UCLA
UCLA Electronic Theses and Dissertations

Title
High Throughput Single Molecule Investigation of Transcription

Permalink
https://escholarship.org/uc/item/3n63v2mr

Author
Kim, Soohong

Publication Date
2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

High Throughput Single Molecule Investigation of Transcription

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Chemistry

By

Soohong Kim

2012

© Copyright by

Soohong Kim

2012

ABSTRACT OF THE DISSERTATION

High Throughput Single Molecule Investigation of Transcription

by

Soohong Kim

Doctor of Philosophy in Chemistry

University of California, Los Angeles, 2012

Professor Shimon Weiss, Chair

RNA Polymerase (RNAP) is a crucial protein molecule that decodes the information stored in DNA as base pair sequences, through a process called transcription. The study of the mechanics and composition of RNAP in a cell is crucial to understanding the complex and arduous process of life. Most biological systems respond to a change by adding heterogeneity to its population states. One of the most effective ways of studying a biological system is to compare the population states of proteins as a function of time or chemical environment. Unfortunately, some of the subpopulations induced by external stimulation are too small and are hidden under ensemble averaging when using traditional methods like gel-electrophoresis. I have developed single-molecule methods capable of resolving subpopulations and monitoring the transition of individual molecules to its various function states. Using these methods we can (1) rapidly screen RNAP activity in a large reagent space and (2) map the transcription factor binding sites on DNA with ultrahigh accuracy (σ < 300base pairs) in the genomic scale with the ability to differentiate single cell variations.

The dissertation of Soohong Kim is approved.

William M. Gelbart

Margot Elizabeth Quinlan

Fuyuhiko Tamanoi

Shimon Weiss, Committee Chair

University of California, Los Angeles

2012

To my lord and savior Jesus Christ who has been my guide and strength throughout graduate school. His presence gave me joy and hope, especially when those words seemed distant. I am so unworthy of your grace, ever so grateful for your blessings.

"Trust in the LORD with all your heart, and lean not on your own understanding; in all your ways acknowledge Him, and He shall direct your paths." (Proverbs 3:5-7)

Table of Contents

List of Tables and Figures

Acknowledgements

Chapter two has been published in 2011. **Kim, S.**, Streets, A. M., Lin, R. R., Quake, S. R., Weiss, S., and Majumdar, D. S. (2011). High-throughput single-molecule optofluidic analysis. Nature Methods 8, 242–245.

Chapter three has been published in 2012. **Kim, S.**, Gottfried, A., Lin, R. R., Dertinger, T., Kim, A. S., Chung, S., Colyer, R. A., Weinhold, E., Weiss, S., and Ebenstein, Y. (2012). Enzymatically Incorporated Genomic Tags for Optical Mapping of DNA-Binding Proteins. Angewandte Chemie International Edition, n/a-n/a.

I would like to thank my advisor Shimon Weiss for all the guidance and support. His mentorship has pushed me and shaped me into the scientist I am today. I enjoyed very much our small discussions that taught me creative and critical thinking.

I thank Dr. William Gelbart for his guidance, moral support, and advices. Not only is he my role model in science, but he also reminds me of the good person I should become.

I appreciate my committee members, Dr. Fuyu Tamanoi and Dr. Margot Quinlan for being on my committee even though my request was at the very last moment. Their helpful questions and constructive criticism during my qualification exam helped me design my experiments more carefully.

Special thanks to Dr. Stephen Quake at Stanford University for his generosity in provision of resources. I really appreciate such hospitality during all my visits to the Quake lab.

Great thanks to Dr. Dylan Taatjes for his kindness and generosity during my visit to the Taatjes lab. I have learned so much from our discussions and brainstorms.

I thank my mentors Devdoot Majumdar and Yuval Ebenstein for their leadership and influence. I have learned so much by them and have always admired their passion in science.

My mentor Ryan Colyer is a God sent. He has always given me moral support during difficult times. Most of my programming skills and hardware knowledge comes from him. I would not be in this position if it wasn't for Ryan.

All the members of my lab have always been an inspiration. I truly appreciate everyone for their patients during conflicting times and support in times of need. I especially thank my brother like figure Ron Lin and Sangyoon Chung for making the lab experience pleasing.

I am very blessed to have met my friends Argyris, Aaron, Terry, Kenny, Jenny, Odo, Mauricio, Zheng, Lufeng, Xiaoming, Ellen, Laura, Peter, and David. I was kept sane by our laughs and comradely through the graduate school moments.

Lastly, I appreciate my family for their faith in me to become better every day.

I offer my sincerest appreciation to everyone for making the graduate school experience memorable.

VITA

Education

2007 — **B.A Chemistry**
Binghamton University, State University of New York
Vestal, NY

Research

2007-2012 — **Shimon Weiss Group, UCLA.** Development of high-throughput single molecule transcription assays

2006-2007 — **Chuan-Jian Zhong Group, SUNY Binghamton University.** Synthesis and anti-body conjugation of gold covered iron oxide nano particles for SERS detection

Teaching Experience

2007-2009 — **Teaching Assistant/Associate**
Department of Chemistry
University of California, Los Angeles
Los Angeles, CA

Awards

2011 — **George Gregory Award for research in Physical Chemistry, UCLA**
Jim & Barbara Tsay Fellowship, UCLA
Audree Fowler Fellowship in Protein Science, UCLA
Best Poster Award, MBI Retreat, UCLA

2006, 2007 — **Undergraduate Research Award, SUNY Binghamton University**
Dean's List, SUNY Binghamton University

Publications

1. **Kim, S.**, Gottfried, A., Lin, R. R., Dertinger, T., Kim, A. S., Chung, S., Colyer, R. A., Weinhold, E., Weiss, S., and Ebenstein, Y. (2012). Enzymatically Incorporated Genomic Tags for Optical Mapping of DNA-Binding Proteins. Angewandte Chemie International Edition, n/a-n/a.

Back Cover: http://onlinelibrary.wiley.com/doi/10.1002/anie.201200628/abstract

2. **Kim, S.**, Streets, A. M., Lin, R. R., Quake, S. R., Weiss, S., and Majumdar, D. S. (2011). High-throughput single-molecule optofluidic analysis. Nature Methods 8, 242–245.
Featured:
Nature Methods: doi:10.1038/nmeth0311-213
Physics Today: http://dx.doi.org/10.1063/1.3580484

3. Ebenstein, Y., Gassman, N., **Kim, S.**, Antelman, J., Kim, Y., Ho, S., Samuel, R., Michalet, X., and Weiss, S. (2009a). Lighting Up Individual DNA Binding Proteins with Quantum Dots. Nano Letters 9, 1598–1603.

4. Ebenstein, Y., Gassman, N., **Kim, S.**, and Weiss, S. (2009b). Combining atomic force and fluorescence microscopy for analysis of quantum-dot labeled protein–DNA complexes. Journal of Molecular Recognition 22, 397–402.

5. Park, H.-Y., Schadt, M. J., Wang, Lim, I.-I. S., Njoki, P. N., **Kim, S. H.**, Jang, M.-Y., Luo, J., and Zhong, C.-J. (2007). Fabrication of Magnetic Core@Shell Fe Oxide@Au Nanoparticles for Interfacial Bioactivity and Bio-separation. Langmuir 23, 9050–9056.

Chapter 1:

Introduction

RNA Polymerase (RNAP) is a crucial protein molecule that decodes the information stored in DNA as base pair sequences, through a process called transcription. Comprehending the mechanics of this enzyme is absolutely essential for understanding the cell cycle. Researchers have studied RNAP with different ensemble methods like gel electrophoresis[1], Southern and Western blots, and Chromatin Immuno-precipitation (ChIP) to answer questions about RNAP's binding affinities and activity. Nonetheless, ensemble methods are unable to capture the fast mechanics of individual molecules, nor can they distinguish variations within a population. Advances in single molecule techniques revealed details about RNAP that were obscured by ensemble averaging. Much is yet to be understood regarding this complex multi-component system. *In vitro* single molecule methods require mimicking the physiological state of RNAP in cellular environments. Screening the components needed to achieve this is an arduous and complicated process, due to the low throughput[2] nature of single molecule techniques**.** Single molecule methods capable of resolving subpopulations while varying the molecular environment with relative ease allow a researcher to monitor transitions of individual molecules into a variety of functional states. Building a tool that aids in understanding these kinds of responses to changes in the presence of certain reagent or proteins would significantly contribute to understanding of the transcriptional process.

Our goal was to develop two techniques: a) Full automation for a single-molecule transcription assay that can screen each component of the transcription system in a high-throughput manner b) localizing RNAP-bound quantum dots (QD) on combed viral genomic DNA to visualize transcription factor binding sites.

High Throughput Single Molecule Transcription Assay

A method capable of automatically varying the environment in which RNAP transcribes coupled with a readout system to compare the efficiency of the enzyme would be a powerful tool with virtually unlimited applications. Having a programmed routine that injects reagents, mixes, incubates, and acquires data with precision superior to human control allows direct comparisons of operations done on any state of the molecule. To achieve this, we collaborated with Stephen Quake's group at Stanford University and merged two techniques: microfluidics and single-molecule fluorescence resonance energy transfer (smFRET). Laminar fluids are manipulated in solution circuits engraved on a two-layered poly-dimethylsiloxane (PDMS) chip. Solutions flow in the bottom layer along the flow channels, while the top layer, which crosses above the bottom flow layer, acts as valves that close and open upon expansion and contraction due to pressure applied[3]. Solutions in seven independently-addressable channels are sent into a mixing area upon completion of mixing for rapid sampling of chemical space. This chip design, called the "formulator," has two on-chip peristaltic pumps, where one is used to inject in precise amount of reagents into the mixing ring, while the other pump is used to circulate the mixture in the ring for homogenization.

While the molecules of interest, labeled with a donor and an acceptor fluorophore, diffuse in a confocal spot focused on the mixing area in the chip, FRET between the dye pair reports the distance between the two dyes. Using a single stranded DNA (ssDNA) that is labeled with a dye pair at both ends, DNA polymer behavior was studied as we titrated different salt concentration levels with the formulator. As the presence of cation increased in the solution, the molecules shifted to a higher FRET state due to screening of the negative backbone charges, which causes the DNA to be in a more collapsed state. The molecule is shifted to a lower FRET state upon

hybridization with its opposing strand as it forms a double stranded DNA (dsDNA) by the rigidification that decreases the persistence length.

