from the SD of at least three biological replicates. See also Figure S2 and Table S3.

reduce the time required for generating new genetic circuits. The first-generation toehold switches fell short of this goal because they displayed large variations in output characteristics (Figure 1E). To determine the design features that yielded high-performance switches, we performed detailed analyses of the first-generation library to uncover sequence-dependent factors that affected switch performance (see sections S9 and S12 of the Extended Experimental Procedures; Figures S1G and S2A; Table S1). We integrated these findings into the design of second-generation toehold switches forward engineered for wide dynamic range.

The forward-engineered switches incorporate four design changes as shown in Figure 3A (see Figure S1A for direct comparisons). Modifications were made to the sequences at the top of the switch RNA stem, the loop of the switch RNA was increased, and the binding site of the trigger RNA was shifted away from the RBS. These three design changes increased the strength of the switch RNA RBS to increase its ON state protein production. The length of the toehold domain was also increased to promote binding between switch and trigger RNAs.

We designed and evaluated a set of 13 of these second-generation toehold switches (see Table S3 for RNA sequences). Figure 3B presents the ON/OFF fluorescence ratios for the forward-engineered systems regulating GFP after 3 hr of induction (see Figure S2B for flow cytometry data). There is a dramatic increase in ON/OFF fluorescence for almost all the systems

tested, with 12 out of 13 switches exhibiting a dynamic range comparable to or higher than the highest performance toehold switch from the initial library (ON/OFF ratio  $290 \pm 20$ ). These systems exhibit an average ON/OFF ratio of 406 compared to 43 for the first-generation design. This mean ON/OFF ratio rivals the dynamic range of protein-based regulators, and it does so using a highly programmable system design without requiring any evolution or large-scale screening experiments. Furthermore, even the lowest performance second-generation switch displayed an ON/OFF ratio of  $33 \pm 4$ , which is still sufficient for many logic operations. Hourly time course measurements reveal activation of forward-engineered switches after 1 or 2 hr of induction (Figure 3B, inset). Furthermore, ON state fluorescence increased steadily over 4 hr, yielding ON/OFF levels well over 600 for the switches. Additional testing confirmed successful forward-engineered switch operation in different strains of *E. coli* and with the endogenous *E. coli* RNA polymerase (see section S10 of the Extended Experimental Procedures, Figure S2C, and Table S3).

We quantified the effectiveness of our rational engineering strategy by calculating the percentage of forward-engineered designs with ON/OFF ratios exceeding a given minimal level and comparing them to the same evaluation performed on the library of 168 first-generation toehold switches (Figure 3C). The yield of high-performance switches is higher for the forward-engineered devices for all ON/OFF ratios tested.

### Thermodynamic Parameter Correlated with Switch Performance

We conducted extensive analyses to identify thermodynamic parameters that were linked to the ON/OFF performance of the toehold switches. Candidate thermodynamic factors were rapidly screened by computing linear regressions against subsets of the first-generation toehold switches satisfying particular sequence constraints (see section S13 of the Extended Experimental Procedures, Figure S2D, and Table S3). The screening process identified a single parameter  $\Delta G_{\text{RBS-linker}}$  that was correlated with the ON/OFF ratios for the subset of 68 first-generation switches with an A-U base pair at the top of their stem. A linear fit of  $\Delta G_{\text{RBS-linker}}$  to the logarithm of their ON/OFF ratios yielded a coefficient of determination  $R^2 = 0.40$  (Figure 3D). This  $\Delta G_{\text{RBS-linker}}$  term is defined as the free energy of the sequence running from the RBS region to the end of the linker (Figure 3D, inset). It reflects the amount of energy required by the ribosome to unwind the RBS and early-mRNA region as it binds and begins translation of the output gene. A fit of  $\Delta G_{\text{RBS-linker}}$  to the ON/OFF ratio for the forward-engineered switches yielded a much stronger correlation with  $R^2 = 0.79$  (Figure 3E). Importantly, we found that this single thermodynamic factor could explain the lone low-performance, forward-engineered toehold switch.

### Toehold Switches Triggered by mRNAs and Endogenous RNAs

The capacity of toehold switches to accept trigger RNAs with arbitrary sequences enables them, in principle, to be activated by functional mRNAs (Figure 4A). However, the fixed sequences of potential mRNA triggers present significant challenges for effective system activation. Unlike synthetic trigger RNAs designed de novo to be single stranded, strong secondary structures abound within mRNAs, complicating toehold binding and decreasing the thermodynamics driving the branch migration process. The cognate toehold sequences defined by the trigger mRNA can also exhibit base pairing both internally and with sequences downstream of the hairpin module and thus pose similar challenges to activation.

To counter these effects, we increased the toehold domain length of the mRNA-sensing switches from 12 or 15 nts to  $\geq 24$  nts. This modification shifted the importance of single-stranded regions for binding initiation from the trigger mRNA to the toehold switch itself. We also incorporated a common sequence element from a high-performance first-generation switch into the sensor designs (see section S14.1 of the Extended Experimental Procedures, Figures S3A–S3C, and Table S4) and adopted a programmed RNA refolding mechanism to decrease the energetic barrier to switch activation (Figures 4A and S3D; see section S14.2 of the Extended Experimental Procedures). Candidate sensors for a given mRNA or endogenous RNA were designed and screened in silico to identify those with the greatest probability of success (see section S14.3 of the Extended Experimental Procedures, and Table S4 for RNA sequences).

We first validated sensors for detecting the *mCherry* mRNA. These sensors express GFP only upon binding to *mCherry* transcripts enabling simultaneous monitoring of sensor output via GFP and transcription of the trigger species through *mCherry* fluorescence. Figure 4B displays the GFP and *mCherry* fluores-

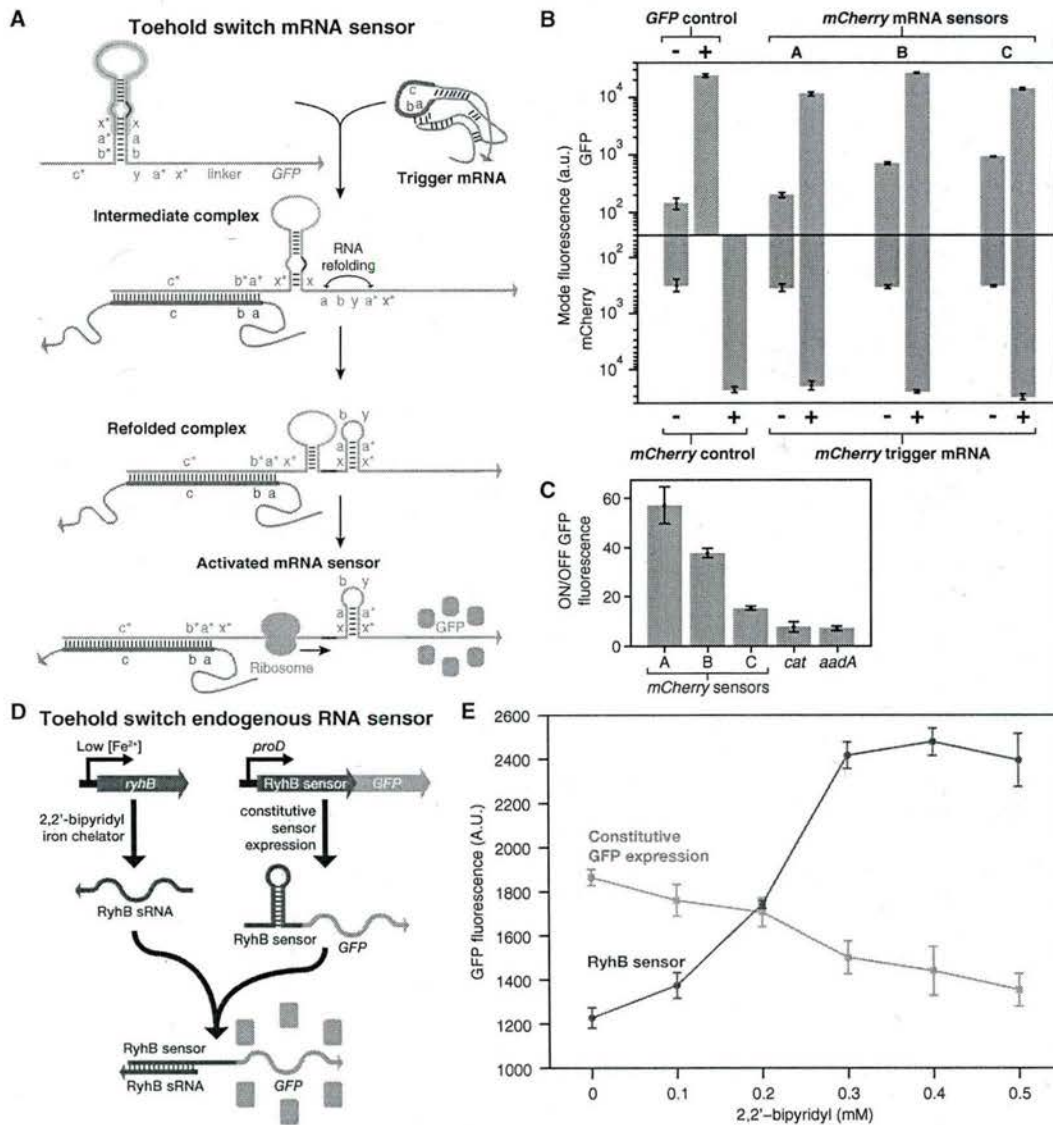
cence levels measured for three *mCherry*-responsive switches in the absence or presence of their cognate trigger mRNA. Fluorescence levels from control measurements are also shown for comparison. Negative controls reflect *E. coli* autofluorescence measured in either fluorescence channel, whereas positive controls were obtained from cells expressing unrepressed GFP and *mCherry* from medium and high copy vectors, respectively. These results show that the switches strongly activate expression of GFP only upon transcription of the *mCherry* mRNA and provide low GFP expression in the absence of the *mCherry* trigger. ON/OFF GFP fluorescence values obtained from the three *mCherry* mRNA sensors are shown in Figure 4C. In addition, we validated two sensors for detecting antibiotic resistance conferring mRNAs: chloramphenicol acetyltransferase (*cat*) and *aadA* (spectinomycin resistance) (Figure 4C).

We also designed a toehold switch sensor to detect the endogenous *E. coli* small RNA (sRNA) RyhB. RyhB is a 90 nt transcript that downregulates iron-associated genes in conditions in which iron levels are low (Massé and Gottesman, 2002). To characterize the sensors, cells were transformed with plasmids constitutively expressing the RyhB-responsive toehold switch regulating GFP (Figure 4D; see section S15 of the Extended Experimental Procedures and Table S4). The RyhB sRNA was induced by adding the iron-chelating compound 2,2'-bipyridyl to the growth media (Figure 4D), which is known to rapidly stimulate expression of RyhB (Massé and Gottesman, 2002). We measured sensor output 1 hr after induction by the chelator using flow cytometry. The sensor transfer function shows a steady increase in GFP expression as 2,2'-bipyridyl levels increase to 0.3 mM, beyond which levels plateau (Figure 4E). In contrast, a GFP-positive control construct demonstrated decreasing output as 2,2'-bipyridyl levels increased.

### Synthetic Regulation of Endogenous Genes

Toehold switches can be integrated into the genome to regulate translation of endogenous genes. Template genome-editing plasmids were constructed that contained a high-performance second-generation switch adjacent to a kanamycin resistance marker flanked by a pair of FRT sites (Figure 5A; see section S16 of the Extended Experimental Procedures, Table S5, and Figure S4). Linear DNA fragments were amplified from these plasmids and were inserted upstream of targeted chromosomal genes using  $\lambda$  Red recombination (Datsenko and Wanner, 2000). The resulting *E. coli* strain retains a functional copy of the targeted gene in its chromosome; however, it is deactivated as a result of the cotranscribed switch RNA module. This repressed gene can be activated posttranscriptionally by expression of a cognate trigger.

We inserted switches upstream of the genes *uidA*, *lacZ*, and *cheY*. The genes *uidA* and *lacZ* produce the enzymes  $\beta$ -glucuronidase and  $\beta$ -galactosidase, respectively. Cells expressing these enzymes can be readily identified by their blue/green color on plates containing the corresponding substrates X-Gluc and X-Gal. We constructed two strains with synthetic *uidA* regulation by integrating switches A and B into the chromosome (*uidA*::Switch A and *uidA*::Switch B, respectively). Figure 5B displays these strains upon expression of different trigger RNAs, as well as a control strain with the wild-type *uidA* genotype. As



**Figure 4. Toehold Switch Activated by mRNA and Endogenous Small RNA Triggers**

(A) Design schematic and putative activation pathway of the toehold switch mRNA sensors. Switch common sequence element is outlined in pink.

(B) Mode GFP and mCherry fluorescence obtained from flow cytometry of three mCherry sensors in their repressed and activated states, as well as positive and negative controls. Error bars are the SD from at least three biological replicates.

(C) ON/OFF GFP fluorescence ratios for a series of toehold switches activated by the mCherry mRNA, and *cat* and *aadA* mRNAs, which confer chloramphenicol and spectinomycin resistance, respectively. Relative errors for the mRNA sensor ON/OFF ratios were obtained by adding the relative errors of the sensor ON and OFF state fluorescence measurements in quadrature. Relative errors for ON and OFF states are from the SD of at least three biological replicates.

(D) Endogenous and synthetic gene networks used for sensing the RyhB sRNA.

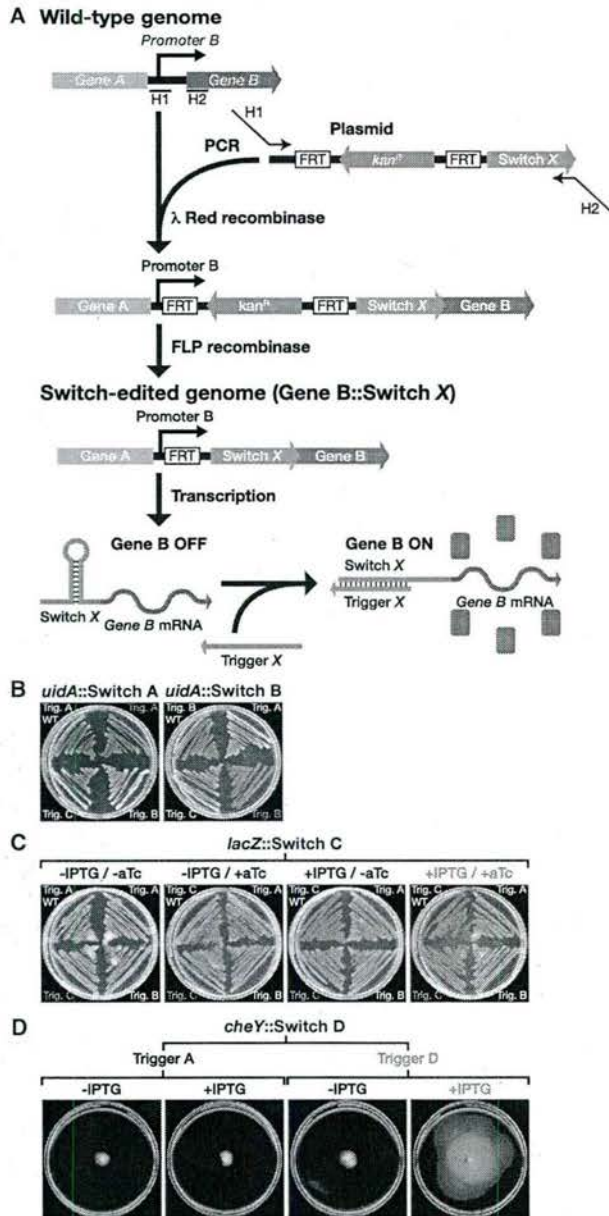
(E) Transfer function for the RyhB sensor (purple curve) as a function of RyhB inducer concentration. Output of a constitutive GFP expression cassette is shown for comparison (green curve). Error bars are the SD from at least three biological replicates.

See also Figure S3 and Table S4.

expected, *uidA*::Switch A only exhibits the blue/green wild-type phenotype upon expression of trigger A. Similarly, *uidA*::Switch B activates  $\beta$ -glucuronidase only with cognate trigger B.

The edited strain *lacZ*::Switch C provides more complicated behavior because the lac operon is regulated at the transcrip-

tional level by lactose or chemical analogs such as IPTG. Thus, *lacZ*::Switch C requires both lactose/IPTG and trigger RNA C to turn on expression of  $\beta$ -galactosidase. This behavior results in a genetic AND circuit combining transcriptional and posttranscriptional regulation. We tested this AND circuit by expressing



**Figure 5. Synthetic Regulation of Endogenous Genes**

(A) Integration of switch modules into the genome upstream of the targeted gene (gene B) at sites H1 and H2 using  $\lambda$  Red recombination. Switch-edited gene B is translationally repressed but can be activated via the cognate trigger RNA. (B) Images of *uidA*::Switch A and *uidA*::Switch B spread onto X-Gluc plates with different trigger RNAs. *uidA* expression like the wild-type (top left) is only observed with cognate trigger RNAs as seen by blue/green color change. (C) Images of *lacZ*::Switch C with different combinations of IPTG and aTc chemical inducers. *lacZ*::Switch C only activates with aTc-induced expression of trigger C in conditions in which *lacZ* transcription is induced by IPTG. Wild-type *lacZ* (top left) is activated whenever IPTG is present. (D) Motility assays for *cheY*::Switch D on soft agar plates. *cheY*::Switch D is only able to move away from the point of inoculation at the plate center when trigger D is induced with IPTG.

See also Figure S4 and Table S5.

different trigger RNAs using inducible promoters responsive to anhydrotetracycline (aTc). Figure 5C provides images of *lacZ*::Switch C transformed with different trigger plasmids for all four combinations of the two chemical inputs (i.e., IPTG and aTc) of the AND circuit. In the absence of IPTG, none of the strains show any change in color because expression of the *lac* operon is strongly repressed. For a plate containing IPTG and no aTc, the wild-type *lacZ* strain (upper left quadrant) becomes blue/green in color, whereas the *lacZ*::Switch C strains with different triggers do not change in color because the trigger RNAs are not being expressed. When the AND condition is satisfied by adding both IPTG and aTc to the plate, *lacZ*::Switch C exhibits the blue/green color change with trigger RNA C as expected, whereas those expressing triggers A and B are unchanged.

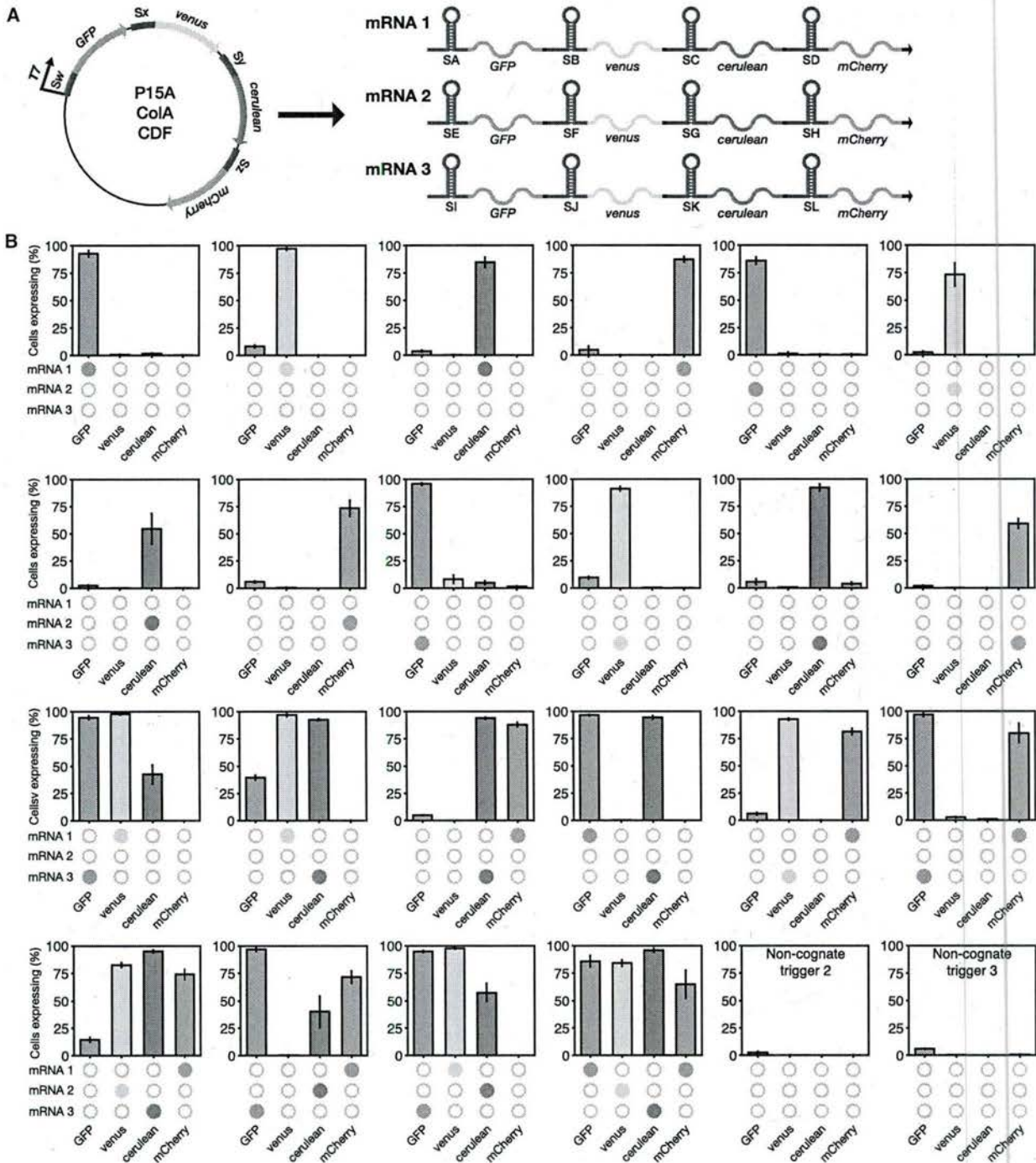
Lastly, we conditionally regulated the *E. coli* chemotaxis gene *cheY*. *CheY*::Switch D was transformed with plasmids that expressed triggers inducibly via IPTG. Only cells expressing trigger RNA D with IPTG demonstrated significant motility, whereas those expressing noncognate trigger RNA A or lacking the IPTG inducer were unable to move from the point of inoculation (Figure 5D).

### Multiplexed Regulation

To demonstrate the full multiplexing capabilities of toehold switches, we expressed 12 toehold switches in the same cell and independently confirmed their activity via flow cytometry. We used four different fluorescent proteins (GFP, venus, cerulean, and mCherry) as reporters and constructed three plasmids to express each of the reporter proteins (Figure 6A; see section S17 of the Extended Experimental Procedures; Table S6). Each of these proteins was regulated using a different switch from the second-generation library. The resulting constructs were expressed from a single T7 promoter as  $\sim 3.4$  kb polycistronic mRNAs resulting in a total synthetic network size of over 10 kb.

Figure 6B presents the outcome of the multiplexing experiments with the output of each reporter represented in terms of the percentage of cells expressing the reporter. A cell was determined to be expressing a given reporter if its fluorescence exceeded a threshold level held constant for all the plots in Figure 6B (see section S17.3 of the Extended Experimental Procedures and Figure S5). The first two rows of Figure 6B display the output from all 12 of the switches activated separately from a single expressed trigger RNA. In all 12 cases, significant expression is only observed from the intended reporter with limited crosstalk in the three other channels.

We also tested activation of all two- and three-color combinations of reporter proteins and observed all the expected output color combinations (Figure 6B). We detected some unintended leakage of cerulean and GFP for trigger combinations activating GFP/venus and venus/cerulean, respectively. Because single trigger measurements of each of these triggers displayed low leakage, we attribute much of the observed leakage in these two-trigger experiments to imperfect compensation of the flow cytometry data caused by the spectral overlap between GFP, venus, and cerulean (see Figure S5A). We also successfully activated all four output proteins by simultaneously expressing four trigger RNAs and observed low system leakage from two non-cognate trigger RNAs (Figure 6B, two bottom-right panels).

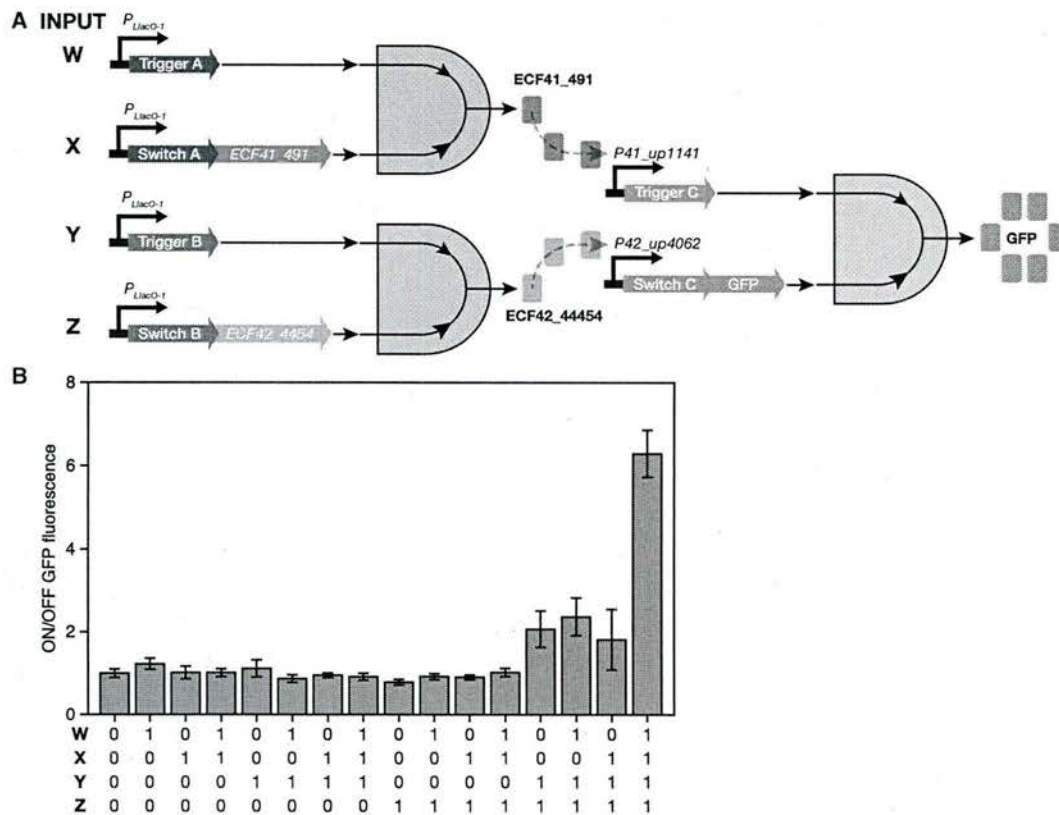


**Figure 6. Simultaneous Regulation of Gene Expression by 12 Toehold Switches**

(A) Schematics of plasmids and ~3.4-kb polycistronic mRNAs used for multiplexing studies. Each reporter has its own switch RNA that can be independently activated by its cognate trigger RNA.

(B) Percentage of cells expressing each of the four reporters for a set of 24 different trigger RNA combinations. Gray and colored circles are used to identify the particular trigger RNA being expressed and the corresponding switch RNA. Error bars are the SD from at least three biological replicates.

See also Figure S5 and Table S6.



**Figure 7. A Layered 4-Input AND Circuit**

(A) Design schematic for the 4-input AND circuit consisting of three 2-input AND gates formed by three toehold switches, two orthogonal transcription factors (ECF41\_491 and ECF42\_4454), and a GFP reporter.

(B) Complete 16-element truth table for the 4-input AND system. GFP expression from the sole logical TRUE output case with all input RNAs expressed (far right) is significantly higher than the logical FALSE output cases where one or more input RNAs is absent. Relative errors for the circuit ON/OFF ratios were obtained by adding the relative errors of the circuit fluorescence measurements in quadrature. The fluorescence level measured from cells not expressing any of the input RNAs was used as the OFF state fluorescence levels in the ON/OFF calculations. Relative errors for different input states are from the SD of at least three biological replicates.

See also Figure S6 and Table S7.

### Implementation of a 4-Input AND Circuit

Toehold switches are readily integrated with existing biological components to build sophisticated genetic programs. We constructed a layered 4-input AND gate consisting of three toehold switches coupled to two orthogonal transcription factors and a GFP reporter (Figure 7A). In this circuit, toehold switch RNA pairs act as two independent input species that must both be present before a 2-input logical AND expression evaluates as TRUE. The first computational layer in the circuit consists of two 2-input AND toehold switch gates, which each produce a transcription factor. A pair of extracytoplasmic function (ECF) sigma factors were used as the transcription factors in the circuit because they had previously been reported to be highly orthogonal (Rhodius et al., 2013). The sigma factors produced from the first layer then activate transcription of the toehold switch RNAs in the second computational layer, which in turn lead to expression of a GFP reporter. Similar layered circuits have previously been constructed using transcription factors

that required a second chaperone protein for full activity (Moon et al., 2012).

To validate the circuit, we constructed plasmids to express all 16 combinations of the four input trigger and switch RNAs (see section S18 of the Extended Experimental Procedures, Table S7, and Figure S6). The full truth table for the 4-input AND computation is shown in Figure 7B with ON/OFF levels calculated by normalizing GFP fluorescence to that measured for the case in which none of the input bits are expressed. GFP output when all four input bits are present is significantly higher than all other input permutations, as expected for a functional 4-input AND computation.

### DISCUSSION

#### Origins of Increased Toehold Switch Performance Compared to Previous Riboregulators

Toehold switches represent a versatile and powerful new platform for regulating translation at the posttranscriptional level.

They combine a high degree of component orthogonality with system dynamic range comparable to protein-based transcriptional regulators (Lutz and Bujard, 1997; Rhodius et al., 2013). Of note, the dynamic range of the toehold switches is significantly greater than that observed for previous engineered riboregulators. Earlier efforts to design riboregulators with higher performance have been hindered by three general misconceptions. (1) Because natural riboregulators have served as the starting point for new riboregulators, it has been implicitly assumed that engineered riboregulators cannot be substantially modified from their parent natural systems without abolishing their functionality. This approach employing small sequence perturbations parallels efforts in engineering new protein-based parts. However, it neglects the considerable advantage in *in silico* simulation and design available for engineering RNAs compared to proteins. This underlying assumption has in turn led to (2) an emphasis on loop-loop and loop-linear RNA interaction mechanisms and (3) the use of RNA binding to the RBS region to repress translation.

Although application of any of these criteria for engineered riboregulators does not necessarily preclude them from functioning well as biological parts, toehold switches demonstrate that each of these design rules is not required for high-performance devices. First, toehold switches do not have any direct natural riboregulator counterparts, and thus they have no natural analogs to narrow the sequence space of potential devices. Relieving this constraint enabled us to generate large libraries of potential devices and explore a more diverse array of sequences compared to previous riboregulators. Furthermore, this new approach allowed RNA sequences to be designed *de novo* with ideal secondary structures for optimal device performance. Thus, trigger RNA transcripts feature single-stranded domains for binding to the switch RNA, and the ensuing trigger-switch RNA complex is engineered to have low secondary structure nearby the start codon to encourage efficient translation by the ribosome to maximize the device ON state.

Second, toehold switch RNA interactions are mediated through linear-linear interactions rather than loop-mediated ones. These reactions can provide faster kinetics than loop-based reactions, and the long  $\geq 12$  nt toehold of the switch RNA provides stronger thermodynamics than the shorter loop domains of earlier riboregulators. These features cause a higher percentage of the total switch RNAs present in the cell to be triggered, thus increasing production of the output protein. We found that the fraction of activated switch RNAs was  $\sim 100\%$  based on comparison measurements with unrepresed versions of the switch RNA (Figures 2A and 2B). In contrast to earlier reports, thermodynamic analyses of toehold switch performance did not reveal significant correlations between riboregulator ON/OFF levels and the free energy of the switch-trigger interaction nor the free energy of toehold-trigger binding (Mutalik et al., 2012). These observations suggest that RNA-RNA interactions for the toehold switches are so strongly thermodynamically favored that other factors are presently limiting device performance in cells.

Third, toehold switches do not rely on direct binding to the RBS to repress translation. Instead, the RBS is enclosed within a loop, and repression is achieved by secondary structures

immediately before and after the start codon. This design feature increased the space of potential trigger RNA sequences, and it provided a straightforward means of enhancing ON state translation levels simply by adding A-rich sequences to the loop upstream of the RBS. Importantly, these modifications did not require corresponding changes to the trigger RNA, which enabled facile interpretation of the effects of modifications to this pre-RBS region. Taken together, all these design features enabled the toehold switch to detect arbitrary trigger RNAs and translate any output protein.

Our results suggest that existing riboregulators can be improved substantially by adopting some of the features of our toehold switches. The length of loop domains in loop-mediated interactions can be increased to improve reaction thermodynamics. Furthermore, the unpaired region upstream of the RBS in translational riboregulators can be lengthened to enhance protein output and, in turn, device dynamic range.

#### Constraints on Toehold Switch Trigger RNAs and Output Proteins

The designs we have chosen for the toehold switches place some restrictions on the sequences of trigger RNAs. Inclusion of the 3 nt bulge at the start codon AUG position of the switch RNA precludes the trigger RNA from having an AUG sequence at positions programmed to hybridize with this bulge. The switch RNA also possesses a 9 bp stem after the start codon, which must not code for an in frame stop codon. This restriction in turn imposes sequence conditions on the trigger, which must unwind this region of the switch RNA stem. The toehold switches also add 11 residues from the switch stem to the linker to the N terminus of the regulated protein, which could be problematic for an output protein sensitive to changes in its N terminus.

In practice, the conditions imposed on the trigger RNA and output protein can be avoided with a few modifications to the toehold switch design. Constraints on the trigger RNA are a by-product of the base-pairing conditions specified for the switch RNA stem and the trigger-switch complex. These particular secondary structures are not strictly required for switch operation. For instance, forward-engineered switch number 5 has only a 1 nt bulge in its stem but provides ON/OFF  $> 400$ . We expect that design modifications that add and subtract base pairs from the switch RNA will allow the toehold switches to modulate gene expression while providing sufficient design flexibility to eliminate stop and start codon constraints on the trigger sequence. The toehold switches can also be modified to incorporate the coding sequence of the output protein directly into the switch RNA stem, thus making them compatible with any protein sensitive to N-terminal modifications.

