Programmable synthesis of DNA nanostructures for spatial and temporal control

Structural and dynamic DNA nanotechnology works have demonstrated the potential of harnessing the predictable and programmable base pairing of DNA to create self-assembled 3D nanostructures, design molecular walkers, and perform logic computation. However, scaling these capabilities has continuously required the development of novel and generalizable molecular mechanisms. We developed the Primer Exchange Reaction (PER) method and related technologies for nanostructure design and assembly that together address these scalability challenges, as well as open up the potential application space for recording and controlling spatio-temporal interactions using DNA circuitry. We showed how PER cascades, which autonomously synthesize single-stranded DNA sequences, can be used to perform computation, record the order of molecular events, synthesize a DNAzyme only in the presence of an input RNA, be used to create strands for a DNA origami nanostructure, and utilized to amplify fluorescence signal for imaging fixed DNA, RNA, and protein targets in situ. These demonstrations show just a few of the many applications of the PER technology for enhancing our ability to manipulate and record biological substrates.
Final report for N00014-16-1-2410

Title: Programmable synthesis of DNA nanostructures for spatial and temporal control

 Principle Investigator: Peng Yin, Harvard University

Table of Contents

<table>
<thead>
<tr>
<th>HEADING</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPROACH</td>
<td>2</td>
</tr>
<tr>
<td>ACCOMPLISHMENTS</td>
<td>2</td>
</tr>
</tbody>
</table>

OVERVIEW | 2
SINGLE-STRANDED SYNTHESIS AND DYNAMIC CIRCUITRY WITH PRIMER EXCHANGE REACTION (PER) CASCADES | 2
RECORDING OF MOLECULAR PROXIMITY REACTIONS WITH AUTO-CYCLING PROXIMITY RECORDING (APR) | 4
DEVELOPMENT AND APPLICATION OF TOEHOLD SWITCHES FOR DETECTION OF RNA AND RIBOCOMPUTATION | 4
CONSTRUCTION OF DNA AND RNA ORIGAMI STRUCTURES FROM SINGLE-STRANDED SEQUENCES | 5
ASSEMBLY OF LARGE DNA BRICK STRUCTURE COMPRISING TENS OF THOUSANDS OF UNIQUE SEQUENCES | 6
CONDITIONALLY ACTIVATED GUIDE RNA (gRNA) CIRCUITS | 6
HIGHLY SPECIFIC AND MULTIPLEXED DETECTION OF miRNA SEQUENCES | 7
AUTOMATED DESIGN OF SINGLE-STRANDED PROBE LIBRARIES FOR FLUORESCENCE IN SITU HYBRIDIZATION (FISH) | 7
APPLICATION OF PER CONCATENERS FOR HIGHLY MULTIPLEXED AND AMPLIFIED FISH AND IF SIGNALS | 8
IN SITU MEASUREMENT OF NANOSCALE DISTANCES WITH PER-BASED MOLECULAR RULERS | 8
RECORDING OF NANOSCALE LANDSCAPES WITH PER-BASED MOLECULAR CRAWLERS | 10
PER-BASED TICKER TAPES FOR TEMPORAL RECORDING OF SIGNALS | 10
HIGHLY SENSITIVE DETECTION OF PROTEIN TARGETS USING PER-BASED SUPER ASSAY | 11
ADDITIONAL CONTRIBUTIONS | 12

MAJOR PROBLEMS / ISSUES | 12
TECHNOLOGY TRANSFER | 12
FOREIGN COLLABORATIONS AND SUPPORTED FOREIGN NATIONALS | 12

PRODUCTIVITY | 13
PUBLISHED ARTICLES | 13
WORKSHOPS AND CONFERENCES | 13
TRANSLATION: PATENTS | 16
AWARDS AND HONORS | 16

AWARDS PARTICIPANTS | 16
UPCOMING PUBLICATIONS | 17
Approach
We proposed to develop the Primer Exchange Reaction (PER) technology as a platform for engineering molecular behaviors including in situ growth of DNA nanostructures (Aim 1), molecular clocks and timers (Aim 2), and environmentally responsive nanomachines (Aim 3). Preliminary results had shown that PER could effectively synthesize arbitrary single-stranded DNA sequences autonomously, and we requested the resources to further develop this exciting technology for a variety of application spaces including biomolecular sensing and imaging.

Accomplishments
Overview
Through the course of the grant period, we were able to accomplish essential aspects and goals for all three aims and demonstrate several key applications of Primer Exchange Reactions (PERs), including: bio-sensing, logic circuits, and environmental recording [1]. We also published work showing a related mechanism, Auto-cycling Proximity Recording (APR) can be utilized to record molecular proximity into DNA records [2]. We further expanded our technological development in several areas related to the grant, including: single-stranded DNA synthesis [3], programmable nanostructure design and construction [4], [5], [6], nucleic acid detection [7], environmentally responsive nanodevices [8], [9], [10], [11], and in situ probe sequence design software [12] and detection [13]. Finally, we combined these innovations to demonstrate highly multiplexed imaging of amplified DNA FISH, RNA FISH, and immunofluorescence signals [14], [15]. In addition to these 15 published works, this grant has also supported a number of ongoing soon-to-be-published projects.