Taking advantage of the two distinct FRET states between ss- & dsDNA, we used this doubly labeled ssDNA as a probe to detect hybridization to the mRNA products transcribed by an *E. Coli* RNAP. As RNAP transcribes nascent mRNA, the ssDNA probes in the solution hybridize to the transcripts. The population of ssDNA that initially exists in high FRET state shifts to a low FRET population that is indicative of a hybridized form of the DNA. We have investigated the enzymatic activity as a function of glutamate concentration using this method and suggested possible transcriptional regulation by osmolytes[4].

Optical Mapping of Protein Binding sites on DNA

Accurate measuring of transcription factor binding sites in a single full genome has always been an interesting question. I, along with collaborators in the Weiss lab, developed a method that demonstrates the use of QD localization to pinpoint the transcription factor binding sites on a T7 viral genome. We targeted the biotinylated T7 RNAP bound on 40 kilo base pair genomic T7 DNA with QDs. After staining the DNA with intercalating dye, we stretch the molecule on a positively treated glass coverslip and image it under a wide-field microscope. Each spot overlapping on a DNA strand, indicating a QD successfully targeted to DNA bound RNAP, is then fit with a Gaussian model. By measuring the distance between one end of the DNA and the peak of the Gaussian fit, a relative position of the RNAP is determined. By comparing the peaks of the fit, instead of the whole point spread function of the emission spot, we have been able to enhance the localization resolution significantly[5]. The molecule was also

measured under atomic force microscopy (AFM) in an AFM-fluorescence combined microscope[6].

A challenge of the method above was that there are 17 binding sites of the RNAP on the T7 DNA. Orientation and the stretching factor become problematic when all 17 binding sites are not occupied. To resolve this issue, the Weiss lab initiated collaboration with Elmar Weinhold in Germany who uses methyltransferases to specifically label DNA at known locations[7]. They modified the co-factor to biotinylate, rather than methylate, restriction sites recognized by the methyltransferase enzyme. The three known positions on the DNA, one at one end and two in the middle, serves as flag sites: the end label tells orientation and the distance between any set of the known position provides the stretching factor. After RNAP binding reaction, we targeted these biotin sites of the DNA with one color QD, and then marked the biotin-RNAP with another color QD and combed the DNA on a surface for imaging. By the use of these flags, we have been able to achieve high localization accuracy and precision ($\sigma < 300$ bps) in mapping transcription factor binding sites on DNA[8].

Conclusion

In vitro studies in biology are a non-ideal necessity. The complexity of the system has to be reconstituted bottom-up for a clear understanding of the molecular mechanics of individual components in a biological system. Screening the components needed to achieve a functional state is non-trivial as the concentration ranges of each component with respect to others are often narrow. Our high throughput single molecule methods will greatly aid our studies of biological systems as the effect of each variable and combinations of variables can be screened with

relative ease while having the sensitivity and resolution to distinguish changes in the heterogeneity of the subpopulations levels.

An Overview of this Dissertation

The development of fluorescence based high throughput single molecule methods for transcriptional studies is described in this thesis. Chapter two details the development of a ALEX-based automated transcription assay equipped with microfluidics. Chapter three describes an optical mapping method for localizing proteins on DNA with high accuracy and precision. The appendix lists all the assays used for testing RNAP activity at different stages of transcription.

Chapter 2:

Biological function is context-dependent, and *in vitro* study of a biological system is often limited to a predefined chemical environment chosen to recapitulate facets of the system's function *in vivo*[9]. As a result, *in vitro* studies may overlook real biological phenomena by undersampling the many chemical environments available in the cell. Here we describe a new method that can be used to easily screen many hundreds of chemical environments to aid in the construction of a 'phase diagram' of biological function in the context of chemical space.

A widely adopted readout of biomolecule conformation (or interaction) is single-molecule fluorescence resonance energy transfer (smFRET), wherein a single donor-acceptor pair reports on the distance between dyes on the 1–10 nm scale[10], a technique used to study an ever-growing list of macromolecules (helicases, polymerases and ribosomes, to name a few). Microfluidic technology enables precise and rapid handling of small liquid volumes and thus facilitates large-scale screening of biological molecules with minimal sample consumption[11,12]. The integration of optics and microfluidics is rapidly becoming a useful tool in single-molecule biophysics: microfluidic systems are used for hydrodynamic focusing in single-molecule studies of protein folding[13–16], as 'gradient generators' in large-scale ensemble FRET measurements[17] and gas-controlled smFRET measurements[18], and as formulators[19] capable of mixing labeled RNA molecules[20] for ensemble fluorescence measurements.

We used a microfluidic formulator to perform large-scale, automated single-molecule measurements across a wide range of chemical conditions (**Fig. 2-1**). The critical fluidic element was a microfluidic mixing ring in which a peristaltic pump made of three integrated valves was used to inject reagents of interest into a ring and in which a second pump circulated and mixed the reagents (**Fig. 2-1a**). The dual-layer polydimethylsiloxane (PDMS) device consists of a control layer that uses 'push-down' valves to manipulate fluid on the flow layer, with nominal

channel width of 100 μm and height of 10 μm. We detected the molecules of interest in the flow layer using confocal microscopy. Using the first peristaltic pump (**Fig. 2-1a**) we injected the contents of seven independently addressable input channels (**Fig. 2-3**) into the mixing ring with precision to tens of picoliters. As reagents are pumped into the ring, the reagent valve, inlet valve and outlet valve are opened and previous contents of the ring are displaced. Calculations of final concentrations must take this volume displacement into account.

The push-down valve configuration permits sample detection directly above the glass coverslip, a benefit for single-molecule spectroscopy. An observation chamber (50 μm tall by 250 μm wide) is situated in the mixing ring to allow optimal confocal detection. The ring is then flushed and prepped for a sequential measurement. We developed software for coordination between microfluidic device control and microscopic data acquisition to enable long-term unassisted data collection with high reproducibility (**Figs.2-4 to -6**).

With this device we could automatically perform reagent titrations to screen multidimensional chemical space for conformational and enzymatic changes in biomolecules. The approach allowed for large sampling of parameter space that could not be easily achieved using ensemble-based methods; whereas the single-molecule assay resolves subpopulations (ssDNA, dsDNA and others), an ensemble measurement 'smears' this information into a single, averaged number. We first used an ssDNA probe consisting of a poly(dT) sequence of 20 nucleotides, flanked by donor (5(6)-carboxytetramethylrhodamine) and acceptor (Alexa Fluor 647) dyes at the 3′ and 5′ ends of the nucleotide backbone (called poly(dT) hereafter). We used smFRET as a readout to resolve changes in DNA polymer conformation resulting from compaction and hybridization.

We first sampled the poly(dT) conformation as a function of ionic strength in the microfluidic device. We performed consecutive smFRET measurements of poly(dT) in automatically titrated salt concentrations. We plotted each measurement as a scatter plot with dimensions of FRET, approximated by the 'proximity ratio', and of dye stoichiometry, measured using alternating laser excitation spectroscopy[21]. In each two-dimensional plot, each point represents a single molecule; donor-acceptor molecules of interest are found in the center half of the stoichiometry axis (stoichiometry of 0.25 to 0.75) (**Fig. 2-1b**). As expected, ssDNA molecules collapsed (lower end-to-end distance, higher FRET) upon ionic strength increase owing to ionic screening and consequent reduction in persistence length (**Fig. 2-1d** and **Fig. 2-7a**). We also studied DNA conformation with five different cations, in which the magnitude of each salt effect followed the order of Hofmeister series (**Fig. 2-7b**).

This approach allowed us to study the ssDNA conformation and hybridization efficiency (to a complementary strand) under the influence of two competing processes: increase in rigidity resulting from hybridization and chain collapse resulting from an increase in ionic strength. We conducted 64 serial measurements comprised of an eight-increment NaCl concentration gradient and an eight-increment complementary strand concentration gradient (**Fig. 2-1**) to analyze the differential effects of hybridization (causing lower FRET) and ionic strength (causing higher FRET) on ssDNA. We observed coexistence of ssDNA and dsDNA subpopulations in some cases (**Fig. 2-1d**). We fit subpopulations to two-dimensional Gaussian distributions and extrapolated concentrations from the volume under each subpopulation's fit; we represented the extent of hybridization as the ratio, R = [dsDNA] / ([dsDNA] + [ssDNA]). At the lowest salt and complementary strand concentrations, hybridization efficiencies were negligibly low (**Fig. 2-1c**), and we observed chain collapse caused by increasing ion content. At complementary strand

concentrations greater than 7 nM, the hybridizing effect (low FRET population) dominated the chain collapse effect (high FRET population) as cation concentration increased. Thus, increase in ionic strength not only promoted ssDNA compaction but also promoted hybridization.

Theoretical estimates for the poly(dT) hybridization midpoint were in agreement with the midpoints we observed (**Fig. 2-8**). We therefore recapitulated a large two-dimensional titration to directly measure the hybridization landscape of ssDNA to its complement in 64 different chemical environments. Screening a large phase-space of physicochemical parameters is crucial, for example, for the design of biologically relevant aptamers, molecular beacons or short interfering RNA constructs. To assess the higher dimensionality accessible with this technique, we titrated two salts, NaCl and MgCl2, against poly(dA). The resulting three-dimensional hybridization phase map (**Fig. 2-9**) revealed the comparative and combinatorial effects of the monovalent and divalent salts on ssDNA compaction and hybridization.

We extended this approach to the study of enzyme activity, exploiting the poly(dT) probe as a biosensor for productive mRNA synthesis by bacterial RNA polymerase (RNAP). Previous reports have established dependence of RNAP activity on glutamate concentration, showing that the osmolyte acts as both a positive and negative regulator of transcription, depending on the absolute glutamate concentration[22]. Glutamate facilitates RNAP promoter escape, enabling productive transcription, by altering the RNAP conformation (likely owing to the Hofmeister effect). Conversely, high concentrations of glutamate are reported to cause considerable decreases in transcription by inhibiting RNAP from binding to the ribosomal promoters. Our optofluidic single-molecule transcription assay allowed thorough exploration of the glutamate response space for RNAP activity.

We designed a DNA template for transcription such that the RNA transcript would hybridize to poly(dT) (**Fig. 2-2a**). Maintaining a constant poly(dT) concentration (250 pM), we titrated RNAP in six increments and potassium glutamate in six increments (**Fig. 2-2b**). Here low-FRET species correspond to DNA-RNA heteroduplexes of poly(dT) hybridized to the mRNA transcripts. The volume under the low-FRET peak therefore scaled with productive mRNA synthesis (**Fig. 2-2b**). Fitting each subpopulation to obtain a hybridization ratio provides a metric to directly quantify RNA transcripts.