#### Engineered Regulators with Very Low System Crosstalk

The capacity of toehold switches to respond to diverse trigger RNAs enabled us to use *in silico* techniques to design large libraries of orthogonal components. Comprehensive evaluations of *in vivo* trigger-switch RNA pairwise interactions revealed a set of 26 toehold switches with 12% crosstalk levels. The largest previously reported orthogonal riboregulator set consisted of seven transcriptional attenuators displaying 20% crosstalk (Takahashi and Lucks, 2013). In this case, crosstalk level of

20% meant that the set of 42 off-target RNA sense-antisense interactions attenuated transcription by at most 20%. Such crosstalk results in an upper bound in overall library dynamic range of 5 (Figure 3C). Earlier orthogonal translational activators and repressors have been limited to sets of four (Callura et al., 2012) and six (Mutalik et al., 2012), respectively, at 20% crosstalk. For proteins, an engineered library of five orthogonal eukaryotic transcription factors with crosstalk of ~30% was reported (Khalil et al., 2012). Recently, part mining was used to engineer a set of four orthogonal chimeric sigma factors with ~10% crosstalk and a subset of three with ~2% crosstalk (Rhodius et al., 2013). Thus, to our knowledge, these toehold switches constitute the largest set of engineered orthogonal regulatory elements, RNA- or protein-based, reported to date.

Crosstalk experiments also revealed smaller sets of orthogonal switches exhibiting substantial improvements in overall dynamic range, including a set of 18 with 50-fold predicted minimum dynamic range (<2% crosstalk) and a set of 5 with >200-fold dynamic range (<0.5% crosstalk). Subsets of orthogonal toehold switches of comparable size to previously reported riboregulator libraries therefore exhibit minimum dynamic range over an order of magnitude larger than the earlier RNA-based systems. At this point, the ultimate size of the orthogonal libraries of toehold switches is limited by throughput of our crosstalk assay, not design features intrinsic to the devices themselves.

#### RNA-Based, Forward-Engineered Components for Synthetic Biology

We anticipate that our large library of orthogonal toehold switches will enable new applications and capabilities for synthetic biology. Their combination of wide dynamic range, low leakage levels, orthogonality, and rapid response are ideal for implementing time-sensitive logic circuitry in living systems. Furthermore, the protein-like dynamic range of the toehold switches is achieved at a lower metabolic cost than protein-based systems because all regulatory components are composed of RNA, and translation only occurs if the switches are activated. As synthetic genetic networks grow in complexity and impose greater burdens on the host, the lower metabolic footprint of the toehold switches will become an increasingly important advantage in their favor. Moreover, their low leakage levels suggest they can be used as constitutively expressed passive monitoring or sensor systems that trigger a downstream response only in desired situations.

Our success in using in silico design tools to reliably produce high-performance, orthogonal toehold switches represents an important step forward for synthetic biology. The process of constructing new synthetic gene networks generally requires multiple design-build-test cycles before the desired network behavior is achieved. Consequently, new devices like the toehold switch that routinely provide functional in-silico-designed components will reduce the number of design cycles required for optimization by improving the components tested with each passing cycle. Improvements in device design can be used for decreasing development time but will also facilitate construction of more complex genetic networks. In particular, in silico design tools currently enable simulation of multiple interacting RNAs, which could be used for carrying out complex logic operations (Qian and Winfree, 2011) from programmed RNA networks. In these higher-order

computations, the programmability of toehold switches provides them with a distinct advantage over protein-based regulators that are currently less amenable to forward engineering. We also expect that additional improvements to toehold switch design performance can be obtained by incorporating the effects of the  $\Delta G_{\text{RBS-linker}}$  term into future in silico design algorithms.

#### Toehold Switches as RNA-to-Protein Signal Transducers

The ability of toehold switches to activate translation in response to arbitrary RNA sequences, including full-length mRNAs, enables them to act as universal RNA-to-protein signal transducers. This new capability provides a generalizable approach to interfacing endogenous RNA networks with artificial genetic networks for synthetic biology and a nondestructive means of monitoring RNAs in living cells for answering fundamental questions in biology. Many future applications in synthetic biology will require genetic circuits that can sense the current status of the cell and operate on this information to modulate their activity. For instance, this capability can be used in metabolic engineering to construct self-optimizing networks that detect cellular stress levels or pathway-associated RNAs to tune network activity up or down for maximizing yields of useful chemical products. For biological studies, toehold switch signal transducers can be used to simultaneously monitor the levels of multiple untagged, endogenous RNA species by producing different output fluorescent proteins.

#### Potential Role of Similar Systems in Nature

The protein-like dynamic range of toehold switches challenges the widely held view that RNA-based systems offer weaker regulation than their protein counterparts (Liu et al., 2012). The performance of these riboregulators also suggests that some of their operating mechanisms may already play a role in natural systems. Although, to our knowledge, analogous systems have yet to be found in nature, our results indicate that a re-examination of bacterial genomes in a directed search for these specific regulatory mechanisms may be in order. Conversely, the absence of well-known natural systems that exploit the design principles of toehold switches could also imply that these synthetic devices are not easy to evolve. In particular, we hypothesize that the loop regions of loop-mediated riboregulators are more insulated from interactions with nearby bases than the unprotected single-stranded regions of toehold switch trigger RNAs. Thus, single-stranded trigger RNAs must sample a greater number of potential sequences to evolve to their intended function than natural loop-mediated trigger RNAs.

#### EXPERIMENTAL PROCEDURES

The Extended Experimental Procedures are available online.

#### Strains and Growth Conditions

The following *E. coli* strains were used in this study: BL21 Star DE3 ( $F^- ompT hsdS_B (r_B^- m_B^-) gal dcm meT31$  [DE3]; Invitrogen); BL21 DE3 ( $F^- ompT hsdS_B (r_B^- m_B^-) gal dcm$  [DE3]; Invitrogen); MG1655Pro ( $F^- \lambda^- ilvG^- rfb-50 rph-1 Sp^R lacR tetR$ ); and DH5 $\alpha$  ( $endA1 recA1 gyrA96 thi-1 glnV44 relA1 hsdR17(r_K^- m_K^-) \lambda^-$ ). All strains were grown in LB medium with appropriate antibiotics at 37°C (see also section S1 of the Extended Experimental Procedures).

### Plasmid Construction

All DNA oligonucleotides were purchased from Integrated DNA Technologies. Plasmids were constructed and purified using standard molecular biology techniques. Table S1 contains the primers and sequences used for experiments (see also section S2 of the Extended Experimental Procedures).

### Flow Cytometry Measurements and Analysis

Flow cytometry was performed using a BD LSRFortessa cell analyzer equipped with a high-throughput sampler. Cells were typically diluted by a factor of ~65 into phosphate buffered saline (PBS) and sampled from 96-well plates. Forward scatter (FSC) was used for trigger, and ~20,000 individual cells were analyzed. Error levels for the fluorescence measurements of ON state and OFF state cells were calculated from the SD of measurements from at least three biological replicates. The relative error levels for the ON/OFF fluorescence ratios were then determined by adding the relative errors of ON and OFF state fluorescence in quadrature (see also sections S3 and S4 of the Extended Experimental Procedures).

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.002>.

### AUTHOR CONTRIBUTIONS

A.A.G. conceived the project, designed the system, designed and performed the experiments, analyzed the data, and wrote the paper. P.A.S. and J.J.C. assisted in experiment design, data interpretation, and manuscript preparation. P.Y. conceived, designed, and supervised the study, interpreted the data, and wrote the paper. All authors commented on and approved the paper.

### ACKNOWLEDGMENTS

The authors acknowledge J. Kim, C. Myhrvold, K. Pardee, and X. Chen for critical reading of the manuscript and helpful discussions and J.M.L. Ho for preliminary experiments, plasmid preparation, and helpful discussions. This work was supported by a DARPA Living Foundries grant (HR001112C0061) to P.A.S., P.Y., and J.J.C.; an NIH Director's New Innovator Award (1DP2OD007292); an NIH Transformative Research Award (1R01EB018659); an Office of Naval Research (ONR) Young Investigator Program Award (N000141110914); ONR grants (N000141010827, N000141310593, and N000141410610); a National Science Foundation (NSF) Faculty Early Career Development Award (CCF1054898); an NSF Expedition in Computing Award (CCF1317291); NSF grant (CCF1162459) and Wyss Institute funds to P.Y.; Defense Threat Reduction Agency grant HDTRA1-14-1-0006 to J.J.C.; ONR MURI grant N000141110725 to J.J.C. and P.A.S.; and the Howard Hughes Medical Institute (J.J.C.). A provisional patent based on this work has been filed.

Received: December 24, 2013

Revised: July 15, 2014

Accepted: September 25, 2014

Published: October 23, 2014

### REFERENCES

Ausländer, S., Ausländer, D., Müller, M., Wieland, M., and Fussenegger, M. (2012). Programmable single-cell mammalian biocomputers. *Nature* 487, 123–127.

Bayer, T.S., and Smolke, C.D. (2005). Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat. Biotechnol.* 23, 337–343.

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., and Blüthgen, N. (2013). Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* 9, 675.

Brantl, S., and Wagner, E.G.H. (2000). Antisense RNA-mediated transcriptional attenuation: an in vitro study of plasmid pT181. *Mol. Microbiol.* 35, 1469–1482.

Callura, J.M., Dwyer, D.J., Isaacs, F.J., Cantor, C.R., and Collins, J.J. (2010). Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proc. Natl. Acad. Sci. USA* 107, 15898–15903.

Callura, J.M., Cantor, C.R., and Collins, J.J. (2012). Genetic switchboard for synthetic biology applications. *Proc. Natl. Acad. Sci. USA* 109, 5850–5855.

Choi, H.M.T., Chang, J.Y., Trinh, A., Padilla, J.E., Fraser, S.E., and Pierce, N.A. (2010). Programmable in situ amplification for multiplexed imaging of mRNA expression. *Nat. Biotechnol.* 28, 1208–1212.

Culler, S.J., Hoff, K.G., and Smolke, C.D. (2010). Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science* 330, 1251–1255.

Daniel, R., Rubens, J.R., Sarpeshkar, R., and Lu, T.K. (2013). Synthetic analog computation in living cells. *Nature* 497, 619–623.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* 97, 6640–6645.

Dirks, R.M., and Pierce, N.A. (2004). Triggered amplification by hybridization chain reaction. *Proc. Natl. Acad. Sci. USA* 101, 15275–15278.

Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338.

Friedland, A.E., Lu, T.K., Wang, X., Shi, D., Church, G., and Collins, J.J. (2009). Synthetic gene networks that count. *Science* 324, 1199–1202.

Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342.

Gulyaev, A.P., Franch, T., and Gerdes, K. (1997). Programmed cell death by *hok/sok* of plasmid R1: coupled nucleotide covariations reveal a phylogenetically conserved folding pathway in the *hok* family of mRNAs. *J. Mol. Biol.* 273, 26–37.

Gutell, R.R., Cannone, J.J., Konings, D., and Gautheret, D. (2000). Predicting U-turns in ribosomal RNA with comparative sequence analysis. *J. Mol. Biol.* 300, 791–803.

Isaacs, F.J., Dwyer, D.J., Ding, C.M., Pervouchine, D.D., Cantor, C.R., and Collins, J.J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.* 22, 841–847.

Khalil, A.S., and Collins, J.J. (2010). Synthetic biology: applications come of age. *Nat. Rev. Genet.* 11, 367–379.

Khalil, A.S., Lu, T.K., Bashor, C.J., Ramirez, C.L., Pyenson, N.C., Joung, J.K., and Collins, J.J. (2012). A synthetic biology framework for programming eukaryotic transcription functions. *Cell* 150, 647–658.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324, 255–258.

Liu, C.C., Qi, L., Lucks, J.B., Segall-Shapiro, T.H., Wang, D., Mutalik, V.K., and Arkin, A.P. (2012). An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. *Nat. Methods* 9, 1088–1094.

Lucks, J.B., Qi, L., Mutalik, V.K., Wang, D., and Arkin, A.P. (2011). Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proc. Natl. Acad. Sci. USA* 108, 8617–8622.

Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/11-12 regulatory elements. *Nucleic Acids Res.* 25, 1203–1210.

Massé, E., and Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 99, 4620–4625.

Moon, T.S., Lou, C., Tamsir, A., Stanton, B.C., and Voigt, C.A. (2012). Genetic programs constructed from layered logic gates in single cells. *Nature* 491, 249–253.

- Mutalik, V.K., Qi, L., Guimaraes, J.C., Lucks, J.B., and Arkin, A.P. (2012). Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol.* **8**, 447–454.
- Purnick, P.E.M., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410–422.
- Qi, L., Lucks, J.B., Liu, C.C., Mutalik, V.K., and Arkin, A.P. (2012). Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. *Nucleic Acids Res.* **40**, 5775–5786.
- Qian, L., and Winfree, E. (2011). Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **332**, 1196–1201.
- Qian, L., Winfree, E., and Bruck, J. (2011). Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368–372.
- Rhodus, V.A., Segall-Shapiro, T.H., Sharon, B.D., Ghodasara, A., Orlova, E., Tabakh, H., Burkhardt, D.H., Clancy, K., Peterson, T.C., Gross, C.A., and Voigt, C.A. (2013). Design of orthogonal genetic switches based on a crosstalk map of  $\sigma$ s, anti- $\sigma$ s, and promoters. *Mol. Syst. Biol.* **9**, 702.
- Rodrigo, G., Landrain, T.E., and Jaramillo, A. (2012). De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells. *Proc. Natl. Acad. Sci. USA* **109**, 15271–15276.
- Siuti, P., Yazbek, J., and Lu, T.K. (2013). Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31**, 448–452.
- Takahashi, M.K., and Lucks, J.B. (2013). A modular strategy for engineering orthogonal chimeric RNA transcription regulators. *Nucleic Acids Res.* **41**, 7577–7588.
- Win, M.N., and Smolke, C.D. (2008). Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460.
- Winkler, W., Nahvi, A., and Breaker, R.R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956.
- Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R., and Benenson, Y. (2011). Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* **333**, 1307–1311.
- Yin, P., Choi, H.M.T., Calvert, C.R., and Pierce, N.A. (2008). Programming biomolecular self-assembly pathways. *Nature* **451**, 318–322.
- Yurke, B., Turberfield, A.J., Mills, A.P., Jr., Simmel, F.C., and Neumann, J.L. (2000). A DNA-fuelled molecular machine made of DNA. *Nature* **406**, 605–608.
- Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M., and Pierce, N.A. (2011). NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173.

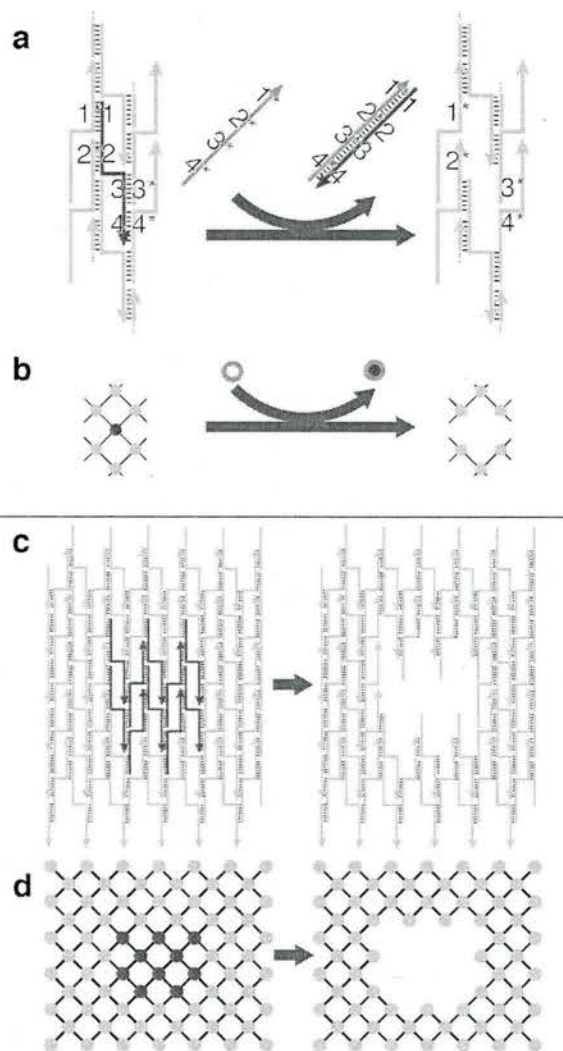
# Complex Reconfiguration of DNA Nanostructures\*\*

Bryan Wei,\* Luvena L. Ong, Jeffrey Chen, Alexander S. Jaffe, and Peng Yin\*

**Abstract:** Nucleic acids have been used to create diverse synthetic structural and dynamic systems. Toehold-mediated strand displacement has enabled the construction of sophisticated circuits, motors, and molecular computers. Yet it remains challenging to demonstrate complex structural reconfiguration in which a structure changes from a starting shape to another arbitrarily prescribed shape. To address this challenge, we have developed a general structural-reconfiguration method that utilizes the modularly interconnected architecture of single-stranded DNA tile and brick structures. The removal of one component strand reveals a newly exposed toehold on a neighboring strand, thus enabling us to remove regions of connected component strands without the need to modify the strands with predesigned external toeholds. By using this method, we reconfigured a two-dimensional rectangular DNA canvas into diverse prescribed shapes. We also used this method to reconfigure a three-dimensional DNA cuboid.

The self-assembly of nucleic acids (DNA and RNA) has produced diverse synthetic structures.<sup>[1–29]</sup> In particular, DNA origami<sup>[9,12–15,19,20,23]</sup> and single-stranded tiles (SSTs) and bricks<sup>[21,22,24,25]</sup> have enabled the construction of megadalton discrete structures with arbitrarily prescribed shapes. In parallel, researchers have used strand displacement to demonstrate the construction of dynamic systems,<sup>[30]</sup> such as

switches,<sup>[31]</sup> walkers,<sup>[6,32,33]</sup> circuits,<sup>[34,32,35]</sup> and triggered assembly systems,<sup>[36,32]</sup> which can go through multiple states of different configurations, either in a directed or in an autonomous fashion.



**Figure 1.** Schematic representation of structural reconfiguration from an SST canvas. a) Strand diagram of strand-displacement-based SST structural reconfiguration and b) the associated interaction graph. The strand/node to be displaced is highlighted in red with four domains 1, 2, 3, and 4 complementary to domains 1\*, 2\*, 3\*, and 4\*, respectively, of neighboring strands. When introduced into the system, a full complementary strand 4\*–3\*–2\*–1\* (in salmon) forms a duplex with the red target component strand to displace the target from the canvas. c) Strand diagram of strand displacement for a 12 helix (H) × 10 turns (T) canvas and d) the associated interaction graph. Strands/nodes highlighted in red on the left depict the subset of component strands to be displaced. The carved structure is shown on the right.

[\*] Dr. B. Wei, L. L. Ong, J. Chen, A. S. Jaffe, Prof. Dr. P. Yin  
Wyss Institute for Biologically Inspired Engineering  
Harvard University, Boston, MA 02115 (USA)  
E-mail: bryan.wei@wyss.harvard.edu  
py@hms.harvard.edu  
Homepage: <http://molecular-systems.net>  
<http://yin.hms.harvard.edu>

Dr. B. Wei, Prof. Dr. P. Yin  
Department of Systems Biology, Harvard Medical School  
Boston, MA 02115 (USA)

L. L. Ong  
Harvard-Massachusetts Institute of Technology (MIT)  
Division of Health Sciences and Technology, MIT  
Cambridge, MA 02139 (USA)

[\*\*] We thank David Zhang and Yonggang Ke for discussions and Michelle K. Vhudzijena for technical assistance. This research is supported by an Office of Naval Research (ONR) Young Investigator Program Award (N000141110914), ONR grants (N000141010827, N000141310593), a National Science Foundation (NSF) Faculty Early Career Development Award (CCF1054898), an NSF Expedition in Computing Award (CCF1317291), NSF grants (CCF1162459, CMM11333215), a National Institutes of Health (NIH) Director's New Innovator Award (1DP2OD007292), and a Wyss Institute for Biologically Inspired Engineering Faculty Startup Fund to P.Y. L.L.O. acknowledges support from the NSF through a Graduate Research Fellowship.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/ange.201402437>.

Researchers have combined techniques for DNA strand displacement with methods for DNA structural assembly to create reconfigurable and/or reversible structures. For example, DNA origami boxes<sup>[13]</sup> and clamshells<sup>[20]</sup> were reconfigured from a closed to an open state through the toehold-mediated strand displacement of a few component strands in the structure. More complex reconfiguration methods have placed single-stranded toeholds at selected sites on the structure for shape transformation,<sup>[37–40]</sup> including the formation of a catenane derived from a DNA-origami Möbius strip<sup>[39]</sup> and the changing of fractal patterns in origami structures.<sup>[40]</sup>

However, it still remains challenging to develop a general framework for complex structural reconfiguration in which a structure changes from a particular starting shape to another arbitrarily prescribed shape. To address this challenge, we have developed a method based on the modularly interconnected architecture of single-stranded DNA tile and brick structures. The removal of one component strand reveals a newly exposed toehold on a neighboring strand, thus enabling us to remove connected regions of component strands without the need to modify them with predesigned external toeholds. This method applies only to SST/brick-based structures and not to origami-based structures because successful reconfiguration relies on the modular architecture.

By using this method, we demonstrate that a two-dimensional rectangular DNA SST canvas<sup>[21,25]</sup> can be reconfigured into many different shapes, including two full sets of the alphabet (one carved in intaglio, with the cavity forming the letters, and the other in relief, with DNA forming the letters). We also show that individual carved pieces can be reassembled to form the original canvas and be subjected to a second round of reconfigurations. Finally, this molecular-carving concept was also applied to a three-dimensional DNA-brick cuboid structure.<sup>[22]</sup> Overall, we demonstrate that SST and brick reconfiguration is a robust and modular method for engineering complex structural reconfiguration, with resolution at the scale of the component strand (e.g.  $3 \times 7 \text{ nm}^2$  for a typical 42 nt SST<sup>[21]</sup>).

In an SST structure,<sup>[21,25]</sup> each strand typically has four binding domains that are complementary to four different neighboring strands, as depicted in Figure 1 a,c. A self-

assembled rectangular SST structure can be viewed as a “molecular canvas” and depicted as an interaction graph (Figure 1 b,d): each node represents a component strand, and each edge represents the binding interaction between two strands or nodes. A component strand (red) can be removed (“carving”) by introducing a “carving strand” (salmon) that is fully complementary to the component strand (Figure 1 a,b). Unlike previous strand-displacement approaches, our method does not involve the use of an external toehold<sup>[31]</sup> (the detailed molecular mechanism is discussed later). By the displacement of multiple component strands with corresponding carving strands, the canvas can be reconfigured into a prescribed shape (Figure 1 c,d). Moreover, because each component strand can be removed modularly, it is possible to create a combinatorially large number of distinct shapes.

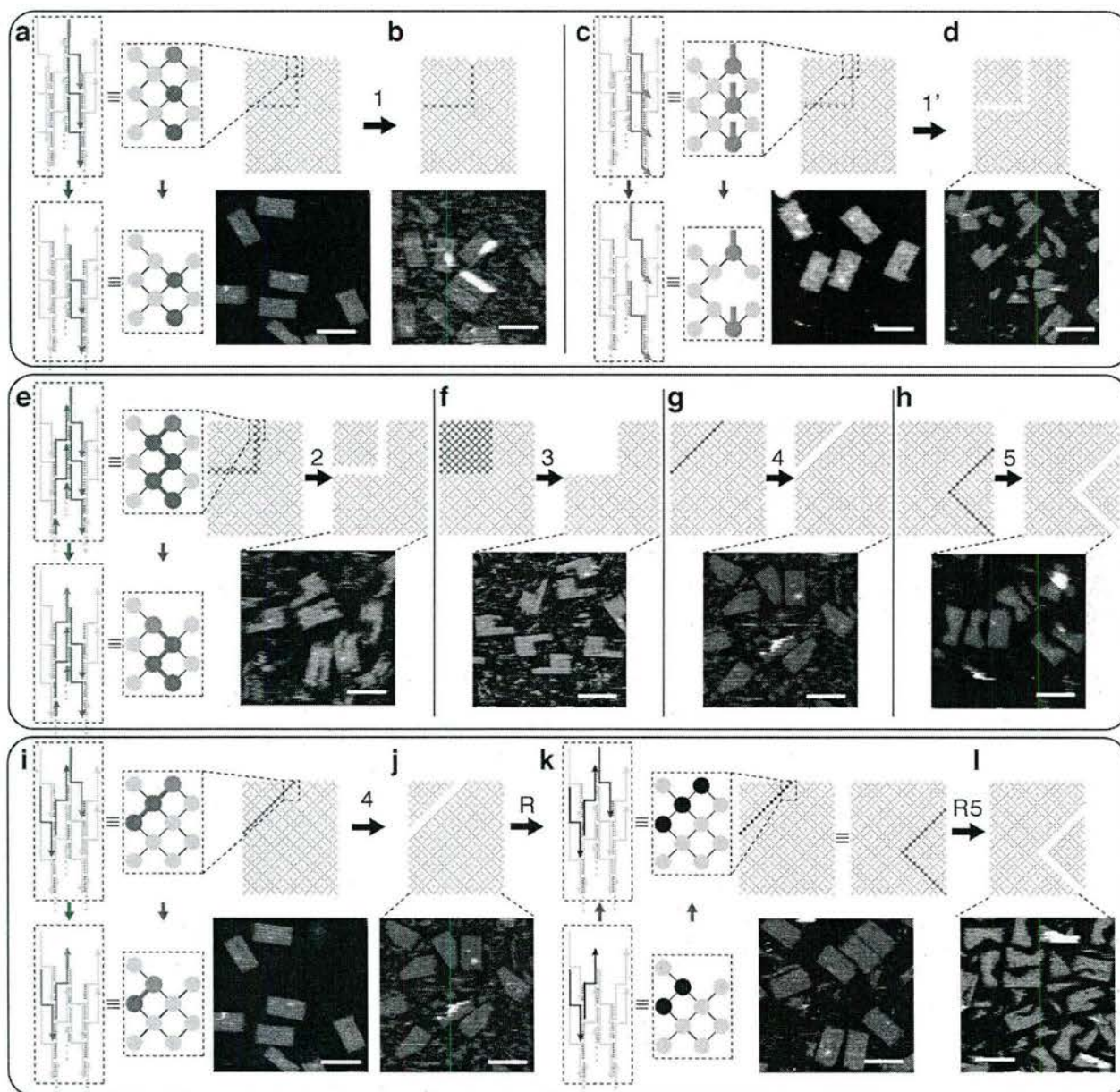
The canvas<sup>[25]</sup> was made by the self-assembly of 375 distinct component strands in 15 mM  $\text{Mg}^{2+}$  buffer at 48 °C overnight with 30% yield, as determined by 2% native gel electrophoresis (see Figure S4 in the Supporting Information). The structures showed the expected morphology and dimensions under atomic force microscopy (AFM) imaging (Figure 2 a, bottom left).

To reduce the overall time and potential human errors in the picking and mixing of carving strands, we wrote a computer program to design target shapes. The program takes the desired shape as input and then directs a robotic liquid handler to select the appropriate subset of strands from a master library of 375 carving strands. These carving strands were then applied to the canvas solution in an equimolar ratio for overnight incubation at 35 or 45 °C to produce the carved shapes. Figure 2 a shows an example of carving a corner off the canvas (the detailed carving mechanism is discussed later). Figure 2 b shows AFM images of the carved shapes of two full sets of the alphabet in intaglio and relief. The yield of the reconfigured structure in Figure 2 a was determined by gel electrophoresis to be 50% (see Figure S17) and by AFM to be 99% ( $N = 202$ ; see Figure S30). Such a high AFM yield is not typical for the more complex alphabet structures that we carved (see Figures S4 and S5 (gels) and Figures S6–S15 (AFM images)).

Unlike most previous strand-displacement-based dynamic systems, our carving scheme does not use predesigned



**Figure 2.** Alphabet sets reconfigured from a rectangular SST canvas. a) Interaction graphs (top) and AFM images (bottom) of the 24 H × 29 T canvas used in this study (left) and its reconfiguration into a rectangle with a missing corner (right). Scale bars: 100 nm. b) AFM images of letters of the alphabet carved in intaglio (top) and relief (bottom). Each image is  $150 \times 150 \text{ nm}^2$  in size. See Figures S4 and S5 in the Supporting Information for the results of agarose gel electrophoresis.



**Figure 3.** Diagrams and AFM images of a mechanistic study. In each case, interaction graphs with the carving pattern highlighted in red and blue are shown at the top, and AFM images are shown at the bottom (scale bars: 100 nm). a–d) Carving of pattern 1 without predesigned external toeholds (a, b) and pattern 1' with predesigned external toeholds (c, d). e–h) Carving of patterns 2–5. i–l) Experiments demonstrating the reversibility of carving: reconfiguration of the canvas to give pattern 4, reassembly, and generation of pattern R5 (identical to pattern 5). The strand diagram and interaction graph boxed by the dashed lines show the reconfiguration mechanism at the zoomed-in fraction of the canvas (in a, c, e, i, and k). Gray depicts common components; red depicts the strands without external toeholds to be carved; blue marks the presence of an exposed single-stranded toehold; black indicates the introduction of strands for reassembly of the carved canvas pieces. See the Supporting Information for a detailed study of carving yields.

external toeholds to initiate strand displacement. We thus conducted a set of experiments to study the effects of external toeholds on carving (Figure 3a–d, patterns 1 and 1' highlighted in red). We designed a different canvas in which the component strands in the carving pattern had external toeholds (blue; Figure 3c,d, pattern 1'). After overnight carving at 45 °C, gel electrophoresis showed minimal carving for the toehold-free canvas (Figure 3a,b, pattern 1; yield around 0%; see Figure S16) but significant carving for the canvas with external toeholds (Figure 3c,d; 19–39%

yield; see Figure S16). AFM images were consistent with the gel results: carving a canvas without predesigned external toeholds rarely reached completion (ca. 14% yield; see Figure S27), whereas carving a canvas with external toeholds led to higher conversion (ca. 81% yield; see Figure S28). The observation of higher AFM yields than gel yields suggests that partially carved structures may have comigrated with the intact canvas on the gel, but may have broken apart under AFM (e.g. during the deposition or imaging process).

We next designed a second set of experiments to test five distinct connection patterns of component strands to be carved (carving patterns) in a canvas without external toeholds. The carving samples ([carving strands]/[component strands]: 1:1, 2:1, or 3:1) from patterns 1–5 were incubated at 45 °C overnight before agarose gel electrophoresis and AFM imaging. When the sample from carving pattern 1 was compared with those from patterns 2–5, significantly lower carving yields were observed for pattern 1 by gel electrophoresis and AFM (see the Supporting Information for a detailed yield study).

A component strand to be carved is denoted by a red or blue node. A blue node indicates the presence of an unpaired single-stranded domain. Each node in carving pattern 1 (Figure 3 a,b) is disjoint from any other node in the pattern. In contrast, in each of patterns 2–5 (Figure 3 e–h), the nodes to be carved are fully connected, such that a path consisting only of nodes to be carved exists between any two nodes within the carving pattern. Additionally, each of these patterns contains at least one blue node representing a component strand at the end of a constituent helix with exposed poly-T domain(s) that can serve as an external toehold (Figure 3 e–h).

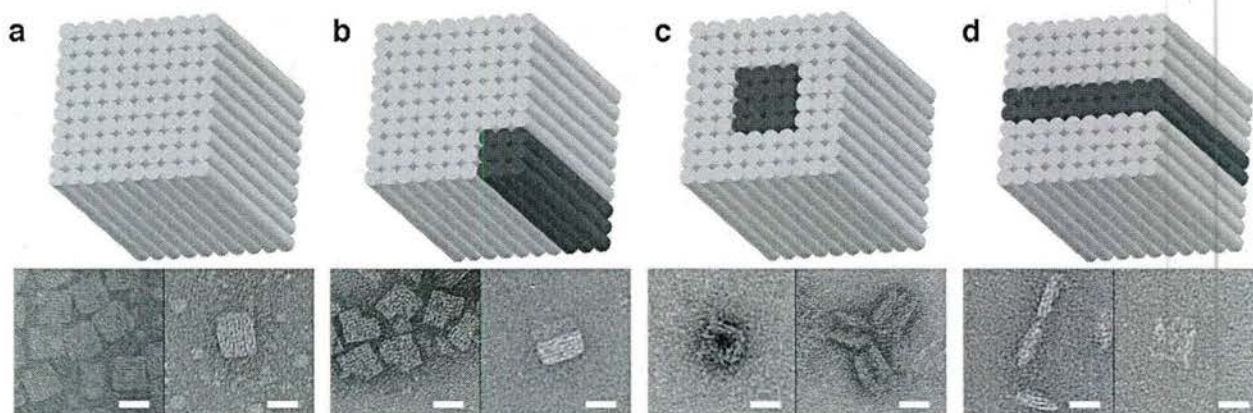
The above experiments on the carving patterns demonstrate that appending an external toehold on each individual node in a carving pattern composed of disjoint nodes can enable effective carving (e.g. pattern 1') of an initially hard-to-carve pattern (e.g. pattern 1). In contrast, effective carving can be obtained in a fully connected carving pattern (e.g. patterns 2–5) composed mostly of toehold-free nodes if the pattern also contains an initial blue node (e.g. the strand with an exposed poly-T domain at the helix end). In the latter case, the carving reaction can be initiated from an external toehold in a blue node, and followed by a cascade of sequential exposure of initially hidden toeholds. Detailed schematics of this mechanism can be found in the zoomed-in diagrams of Figure 3 e,i. For example, in the zoomed-in diagrams of Figure 3 e, the strand with initial toeholds of poly-T domains (top, shown as a blue node and as blue domains in the strand diagram) can be displaced by a carving strand. The displace-

ment of this strand will reveal a newly exposed toehold on its neighboring strand (bottom, shown as a blue node and as a blue domain in the strand diagram), allowing this neighbor to be displaced next. Following this mechanism, a cascade of new toeholds will be sequentially generated, enabling the removal of all strands in the carving pattern. Note that such a consecutive toehold exposure is not possible in the disjoint pattern (e.g. pattern 1, see zoom-in of Figure 3 a). In the case of carving pattern 1', where each node contains a pre-designed external toehold, each component can be displaced independently in parallel (zoom-in of Figure 3 b).