Below, we describe each of these key technical advancements in detail.

Single-stranded synthesis and dynamic circuitry with Primer Exchange Reaction (PER) cascades
The key technology we developed for this grant was the Primer Exchange Reaction (PER) method, which can be used for the autonomous synthesis of single-stranded sequences. The core technology was published in Nature Chemistry in 2017, and several follow-up and ongoing studies utilizing PER are described in detail in the next sections [1]. Below, the key accomplishments of PER in the initial publication are described in the context of our original proposed aims.

Aim 1: Triggered growth of complex nanostructures in situ
Software. We developed more advanced software to design our primer sequences for PER. It takes into account features such as single-strandedness of intermediate synthesis products, in addition to binding energies and reduced binding probabilities between primers and non-cognate hairpins.

**4-letter code synthesis.** We implemented our first system for synthesizing a 4-letter code sequence (A/T/G/C), the details of which can be seen in our published paper [1]. This represents a major advancement in PER cascades, which we had previously only implemented with sequences comprising 3-letters (typically A/T/C). The 3-letter code allows for the fourth letter to serve as the requisite “stop” sequence in the hairpin, if that letter (e.g. dGTP) is excluded from the reaction mixture. However, this strategy severely limited the set of sequences that could be synthesized and therefore the biological relevance of the technology, as interesting and functional DNA sequences typically have four letters. We determined experimentally that the synthetic iso-dG/iso-dC nucleotides, which pair with each other but not well with other bases, were more effective at stopping the polymerase over several other candidate sequences. We used this synthetic base pair as a stop sequence in hairpins used to construct an environmentally responsive nanodevice using a 4-letter code, described further under Aim 3 accomplishments below.

**Origami from 80 PER reactions in parallel.** As a first step toward scaling up PER reactions, as well as creating nanostructures from PER-synthesized components, we designed a system in which 40 staple strands for a DNA origami structure were synthesized each in two steps for a total of 80 PER synthesis reactions happening together in a single test tube. After synthesis, these staples were annealed with a long scaffold strand to form an origami nanostructure.

Further work on the assembly of PER-based structures in situ is described in the “Application of PER concatemers for highly multiplexed and amplified FISH and IF signals” section below.

**Aim 2: Implement molecular clocks and timers**

**RNA signal detection modules.** In our proposal (Figure 14, Aim 2.1), we proposed a molecular mechanism for detecting arbitrary RNA sequences and using this conditional detection to conditionally activate PER synthesis. We experimentally developed this detection, based on a mechanism known as toehold exchange, and several of the examples below use the modules to implement various molecular behaviors (see in situ signal amplifier, molecular logic circuits, nanodevice, and event recorder below). More detailed descriptions of the module are described in the section on the section for the in situ signal amplifier of the paper [1].

**In situ signal amplifier.** We implemented a single-hairpin system that synthesizes long strands of repeated sequence domains, which we call a synthetic “telomerase”, and then used this construct as a signal amplifier which grows fluorescent concatemers upon detection of a particular miRNA input. By designing the repeated sequence to bind a particular fluorescent dye, we were able to make these synthesized telomeres fluorescent, allowing for direct visualization of the result under a blue light.
Molecular logic circuits. We proposed to implement molecular logic circuits using signal-dependent primer exchange reactions (Aim 2.2), which we succeeded in doing. Several different circuits were implemented, including an AND gate, an OR gate, and a NOT gate that took biological RNA sequences as signal inputs. A more complex circuit, (miR-19a OR TWT) AND (NOT miR-21).

Aim 3: Construct environmentally responsive nanomachines

Nanodevice. We constructed a PER nanodevice for conditional RNA degradation. More specifically, the nanodevice detects the oncogenic miR-19a signal and subsequently synthesizes a functional DNAzyme that is programmed to cut an independent RNA substrate and has been shown to promote apoptosis in vivo. In this example, we used a PER cascade to synthesize a 4-letter biologically constrained sequence, which represents a large advancement in the types of sequences we can synthesize (previously just 3-letter de novo designed sequences). We also demonstrated how synthesized sequences can serve genetically functional purposes such as RNA cleavage, a capability with potentially very interesting future applications of the PER technology.

Molecular event recorder. We created a temporal molecular event recording system that encodes the sequential order in which two RNA targets are witnessed into dynamically synthesized transcripts. Depending on the order in which the two RNA signals are introduced, primers undergo one of two elongation pathways. This differentiation effectively records temporal information about the molecular environments of the primers and represents the first step toward implementing the ticker tapes proposed in Aim 3 of our proposal.