Although changes to glutamate concentration affect ssDNA conformation, hybridized and unhybridized subpopulations were clearly resolved. Increases in hybridized transcript were apparent with increasing RNAP and glutamate concentrations until excessive concentrations of glutamate (514 mM) were reached (**Fig. 2-2**), consistent with previous observations of glutamate inhibition at concentrations >500 mM[22]. Although the contribution of ionic strength to hybridization efficiency could, in principle, complicate this readout, this effect appears to be minimal when measured directly (**Fig. 2-10**).

In contrast to traditional approaches to transcript detection, the single-molecule approach can be used to detect very low concentrations of RNA; for instance at 50% hybridization, we expected only ~100 pM RNA transcript in only a few nanoliters. Using this approach, we sampled enzyme activity in a large physiochemical space. The possibilities for this type of multiparameter inquiry may help to understand the complex and multifactorial mechanisms underlying transcriptional regulation and other similarly complex enzymatic systems.

The ability to observe multiparameter physiochemical landscapes without having to manually dilute and load the sample on the microscope substantially simplifies the process and dramatically reduces variability during titration. Whereas typical pipetting error can be around

0.5%, or 50 nl for microliter reactions, the ring injection error is ~0.0001%[19]. A possible enhancement to the approach presented here would take advantage of emerging multispot photon-counting detectors[23].

The high-throughput analysis of conformational landscapes may considerably enrich studies of protein regulation, interaction networks, allostery and folding. Moreover, in its ability to quantify small amounts of mRNA, this technique represents meaningful progress toward single-cell gene expression profiling.

Methods

DNA. All oligonucleotides were synthesized and purified by high-performance liquid chromatography (IDT Technologies) (**Table 2-1**).

Buffers. All measurements (except those for RNAP) were conducted in 10 mM Tris (pH 8.0) and 50 mM NaCl. RNAP assays were conducted in 40 mM HEPES, 10 mM MgCl2, 1 mM DTT, 100 μg ml−1 BSA, 5% glycerol and 1 mM β-mercaptoethylamine (pH 7.0) containing 2.8 mM NTP mixture.

RNAP assays. We incubated 1.6 μl of commercially purified 1.111 μM RNAP holoenzyme (Epicentre) with 0.6 μl of 1 μM template DNA at 37 °C in 26 μl of 50 mM Tris HCl, 100 mM KCl, 10 mM MgCl2, 1 mM DTT, 100 μg ml−1 BSA and 5% glycerol (pH 8.0) for 15 min to form the RNAP open complex (RPO). This mixture was challenged with heparin sepharose resin as described previously[24] to eliminate nonspecific complexes. The RPO solution was diluted to a final concentration of 2 nM and loaded on the sample channel of the PDMS chip. Assays were performed at room temperature (19 ℃) as productive transcription occurs well below the optimal

37 ℃. Samples were incubated in the mixing ring for 15 min before measurement to facilitate transcription. As anticipated, negligible transcript was detected in the absence of RNAP; this was also the case in the presence of RNAP inhibitors or in the absence of DNA template or proper nucleotides (data not shown). RNAP stability was measured over the timescale of several hours and shown to yield identical results within experimental error (s.d. = 0.04) (**Fig. 2-6**).

Microfluidic device fabrication. Standard soft lithography techniques were used to fabricate microfluidic devices with integrated valves and pumps[25]. The coverslip (number 1, ~0.15-μm thickness) and PDMS cast were irradiated for 30 s in a plasma chamber and subsequently placed upon one another and kept at 80 ℃ for 10 min. Instructions and parts for building the micro-fluidics setup and the valve controller software are available at http://www.stanford.edu/group/foundry/. Software for hardware control and acquisition is available as **Supplementary Software** and **Supplementary Note** and at http://sourceforge.net/projects/uformulator/files/.

Microfluidic device characterization and calibration. For each channel, peristaltic pump injections were calibrated using fluorescence readouts (Alexa Fluor 488, free dye). The precise amount of injected volume per slug or injection cycle was determined by measured fluorescence intensity in the mixing ring from a single slug normalized to the intensity of 100% Alexa Fluor 488 solution in the mixing ring. For all channels, a single injection slug was 0.4% of the mixing ring volume or a 243-fold dilution of the contents of the ring; calibration curves were generated for device characterization (**Fig. 2-4**). The relatively high observation chamber in the mixing ring creates a slightly thicker flow layer; to compensate, pressure of the control layer was maintained at 40 pounds per square inch (psi) to ensure complete valve closing. Flow pressure was

maintained at 10 psi. To ensure proper mixing, the transient response, as reported by fluorescence intensity, was measured at a fixed point in the ring. The time to establish steady-state mixing was characterized and used in all subsequent measurements (**Fig. 2-11**). Reproducibility of measurement was assessed by conducting a 4 × 4 hybridization-by-salt measurement, performing the serial titrations in both the forward and reverse sequence, which yielded identical results, as indicated by the heatmaps of respective hybridization (**Fig. 2-5**). To prevent cross-contamination and sample loss from surface adsorption, flow channels were incubated with 0.5% of Pluronic F-127 for 1 h.

Estimation of DNA Hybridization Midpoints. The melting temperature (T_M) at 1M NaCl was estimated from the nearest-neighbor two-state-model[26]:

$$T_M(°C) = \frac{\Delta H°}{\Delta S° - Rln[Oligo]} - 273.15$$

with R, the ideal gas constant (1.987 cal K^{-1} $mole^{-1}$), and the thermodynamic parameters $\Delta H°$ (enthalpy) and $\Delta S°$(entropy)[27].

The theoretical monovalent ion (Na^+) concentration at 50% hybridized form at fixed [Oligo] was extrapolated from the following equation[28]:

$$\frac{1}{T_M(Na^+)} = \frac{1}{T_M(1M\ Na^+)} + [(4.29f_{GC} - 3.95)ln[Na^+] + 0.940ln^2[Na^+]] \times 10^{-5}$$

where f_{GC} is the fraction of the GC content in the oligo (in our case, 0).

RNA quantification. The free energy of hybridization of 20-base-pair dsDNA is large enough (about −18 kcal mol−1 at 1 M NaCl) that, as experimental measurements show, essentially all

single strands hybridize at a 1:1 ratio of d(A)20 to d(T)20 (**Fig. 2-12**). An RNA-DNA heteroduplex is known to have similar free energy of hybridization (~−16 kcal mol−1 at 1 M NaCl)[29], and so we consider it reasonable to assume complete hybridization and quantify mRNA transcripts by multiplying the measured *R* by the amount of probes in each experiment (250 pM). Furthermore, we can conclude that our assay is a single-round, run-off transcription assay because the area under the dsDNA population curve did not change with increasing incubation time (data not shown). Other studies have reported the timescale for complete hybridization of 20 base pairs to be considerably shorter than our experimental timescales[30]. Because an RNAP complex can only produce a single mRNA, we can relate the amount of mRNA transcripts to promoter clearance of RNAP. For instance, at 569 pM RNAP and 143 mM glutamate with 250 pM probes (**Fig. 2-2a**), $R = 0.75$. This corresponds to 187.5 pM mRNA or 33% 'active' RNAP complexes in this condition.

Single-molecule spectroscopy. Single-molecule measurements were performed as described previously[31] with slight alterations to microscope setup construction. Briefly, a home-built confocal microscope setup based around an Olympus IX-71 body was used in this study with Plan-Apo 1.4 numerical aperture (NA) 63× oil-immersion objective (Zeiss). Alternating excitation of fluorophores was accomplished using acousto-optic modulator (NEOS Technologies) and continuous-wave 532 nm (CNI) and 635 nm (Coherent) lasers. Confocal detection through a 100 μm pinhole used avalanche photodiodes (Perkin Elmer); a 635 DCXR dichroic mirror (Chroma) and two filters (580 DF60 for donor and 665LP for acceptor (Chroma)) were used to spectrally separate emitted photons. The observation chamber (50 μm) permitted focusing the objective lens 20 μm into the sample preventing stray light from scattering off the coverslip or channel walls. The larger observation chamber also allowed the point spread

function of the oil-immersion objective to fit in the flow channel. To avoid evaporation of the medium between the coverslip and the objective during our long experiments, an oil objective was preferred over a water objective. These data collection parameters yielded detection of single-molecule bursts identical to those collected under similar conditions in the absence of any microfluidic device. Data were collected for 20 min per measurement; a LabView routine was written to permit communication between spectroscopic data acquisition software and the microfluidic controller apparatus for full automation. Each molecule was detected as a 'burst' of fluorescence, from which two ratiometric values were extracted: FRET (reporting on distance between dyes) and stoichiometry ('S', reporting on whether the molecule contains both donor and acceptor (S = 0.5) or donor only or acceptor only (S = 1, S = 0, respectively). FRET and stoichiometry determination was performed as described previously23; but, as this study was not concerned with precise determination of distances but ratio of subpopulations, the FRET measurement was defined simply as the 'proximity ratio' ((acceptor photons)/(all emitted photons) during donor excitation).

Table 2-1. Table of Primers Used In This Study

Oligo#1 (20 dT Probe)	ALEXA647-TTTTTTTTTTTTTTTTTTT-TAMRA'
Oligo#2 (Poly 10 dA)	5’-AAAAAAAAAA-3’
Oligo#3 (Poly 20 dA)	5’-AAAAAAAAAAAAAAAAAAAA-3’
DNA Template for Transcription (non-Template Strand)	5’- AGGCTTGACACTTTATGCTTCGGCTCGTATAATGTGTG GAATTGTGAGAGCGGAAAAAAAAAAAAAAAAAAAA-3’

Figure 2-1. A microfluidic formulator for high-throughput single-molecule FRET measurements. (**a**) Device image (left) with the mixing ring highlighted (arrow). Scale bar, 5 mm. The schematic (right) depicts critical features of the control and flow layer (control and flow channels). (**b**) A schematic plot representing smFRET measurements as a two-dimensional histogram of FRET versus stoichiometry. The subpopulations of interest are hybridized poly(dT) (low FRET population; dsDNA) and unhybridized poly(dT) (high FRET population; ssDNA). (**c**) Heat map of hybridization efficiency (ratio of dsDNA to total DNA) for various concentrations of NaCl and complementary strand. (**d**) Matrix of FRET-stoichiometry scatter plots with contour overlay of fits used to generate heat map shown in **c**.