In the cases of the intaglio alphabet and the cavities in the relief set where the carving patterns contain no initial external toehold, the cascading could be initiated from the transient dissociation of a few bases on an internal component, or an unpaired domain from empty neighboring components (such a defect rate is estimated as 5–10% for a certain component in a DNA origami structure.<sup>[41]</sup>

We tested the reversibility of this structural reconfiguration by adding back the displaced components following the carving of pattern 4 (Figure 3 j,k, pattern R). This overnight reaction at 45 °C was sufficient to glue back the two carved out pieces almost seamlessly with a yield of around 89% by gel electrophoresis and 78% by AFM (Figure 3 k; see Figures S18 and S33). The reassembled canvases were then subjected to a second round of carving (pattern R5; Figure 3 l), which resulted in yields similar to those obtained directly from carving the original canvas (gel yield of 58% and AFM yield of 72%; Figures S18 and S34).

We then applied this carving method to 3D DNA-brick cuboids. By using a 10 helix × 10 helix × 80 base-pair structure (Figure 4 a, top) that we reported previously,<sup>[22]</sup> we tested multiple carving patterns. We used a reaction temperature (28 °C) lower than that for the 2D carving, since 3D structures contain 8 nt binding domains and are less thermally stable than the 2D structures, which contain 10 or 11 nt binding domains. The successful results of carving a corner off (Figure 4 b, top) or a tunnel through a cuboid (Figure 4 c, top) and carving the cuboid into two halves (Figure 4 d, top) were shown by agarose gel electrophoresis (see Figure S37)



**Figure 4.** Structural reconfiguration from a 3D cuboid. Top: cylinder model of the cuboid (red cylinders denote those to be displaced); bottom: TEM images (scale bars: 20 nm). a) Cuboid before structural reconfiguration. b) Carving of a corner from the cuboid. c) Carving of a tunnel through the cuboid. d) Carving of the cuboid into two halves.

and transmission electron microscopy (TEM; Figure 4, bottom). Because of the limited thermal stability and the limited accessibility of the majority of the component strands, the structural reconfiguration of 3D DNA structures is much more difficult than that of the 2D counterparts, and it is even more challenging to obtain fine features. Since it is difficult to find a temperature high enough for fast strand displacement but low enough ( $< 30^{\circ}\text{C}$ ) to maintain structural stability, we were only able to demonstrate a few cases of successful carving with coarse resolution.

Carving may not appear to be thermodynamically favorable, since a component strand in the canvas structure that is fully paired with its neighbors has the same designed number of hybridized bases as when it is bound to its fully complementary carving strand (after carving). However, the carving off of a component strand may help to alleviate the electrostatic repulsion that results from the closely packed neighboring DNA duplexes in the canvas structure<sup>[9,21]</sup> and release the mechanical stress that might be accrued at the cross-over points,<sup>[42]</sup> thereby favoring the reaction.

In a similar way to typical toehold-mediated strand-displacement circuits, our carving system shows leakage (carving can be initiated in a connected region even when no external toehold is present), which we utilize to initiate carving. However, it is conceivable that under other reaction conditions (e.g. if the carving reaction is carried out at room temperature over a shorter period of time rather than at an elevated temperature ( $35\text{--}45^{\circ}\text{C}$ ) overnight as in this study), such leakage could be mitigated, and carving may only start at a site with an external toehold. Such a system would be analogous to a strand-displacement circuit embedded in a nanostructure (the canvas), in which case the strand-displacement cascade would be directly coupled with the structural change at single-strand resolution. This system would provide a platform to explore the rich interplay between structure, dynamics, and computation. For example, it would be interesting to design spatial logic gates and circuits: A multi-input AND gate could be designed as a carving path with the component strands along the path serving as the inputs and the strand at the end of the path serving as the output, or a two-input OR gate could be designed as two converging paths with one strand on each path serving as the inputs and the strand at the merging point serving as the output. In the latter case, when multiple strands along each path are used as inputs, these two converging paths would together implement a disjunctive normal form formula ( $[A_1 \text{ AND } A_2 \text{ AND } \dots \text{ AND } A_m] \text{ OR } [B_1 \text{ AND } B_2 \text{ AND } \dots \text{ AND } B_n]$ ).

Received: February 14, 2014

Revised: March 22, 2014

Published online: ■■■■■, ■■■■■

**Keywords:** DNA bricks · nanostructures · single-stranded tiles · strand displacement · structural reconfiguration

[1] N. C. Seeman, *J. Theor. Biol.* **1982**, *99*, 237–247.

[2] J. Chen, N. C. Seeman, *Nature* **1991**, *350*, 631–633.

- [3] T. J. Fu, N. C. Seeman, *Biochemistry* **1993**, *32*, 3211–3220.
- [4] E. Winfree, F. Liu, L. A. Wenzler, N. C. Seeman, *Nature* **1998**, *394*, 539–544.
- [5] H. Yan, S. H. Park, G. Finkelstein, J. H. Reif, T. H. LaBean, *Science* **2003**, *301*, 1882–1884.
- [6] W. B. Sherman, N. C. Seeman, *Nano Lett.* **2004**, *4*, 1203–1207.
- [7] A. Chworos, I. Severcan, A. Y. Koyfman, P. Weinkam, E. Oroudjev, H. G. Hansma, L. Jaeger, *Science* **2004**, *306*, 2068–2072.
- [8] S. H. Park, C. Pistol, S. J. Ahn, J. H. Reif, A. R. Lebeck, C. Dwyer, T. H. Labean, *Angew. Chem.* **2006**, *118*, 749–753; *Angew. Chem. Int. Ed.* **2006**, *45*, 735–739.
- [9] P. W. K. Rothmund, *Nature* **2006**, *440*, 297–302.
- [10] Y. He, T. Ye, M. Su, C. Zhang, A. E. Ribbe, W. Jiang, C. Mao, *Nature* **2008**, *452*, 198–201.
- [11] P. Yin, R. Hariadi, S. Sahu, H. M. T. Choi, S. H. Park, T. H. LaBean, J. H. Reif, *Science* **2008**, *321*, 824–826.
- [12] Y. Ke, J. Sharma, M. Liu, K. Jahn, Y. Liu, H. Yan, *Nano Lett.* **2009**, *9*, 2445–2447.
- [13] E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf, J. Kjems, *Nature* **2009**, *459*, 73–76.
- [14] S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, W. M. Shih, *Nature* **2009**, *459*, 414–418.
- [15] H. Dietz, S. M. Douglas, W. M. Shih, *Science* **2009**, *325*, 725–730.
- [16] J. Zheng, J. J. Birktoft, Y. Chen, T. Wang, R. J. Sha, P. E. Constantinou, S. L. Ginell, C. Mao, N. C. Seeman, *Nature* **2009**, *461*, 74–77.
- [17] I. Severcan, C. Geary, A. Chworos, N. Voss, E. Jacovetty, L. Jaeger, *Nat. Chem.* **2010**, *2*, 772–779.
- [18] C. J. Delebecque, A. B. Lindner, P. A. Silver, F. A. Aldaye, *Science* **2011**, *333*, 470–474.
- [19] D. Han, S. Pal, J. Nangreave, Z. Deng, Y. Liu, H. Yan, *Science* **2011**, *332*, 342–346.
- [20] S. M. Douglas, I. Bachelet, G. M. Church, *Science* **2012**, *335*, 831–834.
- [21] B. Wei, M. Dai, P. Yin, *Nature* **2012**, *485*, 623–626.
- [22] Y. Ke, L. L. Ong, W. M. Shih, P. Yin, *Science* **2012**, *338*, 1177–1183.
- [23] D. Han, S. Pal, Y. Yang, S. Jiang, J. Nangreave, Y. Liu, H. Yan, *Science* **2013**, *339*, 1412–1415.
- [24] C. Myhrvold, M. Dai, P. A. Silver, P. Yin, *Nano Lett.* **2013**, *13*, 4242–4248.
- [25] B. Wei, M. Dai, C. Myhrvold, Y. Ke, R. Jungmann, P. Yin, *J. Am. Chem. Soc.* **2013**, *135*, 18080–18088.
- [26] C. Lin, Y. Liu, S. Rinker, H. Yan, *ChemPhysChem* **2006**, *7*, 1641–1647.
- [27] W. M. Shih, C. Lin, *Curr. Opin. Struct. Biol.* **2010**, *20*, 276–282.
- [28] N. B. Leontis, A. Lescoute, E. Westhof, *Curr. Opin. Struct. Biol.* **2006**, *16*, 279–287.
- [29] N. C. Seeman, *Annu. Rev. Biochem.* **2010**, *79*, 65–87.
- [30] D. Y. Zhang, G. Seelig, *Nat. Chem.* **2011**, *3*, 103–113.
- [31] B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel, J. L. Neumann, *Nature* **2000**, *406*, 605–608.
- [32] P. Yin, H. M. T. Choi, C. R. Calvert, N. A. Pierce, *Nature* **2008**, *451*, 318–322.
- [33] T. Omabegho, R. Sha, N. C. Seeman, *Science* **2009**, *324*, 67–71.
- [34] G. Seelig, D. Soloveichik, D. Y. Zhang, E. Winfree, *Science* **2006**, *314*, 1585–1588.
- [35] L. Qian, E. Winfree, *Science* **2011**, *332*, 1196–1201.
- [36] R. M. Dirks, N. A. Pierce, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15275–15278.
- [37] H. Yan, X. Zhang, Z. Shen, N. C. Seeman, *Nature* **2002**, *415*, 62–65.
- [38] Z. Li, Y. Ke, C. Lin, H. Yan, Y. Liu, *Chem. Commun.* **2008**, 4318–4320.

- [39] D. Han, S. Pal, Y. Liu, H. Yan, *Nat. Nanotechnol.* **2010**, *5*, 712–717.
- [40] F. Zhang, J. Nangreave, Y. Liu, H. Yan, *Nano Lett.* **2012**, *12*, 3290–3295.
- [41] R. Jungmann, C. Steinhauer, M. Scheible, A. Kuzyk, P. Tinnefeld, F. C. Simmel, *Nano Lett.* **2010**, *10*, 4756–4761.
- [42] P. K. Maiti, T. A. Pascal, N. Vaidehi, W. A. Goddard, *Nucleic Acids Res.* **2004**, *32*, 6047–6056.
-

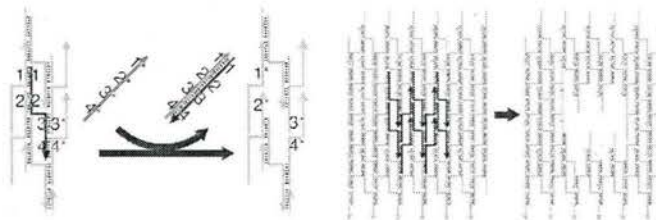
## Zuschriften



## DNA-Architekturen

B. Wei,\* L. L. Ong, J. Chen, A. S. Jaffe,  
P. Yin\* ————— ■■■■-■■■■

Complex Reconfiguration of DNA  
Nanostructures



**Schnitzen von DNA-Nanostrukturen:** Die modular verzahnte Architektur einzelsträngiger Kachel- und Ziegelstein-Strukturen wurde genutzt, um eine allgemeine Methode für die Strukturrekonfiguration zu entwickeln. Das Entfernen eines Stranges (siehe Bild) legt einen neuen

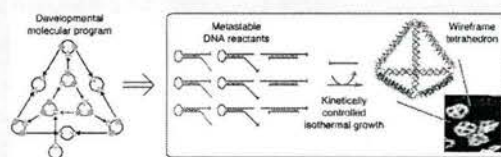
Haltepunkt an einem Nachbarstrang frei, sodass es gelingt, Abschnitte aus verbundenen Strängen auszuschneiden, ohne dass die Stränge mit vorab entworfenen externen Haltepunkten modifiziert werden müssten.

# Developmental Self-Assembly of a DNA Tetrahedron

John P. Sadowski,<sup>†,‡</sup> Colby R. Calvert,<sup>‡</sup> David Yu Zhang,<sup>†,§</sup> Niles A. Pierce,<sup>‡,||</sup> and Peng Yin<sup>†,§,\*</sup>

<sup>†</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, United States, <sup>‡</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States, <sup>§</sup>Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, United States, <sup>||</sup>Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125, United States, and <sup>||</sup>Division of Engineering and Applied Science, California Institute of Technology, Pasadena, California 91125, United States

**ABSTRACT** Kinetically controlled isothermal growth is fundamental to biological development, yet it remains challenging to rationally design molecular systems that self-assemble isothermally into complex geometries *via* prescribed assembly and disassembly pathways. By exploiting the programmable chemistry of base pairing, sophisticated spatial and temporal control have been demonstrated in DNA self-assembly, but largely as separate pursuits. By integrating temporal with spatial control, here we demonstrate the “developmental” self-assembly of a DNA tetrahedron, where a prescriptive molecular program orchestrates the kinetic pathways by which DNA molecules isothermally self-assemble into a well-defined three-dimensional wireframe geometry. In this reaction, nine DNA reactants initially coexist metastably, but upon catalysis by a DNA initiator molecule, navigate 24 individually characterizable intermediate states *via* prescribed assembly pathways, organized both in series and in parallel, to arrive at the tetrahedral final product. In contrast to previous work on dynamic DNA nanotechnology, this developmental program coordinates growth of ringed substructures into a three-dimensional wireframe superstructure, taking a step toward the goal of kinetically controlled isothermal growth of complex three-dimensional geometries.



**KEYWORDS:** molecular programming · DNA nanotechnology · kinetic pathway control · developmental self-assembly

Molecular self-assembly, a fundamental process underlying the development and operation of biological organisms, has emerged as an important engineering paradigm for nanotechnology. Biological development controls molecular arrangement both spatially and temporally to produce a complex organism that robustly responds to its chemical and physical environment. By contrast, it has remained challenging to achieve sophisticated spatial and temporal control in an integrated fashion in rationally designed synthetic biomolecular systems.

There have been many recent advances in the design of sophisticated synthetic nucleic acid systems that enable either spatial or temporal control of molecular self-assembly, but these two capabilities have largely been relegated to separate realms. Past work in structural DNA nanotechnology,<sup>1–3</sup> including the synthesis of ribbons,<sup>4,5</sup> tubes,<sup>5,6</sup> two- and three-dimensional extended crystals,<sup>6–11</sup> and discrete objects,<sup>12–21</sup> has largely had the goal of engineering static target structures, without

an explicit focus on controlling the assembly order and transient dynamics of how individual units come together to produce such a target structure. Conversely, past work in dynamic DNA nanotechnology,<sup>22,23</sup> including demonstrations of reconfigurable devices,<sup>24,25</sup> autonomous logical circuits,<sup>26–29</sup> dynamic self-assembling systems,<sup>28,30,31</sup> and walkers,<sup>28,32–35</sup> has typically focused on either engineering the transient interaction pattern of the individual molecular species to achieve desired computational or kinetic behavior, without explicit concern for the structural properties that the assembling and disassembling molecular species produce, or has introduced limited reconfigurability to an otherwise static structure.

To capture the complexity and robustness of molecular self-assembly demonstrated by the biological developmental process, it is necessary to design the temporal in addition to the spatial order of self-assembly. This would allow direct, molecular-scale kinetic control over the entire assembly pathway, rather than being limited to specification of the final structure only.

\* Address correspondence to py@hms.harvard.edu.

Received for review July 23, 2013 and accepted February 19, 2014.

Published online 10.1021/nn4038223

© XXXX American Chemical Society

Such a system would no longer need a thermal annealing step to initiate the assembly and encourage the dominance of the desired, lowest-energy product; the reaction could instead proceed isothermally and *in situ* with a molecular input to trigger the assembly. This would allow the construction of a "developmental" self-assembly system where a synthetic structure "grows" following prescribed kinetic assembly pathways organized both in series and in parallel, with some reactions occurring sequentially and others occurring simultaneously, ultimately developing into the desired complex target structure.

To address this largely unexplored challenge of integrating temporal control with spatial control in synthetic molecular self-assembly, we demonstrate here the rational design and kinetically controlled isothermal synthesis of a DNA tetrahedron with a well-defined three-dimensional wireframe structure. The assembly is the execution of a developmental molecular program specified using a reaction graph abstraction<sup>28</sup> that describes the kinetic pathways by which the DNA reactants self-assemble and disassemble. This work builds on a previous demonstration of programming molecular self-assembly and disassembly pathways using a versatile DNA hairpin motif, to execute diverse molecular programs including catalytic formation of branched junctions, cross-catalytic circuitry, conditional assembly of dendritic structures, and autonomous locomotion.<sup>28</sup> Unlike these previous demonstrations, the current molecular program yields a well-defined three-dimensional wireframe structure formed from closed rings. In contrast to DNA origami and tile-based approaches to structural DNA nanotechnology, our method is initiated by a catalytic molecular trigger, works isothermally within a wide temperature range, and follows a predetermined kinetic pathway. We envision that our strategy for rationally designing developmental molecular programs can be generalized to more complex three-dimensional wireframe constructions.

## RESULTS AND DISCUSSION

**Kinetic Pathway Design.** The tetrahedron assembles from six hairpins,<sup>28</sup> divided into **A** and **B** groups, and three cooperative hybridization complexes<sup>37</sup> making up the **C** group (Figure 1a). The hairpins are designed with structures that initially keep key sequence domains, called toeholds,<sup>24</sup> inaccessible; each hairpin can then be opened by a specific initiator strand, allowing the newly accessible toeholds to participate in downstream reactions. Cooperative hybridization complexes each bind to two initiators and are used in ring formation reactions. Thus, the reactants as a group are metastable, so that no reaction appreciably proceeds in the absence of the initiator, but in the presence of the initiator the assembly reaction happens autonomously and follows prescribed kinetic assembly

pathways without the need for external intervention. The final product is a three-dimensional tetrahedron with edges 18 nm long, each containing five turns of DNA (Figure 1b).

The assembly of the tetrahedron represents the execution of a prescribed molecular program. The program is depicted as a reaction graph using a nodal abstraction that concisely describes the kinetic pathways by which DNA reactants self-assemble and disassemble.<sup>28</sup> The reaction graph emphasizes the functional relationship between reactant complexes rather than the detailed structures of each reactant (Figure 1c). Each molecular species is represented by a node depicted as a black ring containing triangles representing input ports, and circles representing output ports. The initial state of a port is either 'accessible' (open symbol) or 'inaccessible' (solid symbol). The ports are functionally connected through an internal logic that toggles their states during the execution of the program, as described below. A developmental molecular program is written as a reaction graph by connecting complementary output and input ports on different nodes *via* either assembly operations (solid arrow) or disassembly operations (dashed arrow).

The reaction graph precisely defines which reactions must occur in series and which reactions may occur in parallel. For the tetrahedron, the reaction can only start with the assembly of **I** to **A1**, because this is the only assembly reaction where both participating ports (the output port on **I** and the input port on **A1**) are initially accessible. The **B1** and **A2** nodes cannot assemble with **A1** until **A1** has been opened (*via* assembly with **I**) and its output ports have become accessible, but these two assembly reactions are not dependent on each other because they are on separate branches downstream of the **A1** assembly. This is in contrast with traditional thermodynamic assemblies, where all parts of each molecule are initially accessible, with no explicit control of assembly order (Figure 1c, right).

The ports each correspond to a physical region on a strand or complex that has a single function during the assembly process. For hairpins, the single input region is represented by an orange triangle, and the two output regions are colored according to their position: the blue ports overlap the loops, while the green ports overlap the tails. The nodal abstraction is capable of representing the state of the molecular program at various points in its execution: in a hairpin assembly reaction (Figure 1d, top), the hairpin input port and the corresponding initiator output port are both initially open, indicating that a reaction is possible. After assembly, a line is drawn between these ports indicating that they are now bound, and, importantly, the two output ports on the hairpin change state from closed to open. This corresponds to the structural change, where the 1\* toehold on the single-stranded initiator



and multimeric products, which are the result of intermolecular interactions outcompeting the desired intramolecular ones (see Supporting Information Section S3.3). Cooperative hybridization complexes contain two toeholds and are assembled into the growing complex only in the presence of both initiator domains (Figure 1d, bottom). This is because the singly bound intermediate is short-lived due to the fact that neither initiator domain can fully displace the protector strand by itself, leading to quick disassembly if the second initiator is not immediately available.<sup>37</sup> In molecular terms, the depicted cooperative complex has two toeholds, 7 and 10. Upon binding to the exposed 7\* toehold in the blue initiator, the protector strand is partially displaced. This initial reaction is by itself reversible; in the absence of the green initiator, the blue initiator will rapidly disassemble. However, if the green initiator with toehold 10 is also present, the 10 and 10\* toeholds may then bind, allowing the protector strand to be completely and irreversibly displaced. In the tetrahedron, all cooperative assembly events are intended to occur intramolecularly, involving two initiator domains on the same molecule, and thus, each such assembly causes the formation of a ring. See Supporting Information Section S4 for additional details on the nodal abstraction and reaction graphs.

Figure 1e depicts an example execution trajectory for the molecular program. The top panel shows the molecular structures and the bottom panel shows the corresponding state of the molecular program using the nodal abstraction. The assembly begins with the sequential opening of the **A1**, **A2**, and **A3** hairpins causing self-assembly of a three-arm junction<sup>28</sup> corresponding to the first vertex of the tetrahedron (reactions 1–3). The initiator **I** is released upon completion of the junction (reaction 3), allowing it to catalyze the assembly of further tetrahedra. Next, each arm of this junction then extends to incorporate the **B1**, **B2**, and **B3** reactants through another set of hairpin opening reactions (reactions 4–6). Finally, the overhanging portions of the **B** strands are involved in three ring forming reactions using a set of cooperative hybridization complexes<sup>37</sup> incorporating the **C1**, **C2**, and **C3** strands (reactions 7–9).

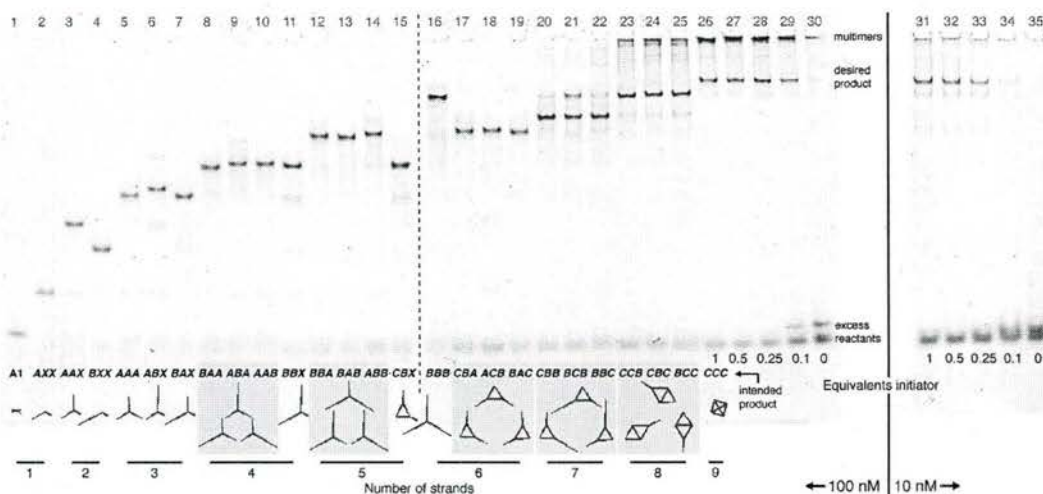
Because the assembly was designed to proceed in a branched fashion, the assembly of the three independent branches is not synchronized. Thus, there are multiple potential paths leading to the desired product, involving a total of 24 on-pathway intermediates (Figure 1f). We identify each intermediate of the tetrahedron formation with a three-letter code where each letter denotes the progress of one of the three branches of assembly. For example, in the intermediate **ABA**, **A1** has been incorporated in the first branch, **A2** and **B2** are in the second branch, and **A3** is in the third branch (note again that **A2** incorporation is a prerequisite of **B2** incorporation). Letter **X** is used to denote

the lack of assembly of the **A** strand along the corresponding branch; if the third letter is **X**, the initiator **I** is still attached (for example, **AAX** contains **A1**, **A2**, and **I**). See Supporting Information Figure S3 for a detailed description of the codes. Note that complexes that are related by cyclic permutations of their abbreviations have congruent secondary structures; these are linked by gray boxes in the figure.

**Gel Electrophoresis Studies.** We studied the formation of the tetrahedron using a gel electrophoresis mobility shift assay (Figure 2), where each intermediate was individually synthesized by mixing the initiator with different subsets of the nine reactants of the tetrahedron assembly. We were able to observe the formation of each of the 24 possible intermediates as a distinct band in a native polyacrylamide gel (lanes 2–25). Intermediates with greater numbers of incorporated strands migrated more slowly, with the exception of the ring formation transitions, which tended to have unpredictable effects (e.g., **BBX** → **CBX** and **BBA** → **CBA** showed little mobility change, and **BBB** → **CBB** showed increased mobility). This is likely because the first ring formation reduces the angle between two of the arms from its natural angle to  $\approx 60^\circ$ , making the complex more compact. Groups of intermediates with congruent structures, linked by the gray boxes in the figure, were observed to have nearly identical mobility, consistent with our expectations. The band identified as the assembled tetrahedron (lane 26) had a mobility distinct from any of the intermediates. The lower-mobility bands above the tetrahedron band are hypothesized to be multimers of the tetrahedron structure, formed when assembly of the cooperative complexes cause the intermolecular joining of two different growing tetrahedra rather than the desired intramolecular ring forming reactions (Supporting Information Figure S4).

When all reactants except **I** were incubated together (lane 30), only a small amount of unintended “leakage” products was observed: 3.5% of the signal was in the product band and 80% was in the reactants band. When at least 0.25 equiv of **I** was included (lanes 26–28), the reaction proceeded to completion (that is, all **A1** hairpins were consumed), indicating a catalytic turnover of at least 4. Typical yields of the tetrahedron varied between 20 and 40%, depending on concentration (Supporting Information Figure S5). At a higher reactant concentration (100 nM, lanes 1–30), the dominant side products were the higher-weight products, while at a lower concentration (10 nM, lanes 31–35), the dominant side products were intermediates of lower molecular weight. Varying the toehold lengths of the cooperative complexes was not observed to qualitatively improve the reaction yield (Supporting Information Figure S6).

We observed that the reaction yield was primarily bottlenecked by the ring-forming reactions. On the



**Figure 2.** Characterization of the tetrahedron assembly pathway. This is a gel electrophoresis mobility shift assay where lanes 1–25 show all on-pathway intermediates of the tetrahedron assembly, formed by mixing different subsets of reactants with the initiator. Gray boxes mark groups of intermediates that are structurally congruent and expected to have the same mobility, and the structures of the intended products are shown, as in Figure 1f. Lanes 26–35 show the analysis of catalytic turnover at two concentrations, showing reactions containing all nine reactants but varying concentrations of initiator I. These are 6% native polyacrylamide gels of assembly reactions containing 1 equiv of all reactants at the specified concentration, except A1 for which we used 0.9 equiv of a FAM fluorophore-labeled hairpin to observe incorporation yields. The initiator was included at 1 equiv unless otherwise specified in the figure. The assembly reactions were conducted at room temperature in TAE/Mg<sup>2+</sup> buffer containing 12.5 mM Mg<sup>2+</sup> over 19 h. The dotted line separates two gel slabs that were run simultaneously, and the solid line separates gels that were run at different times. The intensity of the FAM fluorescent label is shown in red, and SYBR Gold staining intensity is shown in blue; these channels are separated in Supporting Information Figure S2. See Supporting Information Figure S4 for an agarose gel of these same samples, in which the side products of higher molecular weight are well-resolved.

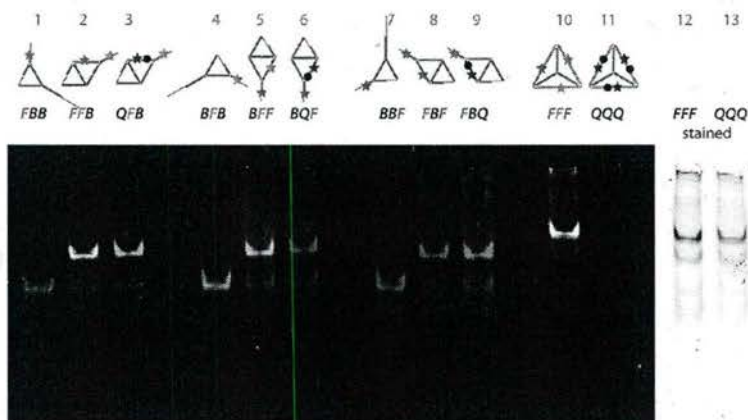
basis of analysis of Figure 2, the yield of **BBB**, the largest intermediate with no rings, was 83% for six hairpin incorporation events. However, the average yield of the seven intermediates containing one **C** strand, and thus one formed ring, dropped to 71%. With two **C** strands, the yield was reduced to 50%, and the yield of the full tetrahedron was only 19%. This drop in yield is likely the result of undesired intermolecular interactions outcompeting the desired intramolecular ring formation during the assembly of the cooperative complexes. This may be exacerbated by steric strain introduced in the ring formations. Although the single-nucleotide spacers and nicks at each junction were meant to relieve this strain, it is not clear if this was sufficient, and the use of longer single-stranded spacers at the vertices might further improve the yield.

Gel electrophoresis studies showed that the assembly is temperature-robust, working isothermally at temperatures in roughly the range 16–41 °C (Supporting Information Figure S7). At room temperature at 100 nM, the assembly reaction was observed to be complete after 9 h (Supporting Information Figure S8).

While the unique mobility of the **CCC** band indicates that all nine strands are incorporated into a single complex, additional assays are required to demonstrate ring formation and distinguish it from, for example, a large floppy three-arm junction. To demonstrate the ring formation, we conducted a

fluorescence-quenching assay (Figure 3), using **C** strands functionalized with a fluorophore and optionally a quencher. The formation of each ring caused the fluorophore from one **C** molecule to become colocalized with the quencher on the neighboring strand, quenching the fluorescent signal. For example, lane 1 contained a **CBB** complex where a red fluorophore was attached to the **C1** strand; as expected, a red fluorescent band was visible on the gel. Lane 2 contained a **CCB** complex where red and green fluorophores were, respectively, attached to the **C1** and **C2** strands; both red and green fluorescence were detected in the target band, as expected. Lane 3 showed another **CCB** complex which additionally had a quencher attached to the **C2** strand; here, green but not red fluorescence signal was detected, indicating that proper ring formation had resulted in the colocalization of the quencher on **C2** with the red fluorophore on **C1**, thus removing the red fluorescence from the product band. In the full tetrahedron, all three fluorophores were fully visible in the absence of quenchers (lane 10) but fully quenched when quenchers were added (lane 11), indicating that all three ring formations were simultaneously successful, and that the tetrahedron's overall structure formed as designed.

**Microscopy Studies.** We further confirmed the correct formation of the geometric structures of the tetrahedron and its key intermediates using atomic force microscopy (AFM). **BBB** appeared as a three-arm



**Figure 3.** Characterization of ring-forming reactions. In this fluorescence-quenching assay, each of the three C strands was functionalized with a different fluorophore at its 5' end (C1—TYE 665, red; C2—TAMRA, green; C3—FAM, blue); F represents a C strand with a fluorophore only, while Q represents a strand with a fluorophore expected to be quenched by a quencher on the 3'-end of a neighboring C strand. Lanes 1–3 show the structures *FBB* (*CCB* with a red fluorophore on the C1 strand), *FFB* (*CCB* with red and green fluorophores on the C1 and C2 strands, respectively), and *QFB* (*CCB* with the same two fluorophores, plus a 3'-quencher on the C2 strand that quenches the red 5'-fluorophore on the C1 strand). Lanes 4–6 and 7–9 show the other two structural permutations. Lane 10 shows *FFF* (the full tetrahedron with all fluorophores), and lane 11 shows *QQQ* (the full tetrahedron with all fluorophores and all quenchers). Lanes 12 and 13 are the same lanes in the same gel as lanes 10 and 11, but after staining with SYBR Gold. This is a 6% native polyacrylamide gel of a 10 nM assembly reaction with 1 equiv of initiator, conducted at room temperature in TAE/Mg<sup>2+</sup> buffer containing 12.5 mM Mg<sup>2+</sup> over 22 h.

junction, *CBB* as a single triangle, *CCB* as a double triangle, and *CCC* as a flattened tetrahedron, all consistent with their designed shapes (Figure 4a). Additionally, digestion at an intentionally engineered endonuclease restriction site on a specific edge converted a full tetrahedron back to a double triangle pattern under AFM, as expected. The double triangle structure contained two noncongruent types of vertices, which were distinguished by attaching a streptavidin molecule to one of them; digestion with either of two endonucleases targeted to two different edges yielded images of structures with the streptavidin on the expected vertex, further confirming the correct formation of the full tetrahedron (Figure 4b).