Several projects below (APR, molecular crawler, and molecular ruler) expand the technical capability of PER to record spatial information into sequences. And further work expanding the temporal recording of PERs is described in the “PER-based ticker tapes for temporal recording of signals” section below.

Recording of molecular proximity reactions with Auto-cycling Proximity Recording (APR)
We implemented a microscope-free “biochemical nanoscopy” method that records nanostructure features in situ and in detail 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. The mechanism was demonstrated on biotin-streptavidin complexes, as well as DNA origami structures of various designs. This work was published in Nature Communications [2].

Development and application of toehold switches for detection of RNA and ribocomputation
In addition to logic circuits with PER cascades, we also implemented logic behavior using the toehold switch technology introduced in Aim 2 of our proposal. We developed a strategy for
constructing RNA-only nanodevices to evaluate complex logic in living cells. Such
'ribocomputing' systems are composed of de-novo-designed parts and operate via predictable
and designable base-pairing rules, allowing for effective in silico design of computing devices
with prescribed configurations and functions in complex cellular environments. These devices
operate at the post-transcriptional level and use an extended RNA transcript to co-localize all
circuit sensing, computation, signal transduction, and output elements in the same self-
assembled molecular complex, which reduces diffusion-mediated signal losses, lowers
metabolic cost, and improves circuit reliability. We demonstrated that ribocomputing devices in
E. coli can evaluate two-input logic with dynamic range up to 900-fold and scale them to four-
input AND, six-input OR, and a complex 12-input expression (A1 AND A2 AND NOT A1*) OR (B1
AND B2 AND NOT B2*) OR (C1 AND C2) OR (D1 AND D2) OR (E1 AND E2). This work was
published in Nature [8], and a subsequent viewpoint article was published in Biochemistry [9].

While our original toehold switch designs conditionally activated translation of specific genes,
we later demonstrated that modified designs can cause constitutively on genes to be turned off
(repressed). Efforts to construct synthetic biological circuits with more complex functions have
often been hindered by the idiosyncratic behavior, limited dynamic range and crosstalk of
commonly utilized parts. Here, we employ de novo RNA design to develop two high-
performance translational repressors with sensing and logic capabilities. These synthetic
riboregulators, termed toehold repressors and three-way junction (3WJ) repressors, detect
transcripts with nearly arbitrary sequences, repress gene expression by up to 300-fold and yield
orthogonal sets of up to 15 devices. Automated forward engineering is used to improve toehold
repressor dynamic range and SHAPE-Seq is applied to confirm the designed switching
mechanism of 3WJ repressors in living cells. We integrate the modular repressors into biological
circuits that execute universal NAND and NOR logic and evaluate the four-input expression NOT
((A1 AND A2) OR (B1 AND B2)) in Escherichia coli. These capabilities make toehold and 3WJ
repressors valuable new tools for biotechnological applications. This work has been published in Nature Chemical Biology [10].

Construction of DNA and RNA origami structures from single-stranded sequences
Self-folding of an information-carrying polymer into a defined structure is foundational to
biology and offers attractive potential as a synthetic strategy. Although multicomponent self-
assembly has produced complex synthetic nanostructures, unimolecular folding has seen
limited progress. We describe a framework to design and synthesize a single DNA or RNA strand
to self-fold into a complex yet unknotted structure that approximates an arbitrary user-
prescribed shape. We experimentally construct diverse multikilobase single-stranded
structures, including a ~10,000-nucleotide (nt) DNA structure and a ~6000-nt RNA structure.
We demonstrate facile replication of the strand in vitro and in living cells. The work here thus
establishes unimolecular folding as a general strategy for constructing complex and replicable
nucleic acid nanostructures, and expands the design space and material scalability for bottom-
up nanotechnology. This work has been published in Science [5].

In a follow-up work, we further explore the potential of single-stranded structures. Molecular
knots represent one of the most extraordinary topological structures in biological polymers.
Creating highly knotted nanostructures with well-defined and sophisticated geometries and topologies remains challenging. Here, we demonstrate a general strategy to design and construct highly knotted nucleic acid nanostructures, each weaved from a single-stranded DNA or RNA chain by hierarchical folding in a prescribed order. Sets of DNA and RNA knots of two- or three-dimensional shapes have been designed and constructed (ranging from 1700 to 7500 nucleotides), and they exhibit complex topological features, with high crossing numbers (from 9 up to 57). These single-stranded DNA/RNA knots can be replicated and amplified enzymatically in vitro and in vivo. This work establishes a general platform for constructing nucleic acid nanostructures with complex molecular topologies. This work has been published in Nature Communications.