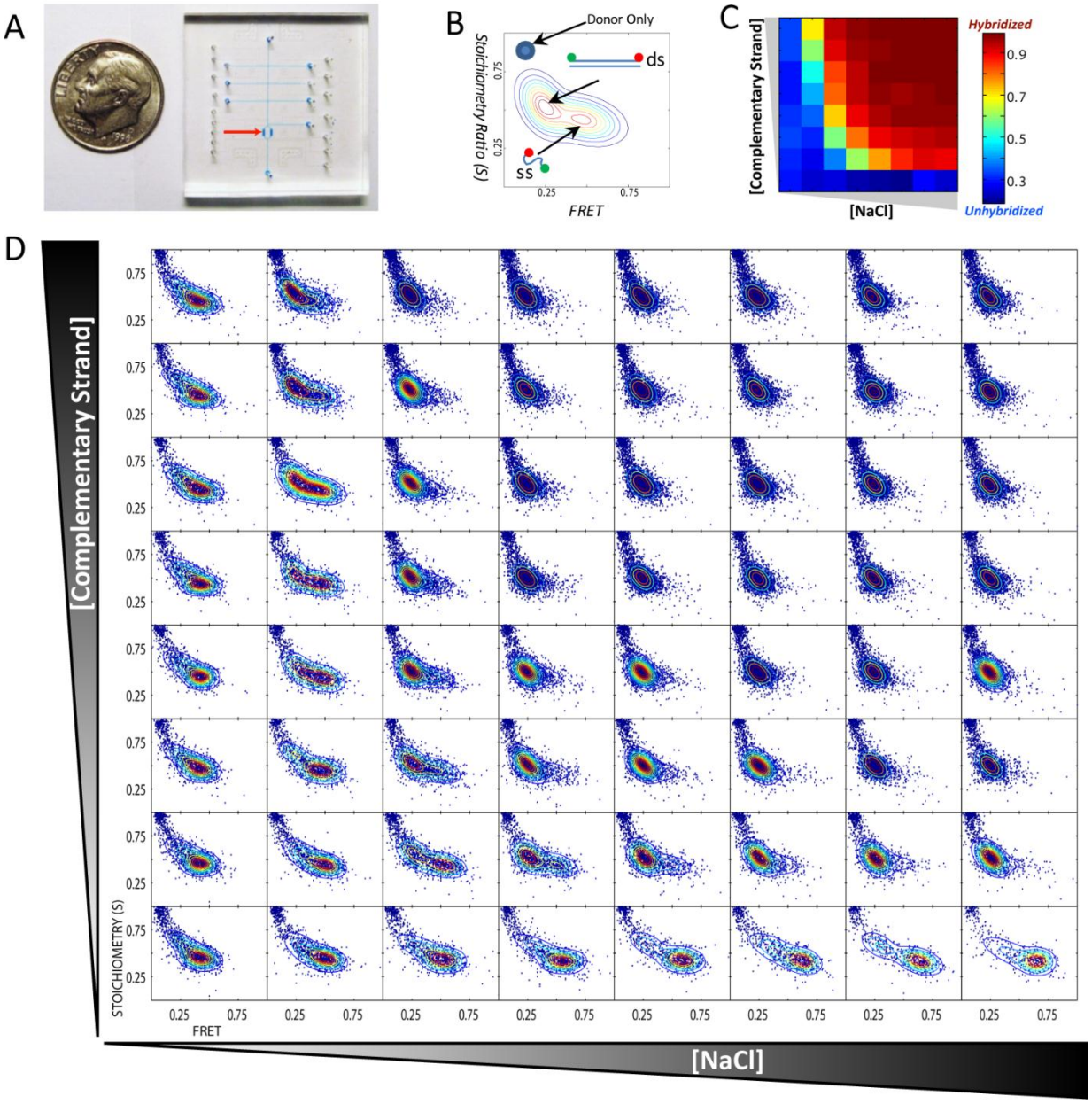

A
B
Donor Only
ds
ss
Stoichiometry Ratio (S)
FRET
0.25
0.75
C
[Complementary Strand]
[NaCl]
Hybridized
Unhybridized
0.9
0.7
0.5
0.3
D
[Complementary Strand]
[NaCl]
STOICHIOMETRY (S)
FRET

Figure 2-2. RNAP activity measured with smFRET. (**a**) A schematic of the assay depicts RNAP transcribing the template, thus producing complementary transcript that hybridizes to the poly(dT) probe. (**b**) A matrix of FRET-stoichiometry scatter plots (as in **Fig.** 1) depicting hybridization of the poly(dT) probe to newly produced transcript upon titrating RNAP and glutamate. (**c**) Heat map showing quantification of the data in **b**. Amount of transcript is plotted for various concentrations of RNAP and glutamate.

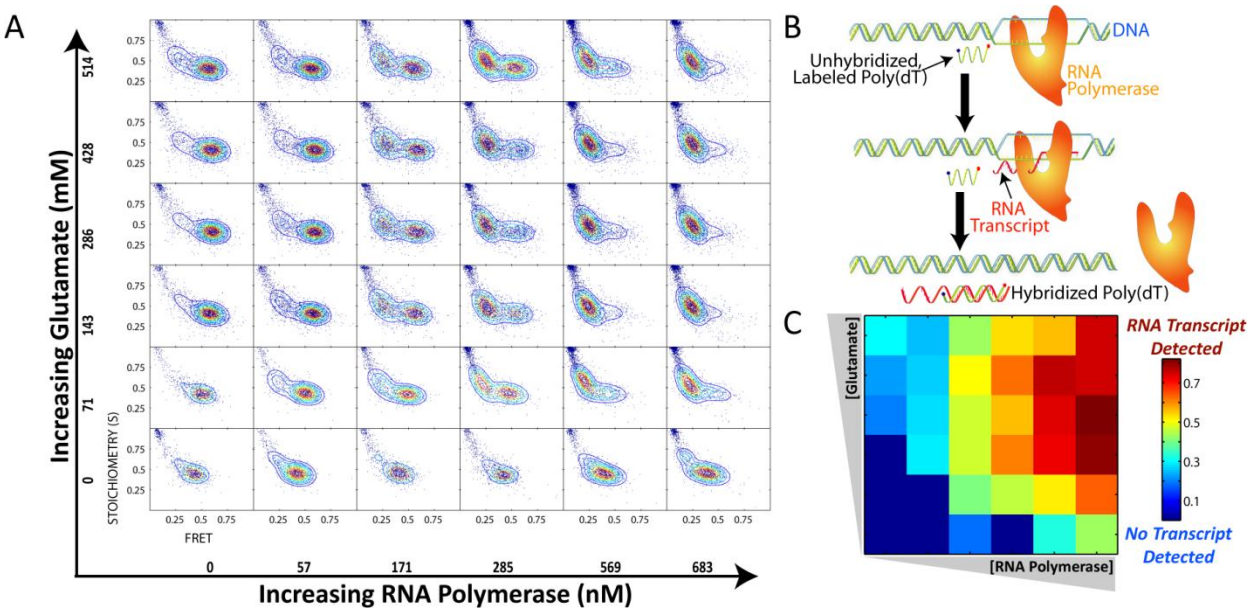

A
Increasing Glutamate (mM)
514
428
286
143
71
0
STOICHIOMETRY (S)
FRET
0.25 0.5 0.75
0 57 171 285 569 683
Increasing RNA Polymerase (nM)
B
DNA
Unhybridized, Labeled Poly(dT)
RNA Polymerase
RNA Transcript
Hybridized Poly(dT)
C
[Glutamate]
[RNA Polymerase]
RNA Transcript Detected
No Transcript Detected
0.7
0.5
0.3
0.1

Figure 2-3. Schematic of the two-layer microfluidic mixing device. Control channels are shown in black and flow channels are shown in red. Locations of the peristaltic injection pump (‘1’), mixing pump (‘2’), and observation chamber (‘3’) are indicated with arrows. (a) Top view of full device. Seven independently addressable reagent inputs are labeled. (b,c) Three-dimensional rendering of mixing ring. Within the observation chamber the confocal detection volume is represented as inverted cones emanating from a microscope objective lens located beneath the chip. Molecules are fed into and mixed into the mixing ring prior to acquisition. Elements of the microfluidic device are to scale, however the objective is not for schematic purposes. The blue cube in the middle serves as a scale bar with 100 um edges.

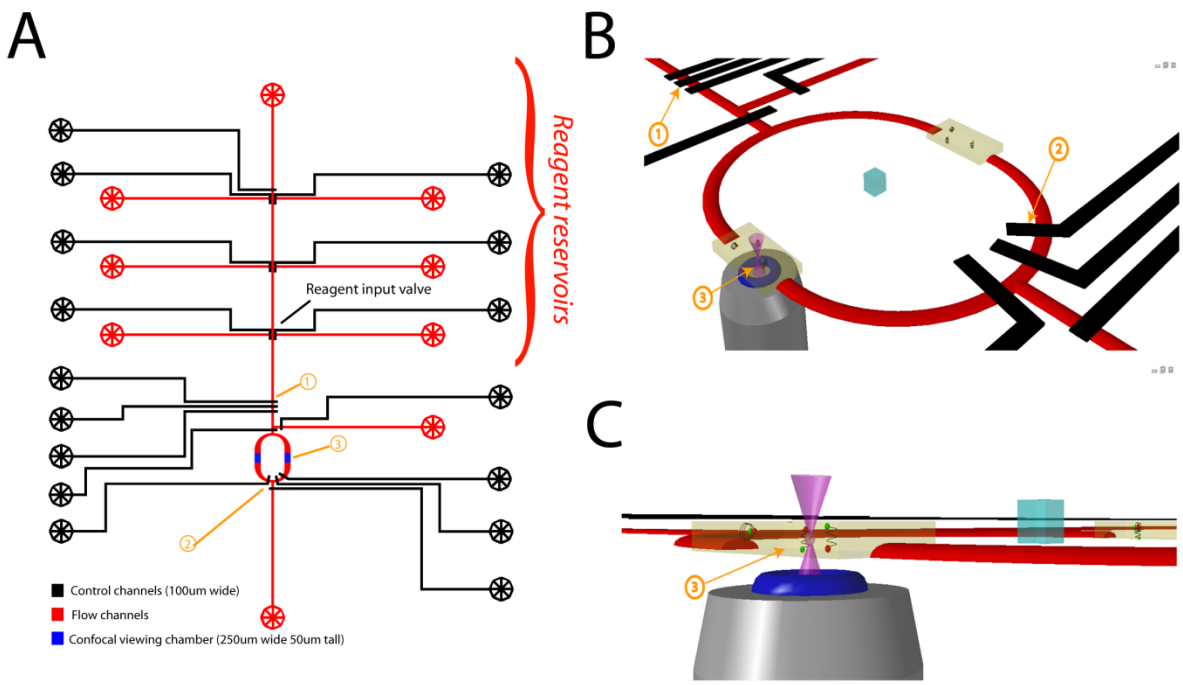

A
Reagent reservoirs
Reagent input valve
1
2
3
Control channels (100um wide)
Flow channels
Confocal viewing chamber (250um wide 50um tall)
B
1
2
3
C
3

Figure 2-4. Calibration of injection pump. "Injection slug" refers to a single cycle of the peristaltic pump, the minimum injection increment. Calibration was conducted with the fluorescent dye Alexa Fluor 488, and relative concentration was determined by normalizing measured fluorescence intensity after mixing to the fluorescence intensity measured in a saturated ring. The linear relationship between injection slugs and relative concentration is used to determine reagent concentrations in ring. Injection from six input channels are calibrated independently.