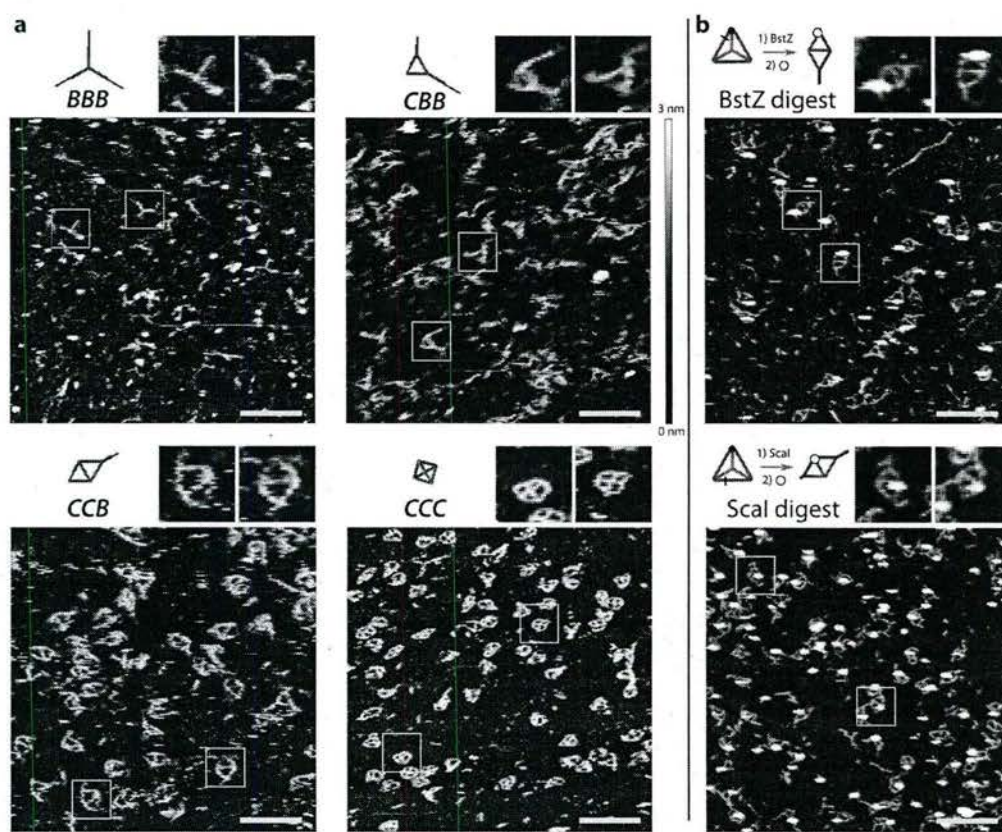
## CONCLUSIONS

We have extended the pathway-controlled DNA self-assembly approach,<sup>28</sup> demonstrating the developmental growth of a well-defined three-dimensional wireframe DNA tetrahedron, formed from metastable reactants conditionally in the presence of a single-stranded trigger and through a kinetically programmed pathway. In particular, this is the first catalytic self-assembly design that incorporates ring forming events, a potentially useful topological primitive for the growth of larger assemblies with well-defined geometries. Ring formation is challenging because the desired intramolecular reaction that forms the ring must compete with off-pathway intermolecular reactions that lead to multimers and other aggregation side products. Using cooperative hybridization complexes for the ring forming reactions appears to avoid significant difficulties encountered with a set of noncooperative designs (Supporting Information Section S3.3). We

believe that even greater improvements in yield could be achieved by designing the geometry of ring forming steps so that they do not require such large conformational changes in the molecule. This would favor the desired intramolecular reaction both thermodynamically and kinetically, because the reaction would induce less strain and would have an increased attempt frequency.

This work represents an advance over previous attempts to control assembly order. One previous approach imposes assembly order by designing different parts of the structure to have distinct melting or formation temperatures, usually by using binding domains of different lengths. The assembly events thus occur in a particular order as the temperature is lowered during the annealing process. This has been used a number of times,<sup>38,39</sup> including with some polyhedra.<sup>40,41</sup> However, such approaches typically require external modulation of the assembly environment in the form of an annealing ramp, coupling the assembly order to bulk environmental conditions such as temperature. Also, while other approaches have recently been demonstrated that work isothermally within narrow ranges of temperatures,<sup>20</sup> no explicit assembly order control is designed in these systems. In contrast, our approach allows direct, molecular-level control of assembly order through complete pathway engineering, is independent of bulk properties like temperature, and operates isothermally over a wide range of temperatures that includes biologically relevant conditions.

Another route for direct isothermal molecular control of assembly order is the algorithmic self-assembly of static DNA tiles.<sup>4,10,42</sup> Here, in a seeded growth system,



**Figure 4.** Atomic force microscopy images of tetrahedron self-assembly. (a) Images of purified samples of the intermediates *BBB*, *CBB*, *CCB*, and the full tetrahedron *CCC*, which, respectively, appeared as a three-arm junction, a single triangle, a double triangle, and a three- or four-lobed structure corresponding to a flattened tetrahedron, each consistent with our design. In each image, the double-stranded edges of each wireframe structure are clearly resolved. (b) Images of the full tetrahedron cut by restriction endonucleases followed by incubation with streptavidin. Digestion with either of two endonucleases targeted to different edges restored the double-triangle structures, with the streptavidin (appearing as small white circular features in the AFM image) appearing at the expected biotin-modified vertices. In the schematic diagrams, the black bar intersecting tetrahedron edge represents the designed restriction site, the black dot is the 5'-biotin, and the yellow circle is the streptavidin. The scale bars of the larger images are 100 nm long; the inset images are 62.5 nm in width, at double the scale of the larger images. These samples were assembled at a concentration of 100 nM and then purified by glycerol gradient ultracentrifugation; for panel b, the restriction and streptavidin binding were performed before purification.

tiles are incorporated sequentially at particular positions, as directed by cooperative binding to neighboring tiles on the growth front. However, this tile-based approach is dependent upon the supersaturated nature of the tile species, and is limited to the specific geometry of tile-based lattices with a growth front. More broadly, this tile-based approach involves “static” monomers incorporated by simple binding to the growing structure without explicitly designed internal state changes. In contrast to “passive” assembly using static tiles, our “active” developmental self-assembly uses reconfigurable hairpins as assembling monomers that are capable of conditional configurational change and can implement both assembly and disassembly. An analytical model has found that methods using active components can in principle be exponentially faster than the corresponding passive methods.<sup>43</sup> Developmental assembly with active components provides a more versatile, expressive molecular programming language

for integrating temporal and spatial control, and a potentially more efficient method for the active construction of molecular structures.

Kinetically controlled developmental self-assembly mimics the well-orchestrated nature of biological reaction cascades, which operate autonomously and without the need for external intervention to maintain their function. The triggered, isothermal nature of our assembly methodology would allow the formation of complex structures to be integrated with nucleic acid computational circuits, allowing the use of structural changes as an output mode of logical computation. These could furthermore be interfaced with non-nucleic acid inputs such as proteins and small molecules through the use of aptamers,<sup>30</sup> which would allow these systems to interact with the larger chemical world and respond in potentially intricate ways to their molecular environment. This methodology is additionally expected to be more amenable to *in vivo*

applications, where thermal annealing is not an available mode of assembly. Thus, synthetic developmental

self-assembly promises to open new doors to bridge computation, chemistry, and biology.

## EXPERIMENTAL METHODS

**Sequence Design.** Motif design was performed by hand using principles described previously<sup>28</sup> and in Supporting Information Section S3.1. Sequence design was performed with Multisubjective (version 1.0.2), a program that identifies and eliminates spurious hybridizations in the full hairpin system (manuscript in preparation), in association with Domain Design (DD) (version 0.2), which uses simple heuristics to evaluate the acceptability of candidate sequence domains.<sup>44</sup> A local copy of NUPACK (version 3.0) was used by Multisubjective to generate base-pairing probabilities for further analysis.<sup>45,46</sup> See Supporting Information Section S3.2 for more sequence design details. See Supporting Information Section S5 for designed sequences.

**Strand Synthesis.** DNA strands were synthesized and purified by Integrated DNA Technologies (IDT), including strands containing phosphates and fluorophores. For the fluorescent C strands used in the fluorescence-quenching assay only, we ordered the strands in two halves, ligated them using 27 U/ $\mu$ L of T4 DNA ligase at a DNA concentration of 27  $\mu$ M at 16 °C for 2 h, and purified the product by denaturing polyacrylamide gel electrophoresis. In all cases, we quantitated the concentrations of DNA stock solutions by measuring the ultraviolet light absorption at 260 nm with the microvolume pedestal of a NanoDrop 2000c spectrophotometer, taking the average of measurements of three samples for each strand, and using extinction coefficients provided by IDT to calculate the concentration of each strand.

**Sample Preparation.** The hairpin strands were separately heated to 95 °C for 5 min and allowed to cool to room temperature over about 20 min. Cooperative hybridization complexes were annealed in the same way, except that 1.5 equiv of the C<sub>p</sub> strands was combined with 1 equiv of the C strand before annealing. The desired reactants were then combined to a final concentration of 100 nM (Figure 2, lanes 1–30 and Figure 4) or 10 nM (Figure 2, lanes 31–35 and Figure 3) of each reactant except A1, for which 0.9 equiv was used instead in order to obviate the effects of inaccuracies in stoichiometry and assist in calculation of the yield. The assembly reactions were performed at room temperature over about 20 h in TAE/Mg<sup>2+</sup> buffer containing 40 mM Tris base, 20 mM acetic acid, 12.7 mM EDTA, and 12.5 mM MgCl<sub>2</sub>.

For the formation assay (Figure 2), an A1 strand incorporating a fluorescent label was used instead of the regular A1 strand. For the fluorescence-quenching assay (Figure 3), the C strands each had a different fluorophore on one end and an optional quencher on the other, as described in the figure caption. For the CCC samples in the atomic force microscopy studies (Figure 4a), we used C strands that contained phosphates on their 5'-ends; this was not expected to affect the assembly. For the streptavidin labeling experiment (Figure 4b), we used an A2 strand containing a 5' biotin.

For the atomic force microscope studies, the samples were purified by ultracentrifugation in a 15–45% glycerol gradient in 1 × TAE/Mg<sup>2+</sup> for 3 h at 50 000 rpm. Fractions of 100  $\mu$ L were taken from the centrifuge tube, and the fractions containing the desired product were identified by native gel electrophoresis. The glycerol was not removed after the purifications, meaning that subsequent manipulations were performed in buffer containing about 20% glycerol.

**Gel Electrophoresis.** The gels in Figures 2 and 3 were precast 6% native polyacrylamide gels run at 100 V for 55 min at 21 °C using 1 × TBE running buffer. For each lane, 4  $\mu$ L of the reaction mixture was mixed with an equal amount of 2 × native loading buffer containing a small amount of bromophenol blue in 2 × TAE/Mg<sup>2+</sup> buffer and 10% glycerol. Then, 4  $\mu$ L of this mixture containing 0.2 pmol of the specified complex was then loaded into the gel.

The gels were imaged with a Typhoon FLA 9000 gel scanner. The gels in Figure 2 and the right side inset of Figure 3 were

stained in SYBR Gold for about 20 min before imaging; all other gels were unstained. For Figure 2, we imaged the gel unstained using the fluorescent label, and then stained it and reimaged the same gel. The two images were then manually overlaid.

We quantitated the yield by dividing the fluorescence intensity of the desired product by the intensity of the entire lane in the fluorescently labeled channel (shown as red) in Figure 2, using ImageQuant TL. We used automatic band detection for the bands representing the desired product, and the reactants band if it was present. We then defined multimers as the lane area with lower mobility than the desired product band, and intermediates as the lane area between the product and reactants band. For background subtraction, we used the rolling ball method with a radius of 500.

**Atomic Force Microscopy.** The samples in Figure 4a were imaged directly after glycerol purification. For the streptavidin labeling experiment (Figure 4b), the appropriate digestion buffer was added after assembly to a final concentration of 1 ×, and the structures were digested with one of the two restriction endonucleases BstZ17I or Scal-HF at a concentration of 1 U/ $\mu$ L at 37 °C for 2 h. The structures were then incubated with a 10 × solution of streptavidin for 30 min and then glycerol purified.

For imaging, we added to a freshly cleaved mica surface, 30  $\mu$ L of filtered 5 × TAE/Mg<sup>2+</sup> buffer, followed by 30  $\mu$ L of a 10 mM solution of NiCl<sub>2</sub> to increase the strength of the DNA–mica binding. After 5 min, we added 10–30  $\mu$ L of the glycerol-purification fraction containing our desired product. AFM images were obtained using a Multimode 8 scanning probe microscope with a Digital Instruments Nanoscope V controller. Images were collected in aqueous phase using tapping mode, using the short and thin cantilevers in the SNL-10 silicon nitride cantilever chip. Within each of the two panels of Figure 4, the images were all generated using the same tip.

**Conflict of Interest:** The authors declare no competing financial interest.

**Acknowledgment.** The authors thank C. Grun and E. Winfree for discussions, and D. Pastuszak for help with draft preparation. This work was funded by Office of Naval Research Grants N000141010827 and N000141310593, Office of Naval Research Young Investigator Program Award N000141110914, NIH Director's New Innovator Award 1DP2OD007292, NSF CAREER Award CCF1054898, NSF Grant CCF1162459, NSF Expedition in Computing Award CCF1317291, and Wyss Institute for Biologically Inspired Engineering Faculty Startup Fund to P.Y., and NIH 5R01CA140759, NIH P50 HG004071, the Molecular Programming Project (NSF-CCF-0832824 and NSF-CCF-1317694), and the Gordon and Betty Moore Foundation (GBMF2809) to N.A.P.; J.P.S. acknowledges a Graduate Research Fellowship from NSF and a J. Marshall & Jane H. Booker Graduate Scholarship from the Buttonwood Foundation, and D.Y.Z. is supported by NIH Transition to Independence Award 1K99EB015331.

**Supporting Information Available:** Expanded versions of figures, additional experimental characterization of the tetrahedron assembly, discussion of design considerations, experimental characterization of alternate tetrahedron structures, discussion of the nodal reaction graph abstraction, and strand sequences. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## REFERENCES AND NOTES

- Seeman, N. C. Nucleic Acid Junctions and Lattices. *J. Theor. Biol.* **1982**, *99*, 237–247.
- Seeman, N. C. DNA in a Material World. *Nature* **2003**, *421*, 427–431.

3. Shih, W. M.; Lin, C. Knitting Complex Weaves with DNA Origami. *Curr. Opin. Struct. Biol.* **2010**, *20*, 276–282.
4. Schulman, R.; Winfree, E. Synthesis of Crystals with a Programmable Kinetic Barrier to Nucleation. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15236–15241.
5. Yin, P.; Hariadi, R.; Sahu, S.; Choi, H. M. T.; Park, S. H.; LaBean, T. H.; Reif, J. H. Programming Molecular Tube Circumferences. *Science* **2008**, *321*, 824–826.
6. Yan, H.; Park, S. H.; Finkelstein, G.; Reif, J. H.; LaBean, T. H. DNA-Templated Self-Assembly of Protein Arrays and Highly Conductive Nanowires. *Science* **2003**, *301*, 1882–1884.
7. Winfree, E.; Liu, F.; Wenzler, L. A.; Seeman, N. C. Design and Self-Assembly of Two-Dimensional DNA Crystals. *Nature* **1998**, *394*, 539–544.
8. Yan, H.; LaBean, T. H.; Feng, L.; Reif, J. H. Directed Nucleation Assembly of DNA Tile Complexes for Barcode-Patterned Lattices. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 8103–8108.
9. Liu, D.; Wang, M.; Deng, Z.; Walulu, R.; Mao, C. Tensegrity: Construction of Rigid DNA Triangles with Flexible Four-Arm DNA Junctions. *J. Am. Chem. Soc.* **2004**, *126*, 2324–2325.
10. Rothmund, P. W. K.; Papadakis, N.; Winfree, E. Algorithmic Self-Assembly of DNA Sierpinski Triangles. *PLoS Biol.* **2004**, *2*, 2041–2053.
11. Zheng, J. P.; Birktoft, J.; Chen, Y.; Wang, T.; Sha, R. J.; Constantinou, P.; Ginell, S.; Mao, C. D.; Seeman, N. C. From Molecular to Macroscopic via the Rational Design of a Self-Assembled 3D DNA Crystal. *Nature* **2009**, *461*, 74–77.
12. Chen, J.; Seeman, N. C. The Synthesis from DNA of a Molecule with the Connectivity of a Cube. *Nature* **1991**, *350*, 631–633.
13. Rothmund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440*, 297–302.
14. He, Y.; Ye, T.; Su, M.; Zhang, C.; Ribbe, A. E.; Jiang, W.; Mao, C. Hierarchical Self-Assembly of DNA into Symmetric Supramolecular Polyhedra. *Nature* **2008**, *452*, 198–201.
15. Douglas, S. M.; Dietz, H.; Liedl, T.; Högberg, B.; Graf, F.; Shih, W. M. Self-Assembly of DNA into Nanoscale Three-Dimensional Shapes. *Nature* **2009**, *459*, 414–418.
16. Dietz, H.; Douglas, S. M.; Shih, W. M. Folding DNA into Twisted and Curved Nanoscale Shapes. *Science* **2009**, *325*, 725–730.
17. Han, D.; Pal, S.; Nangreave, J.; Deng, Z.; Liu, Y.; Yan, H. DNA Origami with Complex Curvatures in Three-Dimensional Space. *Science* **2011**, *332*, 342–346.
18. Wei, B.; Dai, M.; Yin, P. Complex Shapes Self-Assembled from Single-Stranded DNA Tiles. *Nature* **2012**, *485*, 623–626.
19. Ke, Y.; Ong, L.; Shih, W.; Yin, P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* **2012**, *338*, 1177–1183.
20. Sobczak, J.; Martin, T.; Gerling, T.; Dietz, H. Rapid Folding of DNA into Nanoscale Shapes at Constant Temperature. *Science* **2012**, *338*, 1458–1461.
21. Han, D.; Pal, S.; Yang, Y.; Jiang, S.; Nangreave, J.; Liu, Y.; Yan, H. DNA Gridiron Nanostructures Based on Four-Arm Junctions. *Science* **2013**, *339*, 1412–1415.
22. Zhang, D. Y.; Seelig, G. Dynamic DNA Nanotechnology Using Strand-Displacement Reactions. *Nature Chem.* **2011**, *3*, 103–113.
23. Bath, J.; Turberfield, A. J. DNA Nanomachines. *Nat. Nanotechnol.* **2007**, *2*, 275–284.
24. Yurke, B.; Turberfield, A. J.; Mills, A. P.; Simmel, F. C.; Neumann, J. L. A DNA-Fuelled Molecular Machine Made of DNA. *Nature* **2000**, *406*, 605–608.
25. Yan, H.; Zhang, Z.; Shen, X.; Seeman, N. C. A Robust DNA Mechanical Device Controlled by Hybridization Topology. *Nature* **2002**, *415*, 62–65.
26. Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. Enzyme-Free Nucleic Acid Logic Circuits. *Science* **2006**, *314*, 1585–1588.
27. Zhang, D.; Turberfield, A.; Yurke, B.; Winfree, E. Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA. *Science* **2007**, *318*, 1121–1125.
28. Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A. Programming Biomolecular Self-Assembly Pathways. *Nature* **2008**, *451*, 318–322.
29. Qian, L.; Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **2011**, *332*, 1196–1201.
30. Dirks, R.; Pierce, N. Triggered Amplification by Hybridization Chain Reaction. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15275–15278.
31. Venkataraman, S.; Dirks, R.; Rothmund, P.; Winfree, E.; Pierce, N. An Autonomous Polymerization Motor Powered by DNA Hybridization. *Nat. Nanotechnol.* **2007**, *2*, 490–494.
32. Sherman, W. B.; Seeman, N. C. A Precisely Controlled DNA Biped Walking Device. *Nano Lett.* **2004**, *4*, 1203–1207.
33. Shin, J.-S.; Pierce, N. A Synthetic DNA Walker for Molecular Transport. *J. Am. Chem. Soc.* **2004**, *126*, 10834–10835.
34. Yin, P.; Yan, H.; Daniell, X.; Turberfield, A. J.; Reif, J. A Unidirectional DNA Walker That Moves Autonomously along a Track. *Angew. Chem. Int. Ed.* **2004**, *43*, 4906–4911.
35. Omabegho, T.; Sha, R.; Seeman, N. C. A Bipedal DNA Brownian Motor with Coordinated Legs. *Science* **2009**, *324*, 67–71.
36. Nanorex, Inc. Nanoengineer-1. <http://nanoengineer-1.com/> The original website is no longer available as of this writing; an archived version is available at <http://web.archive.org/web/20121218/http://www.nanoengineer-1.com/content/>, accessed April 8, 2014.
37. Zhang, D. Y. Cooperative Hybridization of Oligonucleotides. *J. Am. Chem. Soc.* **2011**, *133*, 1077–1086.
38. Kumara, M. T.; Nykypanchuk, D.; Sherman, W. B. Assembly Pathway Analysis of DNA Nanostructures and the Construction of Parallel Motifs. *Nano Lett.* **2008**, *8*, 1971–1977.
39. Snyder, T. M.; Liu, D. R. Ordered Multistep Synthesis in a Single Solution Directed by DNA Templates. *Angew. Chem. Int. Ed.* **2005**, *44*, 7379–7382.
40. Shih, W.; Quispe, J.; Joyce, G. A 1.7-Kilobase Single-Stranded DNA That Folds into a Nanoscale Octahedron. *Nature* **2004**, *427*, 618–621.
41. Goodman, R.; Schaap, I.; Tardin, C.; Erben, C.; Berry, R.; Schmidt, C.; Turberfield, A. Rapid Chiral Assembly of Rigid DNA Building Blocks for Molecular Nanofabrication. *Science* **2005**, *310*, 1661–1665.
42. Barish, R. D.; Schulman, R.; Rothmund, P. W. K.; Winfree, E. An Information-Bearing Seed for Nucleating Algorithmic Self-Assembly. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6054–6059.
43. Woods, D.; Chen, H.-L.; Goodfriend, S.; Dabby, N.; Winfree, E.; Yin, P. Active Self-Assembly of Algorithmic Shapes and Patterns in Polylogarithmic Time. In *ITCS 2013: Innovations in Theoretical Computer Science*; Association for Computing Machinery: Berkeley, CA, 2013; pp 353–354.
44. Zhang, D. Y. Towards Domain-Based Sequence Design for DNA Strand Displacement Reactions. *Lect. Notes Comput. Sci.* **2011**, *6518*, 162–175.
45. Dirks, R. M.; Bois, J. S.; Schaeffer, J. M.; Winfree, E.; Pierce, N. A. Thermodynamic Analysis of Interacting Nucleic Acid Strands. *SIAM Rev.* **2007**, *49*, 65–88.
46. Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. NUPACK: Analysis and Design of Nucleic Acid Systems. *J. Comput. Chem.* **2011**, *32*, 170–173.

# Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT

Ralf Jungmann<sup>1,2,6</sup>, Maier S Avendaño<sup>1,2,6</sup>, Johannes B Woehrstein<sup>1,6</sup>, Mingjie Dai<sup>1,3</sup>, William M Shih<sup>1,4,5</sup> & Peng Yin<sup>1,2</sup>

Super-resolution fluorescence microscopy is a powerful tool for biological research, but obtaining multiplexed images for a large number of distinct target species remains challenging. Here we use the transient binding of short fluorescently labeled oligonucleotides (DNA-PAINT, a variation of point accumulation for imaging in nanoscale topography) for simple and easy-to-implement multiplexed super-resolution imaging that achieves sub-10-nm spatial resolution *in vitro* on synthetic DNA structures. We also report a multiplexing approach (Exchange-PAINT) that allows sequential imaging of multiple targets using only a single dye and a single laser source. We experimentally demonstrate ten-color super-resolution imaging *in vitro* on synthetic DNA structures as well as four-color two-dimensional (2D) imaging and three-color 3D imaging of proteins in fixed cells.

Far-field fluorescence microscopy has undergone major advances since the advent of methods circumventing the classical diffraction limit, i.e., super-resolution microscopy<sup>1,2</sup>. Most implementations 'switch' molecules between fluorescence on- and off-states to obtain subdiffraction image resolution. This switching is traditionally obtained in two ways: targeted switching actively confines the fluorescence excitation to an area smaller than the diffraction limit of light (for example, in stimulated emission depletion, or STED, microscopy<sup>3</sup>), whereas stochastic switching uses photoswitchable proteins or photoswitchable organic dyes (for example, in photoactivated localization microscopy (PALM)<sup>4</sup> and stochastic optical reconstruction microscopy (STORM)<sup>1</sup>). Although these methods offer unprecedented spatial resolution, they tend to be technically involved to implement, and multiplexing for a large number of distinct targets is generally challenging.

PAINT<sup>5-7</sup> provides an alternative stochastic super-resolution imaging method. Here imaging is carried out using diffusing fluorescent molecules that interact transiently with the sample. This method is straightforward to implement and does not require specialized equipment or conditions to obtain photoswitching, thus making it more accessible than STED or STORM for laboratories with standard instrumentation and sample-preparation

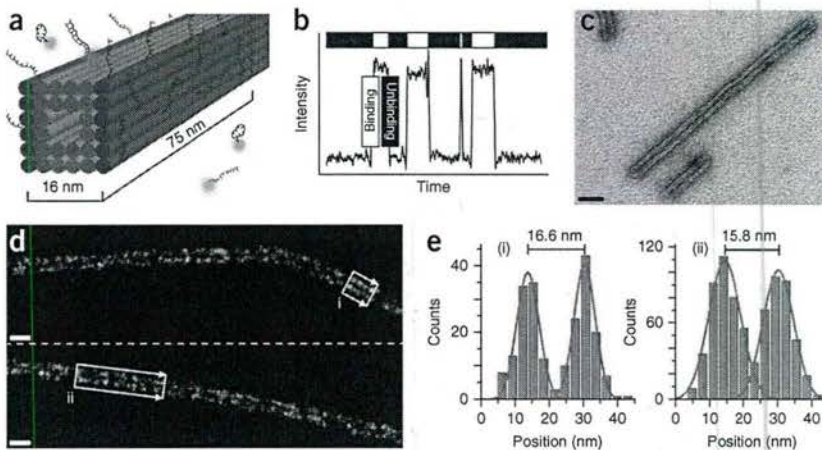
capabilities. Initially, PAINT was applied to obtain super-resolved images of cell membranes<sup>5</sup> and artificial lipid vesicles<sup>5</sup>. However, a key limitation of PAINT's original formulation is that dyes interact with the sample via electrostatic coupling or hydrophobic interactions. This limits the availability of PAINT-compatible dyes, making it hard to simultaneously image specific biomolecules of interest. A recent implementation of PAINT has involved continuously and stochastically labeling specific membrane biomolecules with fluorescent ligands (such as antibodies)<sup>6</sup>. The approach, termed universal PAINT, achieves specific dye-sample interactions but still lacks the ability to specify interactions with programmable kinetics. Similarly to PAINT, binding of DNA intercalating dyes has also been used to obtain super-resolved images of DNA<sup>8,9</sup>.

To achieve programmable dye interactions and to increase the specificity and the number of usable fluorophores, DNA-PAINT was developed<sup>10</sup>. Here stochastic switching between fluorescence on- and off-states is implemented via repetitive, transient binding of fluorescently labeled oligonucleotides ('imager' strands) to complementary 'docking' strands on DNA nanostructures (Fig. 1a,b). In the unbound state, only background fluorescence from partially quenched<sup>10</sup> imager strands is observed (Fig. 1a). However, upon binding and immobilization of an imager strand, fluorescence emission is detected using total-internal-reflection (TIR) or highly inclined and laminated optical sheet (HILO) microscopy<sup>11</sup>. DNA-PAINT enhances PAINT's simplicity and ease of use with the programmability and specificity of DNA hybridization. Notably, it enables a wide range of fluorescence on- and off-times; these can be adjusted by tuning the binding strength and concentration of the imager strand<sup>10</sup>. DNA-PAINT has been used to obtain multicolor subdiffraction images of DNA nanostructures<sup>10,12-15</sup> with ~25-nm spatial resolution<sup>14</sup>. Spectral multiplexing is straightforward as no external photoswitching of dyes is necessary, and imaging specificity is obtained through orthogonality of DNA sequences coupled to spectrally distinct dyes<sup>13</sup>.

By linking DNA-PAINT docking strands to antibodies, we have extended the DNA-PAINT method to enable multiplexed 2D and

<sup>1</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts, USA. <sup>2</sup>Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA. <sup>3</sup>Program in Biophysics, Harvard University, Cambridge, Massachusetts, USA. <sup>4</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA. <sup>5</sup>Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. <sup>6</sup>These authors contributed equally to this work. Correspondence should be addressed to P.Y. (py@hms.harvard.edu).

**Figure 1 | DNA-PAINT.** (a) A microtubule-like DNA origami polymer (cylinders represent DNA double helices) is decorated with single-stranded extensions (docking strands) on two opposite faces (red) spaced  $\sim 16$  nm apart. Complementary fluorescent imager strands transiently bind from solution to docking strands. Biotinylated strands (present on orange helices) immobilize the structures to glass surfaces for fluorescence imaging. (b) Transient binding between imager and docking strands produces fluorescence blinking, allowing stochastic super-resolution imaging. (c) Transmission electron microscopy image of origami polymers with a measured width of  $16 \pm 1$  nm (mean  $\pm$  s.d.). (d) DNA-PAINT super-resolution images obtained using Cy3b-labeled imager strands (15,000 frames, 5-Hz frame rate). Two distinct lines are visible. (e) Cross-sectional histograms of the boxed areas i and ii in d (arrows denote histogram direction) both reveal a distance of  $\sim 16$  nm as designed (FWHM of each distribution is  $\sim 7$ – $10$  nm). Scale bars, 40 nm.



3D super-resolution imaging of protein components in fixed cells. We report sub-10-nm lateral imaging resolution *in vitro* of synthetic DNA structures without the use of a sophisticated setup (in contrast with STED<sup>3</sup> or dual-objective STORM<sup>16</sup>) or specialized experimental conditions such as dye-caging approaches<sup>17</sup>. We used the unique programmability of DNA molecules to perform sequential multiplexing (with Exchange-PAINT) using only a single fluorescent dye and obtained the first ten-color *in vitro* super-resolution image on DNA nanostructures. We also show the applicability of Exchange-PAINT to cellular imaging by demonstrating four-color imaging of protein targets in fixed cells and three-color 3D imaging.

## RESULTS

### Sub-10-nm *in vitro* imaging with DNA-PAINT

First we optimized the spatial resolution of DNA-PAINT relative to that of earlier DNA-PAINT studies<sup>10,13–15</sup> through enhanced drift correction and higher localization accuracy. We achieved the latter by collecting more emitted photons per binding event through optimization of fluorescence on-times and camera integration time and by increasing laser excitation intensities. Extracting more photons per binding event was also facilitated by the fact that DNA-PAINT imaging is not prone to photobleaching, as imager strands are continuously replenished from solution<sup>10</sup>. This replenishing also allows nearly 100% imaging efficiency of all docking sites<sup>10</sup>. This imaging efficiency of docking sites does not necessarily translate to imaging efficiency of targets, as the labeling efficiency of targets with docking strands may not be 100%. However, if present and accessible, every docking site should eventually be imaged during data acquisition<sup>10</sup>.

We evaluated imaging performance by visualizing DNA origami structures that mimic *in vitro*-assembled microtubules (Fig. 1a). In DNA origami, a long ssDNA molecule (the ‘scaffold’) is ‘folded’ into a desired shape or pattern by the sequence-specific binding of hundreds of short oligonucleotides (the ‘staple’ strands) to designated regions on the scaffold<sup>18,19</sup> (Supplementary Fig. 1). The DNA origami monomer is a tunnel-like structure (with dimensions of  $\sim 16$  nm  $\times$   $\sim 16$  nm  $\times$   $\sim 75$  nm). We linked monomers using connector strands to form a homopolymer mimicking microtubules (Fig. 1c), which formed with high yield

(Supplementary Fig. 2). To allow for DNA-PAINT imaging, we extended staple strands on two opposite faces of the structure by adding single-stranded docking sites at the 3’ end. We bound assembled polymers to a BSA-biotin-streptavidin glass surface using biotinylated staple strands extruding from the bottom of the structure<sup>10,13,14,20</sup>. DNA-PAINT was then performed using Cy3b-labeled imager strands.

The super-resolved image revealed two adjacent lines spaced  $\sim 16$  nm apart, a geometry matching that of the designed microtubule-like origami (Fig. 1d). The cross-sectional profiles of two regions of interest possessed well-separated peaks at the designed distance with a full width at half maximum (FWHM) of  $\sim 7$ – $10$  nm (Fig. 1e and Supplementary Fig. 3). We note that this resolution was obtained in standard DNA hybridization buffer without the use of oxygen scavengers<sup>21</sup>, triplet-state quenchers<sup>22</sup> or redox systems<sup>23</sup>.

Straightforward extension to multicolor imaging was obtained by coupling spectrally distinct dyes to orthogonal imager strand sequences as reported previously<sup>13</sup> (Supplementary Fig. 4). The orthogonality of the imager strand sequences prevents cross-talk between different color channels (Supplementary Fig. 4b–d).

### Multiplexed cellular imaging with DNA-PAINT

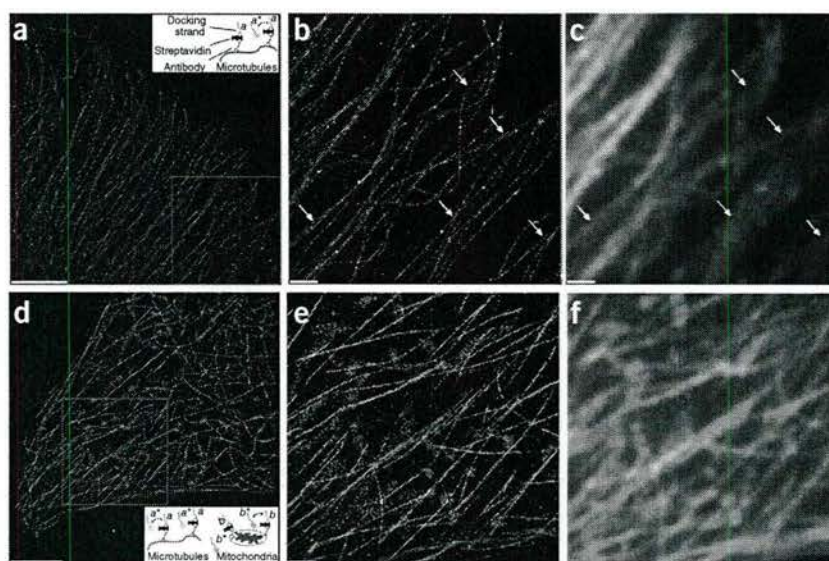
To image cellular components with DNA-PAINT, we specifically labeled protein targets using antibodies conjugated with DNA docking strands (Fig. 2). We formed the conjugates by first reacting biotinylated docking strands with streptavidin and then incubating with biotinylated antibodies. We first immunostained fixed HeLa cells using a preassembled antibody-DNA conjugate against  $\beta$ -tubulin. Atto 655-labeled imager strands were then introduced, and imaging was performed using HILO microscopy<sup>11</sup>. The resulting super-resolution images showed a clear increase in spatial resolution over that of the diffraction-limited representation (Fig. 2a–c). A cross-sectional profile yielded a distance of  $\sim 79$  nm between two adjacent microtubules with an apparent width of  $\sim 47$  and  $\sim 44$  nm for each (Supplementary Fig. 5), dimensions consistent with those in earlier reports<sup>24</sup>. We observed little to no nonspecific binding of imager strands to nonlabeled cellular components (see below for quantitative characterization and discussion of nonspecific binding).