Assembly of large DNA brick structure comprising tens of thousands of unique sequences Nucleic acids (DNA and RNA) are widely used to construct nanometre-scale structures with ever increasing complexity, with possible application in fields such as structural biology, biophysics, synthetic biology and photonics. The nanostructures are formed through one-pot self-assembly, with early kilodaton-scale examples containing typically tens of unique DNA strands. The introduction of DNA origami, which uses many staple strands to fold one long scaffold strand into a desired structure, has provided access to megadalton-scale nanostructures that contain hundreds of unique DNA strands. Even larger DNA origami structures are possible, but manufacturing and manipulating an increasingly long scaffold strand remains a challenge. An alternative and more readily scalable approach involves the assembly of DNA bricks, which each consist of four short binding domains arranged so that the bricks can interlock. This approach does not require a scaffold; instead, the short DNA brick strands self-assemble according to specific inter-brick interactions. First-generation bricks used to create three-dimensional structures are 32 nucleotides long, consisting of four eight-nucleotide binding domains. Protocols have been designed to direct the assembly of hundreds of distinct bricks into well formed structures, but attempts to create larger structures have encountered practical challenges and had limited success. Here we show that DNA bricks with longer, 13-nucleotide binding domains make it possible to self-assemble 0.1–1-gigadalton, three-dimensional nanostructures from tens of thousands of unique components, including a 0.5-gigadalton cuboid containing about 30,000 unique bricks and a 1-gigadalton rotationally symmetric tetramer. We also assembled a cuboid that contains around 10,000 bricks and about 20,000 uniquely addressable, 13-base-pair ‘voxels’ that serves as a molecular canvas for three-dimensional sculpting. Complex, user-prescribed, three-dimensional cavities can be produced within this molecular canvas, enabling the creation of shapes such as letters, a helicoid and a teddy bear. We anticipate that with further optimization of structure design, strand synthesis and assembly procedure even larger structures could be accessible, which could be useful for applications such as positioning functional components. This work has been published in Nature [4].

Conditionally activated guide RNA (gRNA) circuits We have developed nucleic acid-sensing CRISPR-Cas9 guide RNAs (gRNA): switchable gRNAs which switch on or off to either enable or disable Cas9 activity depending on the presence of a
specific nucleic acid sequence profile in its environment. We have demonstrated designs that can be conditionally activated and conditionally deactivated in vitro, varying Cas9 activity by approximately an order of magnitude depending on the sequence profile present, as well as perform Boolean logical computations on the sequence profile present, including NAND, NOR, AND, and OR. These designs have no sequence dependencies between the sequences to be sensed and the sequences targeted by Cas9, and can thus in theory be programmed to sense any arbitrary sequence profile. These results were published in ACS Synthetic Biology in July 2019 [11].

Highly specific and multiplexed detection of miRNA sequences
MicroRNA (miRNA) expression profiles hold promise as biomarkers for diagnostics and prognosis of complex diseases. Here we present a super-resolution fluorescence imaging-based digital profiling method for specific, sensitive, and multiplexed detection of miRNAs. In particular, we applied DNA-PAINT (Point Accumulation for Imaging in Nanoscale Topography) method to implement a super-resolution geometric barcoding scheme for multiplexed single-molecule miRNA capture and digital counting. Using synthetic DNA nanostructures as programmable miRNA capture “nano-array”, we demonstrated high-specificity (single nucleotide mismatch discrimination), multiplexed (8-plex, 2 panels) and sensitive measurements on synthetic miRNA samples, as well as applied one 8-plex panel to measure endogenous miRNAs levels in total RNA extract from HeLa cells. This work has been published in Angewandte Chemie [7].

Automated design of single-stranded probe libraries for Fluorescence in situ hybridization (FISH)
Oligonucleotide (oligo)-based FISH has emerged as an important tool for the study of chromosome organization and gene expression and has been empowered by the commercial availability of highly complex pools of oligos. However, a dedicated bioinformatic design utility has yet to be created specifically for the purpose of identifying optimal oligo FISH probe sequences on the genome-wide scale. Here, we introduce OligoMiner, a rapid and robust computational pipeline for the genome-scale design of oligo FISH probes that affords the scientist exact control over the parameters of each probe. Our streamlined method uses standard bioinformatic file formats, allowing users to seamlessly integrate new and existing utilities into the pipeline as desired, and introduces a method for evaluating the specificity of each probe molecule that connects simulated hybridization energetics to rapidly generated sequence alignments using supervised machine learning. We demonstrate the scalability of our approach by performing genome-scale probe discovery in numerous model organism genomes and showcase the performance of the resulting probes with diffraction-limited and single-molecule super-resolution imaging of chromosomal and RNA targets. We anticipate that this pipeline will make the FISH probe design process much more accessible and will more broadly facilitate the design of pools of hybridization probes for a variety of applications. This work has been published in the Proceedings of the National Academy of Sciences [12].
Application of PER concatemers for highly multiplexed and amplified FISH and IF signals

Two works published in quick succession introduced the SABER signal amplification for use in imaging DNA or RNA (SABER-FISH) or proteins (Immuno-SABER) in fixed cell and tissue samples. These are described further below.