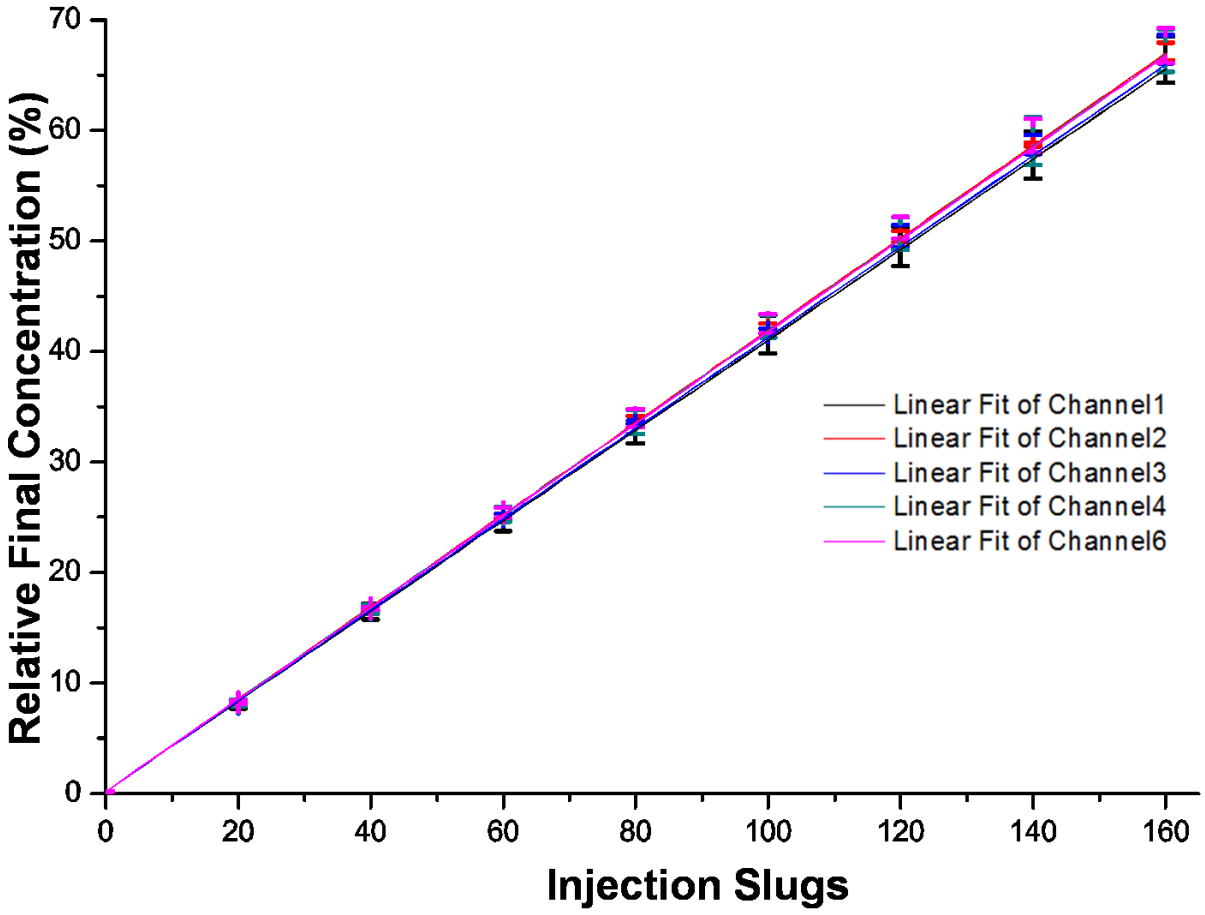
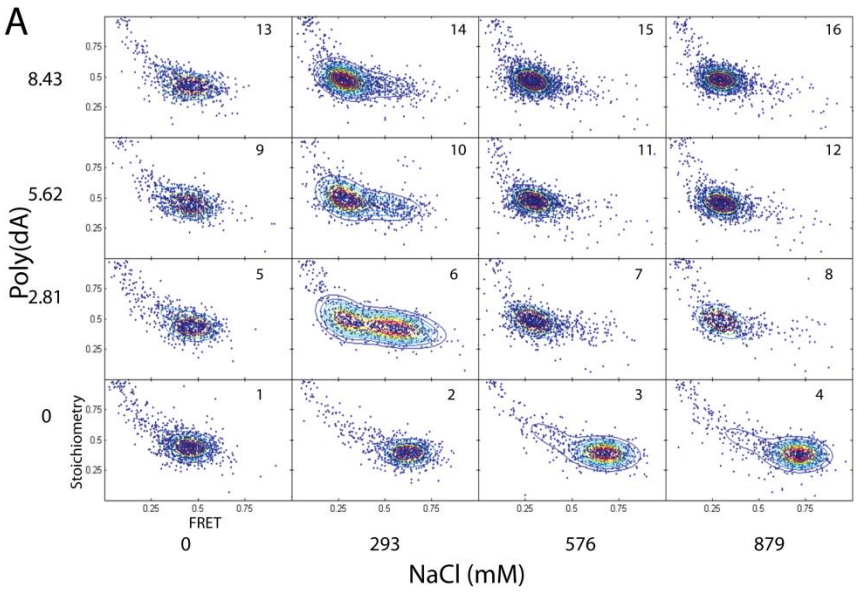
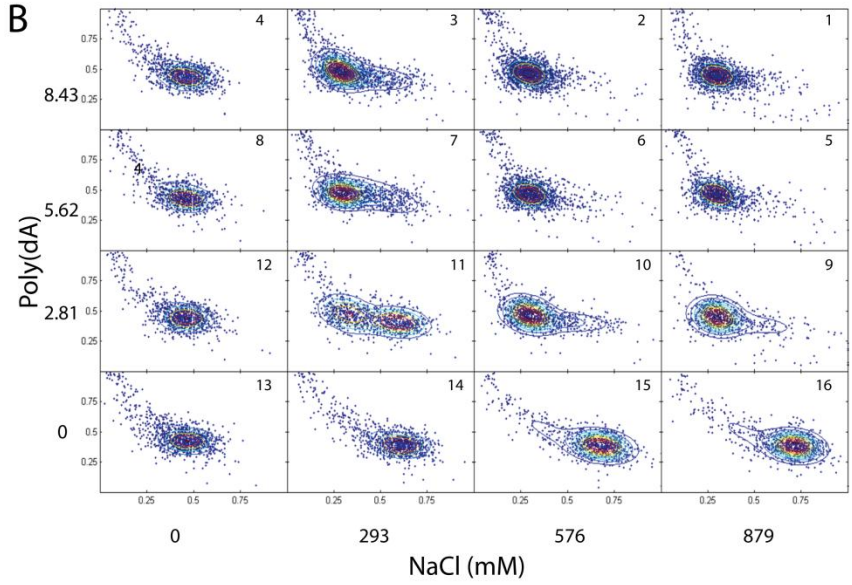
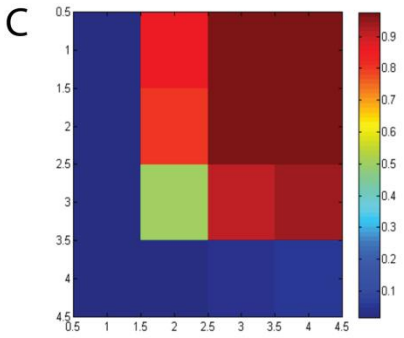

Relative Final Concentration (%)
Injection Slugs
0
10
20
30
40
50
60
70
0
20
40
60
80
100
120
140
160
Linear Fit of Channel1
Linear Fit of Channel2
Linear Fit of Channel3
Linear Fit of Channel4
Linear Fit of Channel6

Figure 2-5. Reproducibility with altered injection sequences. This demonstrates that the reproducibility of the microfluidic device is orthogonal to the order of titrations. This is shown using identical experiments carried out in forward (**a**) and reverse (**b**) order of fluid mixing with corresponding heatmaps of hybridization (**c**) and (**d**), respectively. Experiment carried out in forward and reverse order titrations of 2 M NaCl and 45 nM 10 dA on doubly labeled 20 dT DNA were done in increasing concentrations and repeated in reverse order. The order of the each acquisition sequence is indicated in the top right of each E-S histogram.


A
8.43
5.62
2.81
0
Poly(dA)
Stoichiometry
FRET
0
293
576
879
NaCl (mM)


B
8.43
5.62
2.81
0
Poly(dA)
0
293
576
879
NaCl (mM)


C

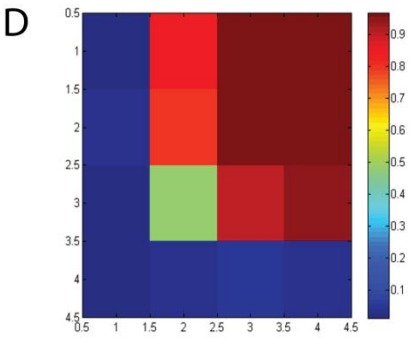

D

Figure 2-6. Sample integrity does not change over the timescale of microfluidic experiments. 227.6 pM of RNA polymerase open complex was diluted into the ring with 250 pM of probe DNA to transcribe under 40 mM of KGlu. This transcription reaction was repeated 15 times in a period of 1120 minutes (320 minutes after the RNAP open complex assembly, when measurement would typically begin, to 1440 minutes) with identical conditions. Throughout the 24 hour period, the transcription efficiency remains constant within a standard deviation of 4%. This demonstrates that the protein/DNA complex and the dyes are stable throughout the course of the experiment, and that the experimental results are highly reproducible.