**Figure 2** | Spectrally multiplexed DNA-PAINT super-resolution imaging of microtubules and mitochondria inside fixed cells. (a) DNA-PAINT super-resolution image of microtubules inside a fixed HeLa cell using Atto 655-labeled imager strands (10,000 frames, 10-Hz frame rate). Inset, labeling and imaging schematic for DNA-PAINT in a cellular environment.

Microtubules are labeled with a preassembled antibody-DNA conjugate, which is formed between a biotinylated anti-tubulin antibody and a biotinylated DNA docking strand (a) using a streptavidin bridge. *a\** indicates the complementary imager strand. (b) Zoomed-in view of the boxed area in a. (c) Diffraction-limited representation of the area in b. Arrows in b and c highlight positions where the increase in resolution of the DNA-PAINT image is clearly visible. Adjacent microtubules with an apparent width of ~46 nm at position i are separated by ~79 nm (see also **Supplementary Fig. 5**).

(d) Dual-color DNA-PAINT super-resolution image (15,000 frames, 10-Hz frame rate) of microtubules and mitochondria inside a fixed HeLa cell obtained using Cy3b-labeled imager strands (*a\**) for microtubules (green) and orthogonal

Atto 655-labeled imager strands (*b\**) for mitochondria (purple). Inset, labeling and imaging schematic. (e) Zoomed-in view of the boxed area in d. (f) Diffraction-limited image of the area in e. Scale bars, 5  $\mu\text{m}$  (a,d) and 1  $\mu\text{m}$  (b,c,e,f).



We achieved multicolor imaging by using orthogonal imager strands coupled to spectrally distinct dyes. We labeled microtubules in a fixed HeLa cell with a preassembled antibody-DNA conjugate carrying a docking sequence for Cy3b-labeled imager strands, and we stained mitochondria using a second antibody linked to an orthogonal sequence for Atto 655-labeled imager strands (Fig. 2d). Although Cy3b- and Atto 655-labeled imager strands coexisted in solution, imaging was performed sequentially. Images were drift corrected, and we aligned different color channels using gold nanoparticles as fiducial markers (Online Methods). The resulting super-resolution images showed a clear increase in spatial resolution as compared to the diffraction-limited representation, and no cross-talk between colors was observed (Fig. 2d-f).

### Ten-color super-resolution imaging with Exchange-PAINT

As imager strands only transiently bind to the docking strands, DNA-PAINT allows for a new multiplexing approach wherein orthogonal imager strands are sequentially applied to the same sample. This approach, which we call Exchange-PAINT, is depicted in **Figure 3a**. Initially, different target species are labeled with orthogonal docking strands. Once all components are labeled, the first imager strand species  $P1^*$  (complementary to docking strands  $P1$ ) is introduced and a DNA-PAINT image is acquired only for the targets labeled with  $P1$ . In a subsequent washing step, imager strands  $P1^*$  are removed and imager strands  $P2^*$  are introduced. Another image for only  $P2$  is then acquired. In each imaging step, the respective docking sites are super-resolved and a unique pseudocolor is assigned. Washing and imaging steps are repeated until all desired targets are imaged. These images are then aligned and combined to produce the final multicolor image for the entire sample.

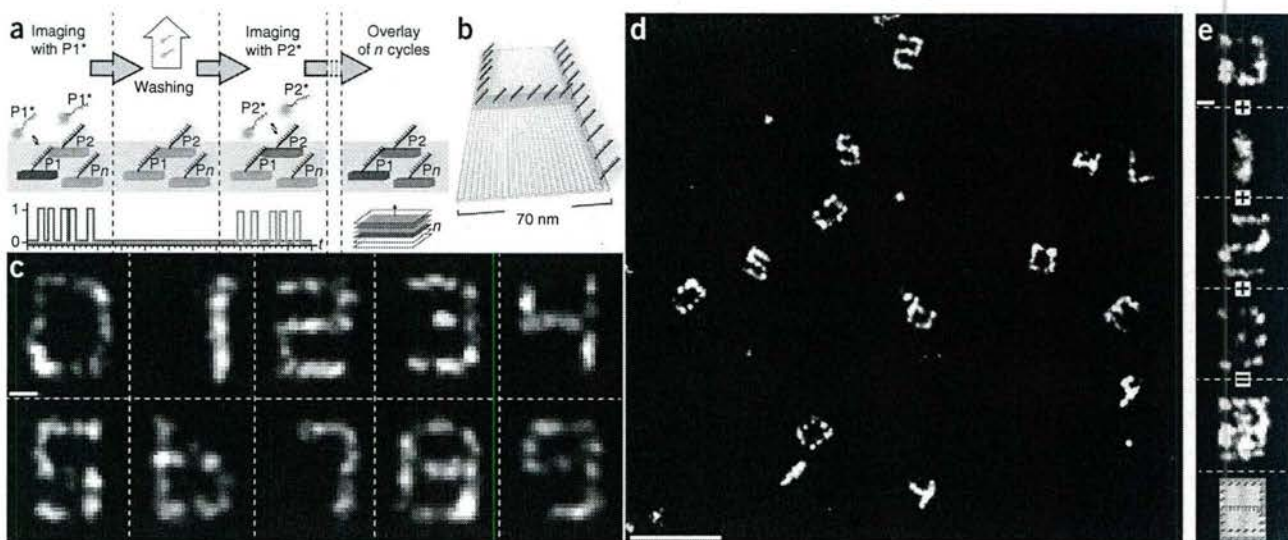
In contrast with the spectral multiplexing approach described in the previous section, in Exchange-PAINT, the same dye—and, hence, the same laser—is used for all the target species. As such, multiplexing is limited only by the number of possible orthogonal

DNA docking sequences instead of the number of spectrally distinct dyes (a typical limit for most previous fluorescence imaging methods).

To demonstrate ten-color super-resolution imaging of DNA structures using Exchange-PAINT, we designed ten unique rectangular DNA origami shapes<sup>18</sup>, each displaying a distinct pattern of orthogonal docking strands that resembles a digit between 0 and 9 (Fig. 3b). After surface immobilization of all ten structures, we performed sequential imaging using a custom-made fluidic chamber (**Supplementary Fig. 6a**) for easy liquid handling. Ten orthogonal imager strands ( $P1^*$  to  $P10^*$ ), all labeled with Cy3b, were used to perform Exchange-PAINT. The resulting digits from all ten imaging rounds are shown in **Figure 3c**. Each target was resolved with high spatial resolution. Cross-sectional histograms along the bars of the digits showed sub-10-nm FWHM of the distributions (data not shown). Note that high resolution was maintained for all digits, as the same optimized dye (Cy3b) and imaging conditions were used in each cycle.

A combined image of all ten rounds (Fig. 3d and **Supplementary Fig. 7**) demonstrates specific interaction of imager strands with respective targets with no observable cross-talk between cycles. An apparent green digit 5 instead of 2 was observed (Fig. 3d). This is likely not a falsely imaged digit 5 from cross-talk but rather a mirrored digit 2 (**Supplementary Fig. 8**). A mirrored image likely results from an origami immobilized upside down<sup>25</sup>, with docking strands trapped underneath yet still accessible to imager strands.

The fluidic setup is designed to minimize sample movement by decoupling the fluid reservoir and syringe from the actual flow chamber via flexible tubing. To avoid sample distortion, we took special care to ensure gentle fluid flow during washing steps. To verify that the sample indeed exhibited little movement and little to no distortion, we performed a ten-round Exchange-PAINT experiment. We imaged the DNA origami for digit 4 in the first round and reimaged it after ten rounds of buffer exchange. The total sample movement (physical movement of the fluidic



**Figure 3** | Exchange-PAINT. (a) Exchange-PAINT schematic showing sequential imaging of multiple targets using imager strands labeled with the same fluorophore. (b) Schematic of a DNA origami pattern (70 nm × 100 nm) displaying docking strands that resemble the digit “4.” (c) Pseudocolor images of ten different origami structures displaying digits 0–9 in one sample with high resolution (FWHM of bar-like features <10 nm) and specificity. The images were obtained using only one fluorophore (Cy3b) through ten imaging-washing cycles (imaging: 7,500 frames per cycle, 5-Hz frame rate; washing: 1–2 min per cycle). (d) Combined overview image of all ten Exchange-PAINT cycles, demonstrating specific interaction with the respective target with no cross-talk between imaging cycles. (e) Four-color image of digits 0–3 that are all present on the same DNA origami (10,000 frames each, 5-Hz frame rate; schematic at the bottom). Scale bars, 25 nm (c,e) and 250 nm (d).

chamber with respect to the objective) was less than 2  $\mu\text{m}$ , which could easily be corrected using fiducial markers. Normalized cross-correlation analysis for select structures produced a correlation coefficient of 0.92 (Supplementary Fig. 9), a result indicative of almost no sample distortion (see also the Discussion below).

Finally, using Exchange-PAINT, we successfully imaged four different digit patterns on the same DNA origami structure (Fig. 3e and Supplementary Fig. 10). Thus, Exchange-PAINT is not limited to spatially separate species and can resolve subdiffraction patterns on the same structure with no observable cross-talk or sample distortion. Aligning images from different Exchange-PAINT rounds is straightforward using DNA origami-based drift markers. Additionally, because imaging is performed using the same dye, no chromatic aberration needs to be corrected between imaging rounds.

### Multiplexed cellular imaging with Exchange-PAINT

We next demonstrated multiplexed *in situ* imaging in a fixed HeLa cell with Exchange-PAINT. Figure 4a shows four-color super-resolution images obtained by sequential imaging using only a single dye (Atto 655). Using custom-made fluidic chambers (Supplementary Fig. 6b), we obtained super-resolution images of  $\beta$ -tubulin in microtubules, COX IV in mitochondria, TGN46 in the Golgi complex and PMP70 in peroxisomes. Imaging and washing was performed in a similar fashion as for DNA structures. Again, we saw little to no nonspecific binding of the imager strands to nonlabeled components (Supplementary Fig. 11).

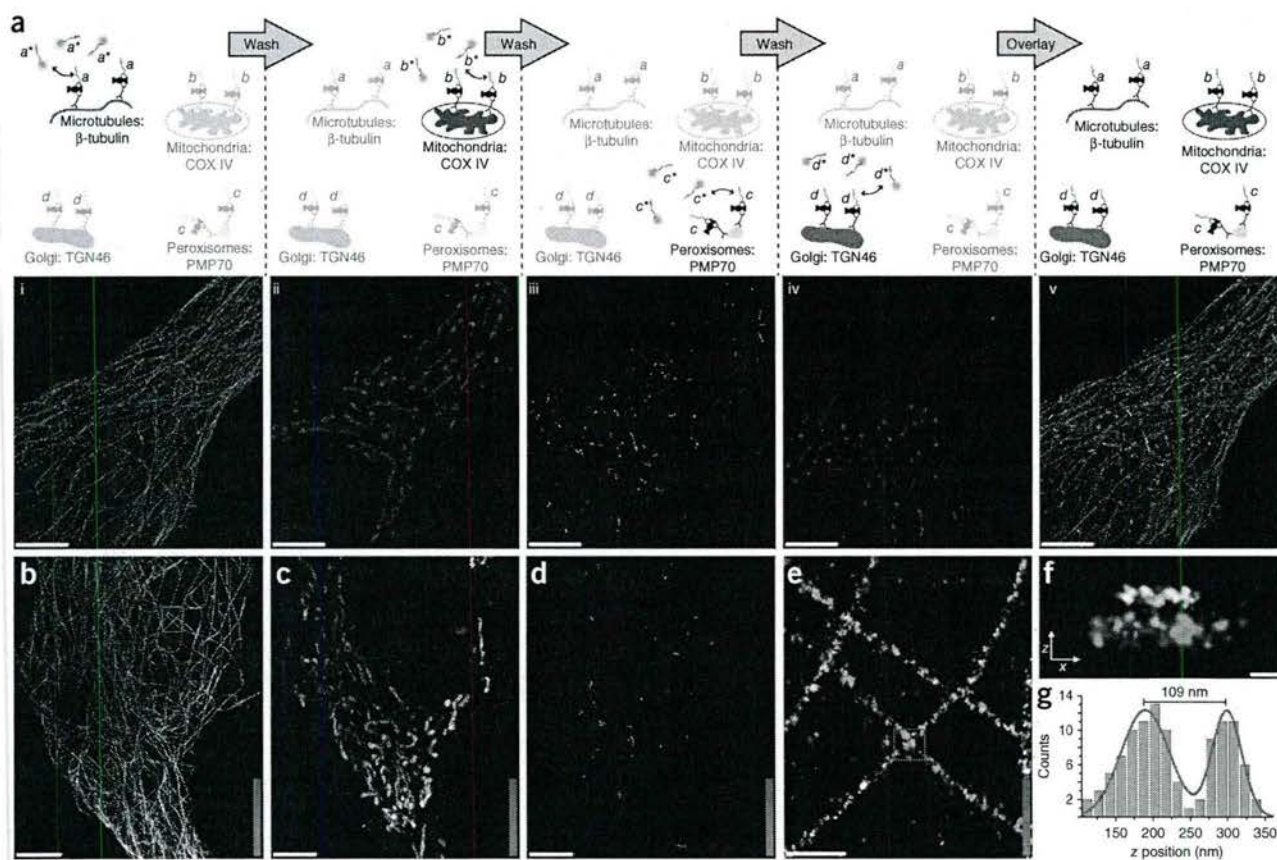
To quantify possible nonspecific interactions of the imager strands with cellular components, we performed Exchange-PAINT experiments in which no DNA-PAINT docking strands were present on the antibody-streptavidin conjugates, but otherwise we performed the labeling and imaging process as described above (Supplementary Fig. 12). We observed minimal nonspecific

interaction of the imager strands (Atto 655 or Cy3b) with the cellular components including genomic DNA. Nonspecific interactions can be excluded by analyzing their blinking behavior in an intensity-versus-time trace<sup>13</sup>: such interactions lead to nonrepeating localization events and therefore show an easily identifiable blinking signature (a non-exponential distribution of on- and off-times) that differs from that of specific DNA hybridization interactions (Supplementary Fig. 12).

To quantify potential sample distortion in cellular Exchange-PAINT imaging, we performed a study similar to that for the *in vitro* case. Here we imaged microtubules and mitochondria in four rounds of Exchange-PAINT by imaging microtubules first and then mitochondria and then repeating the process. We then selected a region of interest in each image and performed a normalized cross-correlation analysis. We obtained cross-correlation coefficients of 0.80 and 0.96 for the microtubule and mitochondria images, respectively (Supplementary Fig. 13), values indicative of minimal sample distortion.

We note that, even in a super-resolution image without liquid exchange, one cannot expect 100% correlation between two consecutive images of the same region in a sample owing to the stochastic nature of the image formation. To see this effect, one can simply split a super-resolution raw data set into two parts of equal length, perform a stochastic reconstruction and calculate the normalized cross-correlation coefficient for these two supposedly equal images. We performed this analysis for a subregion of the microtubule image in Figure 4b and obtained a coefficient of 0.88, similar to the Exchange-PAINT case of 0.80 (Supplementary Fig. 13).

Finally, we used optical astigmatism imaging<sup>26,27</sup> to demonstrate 3D Exchange-PAINT super-resolution imaging in a fixed HeLa cell. We labeled microtubules, mitochondria and peroxisomes and obtained three-color 3D super-resolution Exchange-PAINT



**Figure 4** | Multiplexed 2D and 3D Exchange-PAINT super-resolution imaging in fixed cells. Each target is labeled with an antibody carrying a unique DNA-PAINT docking sequence. 2D and 3D imaging are performed using imager strands labeled with Atto 655 and Cy3b, respectively. (a, i) Only imager strands  $a^*$  are present, and microtubules are specifically imaged. (a, ii) After imager strands  $a^*$  are washed out, imager strands  $b^*$  are introduced and COX IV proteins in mitochondria are imaged. TGN46 proteins in the Golgi complex (a, iii) and PMP70 proteins in peroxisomes (a, iv) are imaged in the same way (imaging: 15,000 frames per cycle, 10-Hz imaging rate; washing: 1–2 min per cycle). (a, v) Overlay of all four targets. (b–g) Three-color 3D Exchange-PAINT images using Cy3b-labeled imager strands. (b) 3D image of microtubules; color indicates height on a scale from 0 nm (green) to 800 nm (purple). (c) 3D image of mitochondria; the height scale is 40–1,100 nm. (d) 3D image of peroxisomes; the height scale is 0–700 nm. (e) Zoomed-in view of the boxed area in b; the height scale is 0–700 nm. (f)  $xz$  profile of the boxed area in e. (g) A two-component Gaussian fit reveals a distance of  $\sim 109$  nm in  $z$  of two adjacent microtubules. For 3D imaging: 50,000 frames per cycle, 16-Hz imaging rate; washing: 1–2 min per cycle. Scale bars: 5  $\mu\text{m}$  (a–d), 500 nm (e) and 100 nm (f).

images using Cy3b (Fig. 4b–d and Supplementary Fig. 14). A magnified image revealed separate microtubules crossing each other in different  $z$  planes (Fig. 4e,f). A cross-sectional histogram in  $z$  yielded a distance of  $\sim 109$  nm, which is well below the diffraction limit (Fig. 4g).

## DISCUSSION

Sequential multiplexing, as that achieved with Exchange-PAINT, provides a complementary approach to previous spectral and geometric multiplexing<sup>13,28</sup>. Compared to geometric multiplexing<sup>13,28</sup>, Exchange-PAINT uses a more compact labeling entity (an  $\sim 9$ -nt DNA strand), but this should allow up to several hundred noninteracting sequences within tight bounds for dissociation kinetics. Exchange-PAINT may enable a substantial increase in multiplexing for both super-resolution and diffraction-limited imaging. In contrast with spectral multiplexing<sup>28,29</sup>, Exchange-PAINT uses only a single fluorophore, thereby allowing the selection of an optimal dye with respect to its photophysical properties (such as the number of emittable photons) for

super-resolution imaging. Unlike previous multiplexing approaches for diffraction-limited imaging based on DNA strand exchange reactions<sup>30</sup> and *in situ* sequencing<sup>31</sup>, Exchange-PAINT does not involve labeling and erasing steps, as imager strands bind to docking strands only transiently. Exchange-PAINT thus permits simpler experimental procedures and faster image acquisition ( $\sim 1$ – $2$  min of washing between imaging cycles) while preserving the intrinsic super-resolution capability of DNA-PAINT.

Fully translating the *in vitro* imaging capability of DNA-PAINT and Exchange-PAINT to *in situ* imaging of cellular components will benefit from strategies to label diverse cellular targets with high specificity and efficiency. One way to facilitate this goal, as noted by other researchers<sup>17,24,32</sup>, is reducing the label size. Our current antibody-DNA conjugation method based on streptavidin bridging, though providing a simple and modular approach, results in a rather bulky conjugate with large linker distances that preclude obtaining the sub-10-nm resolution we demonstrated by direct labeling of DNA nanostructures with imager strands. Thus a next logical step would be direct coupling of primary

antibodies to DNA strands<sup>33</sup> without a streptavidin intermediate. Additionally, to further decrease the label size, it would be advantageous to use nanobody- or aptamer-based labeling strategies<sup>24,32</sup>, with the latter serving as a natural extension to DNA-PAINT imaging as it already uses nucleic acid-based interactions. In addition to its use in protein imaging, our method should be directly applicable to DNA and RNA imaging in fixed cells: for example, by using FISH methods<sup>28,34</sup>.

DNA-PAINT and Exchange-PAINT provide a simple and robust method for highly multiplexed super-resolution imaging. With further development, we anticipate it will become a useful and standard tool for studying complex biomolecular systems.

## METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

## ACKNOWLEDGMENTS

We thank J. Nicoludis and M.T. Strauss for help with DNA origami design, T. Schlichthaerle for transmission electron microscopy imaging support and M. Zhang for help with DLD1 cells. We thank C. Steinhauer for help with DNA-PAINT software development and fruitful discussions. We thank R.D. Barish for critical reading and commenting on the manuscript. This work is supported by a US National Institutes of Health (NIH) Director's New Innovator Award (1DP20D007292), an NIH Transformative Research Award (1R01EB018659), an NIH grant (5R21HD072481), an Office of Naval Research (ONR) Young Investigator Program Award (N000141110914), ONR grants (N000141010827 and N000141310593), a US National Science Foundation (NSF) Faculty Early Career Development Award (CCF1054898), an NSF grant (CCF1162459) and a Wyss Institute for Biologically Engineering Faculty Startup Fund to P.Y., and an NIH Director's New Innovator Award (1DP20D004641) and a Wyss Institute for Biologically Inspired Engineering Faculty Award to W.M.S. R.J. acknowledges support from the Alexander von Humboldt-Foundation through a Feodor-Lynen Fellowship. M.S.A. and M.D. acknowledge support from Howard Hughes Medical Institute International Student Research Fellowships.

## AUTHOR CONTRIBUTIONS

R.J., M.S.A. and J.B.W. contributed equally to this work. R.J. and M.S.A. conceived of the study, designed and performed the experiments, analyzed the data and wrote the manuscript. J.B.W. designed and performed the experiments, analyzed the data and wrote the manuscript. M.D. performed the experiments, analyzed the data and developed the drift correction software. W.M.S. supervised the project, discussed the results and critiqued the paper. P.Y. conceived of, designed and supervised the study, interpreted the data and wrote the manuscript. All authors reviewed and approved the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Rust, M.J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–795 (2006).
- Hell, S.W. Microscopy and its focal switch. *Nat. Methods* **6**, 24–32 (2009).
- Hell, S.W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).
- Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
- Sharonov, A. & Hochstrasser, R.M. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proc. Natl. Acad. Sci. USA* **103**, 18911–18916 (2006).
- Giannone, G. *et al.* Dynamic superresolution imaging of endogenous proteins on living cells at ultra-high density. *Biophys. J.* **99**, 1303–1310 (2010).
- Lew, M.D. *et al.* Three-dimensional superresolution colocalization of intracellular protein superstructures and the cell surface in live *Caulobacter crescentus*. *Proc. Natl. Acad. Sci. USA* **108**, E1102–E1110 (2011).
- Flors, C., Ravarani, C.N. & Dryden, D.T. Super-resolution imaging of DNA labelled with intercalating dyes. *ChemPhysChem* **10**, 2201–2204 (2009).
- Schoen, I., Ries, J., Klotzsch, E., Ewers, H. & Vogel, V. Binding-activated localization microscopy of DNA structures. *Nano Lett.* **11**, 4008–4011 (2011).
- 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).
- Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* **5**, 159–161 (2008).
- Jungmann, R., Scheible, M. & Simmel, F.C. Nanoscale imaging in DNA nanotechnology. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **4**, 66–81 (2012).
- Lin, C. *et al.* Submicrometre geometrically encoded fluorescent barcodes self-assembled from DNA. *Nat. Chem.* **4**, 832–839 (2012).
- Derr, N.D. *et al.* Tug-of-war in motor protein ensembles revealed with a programmable DNA origami scaffold. *Science* **338**, 662–665 (2012).
- Johnson-Buck, A. *et al.* Super-resolution fingerprinting detects chemical reactions and idiosyncrasies of single DNA pegboards. *Nano Lett.* **13**, 728–733 (2013).
- Xu, K., Babcock, H.P. & Zhuang, X. Dual-objective STORM reveals three-dimensional filament organization in the actin cytoskeleton. *Nat. Methods* **9**, 185–188 (2012).
- Vaughan, J.C., Jia, S. & Zhuang, X. Ultrabright photoactivatable fluorophores created by reductive caging. *Nat. Methods* **9**, 1181–1184 (2012).
- Rothmund, P.W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
- Douglas, S.M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
- Steinhauer, C., Jungmann, R., Sobey, T.L., Simmel, F.C. & Tinnefeld, P. DNA origami as a nanoscopic ruler for super-resolution microscopy. *Angew. Chem. Int. Ed. Engl.* **48**, 8870–8873 (2009).
- 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).
- Rasnik, I., McKinney, S.A. & Ha, T. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat. Methods* **3**, 891–893 (2006).
- Vogelsang, J. *et al.* A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angew. Chem. Int. Ed. Engl.* **47**, 5465–5469 (2008).
- Ries, J., Kaplan, C., Platonova, E., Eghlidi, H. & Ewers, H. A simple, versatile method for GFP-based super-resolution microscopy via nanobodies. *Nat. Methods* **9**, 582–584 (2012).
- Wu, N. *et al.* Molecular threading and tunable molecular recognition on DNA origami nanostructures. *J. Am. Chem. Soc.* **135**, 12172–12175 (2013).
- Kao, H.P. & Verkman, A.S. Tracking of single fluorescent particles in three dimensions: use of cylindrical optics to encode particle position. *Biophys. J.* **67**, 1291–1300 (1994).
- Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810–813 (2008).
- Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods* **9**, 743–748 (2012).
- Bates, M., Dempsey, G.T., Chen, K.H. & Zhuang, X. Multicolor super-resolution fluorescence imaging via multi-parameter fluorophore detection. *ChemPhysChem* **13**, 99–107 (2012).
- Schweller, R.M. *et al.* Multiplexed *in situ* immunofluorescence using dynamic DNA complexes. *Angew. Chem. Int. Ed. Engl.* **51**, 9292–9296 (2012).
- Ke, R. *et al.* *In situ* sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
- Opazo, F. *et al.* Aptamers as potential tools for super-resolution microscopy. *Nat. Methods* **9**, 938–939 (2012).
- Kazane, S.A. *et al.* Site-specific DNA-antibody conjugates for specific and sensitive immuno-PCR. *Proc. Natl. Acad. Sci. USA* **109**, 3731–3736 (2012).
- Beliveau, B.J. *et al.* Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. USA* **109**, 21301–21306 (2012).

## ONLINE METHODS

**Materials.** Unmodified DNA oligonucleotides were purchased from Integrated DNA Technologies. Fluorescently modified DNA oligonucleotides were purchased from Biosynthesis. Biotinylated monoclonal antibodies against  $\beta$ -tubulin (9F3; catalog number: 6181) and COX IV (3E11; catalog number: 6014) were purchased from Cell Signaling. Anti-PMP70 (catalog number: ab28499) was purchased from Abcam. Anti-TGN46 (catalog number: NBPI-49643B) was purchased from VWR. Streptavidin was purchased from Invitrogen (catalog number: S-888). Bovine serum albumin (BSA) and BSA-biotin obtained from Sigma-Aldrich (catalog number: A8549). Glass slides and coverslips were purchased from VWR. Lab-Tek II chambered coverglass was purchased from Thermo Fisher Scientific. M13mp18 scaffold was obtained from New England BioLabs. p8064 scaffold for microtubule-like DNA origami structures was prepared as described before<sup>19</sup>. Freeze 'N Squeeze columns were ordered from Bio-Rad. TetraSpeck Beads were purchased from Life Technologies. Paraformaldehyde, glutaraldehyde and TEM grids (FORMVAR 400 mesh copper grids) were obtained from Electron Microscopy Sciences.

Three buffers were used for sample preparation and imaging: buffer A (10 mM Tris-HCl, 100 mM NaCl, 0.05% Tween 20, pH 7.5), buffer B (5 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM EDTA, 0.05% Tween 20, pH 8) and buffer C (1× PBS, 500 mM NaCl, pH 8).

**Optical setup.** Fluorescence imaging was carried out on an inverted Nikon Eclipse Ti microscope (Nikon Instruments) with the Perfect Focus System, applying an objective-type TIRF configuration using a Nikon TIRF illuminator with an oil-immersion objective (CFI Apo TIRF 100×, numerical aperture (NA) 1.49, oil). For 2D imaging an additional 1.5× magnification was used to obtain a final magnification of ~150-fold, corresponding to a pixel size of 107 nm. Three lasers were used for excitation: 488 nm (200 mW nominal, Coherent Sapphire), 561 nm (200 mW nominal, Coherent Sapphire) and 647 nm (300 mW nominal, MBP Communications). The laser beam was passed through cleanup filters (ZT488/10, ZET561/10 and ZET640/20, Chroma Technology) and coupled into the microscope objective using a multiband beam splitter (ZT488rdc/ZT561rdc/ZT640rdc, Chroma Technology). Fluorescence light was spectrally filtered with emission filters (ET525/50m, ET600/50m and ET700/75m, Chroma Technology) and imaged on an electron-multiplying charge-coupled device (EMCCD) camera (iXon X3 DU-897, Andor Technologies).

**DNA origami self-assembly.** The microtubule-like DNA origami structures were formed in a one-pot reaction with a 40- $\mu$ l total volume containing 10 nM scaffold strand (p8064), 500 nM folding staples and biotin handles, 750 nM biotin anti-handles and 1.1  $\mu$ M DNA-PAINT docking strands in folding buffer (1× TAE buffer with 20 mM MgCl<sub>2</sub>). The solution was annealed using a thermal ramp<sup>13</sup> cooling from 80 °C to 14 °C over the course of 15 h. After self-assembly, monomeric structures were purified by agarose gel electrophoresis (1.5% agarose, 0.5× TBE, 10 mM MgCl<sub>2</sub>, 1× SybrSafe) at 4.5 V/cm for 1.5 h (see **Supplementary Fig. 2**). Gel bands were cut, crushed and filled into a Freeze 'N Squeeze column and spun for 5 min at 1,000g at 4 °C. Polymerization was carried out at 30 °C for 48 h with a fivefold excess of polymerization

staples in folding buffer. Polymerized structures were used for imaging without further purification.

DNA origami drift markers were self-assembled in a one-pot reaction (40- $\mu$ l total volume, 20 nM M13mp18 scaffold, 100 nM biotinylated staples, 530 nM staples with DNA-PAINT docking sites, 1× TAE with 12.5 mM MgCl<sub>2</sub>). Self-assembled structures were purified as described before.

DNA origami structures for the four-color *in vitro* Exchange-PAINT demonstration were self-assembled in a one-pot reaction (40- $\mu$ l total volume, 30 nM M13mp18 scaffold, 470 nM biotinylated staples, 400 nM staples with docking sites for number imaging, 370 nM core structure staples, 1× TAE with 12.5 mM MgCl<sub>2</sub>). Self-assembled structures were purified as described before.

DNA origami structures for the ten-color *in vitro* Exchange-PAINT demonstration were self-assembled in a one-pot reaction (40- $\mu$ l total volume, 30 nM M13mp18 scaffold, 36 nM biotinylated staples, 750 nM staples with docking sites for number imaging, 300 nM core structure staples, 1× TAE with 12.5 mM MgCl<sub>2</sub>). Structures were not purified. Excessive staples were washed out of the sample after immobilization of the structure on the surface.

DNA strand sequences for the microtubule-like DNA origami structures can be found in **Supplementary Table 1**. DNA strand sequences for DNA origami drift markers can be found in **Supplementary Table 2**. DNA strand sequences for DNA origami structures for the ten-color *in vitro* Exchange-PAINT demonstration can be found in **Supplementary Tables 3 and 4** for odd and even digits, respectively. DNA strand sequences for DNA origami structures for *in vitro* Exchange-PAINT demonstration (digits 0–3) can be found in **Supplementary Table 5**. The scaffold sequence for p8064 and M13mp18 can be found in **Supplementary Tables 6 and 7**, respectively.

DNA-PAINT imager and docking sequences as well as sequences for surface attachment via biotin are listed in **Supplementary Table 8**.

**Antibody-DNA conjugates.** Antibody-DNA conjugates used to specifically label proteins of interest with DNA-PAINT docking sites were preassembled in two steps. First, 3.2  $\mu$ l of 1 mg/ml streptavidin (dissolved in buffer A) was reacted with 0.5  $\mu$ l biotinylated DNA-PAINT docking strands at 100  $\mu$ M and an additional 5.3  $\mu$ l of buffer A for 30 min at room temperature (RT) while gently shaking. The solution was then incubated in a second step with 1  $\mu$ l of monoclonal biotinylated antibodies at 1 mg/ml against the protein of interest for 30 min at RT. Filter columns (Amicon 100 kDa, Millipore) were used to purify the preassembled conjugates from unreacted streptavidin-oligo conjugates.

**Cell immunostaining.** HeLa and DLD1 cells were cultured with Eagle's minimum essential medium fortified with 10% FBS with penicillin and streptomycin and were incubated at 37 °C with 5% CO<sub>2</sub>. At approximately 30% confluence, cells were seeded into Lab-Tek II chambered coverglass 24 h before fixation. Microtubules, mitochondria, Golgi complexes and peroxisomes were immunostained using the following procedure: washing in PBS; fixation in a mixture of 3% paraformaldehyde and 0.1% glutaraldehyde in PBS for 10 min; 3× washing with PBS; reduction with ~1 mg/ml NaBH<sub>4</sub> for 7 min; 3× washing with PBS; permeabilization with 0.25% (v/v) Triton X-100 in PBS for 10 min; 3× washing with PBS; blocking with 3% (w/v) BSA for 30 min

and staining overnight with the preassembled antibody-DNA conjugates against  $\beta$ -tubulin, COX IV, PMP70 or TGN46 (conjugates were diluted to 10  $\mu\text{g}/\text{ml}$  in 5% BSA); 3 $\times$  washing with PBS; postfixation in a mixture of 3% paraformaldehyde and 0.1% glutaraldehyde in PBS for 10 min; and 3 $\times$  washing with PBS.

**Super-resolution DNA-PAINT imaging of microtubule-like DNA origami structures.** For sample preparation, a piece of coverslip (no. 1.5, 18  $\times$  18 mm<sup>2</sup>,  $\sim$ 0.17 mm thick) and a glass slide (3  $\times$  1 inch<sup>2</sup>, 1 mm thick) were sandwiched together by two strips of double-sided tape to form a flow chamber with inner volume of  $\sim$ 20  $\mu\text{l}$ . First, 20  $\mu\text{l}$  of biotin-labeled bovine albumin (1 mg/ml, dissolved in buffer A) was flown into the chamber and incubated for 2 min. The chamber was then washed using 40  $\mu\text{l}$  of buffer A. 20  $\mu\text{l}$  of streptavidin (0.5 mg/ml, dissolved in buffer A) was then flown through the chamber and allowed to bind for 2 min. After washing with 40  $\mu\text{l}$  of buffer A and subsequently with 40  $\mu\text{l}$  of buffer B, 20  $\mu\text{l}$  of biotin-labeled microtubule-like DNA structures ( $\sim$ 300 pM monomer concentration) and DNA origami drift markers ( $\sim$ 100 pM) in buffer B were finally flown into the chamber and incubated for 5 min. The chamber was washed using 40  $\mu\text{l}$  of buffer B.