Fluorescence in situ hybridization (FISH) reveals the abundance and positioning of nucleic acid sequences in fixed samples. Despite recent advances in multiplexed amplification of FISH signals, it remains challenging to achieve high levels of simultaneous amplification and sequential detection with high sampling efficiency and simple workflows. Here we introduce signal amplification by exchange reaction (SABER), which endows oligonucleotide-based FISH probes with long, single-stranded DNA concatemers that aggregate a multitude of short complementary fluorescent imager strands. We show that SABER amplified RNA and DNA FISH signals (5- to 450-fold) in fixed cells and tissues. We also applied 17 orthogonal amplifiers against chromosomal targets simultaneously and detected mRNAs with high efficiency. We then used 10-plex SABER-FISH to identify in vivo introduced enhancers with cell-type-specific activity in the mouse retina. SABER represents a simple and versatile molecular toolkit for rapid and cost-effective multiplexed imaging of nucleic acid targets. This work was posted on bioRxiv and later published in Nature Methods [14].

Spatial mapping of proteins in tissues is hindered by limitations in multiplexing, sensitivity, and throughput. Here we report immunostaining with signal amplification by exchange reaction (Immuno-SABER), which achieves highly multiplexed signal amplification via DNA-barcoded antibodies and orthogonal DNA concatemers generated by primer exchange reactions (PER). SABER offers independently programmable signal amplification without in situ enzymatic reactions, and intrinsic scalability to rapidly amplify and visualize a large number of targets when combined with fast exchange cycles of fluorescent imager strands. We demonstrated 5–180-fold signal amplification in diverse samples (cultured cells, and FFPE, cryosectioned or whole mount tissues), and simultaneous signal amplification for 10 different proteins using standard equipment and workflows. We also combined SABER with expansion microscopy to enable rapid, multiplexed super-resolution tissue imaging. Immuno-SABER presents an effective and accessible platform for multiplexed and amplified imaging of proteins with high sensitivity and throughput. This work was posted on bioRxiv and later published in Nature Biotechnology [15].

Ongoing work to combine SABER-FISH with our DNA-based super-resolution imaging technique (DNA-PAINT) enables high throughput super-resolution imaging of chromosome targets.

In situ measurement of nanoscale distances with PER-based molecular rulers

(An initial draft of the manuscript for this work can be found in the Appendix.)

The era of nanoscale discovery and engineering has arrived, with examples in nanotechnology, electronics, and biology. The study of any such system benefits from forming an image of it, if possible, with increasingly sophisticated instruments that provide molecular-level detail for
further understanding or validation. Comprehensive visualization can be challenging for two reasons – size and molecular diversity. Current microscopes do not achieve both nanoscale resolution and high discrimination. We propose an alternate technique that captures both the great diversity of targets and provides a nanoscale resolution map of their spatial location. The technique, which we term a DNA nanoscope (Fig. 1C), tags targets with synthetic DNA barcodes, measures distances between many target pairs using biochemical DNA reactions, and then reconstructs a detailed map of the underlying geometry that uniquely identifies every target. The bottom-up, sequencing enabled mechanism of the DNA nanoscope is fundamentally different from top-down microscope-based imaging, and hence offers unique advantages in precision, throughput and accessibility.

We successfully applied the DNA nanoscope technique to nine different patterns and obtained molecular resolution reconstructions. The root-mean-square deviation (RMSD) was used to quantify the average error between the designed and the reconstructed pattern. The RMSDs for the various patterns range from 1.4 nm to 2.6 nm. The points in the most densely packed patterns are merely 6 nm apart and were clearly spatially resolved in the respective reconstructions. We successfully resolved negative space, clusters of segregated points and sparse patterns. The variety of patterns reconstructed demonstrates the robustness of our approach. The biggest patterns are approximately 100 nm wide and 50 nm tall. The highest number of points localized was 135. Apart from spatially localizing the various points of a dense nanoscale pattern, the DNA nanoscope also uniquely distinguishes them by means of their barcode sequence, something that has proven unfeasible for microscopy techniques, which suffer from low-multiplexing capabilities.

The bottom-up DNA nanoscope approach stands in contrast to top-down microscopy methods, and confers unique advantages. First, there is no requirement that the sample be accessible to electromagnetic radiation, only to tiny diffusing DNA molecules that are not restricted to moving in straight lines. Second, the recording interactions of the DNA nanoscope with the sample are via gentle biochemical reactions (DNA hybridization and polymerization) in contrast to high-energy lasers, electron beams or physical probes used in super-resolution microscopy, electron microscopy and scanning probe microscopy respectively. Third, the sample does not need any special preparation, like adhering it to a surface, or freezing it in vitreous ice, that hold it immobile with respect to macro-scale recording instruments. The recording instruments of the DNA nanoscope are molecules themselves, diffusing throughout. This eliminates the need to correct any sample drift with respect to the instrument, which imposes practical and fundamental limits on the resolution of microscopy techniques. Fourth, no capital-intensive purchases are necessary to start using the DNA nanoscope. The recording process costs only as much as a PCR reaction, and is also setup in a similar manner. The read-out is via next-generation sequencing, which is accessible for a $1000 start-up kit available from a commercial source. The per assay cost is currently high, costing about $500 per structure mapped in this work, but is seeing rapid drops in price and increases in throughput as the technology continues to mature. We predict that the DNA nanoscope and related ‘imaging by sequencing’ techniques will gain widespread adoption over the next few years and drive fundamental nanoscale discoveries.
Recording of nanoscale landscapes with PER-based molecular crawlers