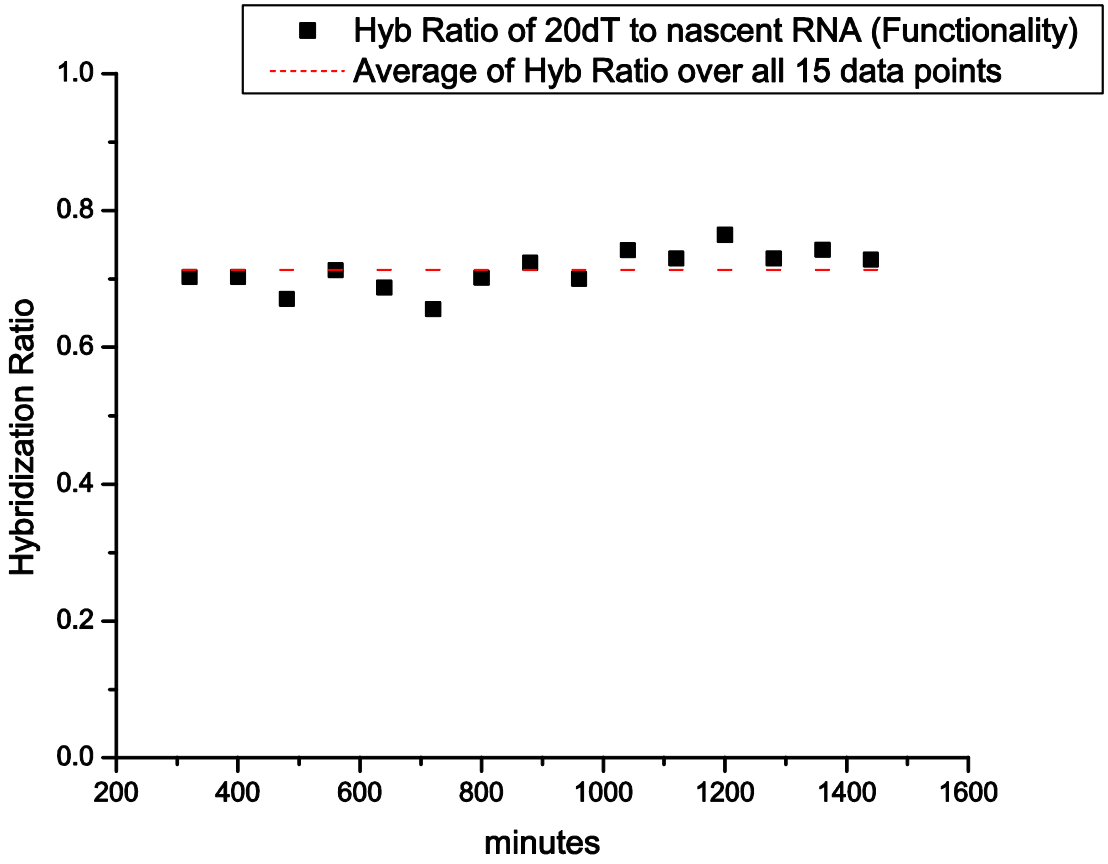

Hyb Ratio of 20dT to nascent RNA (Functionality)
Average of Hyb Ratio over all 15 data points
Hybridization Ratio
1.0
0.8
0.6
0.4
0.2
0.0
200
400
600
800
1000
1200
1400
1600
minutes

Figure 2-7. Effect of ionic strength on ssDNA. (**a**) Increase in ionic strength collapses ssDNA, as reported by FRET histograms (plotted as function of [NaCl]; same data as in bottom row of Figure 1C). (**b**) FRET change as function of salt concentration for 5 different salts (Guanidine HCl, $CaCl_2$, $MGCl_2$, NaCl, and KCl; derived from data as in (a)) Distributions were fitted to Gaussians (assuming normal distributions), peak positions are plotted as a function of salt concentration (injection slugs) to demonstrate the effect of ionic strength, divalent ions, and Hofmeister salts on ssDNA conformation.

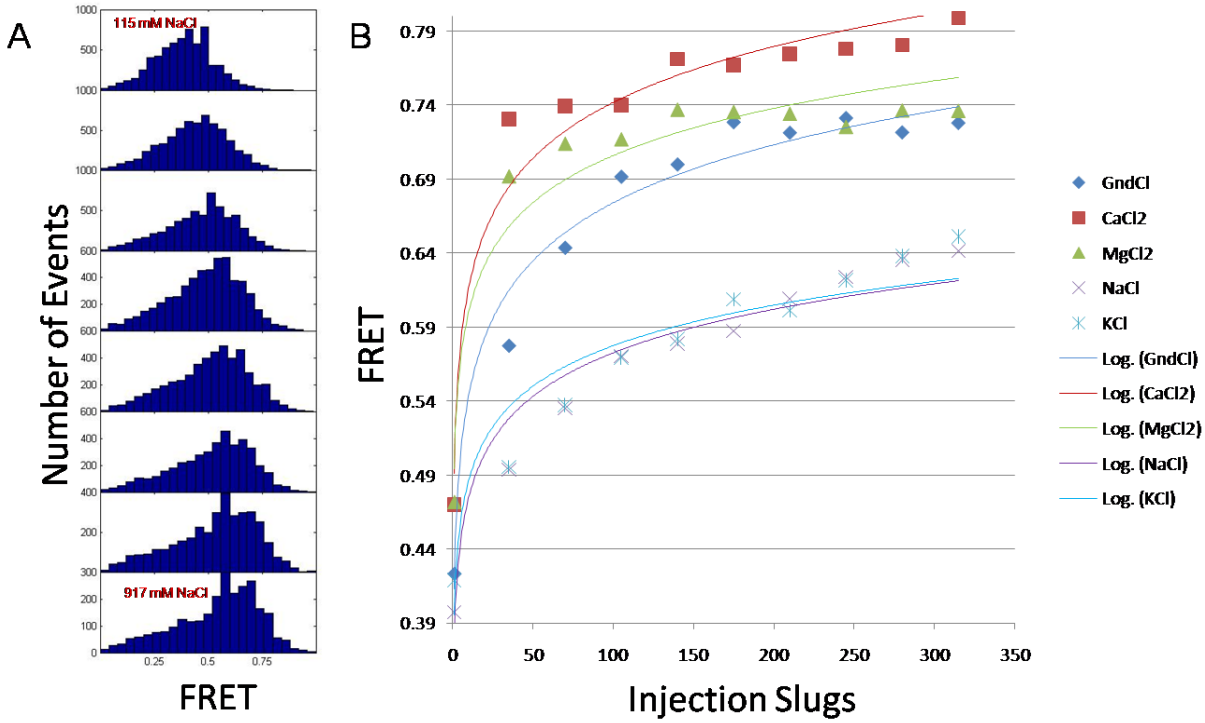

A
115 mM NaCl
917 mM NaCl
Number of Events
FRET
B
FRET
Injection Slugs
GndCl
CaCl2
MgCl2
NaCl
KCl
Log. (GndCl)
Log. (CaCl2)
Log. (MgCl2)
Log. (NaCl)
Log. (KCl)

Figure 2-8. Experimental Tm compared with theoretical Tm. Here, we show Fig. 2-1D, annotated with theoretical hybridization midpoints (blue circles for each row of NaCl titration). Increased sampling as well as elimination of noise in the single molecule data should improve the Gaussian fitting algorithm and further improve agreement between observed and predicted midpoints.

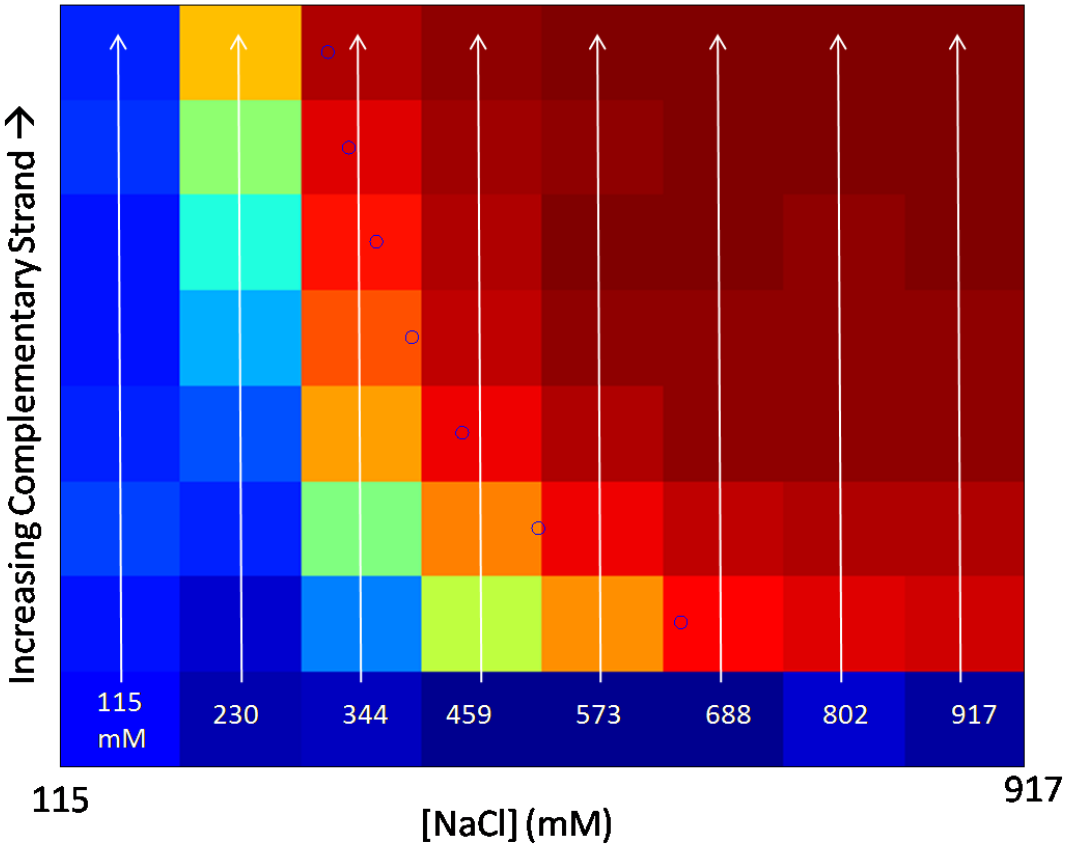

Increasing Complementary Strand →
115 mM
230
344
459
573
688
802
917
115
[NaCl] (mM)
917

Figure 2-9. A three-dimensional (4 x 4 x 4) exploration of physiochemical space (with complementary strand, monovalent and divalent salts). This demonstrates the role of environment in hybridization and conformation of ssDNA. Poly(dT) was combinatorially titrated with 10 dA ssDNA, $MgCl_2$, and NaCl as indicated by the axes. The data represents a three-dimensional landscape of the hybridization efficiency of doubly labeled 20 dT ssDNA. Along with comparing the effects of the different reagents, the additive effects of reagents can be visualized with this method. (a) Raw ALEX data showing FRET and S values for every condition. (b) Processed heatmaps, color represents hybridization ratio as in Figure 2-1 in the main text.