The final imaging buffer solution contained 1.5 nM Cy3b-labeled imager strands in buffer B. The chamber was sealed with epoxy before subsequent imaging. The CCD readout bandwidth was set to 1 MHz at 16 bit and 5.1 pre-amp gain. No electron-multiplying (EM) gain was used. Imaging was performed using TIR illumination with an excitation intensity of 294 W/cm<sup>2</sup> at 561 nm.

**Super-resolution Exchange-PAINT imaging of DNA nanostructures.** For fluid exchange, a custom flow chamber was constructed as shown in **Supplementary Figure 6a**. A detailed preparation protocol can be found in the **Supplementary Protocol**. Prior to the functionalizing of the imaging chamber with BSA-biotin, it was rinsed with 1 M KOH for cleaning. Binding of the origami structures to the surface of the flow chamber was performed as described before. Each image acquisition step was followed with a brief  $\sim$ 1–2 min washing step consisting of at least three washes using 200  $\mu\text{l}$  of buffer B for each. Then the next imager strand solution was introduced. The surface was monitored throughout the washing procedure to ensure complete exchange of imager solutions. Acquisition and washing steps were repeated until all ten targets were imaged. The CCD readout bandwidth was set to 3 MHz at 14 bit and 5.1 pre-amp gain. No EM gain was used. Imaging was performed using TIR illumination with an excitation intensity of 166 W/cm<sup>2</sup> at 561 nm (ten-color Exchange-PAINT with 3 nM Cy3b-labeled imager strands in buffer B; **Fig. 3c,d**) and 600 W/cm<sup>2</sup> at 647 nm (four-color Exchange-PAINT with 3 nM Atto 655-labeled imager strands in buffer B; **Fig. 3e**).

**Super-resolution DNA-PAINT imaging of cells.** All data were acquired with an EMCCD readout bandwidth of 5 MHz at 14 bit, 5.1 pre-amp gain and 255 EM gain. Imaging was performed using HILO illumination<sup>11</sup>. The laser power densities were 283 W/cm<sup>2</sup> at 647 nm in **Figure 2a** and 142 W/cm<sup>2</sup> at 647 nm and 19 W/cm<sup>2</sup> at 561 nm in **Figure 2d**.

Imaging conditions were as follows. For **Figure 2a** we used 700 pM Atto 655-labeled imager strands in buffer C. For **Figure 2d** we

used 600 pM Cy3b-labeled imager strands and 1.5 nM Atto 655-labeled imager strands in buffer C.

**Super-resolution Exchange-PAINT imaging of cells.** A Lab-Tek II chamber was adapted for fluid exchange as shown in **Supplementary Figure 6b**. 2D images (**Fig. 4a, i–iv**) were acquired with an EMCCD readout bandwidth of 5 MHz at 14 bit, 5.1 pre-amp gain and 255 EM gain. 3D images (**Fig. 4b–d**) were acquired with a CCD readout bandwidth of 3 MHz at 154 bit, 5.1 pre-amp gain and no EM gain. Imaging was performed using HILO illumination in both cases. Sequential imaging was done as described for the 2D origami nanostructures, but the washing steps were performed using buffer C. The laser power densities at 647 nm were 257 W/cm<sup>2</sup> in **Figure 4a, i** and 385 W/cm<sup>2</sup> in **Figure 4a, ii–iv**. The laser power densities at 561 nm were 31 W/cm<sup>2</sup> in **Figure 4b–d**.

Imaging conditions were as follows. For **Figure 4a, i** we used 700 pM Atto 655-labeled imager strands in buffer C. For **Figure 4a, ii–iv** we used 2 nM Atto 655-labeled imager strands in buffer C. For **Figure 4b** we used 800 pM Cy3b-labeled imager strands in buffer C. For **Figure 4c,d** we used 2 nM Cy3b-labeled imager strands in buffer C.

**3D DNA-PAINT imaging.** 3D images were acquired with a cylindrical lens in the detection path (Nikon). The N-STORM analysis package for NIS Elements (Nikon) was used for data processing. Imaging was performed without additional magnification in the detection path, yielding 160-nm pixel size. 3D calibration was carried out according to the manufacturer's instructions.

**Imager strand concentration determination.** Optimal imager concentrations were determined empirically according to the labeling density. Generally, a high enough fluorescence off-on ratio has to be ensured in order to guarantee binding of only a single imager strand per diffraction-limited area. Additionally, a sufficient imager strand concentration (and thus sufficiently low fluorescence off-time) is necessary to ensure sufficient binding events and thereby robust detection of every docking strand during image acquisition.

**Super-resolution data processing.** Super-resolution DNA-PAINT images were reconstructed using spot-finding and 2D-Gaussian fitting algorithms programmed in LabVIEW<sup>10</sup> (**Supplementary Software**). A simplified version of this software is available for download at <http://www.dna-paint.net/> or <http://molecular-systems.net/software/>.

**Normalized cross-correlation analysis.** Normalized cross-correlation coefficients were obtained by first normalizing the respective reconstructed grayscale super-resolution images and subsequently performing a cross-correlation analysis in Matlab R2013b (MathWorks).

**Drift correction and channel alignment.** DNA origami structures (**Supplementary Fig. 1**) were used for drift correction and as alignment markers in *in vitro* DNA-PAINT and Exchange-PAINT imaging. Drift correction was performed by tracking the position of each origami drift marker throughout the duration of each image acquisition. The trajectories of all detected drift



markers were then averaged and used to globally correct the drift in the final super-resolution reconstruction. For channel alignment between different imaging cycles in Exchange-PAINT, these structures were used as alignment points by matching their positions in each Exchange-PAINT image.

For cellular imaging, 100-nm-diameter gold nanoparticles (Sigma-Aldrich; 10 nM in buffer C, added before imaging) were used as drift and alignment markers. The gold nanoparticles adsorb nonspecifically to the glass bottom of the imaging chambers. Drift correction and alignment was performed in a fashion similar to that for the origami drift markers. Again, the apparent movement of all gold nanoparticles in a field of view was tracked throughout the image acquisition. The obtained trajectories were then averaged and used for global drift correction of the final super-resolution image. For the dual-color image of mitochondria and microtubules in **Figure 2d–f**, the gold particles were visible in both color channels. The same gold nanoparticles were also used for drift correction and realignment of the different imaging rounds in the *in situ* Exchange-PAINT experiments (**Fig. 4**).

**Transmission electron microscopy (TEM) imaging.** For imaging, 3.5  $\mu$ l of undiluted microtubule-like DNA structures were adsorbed for 2 min onto glow-discharged, carbon-coated TEM

grids. The grids were then stained for 10 s using a 2% aqueous ultrafiltrated (0.2- $\mu$ m filter) uranyl formate solution containing 25 mM NaOH. Imaging was performed using a JEOL JEM-1400 operated at 80 kV.

**Atomic force microscopy imaging.** Imaging was performed using tapping mode on a Multimode VIII atomic force microscope (AFM) with an E-scanner (Bruker). Imaging was performed in TAE/Mg<sup>2+</sup> buffer solution with DNP-S oxide-sharpened silicon nitride cantilevers and SNL sharp nitride levers (Bruker Probes) using resonance frequencies between 7 and 9 kHz of the narrow 100- $\mu$ m, 0.38-N/m-force constant cantilever. After self-assembly of the origami structure, ~20  $\mu$ l of TAE/Mg<sup>2+</sup> buffer solution was deposited onto a freshly cleaved mica surface (Ted Pella) glued to a metal puck (Ted Pella). After 30 s the mica surface was dried using a gentle stream of N<sub>2</sub>, and 5  $\mu$ l of the origami solution was deposited onto the mica surface. After another 30 s, 30  $\mu$ l of additional buffer solution was added to the sample. Imaging parameters were optimized for best image quality while the highest possible set point was maintained to minimize damage to the samples. Images were postprocessed by subtracting a first-order polynomial from each scan line. Drive amplitudes were approximately 0.11 V, integral gains were ~2 and proportional gains were ~4.

## Design Space for Complex DNA Structures

Bryan Wei,<sup>†,‡</sup> Mingjie Dai,<sup>†,§</sup> Cameron Myhrvold,<sup>†,‡</sup> Yonggang Ke,<sup>†,⊥</sup> Ralf Jungmann,<sup>†,‡</sup> and Peng Yin<sup>\*,†,‡</sup>

<sup>†</sup>Wyss Institute for Biologically Inspired Engineering, Harvard Medical School, Boston, Massachusetts 02115, United States

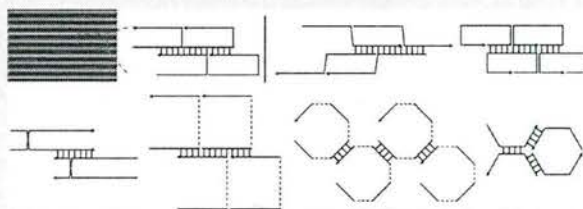
<sup>‡</sup>Department of Systems Biology, Harvard University, Boston, Massachusetts 02115, United States

<sup>§</sup>Program in Biophysics, Harvard University, Boston, Massachusetts 02115, United States

<sup>⊥</sup>Dana Farber Cancer Institute, Boston, Massachusetts 02115, United States

### Supporting Information

**ABSTRACT:** Nucleic acids have emerged as effective materials for assembling complex nanoscale structures. To tailor the structures to function optimally for particular applications, a broad structural design space is desired. Despite the many discrete and extended structures demonstrated in the past few decades, the design space remains to be fully explored. In particular, the complex finite-sized structures produced to date have been typically based on a small number of structural motifs. Here, we perform a comprehensive study of the design space for complex DNA structures, using more than 30 distinct motifs derived from single-stranded tiles. These motifs self-assemble to form structures with diverse strand weaving patterns and specific geometric properties, such as curvature and twist. We performed a systematic study to control and characterize the curvature of the structures, and constructed a flat structure with a corrugated strand pattern. The work here reveals the broadness of the design space for complex DNA nanostructures.



### INTRODUCTION

Self-assembly of nucleic acids (DNA and RNA) provides a powerful approach for constructing sophisticated synthetic molecular structures and devices. By encoding sequence complementarity into component DNA strands, prescribed structures can be assembled under the appropriate formation conditions.<sup>1</sup> After three decades of development, the complexity of synthetic DNA structures has grown from simple branched junctions<sup>1</sup> formed from just a few strands to complex 2D and 3D objects composed of hundreds or even thousands of distinct strands.<sup>2–22</sup> Moreover, researchers have demonstrated the construction of dynamic systems, including switches,<sup>23</sup> walkers,<sup>24–26</sup> circuits,<sup>25,27,28</sup> and triggered assembly systems.<sup>25</sup> Additionally, as diverse functional molecules can be modified onto specific sites of DNA structures,<sup>19,29–32</sup> DNA nanotechnology has enabled applications such as nanofabrication,<sup>33</sup> protein structure determination,<sup>34</sup> fluorescent bioimaging,<sup>31,32</sup> single molecule biophysics,<sup>35</sup> biosynthetic and cell-signaling pathways modulation.<sup>19,36</sup>

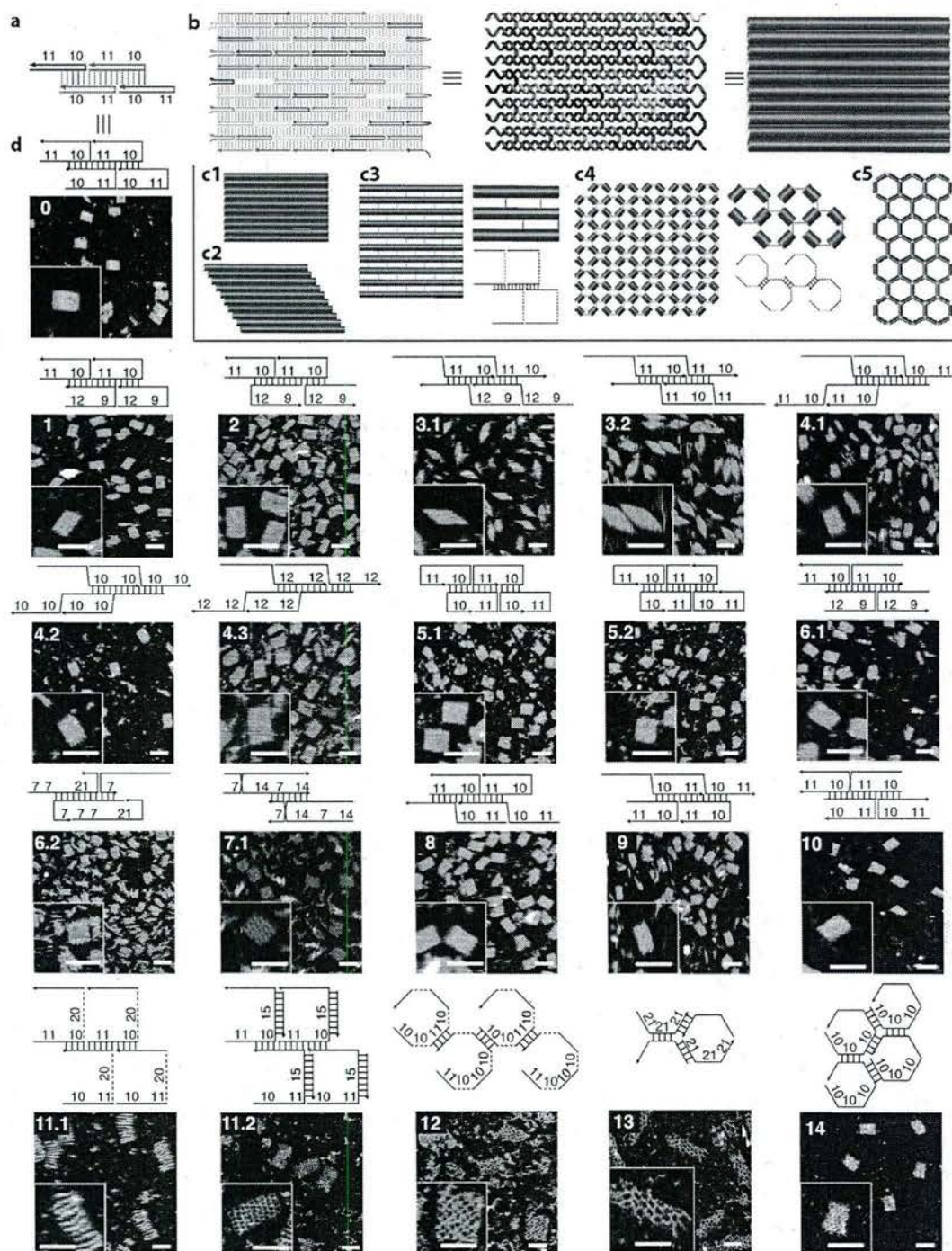
Two methods that are particularly effective for assembling discrete mega-Dalton structures with arbitrarily prescribed shapes are DNA origami,<sup>10,14–18,22</sup> and single-stranded tiles (SSTs)<sup>20</sup> and bricks.<sup>21</sup> In DNA origami, hundreds of short, synthetic DNA strands fold a long scaffold (typically the M13 viral genome) into a desired structure.<sup>10,16–18,22</sup> More recently, researchers have demonstrated finite complex 2D and 3D shapes self-assembled from hundreds to thousands of distinct single-stranded tiles and bricks.<sup>20,21</sup>

Unlike DNA origami, no scaffold strand is required for SST structure, which are composed entirely of short synthetic DNA

strands. The elimination of the scaffold strand makes SSTs effective tools for systematic and rapid study of the geometry, sequence, and structural design space for complex DNA structures. First, as SST-based structures have a modular architecture—each strand can be included, removed, or replaced independently—SSTs enable rapid prototyping of the shape space (more than 100 distinct, sophisticated 2D<sup>20</sup> and 3D<sup>21</sup> shapes were experimentally demonstrated). Second, as the strand sequences are no longer restricted to be derived from the biogenic scaffold sequence, SSTs enable rapid probing of the sequence space (structures were formed from both specifically designed and randomly generated sequences<sup>20,21</sup>). Lastly and importantly, by removing the restriction of scaffold routing through the shape and using diverse structural motifs derived from the canonical SST motif,<sup>20</sup> we will demonstrate here a systematic study of the structural design space for forming complex DNA nanostructures.

To tailor the structures to function optimally for particular applications, a broad design space for DNA nanostructures is desired. Despite the many discrete and extended structures demonstrated in the past few decades, the structural design space has yet to be fully explored. In particular, complex, finite-sized structures formed by origami and SST are primarily based on double- or single-stranded versions of the basic DAE (antiparallel double-crossover molecules with an even number of half-helical turns) motif<sup>3</sup> and their derivatives. Beyond origami and SST, however, a richer motif space has been

Received: June 20, 2013



**Figure 1.** Motif designs and AFM imaging. A representative subset of the motifs studied is shown here; see SI, section SII.3 for the full repertoire. (a) A 2×2 unit (containing four SSTs) of the canonical motif (motif 0). Colors distinguish sequences. (b) Strand diagram (left), helix diagram (middle), and cylinder diagram (right) of a 10 helix × 11 turn (10H×11T) structure formed from motif 0. (c) Cylinder diagrams for structures formed from different motifs; a cylinder depicts a DNA duplex and a line depicts a single-strand DNA segment: (c1) structures from motifs 0–2 and 4–10 have a rectangle shape containing parallel-arranged double-helices; (c2) structures from motifs 3.1 and 3.2 have a parallelogram shape containing parallel-arranged double-helices; (c3) structures from motif 11 have a rectangle shape containing parallel-arranged double-helices connected by single-stranded linkers; (c4) structures from motif 12 have a rectangle shape with a “fishnet” pattern (composed of short duplex segments connected by single-stranded linkers); (c5) structures from motifs 13 and 14 have a honeycomb pattern composed of short DNA duplex segments. (d) Strand diagrams and the corresponding AFM images for structures formed from different motifs. Top panel: a 2×2 repeating unit. Numbers indicate domain or linker lengths in nucleotide. Bottom panel: AFM image. Inset shows a magnified view. Scale bars, 50 nm. See SI, section SII.5 for the measured dimensions and measured gel yields (5–44%).

explored for forming relatively simple discrete and extended structures. This has been achieved using various structural motifs, such as diverse double-crossover motifs,<sup>3,4</sup> PX-based structures,<sup>7,37</sup> structures with flexible linker joints,<sup>11,14,15,29,38</sup> wire-frame structures,<sup>1,2,7,8,11,38–41</sup> and metal–DNA junctions.<sup>42</sup>

The next challenge is to develop a versatile method for creating complex, discrete-sized DNA nanostructures with diverse structural motifs, strand weaving patterns, and desired structural properties. Building on the richness of the previous structural motifs for building DNA nanostructures, and taking advantage of the designability and flexibility of SST-based assembly, we performed a comprehensive study of the design space for constructing complex, finite-sized DNA structures. We tested more than 30 distinct motifs derived from SSTs.<sup>20</sup> Most motifs self-assembled to form structures with diverse strand-weaving patterns and specific geometric properties, such as curvature and twist. Moreover, we performed a systematic study to control and characterize the curvature of the structures, and constructed a flat structure with a corrugated strand pattern. Our success in assembling a myriad of different motifs demonstrates the robustness and versatility of the SST approach. More generally, it reveals the broadness of the design space for constructing complex DNA nanostructures.

## RESULTS AND DISCUSSIONS

**Motif Design Parameters.** We began our study for the structural design space with the canonical U-shaped SST motif (motif 0, Figure 1a,b) reported in our previous work.<sup>12,20</sup> This 42-base SST motif has four consecutive domains (each measuring 10 or 11 nt).<sup>12,20</sup> By pairing up complementary domains, individual tiles self-assemble into DNA lattices composed of parallel-arranged double-helices connected by periodic single-stranded crossover linkages. Note that the linkage is merely a phosphate in the DNA backbone. In the diagrams in Figure 1, the linkage is artificially stretched to more clearly show the strand-weaving pattern. Starting with this simple U-shaped motif, we systematically explored important parameters for motif design. Below, we first give an overview of some important parameters and their putative effects on the geometrical, mechanical, and thermodynamic properties of the structures. In the next section, we describe the motif design and its implementation, in detail.

**Domain Length.** Motif 0 contains 10 and 11 nt domains. For most of the motifs explored in our study (including motif 0), we tested several different domain lengths. Changing domain length could affect (1) the geometrical properties of the assembled structure, such as curvature<sup>17,18</sup> and twist;<sup>17,43,44</sup> (2) thermodynamic properties, such as formation and melting temperature; and (3) mechanical properties, such as structural rigidity and internal strain.<sup>17,45</sup> Domain length is thus a basic design variable that we explored extensively in our study. Changing domain length also directly affects the crossover patterns described below.

**Crossover Type.** Structures formed from motif 0 possess single-stranded, antiparallel crossovers. In our study, we explored four combinations of single-<sup>12,46</sup> vs double-stranded crossovers, with parallel vs antiparallel orientations.<sup>3</sup> Here we use “parallel” (or “antiparallel”) to describe crossovers in which the orientation of the strand forming the crossover is the same (or opposite) on both sides of the crossover.<sup>3</sup> The crossover type directly affects the weaving patterns of the strands and the geometrical properties of the structure, which may in turn affect

their mechanical and thermodynamic properties. For example, geometrically, the parallel or antiparallel crossovers result in structures composed of double-helices that are connected in a parallel or antiparallel fashion, referred to as *parallel-connected double-helices* or *antiparallel-connected double-helices*. Note that we also use the term *parallel-arranged double-helices* to refer to the double-helices connected by either parallel or antiparallel crossovers. In the antiparallel crossover-based structure, all the crossovers display the major grooves on the same side of the structure<sup>47</sup> and as a result, induce an accumulated curvature for the structure. In contrast, the parallel crossover-based structure avoids such asymmetry and curvature. In the antiparallel crossover-based structure, the major groove of one double-helix faces the minor groove of an adjacent double-helix;<sup>3</sup> in contrast, the parallel crossover-based structure has major–major and minor–minor groove pairings. Hence, the former should have less strain and more thermal stability.<sup>3</sup> Compared with double-stranded crossover, single-stranded crossover-based structures may be structurally less rigid, mechanically less strained, and in turn, thermodynamically more stable.

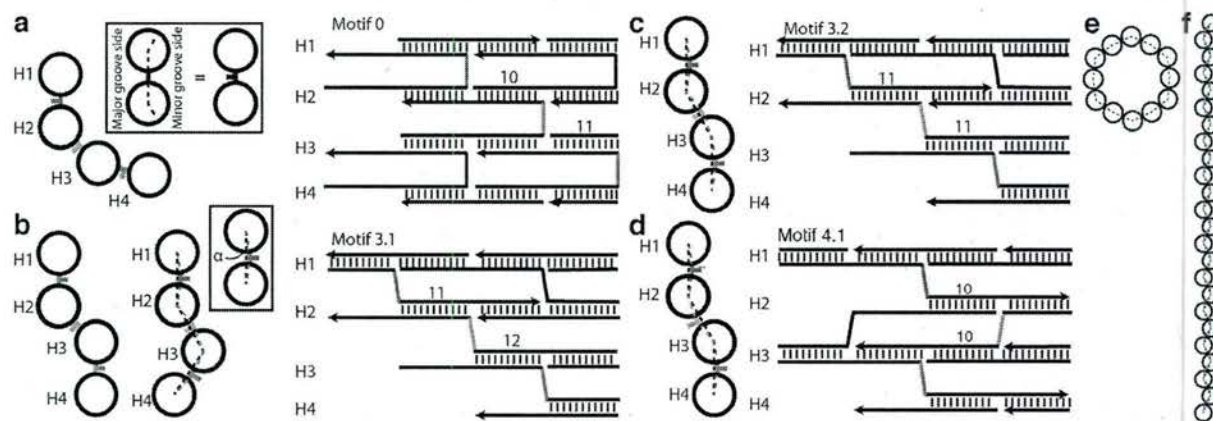
**Crossover Pattern.** Motif 0 has a 21 nt crossover spacing for crossovers between the two adjacent double-helices; along one particular double-helix, a crossover appears every 10 or 11 nt. We systematically changed crossover patterns. Crossover patterns could also lead to a combination of geometric, thermodynamic, and mechanical effects. Geometrically, the curvature of the overall structure (as determined by the dihedral angle formed between three adjacent double-helices) can be tuned by translating all crossover points along the direction parallel to the helical axis while preserving the spacing;<sup>17,18</sup> twist, on the other hand, depends on the crossover spacing.<sup>17,43,44</sup> Mechanically, strain could accumulate in twisted and/or curved structures.<sup>17,43,44</sup> Such mechanical stress in turn could affect the thermodynamic properties of the structure.

**Symmetry.** By adjusting domain length, crossover type, and crossover patterns, we can design structures with more symmetric strand weaving patterns, such as corrugated patterns.<sup>29</sup> Such symmetry can help to eliminate curvature in the structure.

**Linkers.** Structures formed from motif 0 are composed of compact parallel-arranged double-helices connected by periodic single-stranded crossovers (composed of a single phosphate in the DNA backbone). In our study, we included single-stranded linkers of varying lengths between helices and between binding domains. This results in structures that are geometrically and structurally more flexible. Inclusion of linkers should also help to relieve the electrostatic repulsion between adjacent double-helices, hence stabilizing the structure.

**Structures Formed from Diverse Motifs. Overview.** By systematically exploring the above as well as some other design parameters, we tested a total of 36 different motifs, 32 of which formed desired structures (an 89% success rate). A representative subset of all successful motif designs is shown in Figure 1 (see SI, section SI1.3 for the full repertoire). Different morphologies were presented, including compact rectangle and parallelogram of packed DNA duplexes (Figure 1c1,c2), rectangles of spaced DNA duplexes (Figure 1c3), and rectangles of square and hexagonal lattices (Figure 1c4,c5).

**Structure Assembly and Characterization.** The DNA structures were assembled with unpurified DNA strands with randomly designed sequences, and were mixed at roughly equimolar without careful adjustment of stoichiometry. After single-step (one-pot) annealing from 90 to 25 °C over 17 h in



**Figure 2.** Curvature models. (a) Curvature model (left) and corresponding strand diagram (right) of motif 0. Inset depicts the major groove side vs minor groove side. (b) Curvature model of motif 3.1 without (left) and with (right) offset angle  $\alpha$ . (c) Curvature model of motif 3.2 with offset angle  $\alpha$ . (d) Curvature model of motif 4.1. In this corrugated design, the offset angle  $\alpha$  is canceled out so the overall curvature is independent of  $\alpha$ . (e,f) Predicted cross-sectional views of 24-helix structures formed from motifs 0 and 4.1, respectively. See SI, section SI2.1 and Figure SI39 for modeling and calculation details.

0.5 $\times$ TE buffer supplemented with 15 mM  $Mg^{2+}$  (see SI, section SI1.2 for the effects of buffer ion strength and annealing time on the assembly yields), the solution was subjected to 2% native agarose gel electrophoresis and produced one dominant product band, the ratio between the fluorescence intensity of the product band and that of the entire lane was used to approximate the yield of structure formation, which was measured to range from 5% to 44% (Figure SI1).

Annealed samples were subjected to atomic force microscopy (AFM) imaging and showed the expected morphology and dimensions. For samples with relatively low yields (motifs 4.6, 7.2, and 10), structures were gel-purified from the product bands on a 2% agarose gel prior to AFM imaging. See SI, section SI1.3, for larger AFM images and section SI1.5 for measurements of the sizes of the structures.

**Changing Domain Lengths: Motif 1.** First, we adjusted the domain lengths of two adjacent domains within a double-helix from 10 or 11 nt to 9–12 nt (motif 1). This changes the distance between adjacent crossovers along the same double-helix, resulting in the relative change of the dihedral angle formed by the three adjacent double-helices by  $34^\circ$ . This motif was later used to build a larger structure for curvature study.

**Bidirectional Motifs: Motif 2.** Next, we flipped the orientation of the U-shaped motifs in every other row, resulting in a structure composed of U-shaped SSTs with alternating orientations along adjacent rows. Such a modification makes the overall structure more symmetric along the helical direction.

**Z-Shaped Motifs: Motifs 3.1 and 3.2.** Structures formed from U-shaped motifs have antiparallel crossovers. We then designed structures with two Z-shaped motifs and obtained structures containing parallel crossovers. Such parallel crossover weaving patterns are expected to cancel out the curvature resulting from the asymmetric arrangement of major and minor grooves at the crossovers (see Curvature section for details) and was used for curvature study.

**Corrugated Structures: Motif 4.1.** We further changed the orientation of every other row in the Z-shaped motif to form a symmetric structure with a corrugated<sup>29</sup> strand pattern. This corrugation should in principle eliminate any curvature and

result in a perfectly flat structure. This structure was used in the curvature study below.

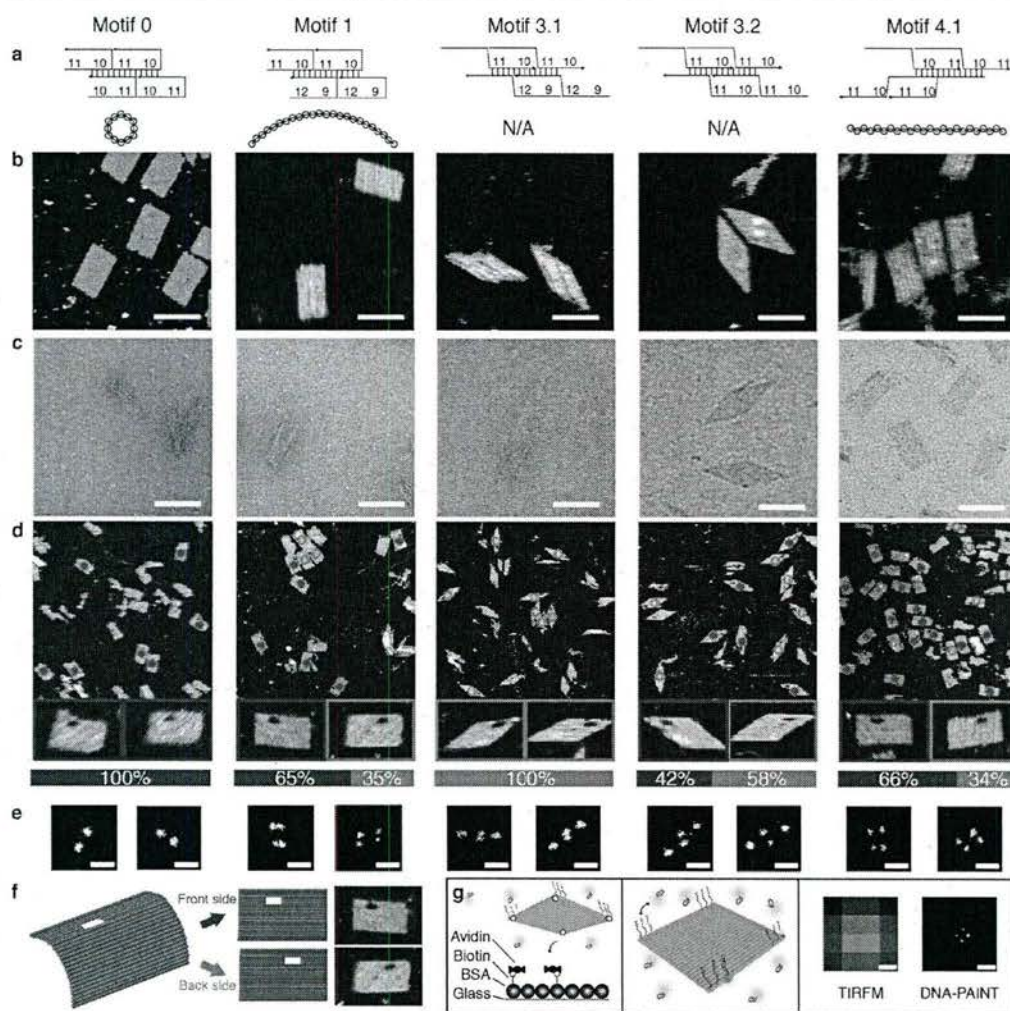
**Twisted Structures: Motifs 4.2–4.6.** Based on the corrugated structure from motif 4.1, we changed the crossover spacing from the canonical two full helical turns (i.e., 21 nt) to 20, 22, 24, 26, 18 nt using motifs 4.2–4.6, respectively. This introduced a significant internal twist to the structure.<sup>17,43,44</sup> The structures with 20 nt crossover spacing (motif 4.2) were used for the twist study and significant twist was observed (see below). The yield of the structure with the shortest spacing (motif 4.6) was relatively low (26%) compared to that of the rest of the group (33–42%).

**Double-Stranded Crossovers: Motifs 5.x, 6.x, and 7.x.** Structures formed from motifs 0–4 contain single-stranded, half crossovers between adjacent double-helices. We next tested structures with double-stranded, full crossovers. We first tried to shift the nicking points in the original U-shaped motif to be positioned away from the crossovers (motifs 5.1–5.3). Then we continued to design H-shaped motifs by increasing crossover spacing from two helical turns to four helical turns (motifs 6.1–6.3). We also designed X-shaped motifs that were derived Z-shaped motifs (motifs 7.1 and 7.2). Structures with full crossovers have nicks on the duplex (rather than at the crossovers), which could permit the use of enzymatic ligation to covalently join the two ends of an SST molecule to form a ring.

**Hybrid Motifs: Motifs 8–10.** We combined different motifs to form hybrid structures with both U-shaped and Z-shaped rows (motifs 8 and 9) or with H-shaped and X-shaped rows (motif 10).

**Linkers at Crossovers: Motifs 11.1–11.4.** At the crossovers, we placed single-stranded DNA linkers with varying lengths (20 nt, 10 nt, and 2 nt for motif 11.1, 11.3, and 11.4 respectively). We also tested to form structures with 15 bp double-stranded linkers (motif 11.2). In addition to the 10HX11T structure, we also used motif 11.1 to form a 24HX29T structure (Figure SI29), demonstrating the scalability of structures formed from linker motifs.