PER allows the programmed growth of DNA strands based on in situ synthesis of components in a programmable and catalytic fashion. Based on the mechanism, we have been developing a molecular motor system that has the ability to inspect and “record” the molecular landscapes while moving along a target. The molecular motors, which we call molecular crawlers, operate by localized cascades of PER reactions along multiple PER probes anchored on a target substrate. In the presence of a trigger (primer), the reaction can initiate, and the localized reaction propagates only in the presence of neighboring probes, generating a concatenated record that copies the information from the probes.

The ability to copy information from molecular targets opens up the possibility of surveying quantitative and spatio-temporal information from molecular targets of interest. The basic mechanisms were demonstrated and optimized on structurally well-defined DNA nanostructures. We applied the crawler system to counting the number of subunits in model molecular complexes such as streptavidin and DNA origami. We also demonstrated that the crawlers can detect multi-valent protein interactions using microtubule-associated proteins as a model system.

To further extend the capability of the crawler system, we are developing a mechanism that measures temporal changes of molecular environment. Furthermore, to realize highly multiplexed single-molecule-level analysis, we have been developing a scheme of unique labeling combined with readout by next-generation sequencing (NGS) methods. This allows the whole sample to be analyzed at once, with a resolution of individual complexes, in a massively parallel fashion. We plan to apply this technology to studies of spatio-temporal changes of protein clusters at the single-cluster level and of single-cell chromosome organizations. The crawler system promises to broadly transform fundamental biological research, drug discovery, and diagnostics, by providing an unprecedented tool for parallel and multiplexed examination of molecular landscapes.

PER-based ticker tapes for temporal recording of signals

We are developing a DNA-based molecular recorder which allows for real-time multiplexed recording of molecular signals into DNA strands. Single-stranded DNA records are generated using Primer Exchange Reaction (PER) technology: multiple PER hairpin species each containing a unique species barcode repetitively extend single-stranded DNA primers to generate the records, which give a sequential readout of barcodes corresponding to molecular events. These records collectively encode real-time information about species concentration changes, which can be retrieved and reconstructed via statistical methods by sequencing the DNA records. One PER hairpin species (clock) present at a constant concentration acts as a molecular clock, while one or more other PER hairpin species (signals) present at variable concentrations tied to the state of a desired molecular recording target provide the molecular events to be recorded. We have successfully recorded and reconstructed several concentration time courses for signal hairpins, and ongoing work seeks to improve the performance of the inference algorithm for
lower numbers of DNA records, and for more complex signals. Recording molecular events nondestructively and noninvasively allows for systems to be interrogated at a scale proportional to volume rather than with surface area and in a massively parallel fashion, expanding prospects for microscope-free data acquisition in biological systems. The fundamental principles and mechanisms underlying the success of these system can be applicable to similar systems involving other materials.

Highly sensitive detection of protein targets using PER-based SUPER assay

Highly sensitive protein detection techniques with minimal sample consumption are of great value in biomedical research and disease diagnostics. Most of current immunoassays are in heterogeneous format which require a wash step to remove excess reporter antibodies after solid-phase capture of targets. Their sensitivity is intrinsically limited by the nonspecific binding of reporter antibodies to the surface of the solid support and a sample volume of 30-100 μl is typically required for a single assay.

The nucleic acid-coupled Proximity Ligation Assay (PLA) and the more recently developed Proximity Extension Assay (PEA) are two promising washing-free methods have demonstrated the potential of sensitive protein detection in very small sample volumes (1 μl). PLA and PEA are developed based on the proximity effect: two affinity agents (e.g. antibodies), against different epitopes of the same target protein, are employed in a reaction that proceeds much faster when the agents are colocalized by a target than when the affinity agents are free-floating in solution. The achievable analytical sensitivity of PLA and PEA are intrinsically limited by the existence of background signal, as reporter sequences are slowly generated in the absence of target in bulk solution. Low probe concentrations result in a lower background signal generation rate, but in turn inevitably lead to diminished target signal due to a low antibody binding yield. Optimized, the achievable Limit of Detection (LoD) is typically above pg/ml.