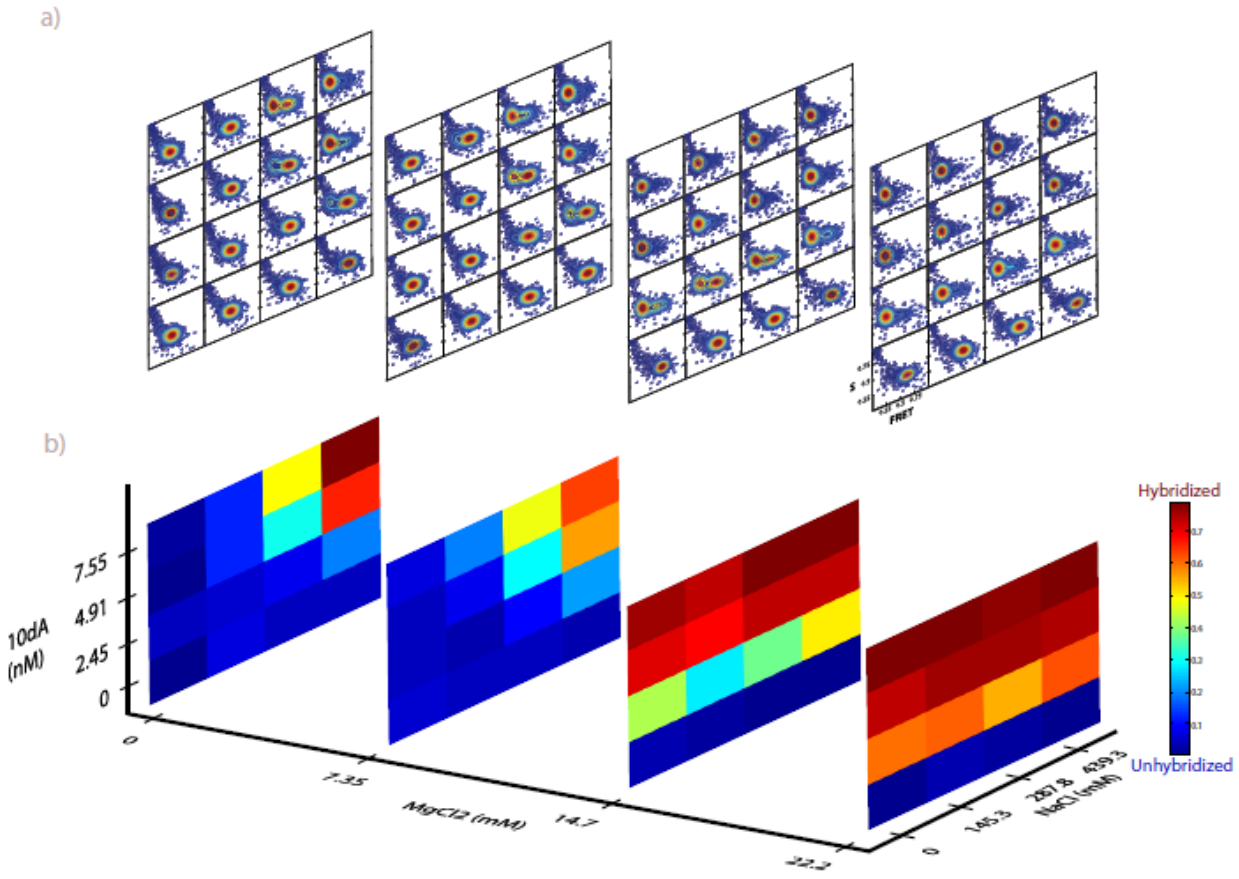

a)
b)
10dA (nM)
7.55
4.91
2.45
0
0
7.35
MgCl2 (mM)
14.7
22.2
0
145.3
287.8
429.3
NaCl (mM)
Hybridized
Unhybridized

Figure 2-10. Effect of glutamate on DNA hybridization vs transcription efficiency. The limited effect of glutamate on hybridization efficiency alone was probed with a 20 - bp length poly-A ssDNA (blue). This is compared with the pronounced effect of glutamate on RNAP transcription (red, data derived from column five of Figure 2-2**a**). Poly(dA) was measured at a concentration of 44 pM and the RNAP protein concentration was 569 pM; as with all other measurements, poly(dT) was measured at 250 pM.

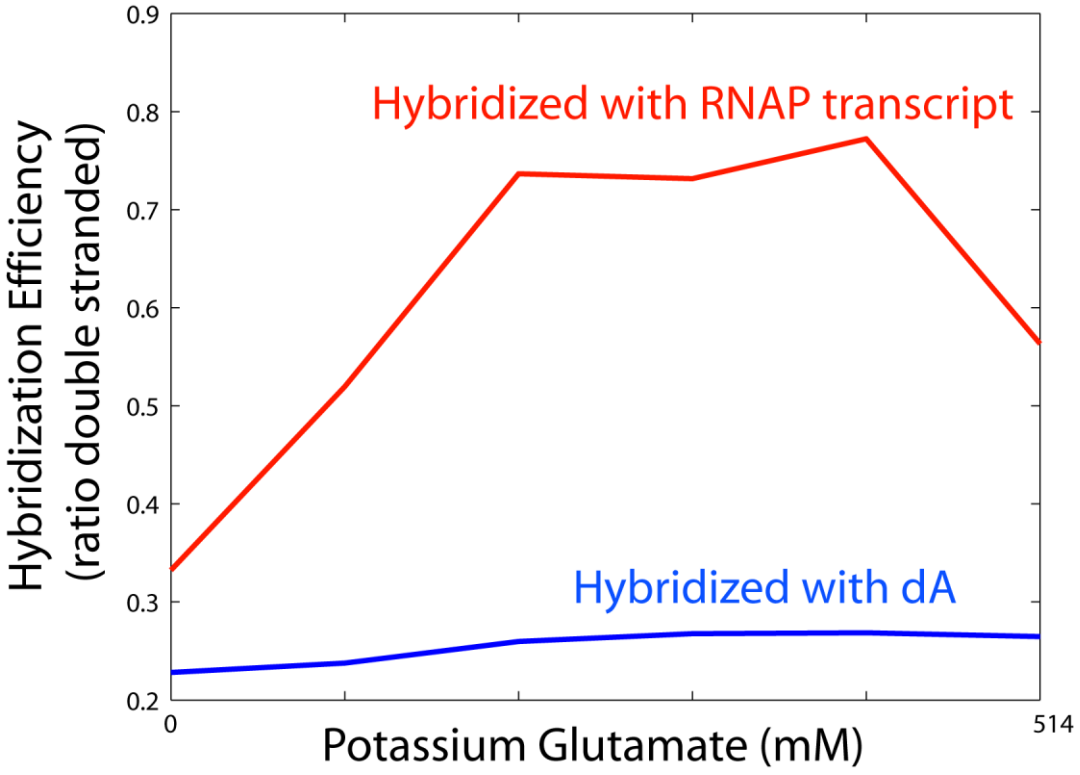

Hybridization Efficiency
(ratio double stranded)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Hybridized with RNAP transcript
Hybridized with dA
0
514
Potassium Glutamate (mM)

Figure 2-11. Mixing curve in the ring. The transient mixing response (reported as fluorescence intensity as function of time) following an initial injection (at ~1200 ms) of dye into the mixing ring of the formulator and subsequent iterations of mixing (in the window 4000 - 8000 ms). Steady state of mixing is achieved at ~10,000 ms (10 s).

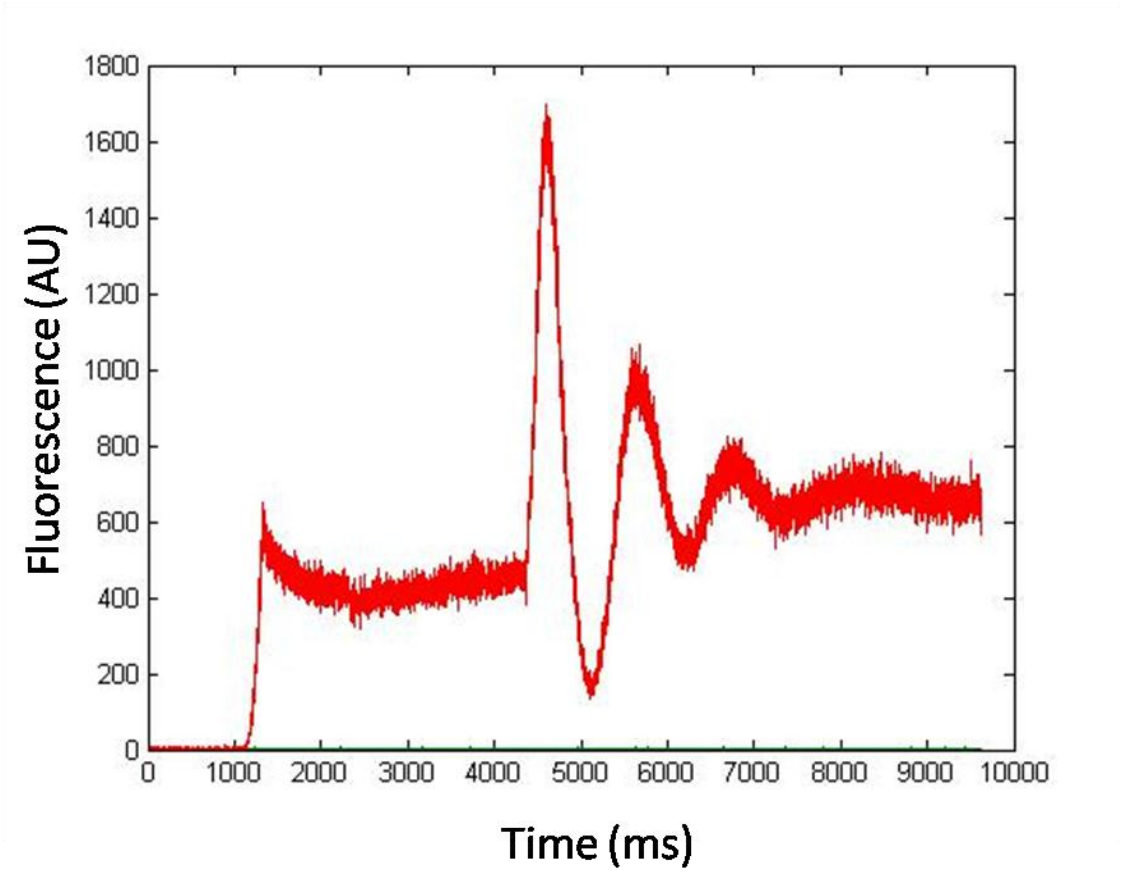
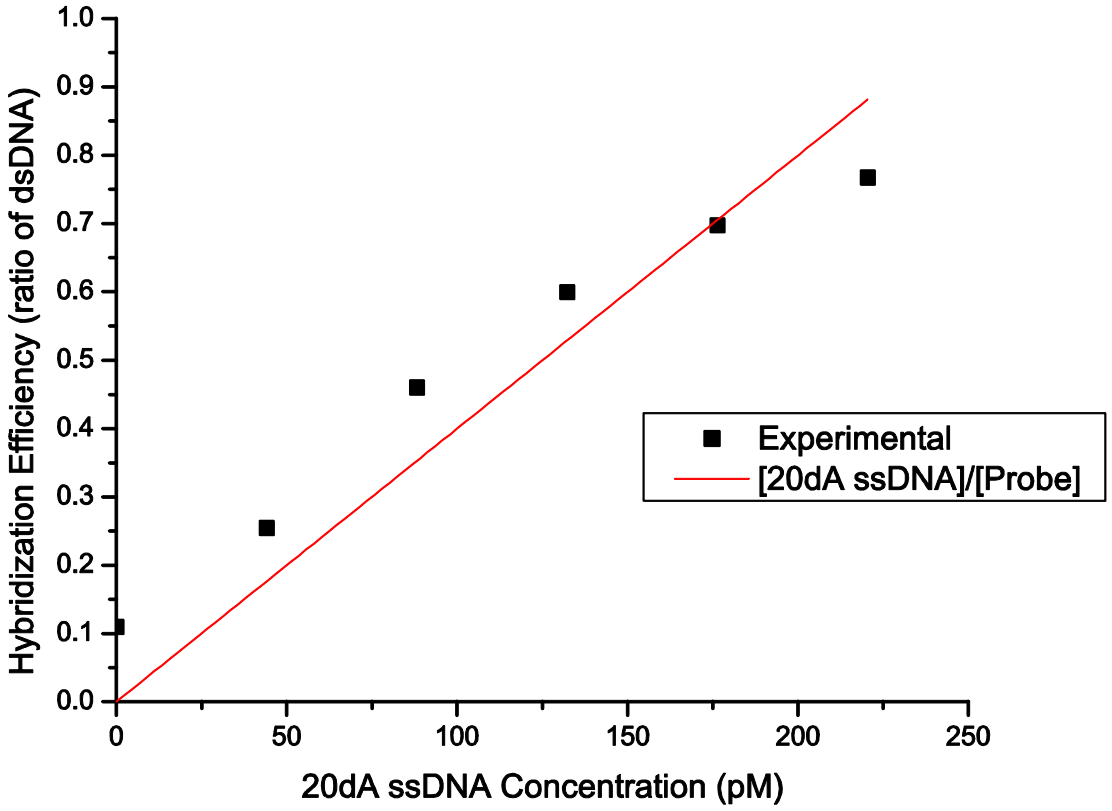