**Linkers between Neighboring Domains: Motif 12.** We formed structures from a motif with linkers between every pair of neighboring domains (motif 12). In the resulting structures, individual 10 or 11 bp double-stranded DNA duplexes were



**Figure 3.** Curvature characterization. See SI, section SI2.1 for curvature calculation and modeling details. (a) Strand diagrams (top) and curvature models in cross section views (bottom). (b) AFM images. Scale bars, 100 nm. See SI, section SI2.3 for larger images. (c) TEM images. Scale bars, 100 nm. See SI, section SI2.4 for larger images. (d) AFM-based landing assay results. A structure that landed with its front side facing up is marked with a blue dot inside a red circle; the one with back side facing up is marked by a hollow red circle. Magnified views show the structures with orientational markers (front side structures in blue boxes and back side structures in red boxes). The ratio between two landing orientations is shown in the bar chart in the bottom panel (blue, front side; red, back side). See SI, section SI2.5 for larger images and more details. (e) Representative DNA-PAINT super-resolution images of different SST structures labeled at four corners (scale bars, 100 nm). Measured point distances for motifs 0, 1, 3.1, 3.2, and 4.1 were 73 nm, 27 and 73 nm (short and long axis), 31 and 154 nm (parallelogram short and long diagonal), 45 and 148 nm (parallelogram short and long diagonal), and 45 and 77 nm (short and long axis), respectively. (f) Schematics of AFM-based landing assay. (g) Schematics of DNA-PAINT super-resolution imaging. Left panel: an SST structure was labeled on the top surface with 9 nt single-stranded "docking" sites for DNA-PAINT imaging strands. It was also labeled on the bottom surface with biotinylated strands for surface attachment. Center panel: imager strands bind transiently to the docking sites in the four corners, producing apparent blinking. Right panel: diffraction-limited TIRFM image (left) and super-resolution image (right). Scale bars, 100 nm.

connected by flexible 10 nt, single-stranded poly-T linkers. The structure was thus likely loosely connected without well-defined morphology. AFM imaging was performed only after the addition of 10 nt poly-A strands, which were used to complement the poly-T linkers, and helped to result in a structure with better defined shape. After poly-A complementation, grid-like structures were observed.

**Wire-frame Lattices: Motifs 13 and 14.** We constructed hexagonal lattices with different mesh sizes: one formed from a motif with three 21-nt domains (motif 13) and another from a motif with six 10-nt domains (motif 14). The hexagon pattern is similar to previous work on hexagon lattice-based tubes.<sup>41</sup> In

contrast to the typical structures composed of parallel-arranged double-helices linked by periodic (half) crossovers, these wire-frame structures<sup>40,41</sup> are composed of short duplex fragments connected with junction points. Interestingly, due to the electrostatic repulsions between adjacent DNA double-helices, a structure composed of parallel-arranged double-helices typically possesses a gap (e.g., 0.5–1.5 nm in DNA origami<sup>10,14,16</sup> and SST structures<sup>2,20,21</sup>) between the adjacent crossovers, and can be viewed as a wire-frame structure with a rectangular mesh pattern. In SI, section SI1.4, we discuss the transformation between parallel pattern and rectangular wire-frame pattern in more detail.

**Curvature. Modeling and Design.** The strand-weaving pattern of the strands can affect the curvature of the structure. Figure 2 summarizes our strategies to control structure curvature. SI, section SI2.1 gives modeling and analysis details.

**Domain Length Adjustment.** Working toward our goal of controlling curvature to design a flat structure, we first attempted to reduce curvature by adjusting the domain length of the canonical U-shaped SST motif. Assuming that the crossover points between adjacent double-helices are collinear with the centers of the adjacent double-helices, our model predicts an average curvature of  $30^\circ$  per double-helix for the motif 0, and  $720^\circ$  for the 24H×29T rectangle (see Figure 2a,e for cross-sectional view of zoomed-in structure and the overall structure; see SI, section SI2.1.1 for detailed calculation). This curvature was confirmed by our experimental data (Figure 3, column motif 0).

By adopting a domain length of 9/12 nt (motif 1) instead of the canonical 10/11 nt in motif 0, we were able to reduce the calculated average curvature per double-helix from  $30^\circ$  to  $-4^\circ$ , and  $-96^\circ$  for the 24H rectangle (see SI, section SI2.1.1 for detailed calculation). This reduced curvature was confirmed by experimental data (Figure 3, column motif 1).

However, a 2D structure (such as that formed by motif 0 or motif 1) designed with antiparallel crossovers will always possess a side that displays minor grooves at crossovers and another side that displays major grooves at crossovers.<sup>47</sup> As such, the curvature will not be perfectly counterbalanced for antiparallel crossover-based structures designed from U-shaped motifs (Figure 2a).

**Parallel Crossover Design.** Next, we hypothesize that the curvature induced by the asymmetric arrangement of major and minor grooves at the crossovers on the two sides of the structures would be negated in parallel crossover-based structures<sup>3</sup> formed from Z-shaped motifs: here, on either side of the structure, the crossovers display major and minor grooves in an alternating fashion. Like the U-shaped motif, we assume the crossover points of adjacent double-helices are collinear with the centers of the adjacent double-helices, and designed a Z-shaped motif (motif 3.1)-based structure that is expected to be perfectly flat (see Figures 2b, left panel; see SI, section SI2.1.2 for detailed calculation). However, our experimental data revealed significant curvature in the structure (Figure 3, column motif 3.1).

The discrepancy between our model and our experiments led us to revise the previous assumption that the crossover points of adjacent double-helices are collinear with the centers of the adjacent double-helices. Although our experimental results from U-shaped motifs (motifs 0 and 1) did not invalidate such an assumption, the results from Z-shaped motifs were clearly incompatible with this collinear assumption. To reconcile this discrepancy between the model and the experimental results, we introduced an offset angle  $\alpha$  formed between a base at the crossover and the two center points of the adjacent double-helices (middle panels of Figures 2b and SI39b). In a structure formed from motif 3.1, this offset angle at the crossover always appears on the same side of the structure, resulting in an accumulative curvature for this otherwise flat structure.

Incidentally, another structure based on a Z-shaped motif with a different domain length (motif 3.2) appeared approximately flat in our experiments (Figure 3, column motif 3.2). Assuming this structure (motif 3.2) has  $0^\circ$  curvature, we back-calculated an approximate value for the offset angle  $\alpha$  ( $8.5^\circ$ , Figure 2c). Using this  $\alpha$  value, the 24H×29T structure

formed from motif 3.1 was expected to have significant curvature (see Figure 2b, middle panel; see SI, section SI2.1.2 for detailed calculation), consistent with experimental observation of a curved structure.

**Corrugation Design.** We next devised a corrugated design (motif 4.1) where the offset angle  $\alpha$  appears alternately on the two sides of the structure, thus canceling out the curvature effect (independent of its value, Figure 2d,f; see SI, section SI2.1.3 for detailed calculation). The predicted flatness of the structure was experimentally verified (Figure 3, column motif 4.1).

**Curvature: Experimental Characterization.** We performed a case study below to test the curvature of 24H×29T rectangle structures of a number of motifs.

**Characterization by AFM and TEM.** To study the curvature of the structures, we constructed 24H×29T structures with each of the above five motifs. Sequence design and structure assembly were conducted in a similar fashion as their 10H×11T counterparts. Then the annealed samples were subjected to 2% native gel electrophoresis (see Figure SI41 for gel data), and the product band was extracted, purified by centrifugation, and imaged with AFM and TEM. Under AFM, all the structures appeared as flat, single-layer structures on mica surface (Figure 3b); no curvature information was revealed. However TEM images more effectively reflected the structure configurations in solution (Figure 3c): the structure generated from motif 0 ( $720^\circ$  curvature expected) appeared as a dense, rolled-up rectangle; the structure from motif 1 ( $-96^\circ$  expected) appeared as a wider rectangle; rolled-up and flat parallelograms were observed for the structure made from motifs 3.1 and 3.2, respectively; and a flat rectangle was observed for structure from motif 4.1 ( $0^\circ$  curvature expected).

**AFM-Based Landing Assay.** The different curvature configurations of the structures were also confirmed by two innovative assays. The first was the AFM-based landing assay<sup>20</sup> in which a small hole was introduced to the top-left corners of the structure as an orientational marker (Figure 3f). If the marker was more likely to show up at the top left corner rather than its mirror image (top right corner), it indicated the structure landed on the mica with its front side facing up, suggesting the structure curved up instead of down (motif 0). On the contrary, if the mirror image (back side facing up) was more likely, the structures had the tendency to curve down (motif 3.1). If the marker showed no bias for either side, the structure would be flat. Our experimental results (Figure 3d) showed a perfect bias for the rolled-up structures in the predicted curving direction [motif 0, 100% front side facing up ( $N = 36$ ); motif 3.1, 100% back side facing up ( $N = 57$ )]. However, possibly due to a limited sample size, the assay did not effectively differentiate between other less curved structures [motif 1, 65%, front side facing up ( $N = 74$ ); motif 3.2, 42% front side facing up ( $N = 45$ ); motif 4.1, 66% front side facing up ( $N = 203$ )]. Note that only well-formed structures (defined as those showing no defects  $>15$  nm in diameter in the expected boundary, nor  $>10$  nm in diameter in the interior of the structure) were counted in our statistics. See SI, section SI2.5 for details.

**Super-Resolution Fluorescence Microscopy Assay.** Super-resolution fluorescence microscopy was the second approach employed to assess the curvature of SST structures (Figure 3g). Here we used DNA-PAINT<sup>48,32,35</sup> to obtain sub-diffraction images of SST structures. DNA-PAINT exploits the repetitive, transient binding of short fluorescently labeled oligonucleotides

("imager" strands) to complementary strands ("docking" sites) on a structure or molecule of interest. Using total internal fluorescence microscopy (TIRFM), binding events of imager strands to docking sites were observed as single-molecule events and the fluorescence emission was fitted to a 2D Gaussian function, yielding sub-diffraction resolution images. SST structures were "labeled" with DNA-PAINT docking sites by extending the three outmost strands in each of the four corners by a 9 nt long sequence at their 3'-end (Figure 3g, center panel). The structures also carried four biotinylated strands pointing toward the opposite side of the docking strands for surface attachment (Figure 3g, left panel). Once bound to a functionalized glass slide, fluorescently labeled DNA-PAINT imager strands were introduced, which transiently bound to the docking strands. The right panel of Figure 3g includes a typical diffraction-limited TIRF image of a DNA nanostructure alongside the DNA-PAINT image after reconstruction. The increase in resolution is clearly visible.

The results of the DNA-PAINT study on the five different SST motifs are in good agreement with the TEM analysis as well as the AFM-based landing assay (Figure 3e). For motif 0, only two apparent points in a distance corresponding to the long side of the rectangular structure were visible, suggesting a "rolled-up" structure where the two corners were placed right next to each other, or only one of the two corners was available to bind the imager strand. In contrast, in the structure from motif 1 we observed four separate points, two of which were in close proximity, suggesting a slightly curved structure. Images for motifs 3.1 and 3.2 were consistent with an almost rolled-up structure (3.1) and a flat parallelogram structure (3.2), respectively. Finally, the image for motif 4.1 (with four visible corners) suggests a flat rectangular structure.

**Twist.** Besides assessing the curvature in different SST designs, we also studied the twist of the SST structures by creating homopolymers of motifs 4.1 and 4.2. We designed the leftmost column of the SST tiles to pair with the rightmost column of the rectangle. As a consequence, multiple units of rectangles align in tandem to form a long polymer with individual rectangles as monomer units (Figure SI58, top). Polymerization helps to identify the small degrees of twist that would be difficult if not impossible to visualize using only monomeric structures.<sup>17,44</sup> AFM images revealed that the homopolymers formed from motif 4.1 yielded ribbons with virtually no global twist (Figure SI58a, bottom). This was expected, as the design used 21 bases between adjacent crossover points, thus obeying 10.5 bp spacing per helical turn as in the natural form of B-DNA. In contrast, polymers formed from motif 4.2 showed a global left-handed twist (Figure SI58b, bottom). This is consistent with the fact that motif 4.2 used 20 bases between adjacent crossover points and was expected to produce a locally overwound double-helix. Our results agree with earlier studies of curvature and twist for DNA origami structures<sup>17,43,44</sup> and suggest that the twist of the SST structures can be modulated by changing crossover spacing distances.

## CONCLUSION

In summary, we evaluated more than 30 SST motifs that form lattice structures with diverse strand weaving patterns. Most of the motifs self-assembled into the designed structures with reasonable yields. Using some of the motifs, a systematic study was conducted to reduce the curvature in the assembled structure. The difference in curvature was characterized by

TEM, an AFM-based landing assay, and super-resolution microscopy. A corrugated weaving pattern was predicted and experimentally verified to produce a flat structure.

The simple and robust nature of SSTs has enabled us to survey the structural space for forming complex structures in a rapid and comprehensive fashion. An unprecedented collection of diverse weaving patterns was demonstrated. We were able to define simple-to-implement design rules that lead to structures with specific geometric properties such curvature and twist. More broadly, the diverse collection of motifs will allow us to modulate the geometrical, mechanical, and thermodynamic parameters. We demonstrated the tuning of curvature and twist with the new motifs. In a related work,<sup>49</sup> we demonstrated that by using the motifs reported here, we can form complex DNA nanostructures isothermally at any prescribed temperature between 15 and 69 °C, and under biocompatible conditions. The work here thus provides a new set of tools for modulating the structural properties of complex DNA structures, and reveals the broad design space for forming such structures.

## METHODS

Finite DNA nanostructures were first designed with different motifs, and then random sequences were generated to fill in the specific structures and populate complementary domains. Without careful adjustment of stoichiometry, unpurified strands were mixed and supplemented with 0.5×TE and 15 mM Mg<sup>2+</sup>. After one-pot annealing from 90 to 25 °C over 17 h, the solution was subjected to native agarose gel electrophoresis. The annealed samples, or the gel-purified samples from extraction of the desired band by centrifugation were imaged with AFM or TEM. The curvature and twist of certain structures were studied by AFM, TEM, and super-resolution fluorescent imaging.

**DNA Sequence Design.** Most DNA sequences were designed with the Uniquimer software by populating the motifs with random sequences<sup>50</sup> while maintaining the required complementarity relations. The following design rules were applied for sequence generation: (1) Nucleotides (i.e., A, C, G, T) are generated one by one randomly. (2) Complementary nucleotides to one generated are matched following the base-pairing rule: A to T, C to G and vice versa. (3) Specific segments (e.g., four consecutive A, C, G, T) are not allowed. When such segments emerge during design, the most recent generated nucleotide will be mutated until disallowance of specific segments is satisfied. Due to the difficulty of designing structures with wire-frame motifs, a separate script was developed following the same design principles. Manual design and/or optimization was used for the linker segment sequences (e.g., motifs 11.1–11.4 and motif 12) and the handle segment sequences (e.g., handle segment sequences for structures of motifs 0, 1, 3.1, 3.2, and 4.1 in super-resolution study).

**Sample Preparation.** DNA strands were synthesized by Integrated DNA Technology, Inc. (<http://www.idtdna.com>). To assemble the structures, DNA strands were mixed to roughly equal molar concentration of 200 nM for smaller structures (all 10H×11T structures, structures with linkers and wire-frame structures) and 100 nM for larger structures (24H×29T and 24H×28T structures for the curvature study) in 0.5×TE buffer (5 mM Tris, pH 7.9, 1 mM EDTA) supplemented with 15 mM MgCl<sub>2</sub>. The mixtures was then annealed in a PCR thermo cycler by cooling from 90 to 25 °C over a period of 17 h with a specific cooling program.<sup>20</sup> The annealed samples were then subjected to 2% agarose gel electrophoresis (gel prepared in 0.5×TBE buffer supplemented with 10 mM MgCl<sub>2</sub> and pre-stained with SYBR safe) in an ice water bath. If necessary, the target gel bands were excised out and put into a Freeze N' Squeeze column (Bio-Rad). The gel pieces were crushed using a microtube pestle in the column and the column was then directly subjected to centrifuge at 438g for 3 min. Purified samples were collected in the eluate, and concentrations were determined by Nanodrop absorption at 260 nm prior to AFM or TEM imaging.

**AFM Imaging.** AFM images were obtained using an SPM Multimode with Digital Instruments Nanoscope V controller (Veeco). A 5  $\mu\text{L}$  droplet (2–10 nM) of annealed (or purified) sample and then a 40  $\mu\text{L}$  drop of 0.5 $\times$ TE/10 mM  $\text{MgCl}_2$  solution were applied to a freshly cleaved mica surface and left for approximately 2 min. Sometimes, additional dilution of the sample was performed to achieve the desired sample density. As for the cases of 10H $\times$ 11T structures, 20  $\mu\text{L}$  of supplemental 10 mM  $\text{NiCl}_2$  was added to increase the strength of DNA–mica binding.<sup>51</sup> The images were taken under the liquid tapping mode, with C-type triangular tips (resonant frequency,  $f_0 = 40\text{--}75$  kHz; spring constant,  $k = 0.24$  N  $\text{m}^{-1}$ ) from the SNL-10 silicon nitride cantilever chip (Bruker Corp.).

**TEM Imaging.** For TEM imaging, a 3.5  $\mu\text{L}$  sample (1–5 nM) was adsorbed onto glow discharged carbon-coated TEM grids for 4 min and then stained for 1 min or a few seconds using a 2% aqueous uranyl formate solution containing 25 mM NaOH. Imaging was performed using a JEOL JEM-1400 TEM operated at 80 kV.

**Super-Resolution Imaging.** Super-resolution studies were performed with DNA PAINT.<sup>48</sup> The samples were diluted in DNA-PAINT imaging buffer (5 mM Tris-HCl (pH 8.0), 10 mM  $\text{MgCl}_2$ , 1 mM EDTA, 0.05% Tween-20) to 50 pM and immobilized to a coverslip in a flow chamber via biotin streptavidin binding interaction. To fix the samples, solutions containing 1 mg/mL BSA-biotin, 2 mg/mL streptavidin, and biotin-labeled DNA nanostructures were flushed into the flow chamber and incubated in the above order. The chamber was then washed with DNA-PAINT imaging buffer containing 20 nM ATTO655-labeled imager strands and sealed with 5 min epoxy before imaging. DNA-PAINT super-resolution images were acquired on an inverted Nikon Ti-E microscope (Nikon Instruments, Melville, NY) with the Perfect Focus System, applying an objective-type TIRF configuration using a Nikon TIRF illuminator with an oil-immersion objective (100 $\times$  Plan Apo, NA 1.49, Oil, Nikon). A 647 nm laser (Agilent MLC400B, 80 mW at the objective) was used for TIRF excitation. The laser beam was filtered with a clean up filter (642/20 Chroma Technologies) and coupled into the microscope objective using a multiband beam splitter (zt405/488-491/561/638rpc, Chroma Technologies). Fluorescence was spectrally filtered with an emission filter (700/75 Chroma Technologies) and imaged on an EMCCD camera (Andor iXon 3, Andor Technologies, North Ireland). A total of 5000 frames were recorded at a frame rate of 10 Hz. Super-resolution images were reconstructed using spot-finding and 2D Gaussian fitting algorithms programmed in LabVIEW (National Instruments) available for download at [www.dna-paint.org](http://www.dna-paint.org).

**Yield Quantification by Gel Electrophoresis.** Yield was estimated by analysis using native agarose gel electrophoresis, pre-stained with SYBR Safe DNA stain. The ratio between the fluorescence intensity of the product band and that of the entire lane was taken as an estimate of the gross yield of structural formation.

**Structure Size Measurement and Statistics.** AFM measurements were obtained using Nanoscope Analysis (version 1.20; Veeco). The “cross-section” function was used to measure distances (lengths and widths of the structures of different sizes). Well-formed structures were chosen for the measurements. TEM images were analyzed using ImageJ (version 1.46r; NIH). The straight line function was used to measure widths of certain structures. Six or ten sample points were collected for each distance measurement and the statistics (e.g., the mean and the standard deviation) were based on those 6 or 10 data points.

## ■ ASSOCIATED CONTENT

### Supporting Information

Experimental details, additional figures, and sequence data for each structure. This information is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

py@hms.harvard.edu

## Notes

The authors declare the following competing financial interest(s): A provisional patent based on this work is pending.

## ■ ACKNOWLEDGMENTS

We thank Wei Sun and Hoang Lu for technical assistance and David Zhang, Erik Winfree, William Sherman, and Robert Barish for discussion. DNA-PAINT data were collected at the Nikon Imaging Center at Harvard Medical School. This work was funded by Office of Naval Research (ONR) Young Investigator Program award N000141110914, ONR grant N000141010827, National Science Foundation (NSF) Faculty Early Career Development Award 1054898, NSF grant 1162459, National Institutes of Health Director’s New Innovator Award 1DP2OD007292, and a Wyss Institute Faculty Startup Fund to P.Y. M.D. acknowledges an international student research fellowship from the Howard Hughes Medical Institute. C.M. acknowledges a graduate research fellowship from the Fannie and John Hertz Foundation. R.J. acknowledges a postdoctoral fellowship from the Alexander von Humboldt Foundation.

## ■ REFERENCES

- (1) Seeman, N. C. *J. Theor. Biol.* **1982**, *99*, 237–247.
- (2) Chen, J.; Seeman, N. C. *Nature* **1991**, *350*, 631–633.
- (3) Fu, T. J.; Seeman, N. C. *Biochemistry* **1993**, *32*, 3211–3220.
- (4) Winfree, E.; Liu, F.; Wenzler, L. A.; Seeman, N. C. *Nature* **1998**, *394*, 539–544.
- (5) Chworos, A.; Severcan, I.; Koyfman, A. Y.; Weinkam, P.; Oroudjev, E.; Hansma, H. G.; Jaeger, L. *Science* **2004**, *306*, 2068–2072.
- (6) Rothmund, P. W. K.; Papadakis, N.; Winfree, E. *PLoS Biol.* **2004**, *2*, 2041–2053.
- (7) Shih, W. M.; Quispe, J. D.; Joyce, G. F. *Nature* **2004**, *427*, 618–621.
- (8) Goodman, R. P.; Schaap, I. A. T.; Tardin, C. F.; Erben, C. M.; Berry, R. M.; Schmidt, C. F.; Turberfield, A. J. *Science* **2005**, *310*, 1661–1665.
- (9) Park, S. H.; Pistol, C.; Ahn, S. J.; Reif, J. H.; Lebeck, A. R.; Dwyer, C.; Labean, T. H. *Angew. Chem., Int. Ed.* **2006**, *45*, 735–739.
- (10) Rothmund, P. W. K. *Nature* **2006**, *440*, 297–302.
- (11) He, Y.; Ye, T.; Su, M.; Zhang, C.; Ribbe, A. E.; Jiang, W.; Mao, C. *Nature* **2008**, *452*, 198–201.
- (12) Yin, P.; Hariadi, R.; Sahu, S.; Choi, H. M. T.; Park, S. H.; LaBean, T. H.; Reif, J. H. *Science* **2008**, *321*, 824–826.
- (13) Zheng, J. P.; Birktoft, J.; Chen, Y.; Wang, T.; Sha, R. J.; Constantinou, P.; Ginell, S.; Mao, C.; Seeman, N. C. *Nature* **2009**, *461*, 74–77.
- (14) Ke, Y.; Sharma, J.; Liu, M.; Jahn, K.; Liu, Y.; Yan, H. *Nano Lett.* **2009**, *9*, 2445–2447.
- (15) Andersen, E. S.; Dong, M.; Nielsen, M. M.; Jahn, K.; Subramani, R.; Mamdouh, W.; Golas, M. M.; Sander, B.; Stark, H.; Oliveira, C. L. P.; Pedersen, J. S.; Birkedal, V.; Besenbacher, F.; Gothelf, K. V.; Kjems, J. *Nature* **2009**, *459*, 73–76.
- (16) Douglas, S. M.; Dietz, H.; Liedt, T.; Hogberg, B.; Graf, F.; Shih, W. M. *Nature* **2009**, *459*, 414–418.
- (17) Dietz, H.; Douglas, S. M.; Shih, W. M. *Science* **2009**, *325*, 725–730.
- (18) Han, D.; Pal, S.; Nangreave, J.; Deng, Z.; Liu, Y.; Yan, H. *Science* **2011**, *332*, 342–346.
- (19) Delebecque, C. J.; Lindner, A. B.; Silver, P. A.; Aldaye, F. A. *Science* **2011**, *333*, 470–474.
- (20) Wei, B.; Dai, M.; Yin, P. *Nature* **2012**, *485*, 623–626.
- (21) Ke, Y.; Ong, L. L.; Shih, W. M.; Yin, P. *Science* **2012**, *338*, 1177–1183.
- (22) Han, D.; Pal, S.; Yang, Y.; Jiang, S.; Nangreave, J.; Liu, Y.; Yan, H. *Science* **2013**, *339*, 1412–1415.

- (23) Yurke, B.; Turberfield, A. J.; Mills, A. P.; Simmel, F. C.; Neumann, J. L. *Nature* **2000**, *406*, 605–608.
- (24) Sherman, W. B.; Seeman, N. C. *Nano Lett.* **2004**, *4*, 1203–1207.
- (25) Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A. *Nature* **2008**, *451*, 318–322.
- (26) Omabegho, T.; Sha, R.; Seeman, N. C. *Science* **2009**, *324*, 67–71.
- (27) Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. *Science* **2006**, *314*, 1585–1588.
- (28) Qian, L.; Winfree, E. *Science* **2011**, *332*, 1196–1201.
- (29) Yan, H.; Park, S. H.; Finkelstein, G.; Reif, J. H.; LaBean, T. H. *Science* **2003**, *301*, 1882–1884.
- (30) Kuzyk, A.; Schreiber, R.; Fan, Z.; Pardatscher, G.; Roller, E.; Högele, A.; Simmel, F. C.; Govorov, A. O.; Liedl, T. *Nature* **2012**, *483*, 311–314.
- (31) Choi, H. M. T.; Chang, J. Y.; Trinh, L. A.; Padilla, J. E.; Fraser, S. E.; Pierce, N. A. *Nat. Biotechnol.* **2010**, *28*, 1208–1212.
- (32) Lin, C.; Jungmann, R.; Leifer, A. M.; Li, C.; Levner, D.; Church, G. M.; Shih, W. M.; Yin, P. *Nat. Chem.* **2012**, *4*, 832–839.
- (33) Jin, Z.; Sun, W.; Ke, Y.; Shih, C.; Paulus, G. L. C.; Wang, Q. H.; Mu, B.; Yin, P.; Strano, M. S. *Nat. Commun.* **2013**, *4*, 1663.
- (34) Berardi, M. J.; Shih, W. M.; Harrison, S. C.; Chou, J. J. *Nature* **2011**, *476*, 109–113.
- (35) Derr, N. D.; Goodman, B. S.; Jungmann, R.; Leschziner, A. E.; Shih, W. M.; Reck-Peterson, S. L. *Science* **2012**, *338*, 662–665.
- (36) Douglas, S. M.; Bachelet, I.; Church, G. M. *Science* **2012**, *335*, 831–834.
- (37) Zhang, X.; Yan, H.; Shen, Z.; Seeman, N. C. *J. Am. Chem. Soc.* **2002**, *124*, 12940–12941.
- (38) Zhang, C.; Su, M.; He, Y.; Zhao, X.; Fang, P. A.; Ribbe, A.; Jiang, W.; Mao, C. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10665–10669.
- (39) Mao, C.; Sun, W.; Seeman, N. C. *J. Am. Chem. Soc.* **1999**, *121*, 5437–5443.
- (40) Malo, J.; Mitchell, J. C.; Venien-Bryan, C.; Harris, J. R.; Wille, H.; Sherratt, D. J.; Turberfield, A. J. *Angew. Chem., Int. Ed.* **2005**, *44*, 3057–3061.
- (41) Wilner, O. I.; Orbach, R.; Henning, A.; Teller, C.; Yehezkeili, O.; Mertig, M.; Harries, D.; Willner, I. *Nat. Commun.* **2011**, *2*, 540.
- (42) Yang, H.; Sleiman, H. *Angew. Chem., Int. Ed.* **2008**, *47*, 2443–2446.
- (43) Ke, Y.; Douglas, S. M.; Liu, M.; Sharma, J.; Cheng, A.; Leung, A.; Liu, Y.; Shih, W. M.; Yan, H. *J. Am. Chem. Soc.* **2009**, *131*, 15903–15908.
- (44) Jungmann, R.; Scheible, M.; Kuzyk, A.; Pardatscher, G.; Castro, C. E.; Simmel, F. C. *Nanotechnology* **2011**, *22*, 275301.
- (45) Kim, D.; Kilchherr, F.; Dietz, H.; Bathe, M. *Nucleic Acids Res.* **2011**, *40*, 2862–2868.
- (46) Liu, H.; Chen, Y.; He, Y.; Ribbe, A.; Mao, C. *Angew. Chem., Int. Ed.* **2006**, *45*, 1942–1945.
- (47) Rothmund, P. W. K.; Ekani-Nkodo, A.; Papadakis, N.; Kumar, A.; Fyngenson, D. K.; Winfree, E. *J. Am. Chem. Soc.* **2004**, *126*, 16344–16353.
- (48) Jungmann, R.; Steinhauer, C.; Scheible, M.; Kuzyk, A.; Tinnefeld, P.; Simmel, F. C. *Nano Lett.* **2010**, *10*, 4756–4761.
- (49) Myhrvold, C.; Dai, M.; Silver, P.; Yin, P. *Nano Lett.* **2013**, *13*, 4242–4248.
- (50) Wei, B.; Wang, Z.; Mi, Y. *J. Comput. Theor. Nanosci.* **2007**, *4*, 133–141.
- (51) Hansma, H. G.; Laney, D. E. *Biophys. J.* **1996**, *70*, 1933–1939.

# Isothermal Self-Assembly of Complex DNA Structures under Diverse and Biocompatible Conditions

Cameron Myhrvold,<sup>†,‡</sup> Mingjie Dai,<sup>‡,§</sup> Pamela A. Silver,<sup>†,‡</sup> and Peng Yin<sup>\*,†,‡</sup>

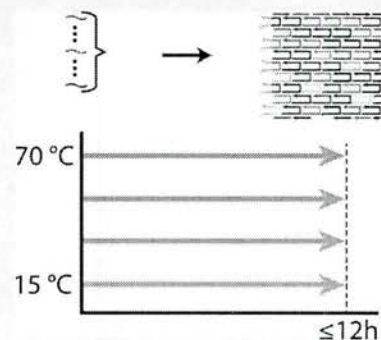
<sup>†</sup>Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, United States

<sup>‡</sup>Wyss Institute for Biologically Inspired Engineering and <sup>§</sup>Program in Biophysics, Harvard University, Boston, Massachusetts 02115, United States

## Supporting Information

**ABSTRACT:** Nucleic acid nanotechnology has enabled researchers to construct a wide range of multidimensional structures in vitro. Until recently, most DNA-based structures were assembled by thermal annealing using high magnesium concentrations and nonphysiological environments. Here, we describe a DNA self-assembly system that can be tuned to form a complex target structure isothermally at any prescribed temperature or homogeneous condition within a wide range. We were able to achieve isothermal assembly between 15 and 69 °C in a predictable fashion by altering the strength of strand–strand interactions in several different ways, for example, domain length, GC content, and linker regions between domains. We also observed the assembly of certain structures under biocompatible conditions, that is, at physiological pH, temperature, and salinity in the presence of the molecular crowding agent polyethylene glycol (PEG) mimicking the cellular environment. This represents an important step toward the self-assembly of geometrically precise DNA or RNA structures in vivo.

**KEYWORDS:** DNA nanotechnology, isothermal assembly, single-stranded tiles, biocompatible assembly, molecular crowding



Nucleic acid self-assembly has proven to be a powerful tool for constructing nanoscale structures due to the precise and predictable relationship between structure geometry and the encoded sequence. Over the past few decades, a wide range of DNA and RNA-based nanostructures have been demonstrated, including a variety of two-dimensional (2D) and three-dimensional (3D) geometries<sup>1–14</sup> and dynamic devices.<sup>15–22</sup> Additionally, such nanostructures can interface with and organize molecules, cells and other materials,<sup>1,12,23–27</sup> thereby enabling biophysical<sup>24,28</sup> and potential biomedical<sup>23,29,30</sup> applications. The self-assembly of individual strands into DNA nanostructures typically involves a thermal gradient in which the system temperature is first raised (e.g., to 80–90 °C) and then gradually lowered to room temperature. This thermal gradient can be replaced with a chemical gradient of a denaturing agent to isothermally assemble precisely controlled DNA origami structures.<sup>31</sup> In addition to gradient-based assembly, researchers have demonstrated the isothermal formation of extended crystals<sup>12,32,33</sup> using DNA and RNA tiles and discrete structures using DNA origami,<sup>34</sup> both under homogeneous conditions.

Isothermal assembly of DNA tiles and origami under homogeneous conditions represents important progress for the assembly of complex DNA nanostructures. Compared with thermal annealing, assembling structures in homogeneous conditions leads to rapid structure formation with higher yield and quality.<sup>34</sup> However, previous work on DNA tile crystals and origami still requires a highly optimized set of

assembly conditions (temperature, salinity, etc.) that are structure-dependent and restricted to a narrow range, thereby limiting the scope of potential applications. In contrast, assembly across a wide range of conditions will increase the range of applications of DNA nanotechnology, especially when the assembly conditions are specified by the application, such as in vivo scaffolding<sup>12</sup> of metabolic enzymes (where the assembly is required to happen under intracellular conditions) or in situ imaging<sup>35</sup> (where the assembly is required to happen under harsh denaturing conditions). Thus, the next important challenge is to devise a general method in which, given a particular homogeneous condition and a target shape, the user can design a system that assembles under that particular condition into the desired shape, e.g. the isothermal assembly of a particular shape under any prescribed temperature over a wide temperature range. Such capability will greatly expand the application scope where an assembly system can be designed to operate under diverse homogeneous conditions specified by the desired application. Here we provide a general solution to this challenge.