To overcome this sensitivity bottleneck of current proximity-based methods, we have developed a new approach named SUPER (Successive Unidirectional Proximity Extension Reaction), which is developed based on our programmable DNA synthesis technology, PER. SUPER enforces a requirement of persistent proximity, such that two affinity probes must be co-localized by a protein antigen for minutes to generate a complete DNA sequence that can be amplified and detected. This greatly reduces background signal from transient interaction of free-floating probes. SUPER offers an fg/ml, digital-ELISA-like LoD, demonstrated with proteins streptavidin and human cytokine IL-4 and a protocol requiring only two mixing steps and droplet-sized sample volume (1μl). we have further demonstrated that the assay is precise (5% intra assay CV and 11% inter assay CV) and can be performed directly in undiluted human serum samples which establish a solid foundation for its clinical utility. We further validate SUPER with biomarkers of low abundance (fg/ml) to better demonstrate the sensitivity, includes IL-2, IL-10, IFNy, IL-8 and IL-5. All the assays have been successfully developed with low-fg/ml LoD and 1ul sample consumption.
Additional contributions

Our grant participants also provided secondary contributions to works from other groups, which resulted in two publications, one related to the production of large amounts of single-stranded DNA and another for super-resolution study of chromosome structure.

There is increasing demand for single-stranded DNA (ssDNA) of lengths >200 nucleotides (nt) in synthetic biology, biological imaging and bionanotechnology. Existing methods to produce high-purity long ssDNA face limitations in scalability, complexity of protocol steps and/or yield. We present a rapid, high-yielding and user-friendly method for in vitro production of high-purity ssDNA with lengths up to at least seven kilobases. Polymerase chain reaction (PCR) with a forward primer bearing a methanol-responsive polymer generates a tagged amplicon that enables selective precipitation of the modified strand under denaturing conditions. We demonstrate that ssDNA is recoverable in ~40–50 min (time after PCR) with >70% yield with respect to the input PCR amplicon, or up to 70 pmol per 100 μl PCR reaction. We demonstrate that the recovered ssDNA can be used for CRISPR/Cas9 homology directed repair in human cells, DNA-origami folding and fluorescent in-situ hybridization. This work was published in Nucleic Acids Research [3].

Chromosome structure is thought to be crucial for proper functioning of the nucleus. Here, we present a method for visualizing chromosomal DNA at super-resolution and then integrating Hi-C data to produce three-dimensional models of chromosome organization. We begin by applying Oligopaint probes and the single-molecule localization microscopy methods of OligoSTORM and OligoDNA-PAINT to image 8 megabases of human chromosome 19, discovering that chromosomal regions contributing to compartments can form distinct structures. Intriguingly, our data also suggest that homologous maternal and paternal regions may be differentially organized. Finally, we integrate imaging data with Hi-C and restraint-based modeling using a method called integrative modeling of genomic regions (IMGR) to increase the genomic resolution of our traces to 10 kb. This work has been published in PloS Genetics [13].

Major problems / Issues
Nothing to report.

Technology Transfer
See section on Patents below.

Foreign Collaborations and Supported Foreign Nationals

Foreign Nationals: Nikhil Gopalkrishnan (postdoc), Youngeun Kim (postdoc), Hiroshi Sasaki (postdoc), Sungwook Woo (postdoc), Piotr Nowak (postdoc), Juanita Lara (technician), Mingjie Dai (postdoc), Weidong Xu (grad student), Feng Xuan (postdoc)

Foreign Collaborations: N/A
Productivity

Published Articles


Workshops and Conferences

Talks:
- Hiroshi Sasaki. 26th Tokyo RNA Club meeting, Tokyo, Japan, December 2019.
- Jocelyn Kishi. 25th International Conference on DNA Computing and Molecular Programming (DNA25), University of Washington, Seattle, WA, August 5th, 2019.
- Peng Yin. Clinical Pathology Conference, Brigham and Women’s Hospital, Boston, MA, May 28th, 2019.
• Peng Yin. Department of Pharmacology and Chemical Biology, Baylor College of Medicine, Houston, TX, April 23rd, 2019.
• Mike Jin. Wyss Institute Molecular Robotics Initiative, Boston, MA, Mar. 14, 2019.
• Peng Yin. Nanoscale Subgroup meeting, Biophysical Society Annual Meeting, Baltimore, MD, March 2nd, 2019
• Peng Yin. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, Feb. 26th, 2019.
• Peng Yin. Department of Biomedical Engineering, Northwestern University, Evanston, IL, Jan. 24th, 2019.
• Peng Yin. Department of Chemistry, Boston College, Chestnut Hill, MA, Nov. 28th, 2018
• Feng Xuan. Wyss Institute Annual Retreat, Boston, MA, Nov. 15, 2019.
• Hiroshi Sasaki. 11th 3R+3C Symposium, Kanazawa, Japan, November 2018.
• Hiroshi Sasaki. 7th Annual Northeast Regional Chromosome Pairing Conference, Sudbury, ON, October 2018.
• Peng Yin. NIH High Risk High Reward Research Symposium, Bethesda, MD, June 7th, 2018.
• Peng Yin. Synthetic Biology USA Congress, Boston, MA, May 10th, 2018.
• Peng Yin. School of Life Sciences, Peking University, China, May 3rd, 2018.
• Peng Yin. Department of Biosystems Science and Engineering at ETH Zurich, Basel, Switzerland, Apr. 10th, 2018.
• Peng Yin. Spring meeting of the Swiss Chemical Society, Neuchatel, Switzerland, Apr. 6th, 2018.
• Peng Yin. Department of Biomedical Engineering, Purdue University, Feb. 27th, 2018.
• Peng Yin. Molecular and Cellular Biology Graduate Program, University of Massachusetts Amherst, Amherst, MA, Jan. 23rd, 2018.
• Peng Yin. Distinguished Seminar Series at Department of Biomedical Engineering, Duke University, Nov. 1st, 2017.
• Peng Yin. Department of Biomedical Engineering, Boston University, Oct. 27th, 2017.
• Sungwook Woo. 15th Annual Conference on Foundations of Nanoscience: Self-Assembled Architectures and Devices (FNANO18), Snowbird, Utah, April 2018.
• Hiroshi Sasaki. 6th Annual Northeast Regional Chromosome Pairing Conference, Bowdoin College, Brunswick, ME, October 2017.
• Jocelyn Kishi. 23rd International Conference on DNA Computing and Molecular Programming (DNA23), University of Texas, Austin, TX, September 2017.
• International Workshop on DNA-based Nanotechnology (dnatecl7), May 29, 2017, Dresden, Germany.
• The second ISH symposium - Applications of In Situ Hybridisation in Research and Disease, Cambridge, UK, May 17th, 2017.
• Department of Biomedical Engineering, Tufts University, Feb. 6th, 2017.
• Elkin Lecture, Emory University, Dec. 2nd, 2016.
• Center for Molecular Bioengineering, TU Dresden, Nov. 21st, 2016.
• RNA Nanotechnology Conference, Berkshire, United Kingdom, Aug. 2nd, 2016.
• The Fourth Annual Workshop on Micro- and Nanotechnologies in Medicine, Cambridge, MA, July 26th, 2016.
• Sino-US Synthetic Biology Workshop, Guangzhou, China, July 1st, 2016.

Posters:
• Mike Jin. 25th International Conference on DNA Computing and Molecular Programming (DNA25), University of Washington, Seattle, WA, August 5th, 2019.
• Mike Jin. Molecular Programming Project Symposium, Pasadena, CA, June 27, 2019.
Translation: Patents
Work on this grant has led to the filing of 15 separate patent applications (see separate Patent Report).

Awards and Honors
Peng Yin was awarded the Rozenberg Tulip Award at the 23rd International Conference on DNA Computing and Molecular Programming (DNA23) for his work in this field.

Jocelyn Kishi won Best Student Speaker award at 23rd International Conference on DNA Computing and Molecular Programming (DNA23) held in Austin, Texas in September 2017, where she presented her work on developing the PER technology. Jocelyn Kishi was awarded a Graduate School of Arts and Sciences Merit Fellowship by Harvard in the Fall of 2018, which she declined due to early graduation. In June 2019, she was awarded a Wyss Technology Development Fellowship from the Wyss Institute to continue her work on this project as a postdoctoral fellow.

Hiroshi Sasaki was a recipient of the Stellar Abstract Awards for the 11th Annual PQG Conference in November 2017.

Awards Participants
Military Personnel: N/A

Salary Support (Foreign Nationals indicated by italics):
- Peng Yin (PI)
- Nikhil Gopalkrishnan (postdoc)
- Youngeun Kim (postdoc)
- Hiroshi Sasaki (postdoc)
- Allen Zhu (technician)
- Sungwook Woo (postdoc)
Upcoming Publications

Below is the abstract for the PER-based molecular ruler project manuscript entitled “A DNA nanoscope that identifies and precisely localizes over a hundred unique molecular features with nanometer accuracy.”

Techniques that can both spatially map out molecular features and discriminate many targets would be highly valued for their utility in studying fundamental nanoscale processes. In spite of decades of development, no current technique can achieve both nanoscale resolution and distinguish hundreds of targets. Here, we developed a DNA nanoscope that both spatially localized and uniquely identified over a hundred densely packed elements on a DNA origami testbed. We successfully resolved elements spaced just 6 nm apart with an average spatial localization accuracy (RMS deviation) of ~ 2 nm. The bottom-up, sequencing enabled mechanism of the DNA nanoscope is fundamentally different from top-down microscope based imaging, and hence offers unique advantages in precision, throughput and accessibility.

All published works are available as pdfs on the Yin lab website:

https://vin.hms.harvard.edu/publications.html