The key innovation in our approach is that we exploit the tunable nature of single-stranded tile (SST) structures and demonstrate structural designs that can assemble isothermally at a wide range of temperature and salt concentrations,

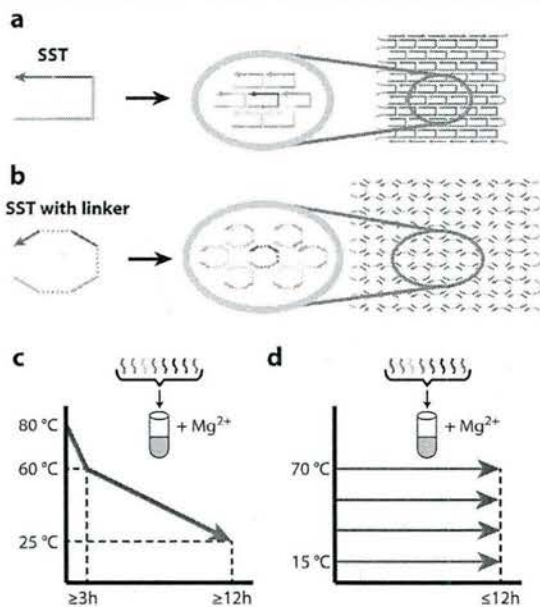
Received: May 28, 2013

Revised: July 4, 2013

Published: August 26, 2013

including conditions mimicking intracellular environments. Compared with DNA origami, which typically uses a fixed scaffold sequence, single-stranded tiles (SSTs) give us the freedom to independently vary the sequence of any tile in the structure. We use this freedom to program the strength of strand–strand interactions, thereby shifting the optimal assembly temperature to a user-specified temperature. In other words, the assembly temperature of a structure is now an explicit design parameter. This will allow one to tune the design of a structure to fit a particular application.

We report the successful isothermal assembly of modified SST structures across a wide range of temperatures. Building upon our previous SST work,<sup>5,13,14,36</sup> we test 11 compact 2D SST structures (Figure 1a) and 17 flexible variants with single-



**Figure 1.** Self-assembly of DNA tiles under diverse and biocompatible conditions. (a,b) We use standard single-stranded tiles (SST; a) and SST tiles containing single-stranded linker regions between domains (b) to assemble 2D rectangular structures. A canonical SST has four domains that bind to four of its nearest neighbors. Each colored segment depicts an individual domain (unique in sequence and between 8 and 21 nt long in our study), a gray solid connection delineates the boundary between two domains, and a gray dotted line depicts a single-stranded linker region (between 1 and 16 nt in our study). Each complete structure (shown on the right) consists of 66 unique SSTs with edges protected by polyT segments to avoid aggregation (light gray segments). (c) A schematic depicting the traditional thermal annealing protocol for structure assembly, where temperature is generally ramped above 60 °C and slowly decreased over at least 12 h. (d) A schematic depicting our isothermal assembly protocol across a wide range of fixed temperatures and conditions for up to 12 h.

stranded polyT linkers between the double-stranded domains (Figure 1b).<sup>36</sup> Instead of the traditional thermal annealing protocol (Figure 1c), all structures were assembled isothermally (Figure 1d). Our approach therefore represents a general method for isothermally assembling complex nanostructures across a wide range of specified temperatures. We generally find higher assembly temperatures for structures with longer domains, and lower temperatures for those with linkers. We

also show that the assembly temperature can be modulated by changing the strand–strand binding energy in several ways, such as altering GC content of binding domains or disrupting continuity of complementary segments.

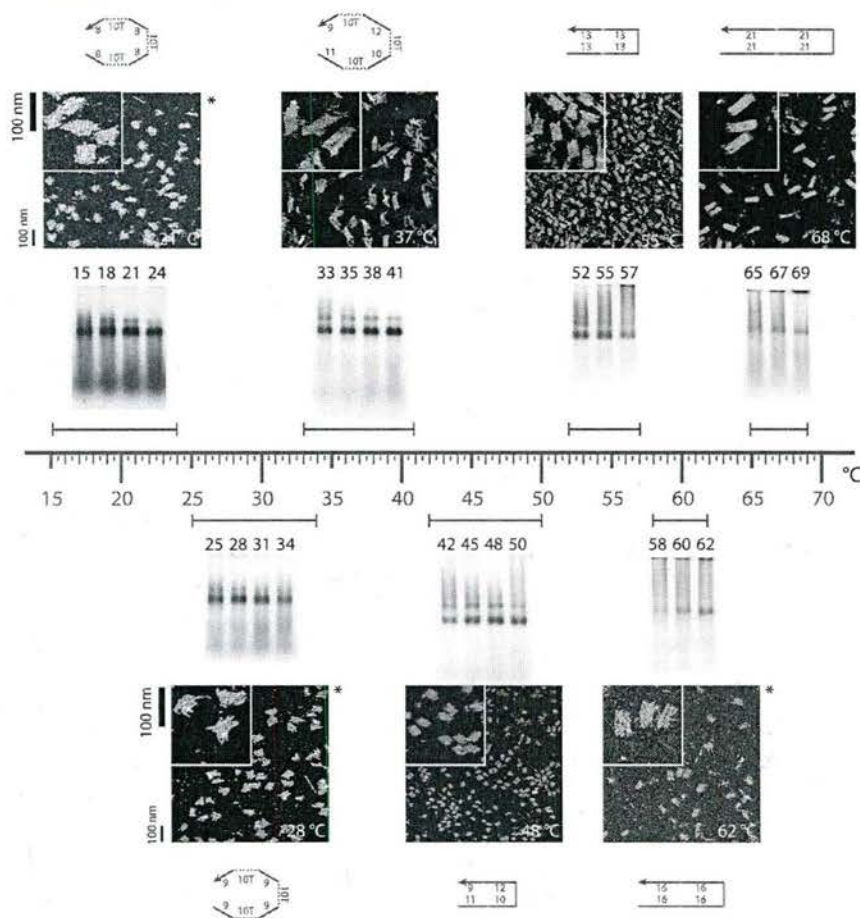
Finally, we report successful assembly of designed DNA nanostructures under biocompatible conditions, that is, at physiological temperature and pH, low salinity, and in the presence of molecular crowding agents. Flexible structures, which assemble at a wider range of conditions than their nonflexible counterparts, assemble well under biocompatible conditions. This represents a significant step toward the *in vivo* assembly of geometrically precise, nucleic acid nanostructures. More generally, this approach may enable additional biological applications of nanostructures in cases where *in situ* delivery of preassembled structures is not an option.

In our single-stranded tile (SST) system, each strand consists of four binding domains, with or without a single-stranded linker region between consecutive domains. In the presence of divalent cations, each strand is bound to its four neighbors via complementary domains, resulting in a self-assembled 2D SST structure.<sup>13</sup> Most of our structures are derived from motifs in our recent work<sup>36</sup> and consist of 66 SSTs that form a 2D rectangular shape. Two categories of motifs are used. The motifs without linkers form 2D rectangles composed of 10 parallel helices (with length ranging from 108 to 252 nt) connected by periodic single-stranded crossovers. Motifs with polyT linkers form flexible rectangular, fish-net patterns containing short segments of DNA helices connected by single-stranded linkers at all junctions. We varied domain length, linker length, and other design parameters. For an overview of the system, see Supporting Information Figure S1 and Text S1.

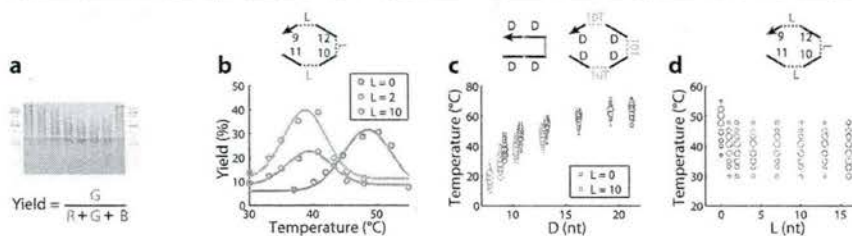
SST structures can isothermally assemble across a wide range of temperatures (Figure 2). We tested a set of U-shaped motifs<sup>13,36</sup> to form 2D rectangular structures. By altering the domain lengths of the SST motifs from 8 to 21 nt, we were able to achieve isothermal assembly across a 54 °C range (from 15 to 69 °C). Structure formation was assayed by gel electrophoresis and atomic force microscopy (AFM). A dominant product band was observed after isothermal assembly for 12 h at the indicated temperatures in the presence of 10 mM Mg<sup>2+</sup>. The formation ranges indicated on the temperature line are a subset of the temperature ranges over which a dominant product band was observed on the gel (see Supporting Information Figure S2 for full range and raw gel data). In some cases, we observed the presence of high molecular weight bands or smears, presumably due to structure aggregation.

Imaging of the unpurified reaction or of the gel-purified product (indicated with an asterisk) revealed the formation of structures with the designed shapes. Structures with polyT linkers between domains were imaged after the addition of complementary polyA strands. They exhibited a wide range of conformations when imaged due to the flexibility inherent in the linker regions.

The assembly yield was quantified using native gel electrophoresis (Figure 3a). We assembled structures isothermally for 1 h (domain length <16 nt) or 12 h (domain length ≥16 nt) at a range of temperatures and then quantified the gel yield in each lane (which corresponds to a particular, fixed assembly temperature) using the TotalLab Quant gel quantification software (Figure 3b). Here, the yield is defined as the ratio of the intensity of the product band divided by the intensity of the entire lane after proper background correction.



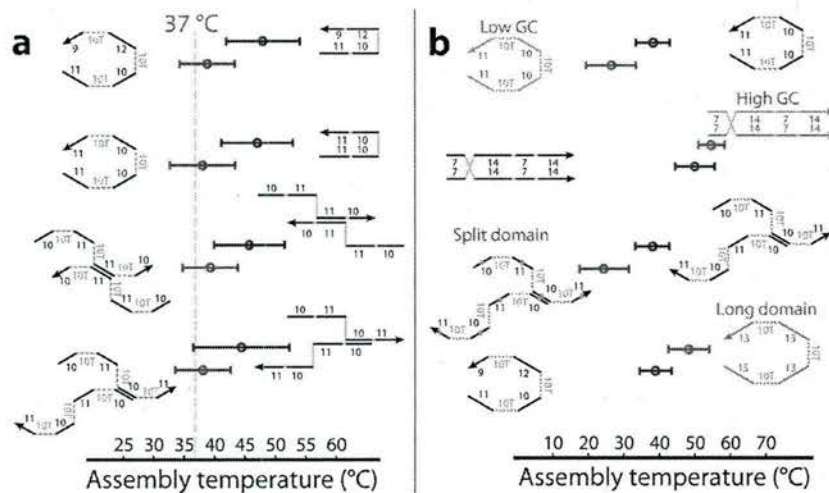
**Figure 2.** SST-based DNA nanostructures can assemble isothermally across a wide range of temperatures. Representative structures that assemble isothermally at different temperatures are shown and collectively cover the entire temperature range from 15 °C (blue) to 69 °C (red). Structures are ordered and displayed close to their optimal formation temperatures. For each structure, strand diagrams (outside), AFM images (middle), and gel electrophoresis results (inside) are shown. A tagged bar indicates part of the range of successful assembly temperatures shown by agarose gel electrophoresis (see Supporting Information Figure S2 for full gel images). The leftmost three structures contain 10T linkers between domains; the other structures do not, as shown in the strand diagrams. Numbers in strand diagrams indicate domain lengths (unit: nt). AFM images show the typical morphology at the indicated temperatures; images with an asterisk contain gel-purified structures. Structures were assembled isothermally for 12 h using 200 nM of each strand and 10 mM  $Mg^{2+}$ .



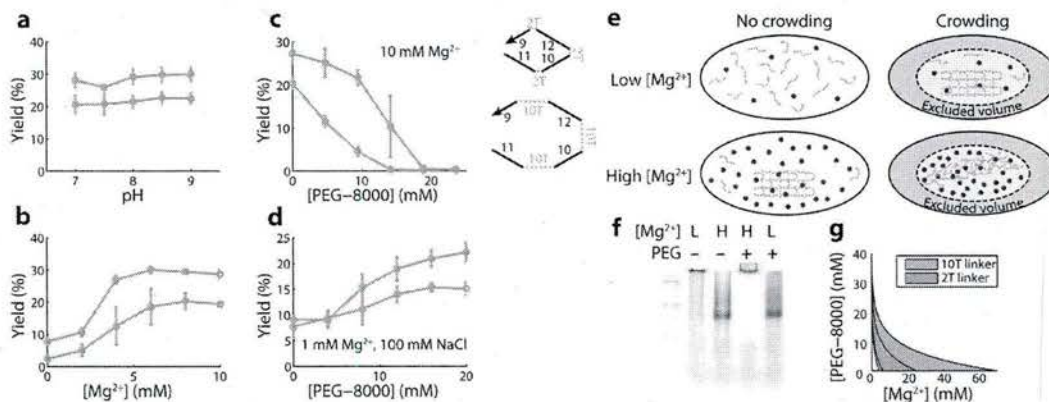
**Figure 3.** Quantification of optimal assembly temperature and effect of domain and linker lengths. Structures were assembled isothermally for 1 h across a temperature range using 200 nM of each strand and 10 mM  $Mg^{2+}$  and assayed using gel electrophoresis. (a) Gel images were quantified using TotalLab Quant software. Gel yield was defined as the integrated intensity of the formation band (green) divided by the total intensity of the lane (red + green + blue). (b) Gel yield of three representative structures ( $L = 0, 2, 10$ ) and Gaussian fits (solid lines) are shown. (c,d) Formation yields as a function of temperature are shown for structures with varying domain lengths ( $D = 8-21$ ; c) and linker lengths ( $L = 0-16$ ; d). Each circle represents a single yield quantification, with its radius proportional to the measured yield; colors indicate linker length from no linker ( $L = 0$ ; green) to 16T linker ( $L = 16$ ; dark orange).

This is only an approximation of the true formation yield<sup>13</sup> but is useful for comparing relative yields for assembly under varying conditions. The yield curves were then fit to a Gaussian

function with a constant background using custom MATLAB software (Figure 3b). The yield as a function of temperature is



**Figure 4.** Optimal assembly temperatures can be rationally designed with linker domains and other methods. (a) Shifting to lower optimal assembly temperatures (blue) by addition of linkers (dashed line: physiological temperature). Strand diagrams of four representative designs with distinct strand geometry are shown. (b) Rational design of optimal assembly temperature (blue, lower; red, higher) with three other methods: altering the GC content of domains to 30% (low GC) or 70% (high GC), splitting domains in half and changing domain lengths. In the case of splitting domains, a single T nucleotide is inserted in the middle of each domain (blue dots) to split each strand into a 5-T-5-10T-6-T-5-10T-5-T-5-10T-5-T-6 sequence pattern. In both panels, numbers in strand diagrams indicate domain lengths (unit: nt), tagged bars delimit the full width at half-maximum (fwhm) of a Gaussian fit to the yield as a function of assembly temperature, and a circle denotes the optimal temperature (see Supporting Information Figure S4 for raw gel data).



**Figure 5.** Assembly of structures under biocompatible conditions. (a–d) Effects of assembly conditions on formation yield of two representative structures (olive, 2T linker; orange, 10T linker; see strand diagrams). Structures were assembled for 1 h at 37 °C using 200 nM of each strand. Yields are calculated as in Figure 3a. Error bars indicate the standard deviation based on  $\geq 3$  replicate experiments. (a–c) Effects of pH,  $[Mg^{2+}]$ , and  $[PEG-8000]$  on formation yield. Each parameter is varied alone while keeping the others at default values (0.5 $\times$  TE buffer, pH 8, 10 mM  $Mg^{2+}$ , no PEG-8000). HEPES buffer was used for the pH 7 condition. (d) Combined effect of low salinity (1 mM  $Mg^{2+}$ , 100 mM NaCl) and  $[PEG-8000]$ . (e) A schematic of the crowding model: formation takes place at in vitro (no crowding, high salinity) or biocompatible (crowding, low salinity) conditions. Orange: component strands. Dark green: magnesium ions. (f) Gel result for the 10T linker structure at the four conditions shown in (e): with or without PEG-8000 (–, 0 mM; +, 20 mM), with low and high salinity (L, 1 mM  $Mg^{2+}$  100 mM Na<sup>+</sup>; H, 10 mM  $Mg^{2+}$ ). (g) A simple mathematical model of crowding, based on effective concentrations (see Supporting Information Text S3 for details), predicts assembly under low  $[Mg^{2+}]$  and high  $[PEG-8000]$  conditions.

well-fit by a Gaussian function, indicating that our structures assemble optimally at a specific isothermal temperature.

We show dependence between the assembly temperature and the domain length (Figure 3c; see Supporting Information Figure S3a for raw gel data). The Gaussian fits obtained from curve fitting to the gel quantification data allowed us to obtain a mean and a full-width at half-maximum (fwhm) for each structure that define an optimal temperature and an optimal temperature range, respectively. The formation range is indicated by the extent of the scatterplot, and the relative

yields at each temperature are indicated by the diameter of the marker. We examined two sets of structures, those with no linker between domains (green) and those with a 10T linker between domains (orange). In both cases, we observed that the formation range was a monotonically increasing function of the domain length with diminishing returns at higher temperatures. It should be noted that since all of our domains have a similar GC content (around 50%) that domain length is a proxy for the strength of strand–strand interactions in these experiments.

Linkers between domains caused an  $\sim 10$  °C downward shift in the assembly temperature (Figure 3d; see Supporting Information Figure S3b for raw gel data). Even a single T nucleotide between domains was sufficient to shift the assembly temperature; additional T nucleotides did not contribute to the temperature shift. As in Figure 3b, we show the formation range using a scatterplot, where the relative yield at each temperature is indicated by the diameter of the marker. The effect of the linkers did not depend on the polyT sequence; a linker consisting of 10 A nucleotides caused a similar downward shift in assembly temperature (see Supporting Information Figure S3c).

We redesigned many structures to assemble at the physiological temperature using polyT linkers (Figure 4a; see Supporting Information Figure S4a for raw gel data). As shown, the effect of polyT linkers was similar for several different strand geometries, causing a downward shift of the assembly temperature of 7–10 °C. In all cases, the redesigned structures formed well at the physiological temperature (37 °C), indicating the generality of this approach for programming the assembly temperature of DNA nanostructures. Furthermore, we demonstrated several additional strategies to shift the assembly temperature (Figure 4b; see Supporting Information Figure S4b for raw gel data): altering the strength of strand–strand interactions via the GC content, changing the length of the complementary domains within a structure, or splitting domains in half by inserting a single T nucleotide. Thus, there exists a wide variety of ways to program the assembly temperature of a DNA nanostructure. This should enable structures with varying geometries and morphologies that assemble at desired temperatures.

Flexible SST DNA structures assembled under biocompatible conditions. A number of structure designs assembled well at the physiological temperature (37 °C; Figures 2, 3, and 4). However, there exist other environmental variables, such as pH, salinity, and molecular crowding, which must be accounted for in order for structure assembly to be biocompatible. We tested two structure designs which assembled well at 37 °C in 1 h (Figure 3b), one with a 2T linker (olive) and one with a 10T linker (orange) by systematically varying other environmental variables in vitro (Figure 5; see Supporting Information Figure S5 for raw gel data). Yields are determined as in Figure 3a.

Physiological pH is compatible with structure assembly (Figure 5a). We varied the pH of the buffer in which the structures were assembled from 7 to 9, which encompassed the physiological pH range. Both structures assembled well across this pH range, indicating that the physiological pH should not prevent SST structures from self-assembling.

Low salinity hindered structure formation (Figure 5b). DNA nanostructures are typically assembled in vitro with a  $Mg^{2+}$  concentration in excess of 10 mM.<sup>4,6,13,37</sup> Recently, it was shown that sodium ions can also be interchanged for magnesium ions in some cases.<sup>38</sup> However, the free concentration of magnesium equivalents in the *E. coli* cytoplasm is approximately 2–4 mM (1–2 mM  $Mg^{2+}$ , 180–200 mM  $Na^+$  or  $K^+$ <sup>39,40</sup>). Our flexible SST designs assembled close to optimally at as low as 6 mM  $Mg^{2+}$  (2T linker, olive) or 4 mM  $Mg^{2+}$  (10T linker, orange), approaching but not quite reaching the biologically relevant salinity.

Crowded conditions decreased the assembly yield (Figure 5c). A crowded environment was simulated using polyethylene glycol with an average molecular weight of 8000 Da (PEG-8000). To test the effect of crowding on structure assembly, the

PEG-8000 concentration was varied from 0 to 23.5 mM. A  $Mg^{2+}$  concentration of 10 mM was used in these experiments. We observed a decline in yield for both structure designs with increasing PEG-8000 concentrations, combined with the appearance of a high molecular weight band indicative of aggregation (see Supporting Information Figure S5c for raw gel data). These results indicate that a combination of crowding and high magnesium decreases assembly yield. Additionally, we have noted that the long linker design (orange) had a consistently higher yield than the short linker design (olive) under the conditions we tested (Figure 5a–d), perhaps due to increased flexibility enhancing the assembly process (e.g., by relieving the electrostatic repulsion between parallel DNA helices).

Molecular crowding compensated for low magnesium, thereby enhancing structure assembly (Figure 5d). When considered separately, both salinity and crowding pose challenges to biocompatible assembly of tile-based DNA nanostructures. However, when we tested the two conditions together by varying the PEG-8000 concentration from 0 to 20 mM with low salinity (1 mM  $Mg^{2+}$ , 100 mM  $Na^+$ ) at pH 7.5, we observed that increased crowding actually improved the assembly yield. This effect was seen for both short (olive) and long linker designs (orange). Together, these results indicate that SST DNA nanostructures can be designed to assemble under biocompatible conditions.

We conjecture that molecular crowding enhances structure assembly by increasing the effective magnesium concentration (Figure 5e). In the absence of crowding, high salinity is required for assembly (top left), but in the presence of crowding high salinity inhibits assembly. This model is consistent with our experimental results; structure assembly (product band) is observed under high salinity with no crowding and low salinity with crowding but not under low salinity with no crowding or under high salinity with crowding (Figure 5f). We used this insight to construct a quantitative model of the effect of molecular crowding on structure assembly and predict a range of conditions under which assembly is likely to occur, indicated by the shaded regions (Figure 5g; see Supporting Information Text S3 and Figure S6 for details).

SST structures can be tuned to assemble isothermally across a range of temperatures from 15 to 69 °C. Additionally, we demonstrated four different ways to modulate the assembly temperature of a structure: changing domain length, altering domain GC content, adding a linker between domains, or splitting a domain in half. This has enabled us to design SST structures that can assemble well under biocompatible conditions. More generally, we now have the capability to design structures that can assemble under a diverse range of specified conditions. We noticed that structures tend to assemble with lower yield at extreme temperatures or conditions. However, the yield at a particular condition, or the range of assembly conditions, can be increased by extending the reaction time. Together, these results indicate that the isothermal SST assembly process is remarkably robust to the assembly temperature and conditions.

Our ability to assemble structures at room temperature can enable the study of structure assembly in real time using atomic force microscopy<sup>22,32</sup> or super-resolution microscopy techniques;<sup>41</sup> at the other end of the spectrum we can assemble structures at temperatures above the typical melting temperature of previous DNA nanostructures. More broadly, our

unique ability to control the assembly temperature of DNA nanostructures allows structured to be tailored to in vivo or in situ applications where control of the ambient temperature may not be feasible. Our success in biocompatible assembly is particularly exciting because it is a key step toward in vivo assembly of geometrically precise nanostructures from many individual DNA or RNA components. Furthermore, biocompatible assembly takes us beyond the paradigm of annealing structures over many hours under precisely controlled reaction conditions. Our approach of rationally modulating the assembly conditions should be generalizable to other types of structures with a few potential constraints. One such constraint is that the component strands should not contain significant secondary structures that hinder their interactions with other strands. For example, the secondary structure of a typical M13 DNA origami scaffold strand may interfere with the designed structure formation at low temperatures. A scaffold with designed sequence could be used to overcome this constraint.

Beyond tunable SST structures, control over the assembly conditions could perhaps be further improved by exploiting toe-hold-mediated<sup>42</sup> strand displacement cascades,<sup>17,43</sup> for example, using the triggered isothermal assembly of reconfigurable hairpins.<sup>18,35,44</sup> Although the resulting hairpin-based structures (e.g., branched junctions,<sup>18</sup> polymers,<sup>35,44,45</sup> or dendrimers<sup>18</sup>) are not geometrically precise and typically involve very few component strands, they can be formed dynamically via the introduction of an initiator strand. Additionally, a simple two-component toe-hold exchange based system has been shown to be robust to temperature, salinity, and concentration.<sup>46</sup> Thus, by integrating robustness of toe-hold based systems with the tunability and geometrical control of the SST systems described here, one may potentially design self-assembly systems that could allow for assembly across a wide range of temperatures and conditions without the need for condition-specific tuning.

A rational next step after biocompatible assembly is to assemble geometrically precise RNA nanostructures in vivo and to use them to organize metabolic enzymes or other functional biomolecules.<sup>12</sup> One could build upon the knowledge and experience from DNA SST assembly to design and rapidly prototype RNA structures that assemble under biocompatible conditions and act as scaffolds. With greater freedom in design geometry and positioning of protein-binding sites, one would be able to design and test interactions between multiple functional molecules across varying distances and geometrical patterns. One can thus survey a wider design space than has been achieved with previous protein<sup>47,48</sup> or RNA<sup>12</sup> scaffolds. DNA or RNA nanostructures could be delivered or expressed as individual strands and assembled directly at the site of interest. Taken together, our results present a new opportunity to assemble precise, programmable structures in biological and cellular environments, which could lead to exciting in vitro and in vivo applications.

## ■ ASSOCIATED CONTENT

### Supporting Information

Supplementary methods, structure naming conventions, notes about experiments, modeling details, and six supplementary figures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Bryan Wei, Dave Zhang, Erik Winfree, and Robert Barish for discussion. This work was funded by a DARPA Living Foundries grant HR0011-12-C-0061 to P.A.S. and P.Y., an Office of Naval Research (ONR) Young Investigator Program award N000141110914 to P.Y., an ONR grant N000141010827 to P.Y., a National Science Foundation (NSF) Faculty Early Career Development Award 1054898 and an NSF grant 1162459 to P.Y. C.M. acknowledges a graduate research fellowship from the Fannie and John Hertz Foundation. M.D. acknowledges an international student research fellowship from the Howard Hughes Medical Institute.

## ■ REFERENCES

- (1) Yan, H.; Park, S. H.; Finkelstein, G.; Reif, J. H.; LaBean, T. H. *Science* **2003**, *301*, 1882–4.
- (2) Chworos, A.; Severcan, I.; Koyfman, A. Y.; Weinkam, P.; Oroudjev, E.; Hansma, H. G.; Jaeger, L. *Science* **2004**, *306*, 2068–72.
- (3) Rothmund, P. W. K.; Papadakis, N.; Winfree, E. *PLoS Biol.* **2004**, *2*, e424.
- (4) Rothmund, P. W. K. *Nature* **2006**, *440*, 297–302.
- (5) Yin, P.; Hariadi, R. F.; Sahu, S.; Choi, H. M. T.; Park, S. H.; Labean, T. H.; Reif, J. H. *Science* **2008**, *321*, 824–6.
- (6) Douglas, S. M.; Dietz, H.; Liedl, T.; Högberg, B.; Graf, F.; Shih, W. M. *Nature* **2009**, *459*, 414–8.
- (7) Dietz, H.; Douglas, S. M.; Shih, W. M. *Science* **2009**, *325*, 725–30.
- (8) Zheng, J.; Birktoft, J. J.; Chen, Y.; Wang, T.; Sha, R.; Constantinou, P. E.; Ginell, S. L.; Mao, C.; Seeman, N. C. *Nature* **2009**, *461*, 74–7.
- (9) Andersen, E. S.; Dong, M.; Nielsen, M. M.; Jahn, K.; Subramani, R.; Mamdouh, W.; Golas, M. M.; Sander, B.; Stark, H.; Oliveira, C. L. P.; Pedersen, J. S.; Birkedal, V.; Besenbacher, F.; Gothelf, K. V.; Kjems, J. *Nature* **2009**, *459*, 73–6.
- (10) Yang, H.; McLaughlin, C. K.; Aldaye, F. A.; Hamblin, G. D.; Rys, A. Z.; Rouiller, I.; Sleiman, H. F. *Nature Chem.* **2009**, *1*, 390–6.
- (11) Han, D.; Pal, S.; Nangreave, J.; Deng, Z.; Liu, Y.; Yan, H. *Science* **2011**, *332*, 342–6.
- (12) Delebecque, C. J.; Lindner, A. B.; Silver, P. A.; Aldaye, F. A. *Science* **2011**, *333*, 470–4.
- (13) Wei, B.; Dai, M.; Yin, P. *Nature* **2012**, *485*, 623–6.
- (14) Ke, Y.; Ong, L. L.; Shih, W. M.; Yin, P. *Science* **2012**, *338*, 1177–1183.
- (15) Sherman, W. B.; Seeman, N. C. *Nano Lett.* **2004**, *4*, 1203–1207.
- (16) Yin, P.; Yan, H.; Daniell, X. G.; Turberfield, A. J.; Reif, J. H. *Angew. Chem., Int. Ed.* **2004**, *43*, 4906–11.
- (17) Zhang, D. Y.; Turberfield, A. J.; Yurke, B.; Winfree, E. *Science* **2007**, *318*, 1121–5.
- (18) Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. *Nature* **2008**, *451*, 318–22.
- (19) Omabegho, T.; Sha, R.; Seeman, N. C. *Science* **2009**, *324*, 67–71.
- (20) 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* **2010**, *465*, 206–10.
- (21) Qian, L.; Winfree, E. *Science* **2011**, *332*, 1196–201.
- (22) Wickham, S. F. J.; Bath, J.; Katsuda, Y.; Endo, M.; Hidaka, K.; Sugiyama, H.; Turberfield, A. J. *Nat. Nanotechnol.* **2012**, *7*, 169–73.
- (23) Douglas, S. M.; Bachelet, I.; Church, G. M. *Science* **2012**, *335*, 831–4.
- (24) Derr, N. D.; Goodman, B. S.; Jungmann, R.; Leschziner, A. E.; Shih, W. M.; Reck-Peterson, S. L. *Science* **2012**, *338*, 662–5.

- (25) Acuna, G. P.; Möller, F. M.; Holzmeister, P.; Beater, S.; Lalkens, B.; Tinnefeld, P. *Science* **2012**, *338*, 506–10.
- (26) Lin, C.; Jungmann, R.; Leifer, A. M.; Li, C.; Levner, D.; Church, G. M.; Shih, W. M.; Yin, P. *Nature Chem.* **2012**, *4*, 832–9.
- (27) Maune, H. T.; Han, S.-P.; Barish, R. D.; Bockrath, M.; Iii, W. A. G.; Rothmund, P. W. K.; Winfree, E. *Nat. Nanotechnol.* **2010**, *5*, 61–6.
- (28) Douglas, S. M.; Chou, J. J.; Shih, W. M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6644–8.
- (29) Lee, H.; Lytton-Jean, A. K. R.; Chen, Y.; Love, K. T.; Park, A. I.; Karagiannis, E. D.; Sehgal, A.; Querbes, W.; Zurenko, C. S.; Jayaraman, M.; Peng, C. G.; Charisse, K.; Borodovsky, A.; Manoharan, M.; Donahoe, J. S.; Truelove, J.; Nahrendorf, M.; Langer, R.; Anderson, D. G. *Nat. Nanotechnol.* **2012**, *7*, 389–93.
- (30) Liu, X.; Xu, Y.; Yu, T.; Clifford, C.; Liu, Y.; Yan, H.; Chang, Y. *Nano Lett.* **2012**, *12*, 4254–9.
- (31) Jungmann, R.; Liedl, T.; Sobey, T. L.; Shih, W.; Simmel, F. C. *J. Am. Chem. Soc.* **2008**, *130*, 10062–3.
- (32) Evans, C. G.; Hariadi, R. F.; Winfree, E. *J. Am. Chem. Soc.* **2012**, *134*, 10485–92.
- (33) Schulman, R.; Yurke, B.; Winfree, E. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 6405–10.
- (34) Sobczak, J. J.; Martin, T. G.; Gerling, T.; Dietz, H. *Science* **2012**, *338*, 1458–61.
- (35) Dirks, R. M.; Pierce, N. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15275–8.
- (36) Wei, B.; Dai, M.; Myhrvold, C.; Ke, Y.; Jungmann, R.; Yin, P. *J. Am. Chem. Soc.*, in press.
- (37) Winfree, E.; Liu, F.; Wenzler, L. a; Seeman, N. C. *Nature* **1998**, *394*, 539–44.
- (38) Martin, T. G.; Dietz, H. *Nat. Commun.* **2012**, *3*, 1103.
- (39) Alatossava, T.; Jütte, H.; Kuhn, A.; Kellenberger, E. *J. Bacteriol.* **1985**, *162*, 413–9.
- (40) Shabala, L.; Bowman, J.; Brown, J.; Ross, T.; McMeekin, T.; Shabala, S. *Environ. Microbiol.* **2009**, *11*, 137–48.
- (41) Jungmann, R.; Steinhauer, C.; Scheible, M.; Kuzyk, A.; Tinnefeld, P.; Simmel, F. C. *Nano Lett.* **2010**, *10*, 4756–61.
- (42) Yurke, B.; Turberfield, A. J.; Mills, A. P., Jr.; Simmel, F. C.; Neumann, J. L. *Nature* **2000**, *406*, 605–608.
- (43) Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. *Science (New York, N.Y.)* **2006**, *314*, 1585–8.
- (44) Choi, H. M. T.; Chang, J. Y.; Trinh, L. A.; Padilla, J. E.; Fraser, S. E.; Pierce, N. A. *Nat. Biotechnol.* **2010**, *28*, 1208–12.
- (45) Venkataraman, S.; Dirks, R. M.; Rothmund, P. W. K.; Winfree, E.; Pierce, N. A. *Nat. Nanotechnol.* **2007**, *2*, 490–4.
- (46) Zhang, D. Y.; Chen, S. X.; Yin, P. *Nature Chem.* **2012**, *4*, 208–14.
- (47) Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J.; Keasling, J. D. *Nat. Biotechnol.* **2009**, *27*, 753–9.
- (48) Whitaker, W. R.; Davis, S. A.; Arkin, A. P.; Dueber, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 18090–5.