Fluorescence (AU)
1800
1600
1400
1200
1000
800
600
400
200
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Time (ms)

Figure 2-12. 20dT hybridization efficiency to 20dA. The experimentally obtained dsDNA ratios, *R*, at a range of 20 dA ssDNA concentrations (black dots) are plotted with the with the ideal ratio, *Ri*, assuming complete hybridization (red line). *Ri* is determined by dividing experimental 20 dA ssDNA concentrations by the total 20 dT present (250 pM).


Hybridization Efficiency (ratio of dsDNA)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
50
100
150
200
250
20dA ssDNA Concentration (pM)
Experimental
[20dA ssDNA]/[Probe]

Chapter 3:

Affordable DNA sequencing is revolutionizing genetic research and is enabling multiple novel biomedical applications. Among the inherent properties of today's high-throughput sequencing technologies is the fact that it compiles long-range sequences from the assembly of numerous short-read data[32]. This leads to two fundamental limitations: loss of long-range contextual information on the single-genome level and difficulties coping with repetitive or variable genomic regions. Optical mapping and its variants[33–41] rely on the visualization of individual, long (50 kb–1000 kb) DNA molecules and extraction of genomic information by fluorescent labeling of the DNA. These techniques lack the resolution of sequencing but offer genomic context and therefore are attractive both in combination with sequencing to aid in sequence assembly[42–44] and for investigation of genomic structural variations on the individual chromosome level[45,46]. Such variations include deletions, duplications, copy-number variants (CNVs), insertions, inversions, and translocations, all of which have a major impact on the phenotypic variations within a population (or somatic mutations, important in cancer progression). In addition, the available information content of the genome extends beyond the sequence, and the long-range data offered by optical mapping may provide crucial information regarding the distribution of DNA-binding proteins such as transcription factors and histones along the genome.

In previous work we have shown that optical mapping can accurately reconstruct the full promoter map of the T7 bacteriophage genome[5,6]. We used fluorescent quantum dots (QDs) to tag individual RNA polymerase (RNAP) enzymes occupying their genomic binding sites and showed that all 17 promoters could be reliably detected. However, mapping relied on the fact that we were investigating intact viral genomes where the ends of the double-stranded DNA molecules served as reference points for the assignment of the RNAP to its genomic binding

sites. To apply similar optical mapping technology to more complex systems, ranging from bacterial artificial chromosomes (BACs) to bacterial and mammalian genomes, it is essential to create genomic reference tags (refTags) that are independent of the DNA extremities[47]. We utilize sequence-specific methyltransferase-induced labeling of DNA (SMILing DNA) to create a unique, sequence-specific fluorescence pattern[48]. Neely et al. used a similar approach to generate a simple but long-range representation of the lambda phage DNA sequence[41]. The conceptual difference in our case is that rather than labeling the DNA with the objective of sequencing it, we use sequence-specific labeling as a reference overlay to aid in the mapping of information not necessarily encoded in the DNA sequence. In this case we use this pattern to genetically identify the region of DNA under observation and to determine the genomic location of RNAP bound to the DNA and labeled with different color probes.

We demonstrate the utility of this approach by localizing T7-RNAP on one of its target genes, but the technique can similarly be applied to mapping other DNA-binding proteins such as transcription factors or histones (see the Supporting Information), as well as for the long-range mapping of DNA methylation patterns through direct labeling of methylated DNA (with antibodies against 5-methyl cytosine for example).

We show that by choosing a DNA methyltransferase (MTase) that modifies a rare sequence in the genome we can engineer a unique "barcode" that identifies the orientation and exact identity of the observed DNA, and at the same time serves to calibrate the DNA stretching factor (by converting the observed distance between two refTags into the known distance in base pairs). Finally, the genomic location of any observable lying in between two such refTags can accurately be mapped. In fact, we observe a fivefold higher precision in the assignment of the binding site tested in this work when using the genomic markers relative to assigning the

promoter location based on the DNA ends. A schematic representation of the experimental concept is depicted in Figure 3-1.

We used the T7 bacteriophage genome as a model system to demonstrate the utility of this approach to the assignment of T7-RNAP binding sites. Our objective was to generate a unique fluorescent pattern along the genome that would be sparse enough to not produce overlapping fluorescence signals and would also not interfere with RNAP binding. The labeling is performed by DNA MTases that naturally use the cofactor *S*-adenosyl-L-methionine (AdoMet or SAM) to introduce a methyl group onto an adenine or cytosine residue within their recognition sequence. However, upon feeding the enzyme with a modified cofactor, various chemical moieties can be covalently attached to the DNA backbone and the variety of DNA MTases from different species allows modification of a large array of possible recognition sequences including rare sequences in the human genome. We have chosen the bacterial DNA MTase M.BseCI because its recognition sequence (5′-ATCGAT-3′) appears only three times in the T7 genome and creates a well-defined, asymmetric pattern. We have used the aziridine-based AdoMet analogue 6BAz[49,50] to introduce an exposed biotin at three defined positions along the 40 kb genome (see Figure 3-6). These reference points are additionally labeled with streptavidin-coated quantum dots (SAV-QDs) emitting at λ=625 nm. The labeled DNA was extended by capillary flow between two coverslips as reported previously[5] and imaged on a fluorescence microscope. Results from this experiment are presented in Figure 3-3.

The efficiency of the enzymatic reaction is generally extremely high[48] and a lower limit of 95 % can be estimated by a restriction enzyme assay and subsequent agarose gel electrophoresis (see Figure 3-7). However the binding efficiency of the SAV-QDs was lower, with only around 40 % of the 1700 tested DNA molecules displaying two or more QDs, and was

most likely caused by the presence of free streptavidin in the QD solution (saturating biotin sites on the DNA). Nevertheless, this degree of binding was sufficient to generate the binding histograms shown in Figure 3-2 and Figure 3-10. The three reference positions can be easily distinguished from one another and therefore may be independently assigned to their genomic loci. Encouraged by these results we then examined how refTags can aid in the mapping of DNA binding proteins such as RNAP. Recombinant T7-RNAP, containing a biotin tag on the N terminus[5], was reacted with SAV-QDs emitting at λ=700 nm at a 1:2 ratio, thus resulting in a majority of the RNAP labeled with a single QD. The enzyme was incubated with T7 genomic DNA which was prelabeled with the reference QDs. To stabilize the RNAP–DNA binding we initiated transcription by feeding RNAP with a limited set of nucleotides (G/A/UTPs), thus causing transcription to stall as a result of the lack of CTP. A detailed description of all experimental procedures is available in the Supporting Information. After staining the DNA with the intercalating dye YOYO-1, the sample was extended by flow as before and mounted on a fluorescence microscope for imaging. Fluorescence from the sample was recorded in three channels corresponding to the three emission colors from the RNAP–DNA complexes (DNA: λ=500 nm, refTags: λ=625 nm, RNAP: λ=700 nm).

The sample was excited by a single blue excitation band using a Xenon lamp and an excitation bandpass filter (470/40). The three emission wavelengths were recorded on an EMCCD (Andor DU897) by rotating a filter wheel containing the appropriate emission filters (530/50, 620/40, 665 long pass). To correct for image shifts caused by the rotation of the filter wheel between the two QD channels we used a sample of QDs, emitting at λ=655 nm, dispersed upon a clean coverslip. These QDs could be imaged through both channel filters and were used to create a transformation matrix that defined the shift between the channels at every pixel in the

image. The calibration procedure is described in detail in the methods section. On average, the positional error between channels was reduced from over 150 nm to below 30 nm after applying the transformation. Figure 3-3 shows color overlay images of selected genomes carrying both RNAPs (green) and refTags (red). RNAP signals can be traced to various binding sites along the genome. However, we found that mapping precision improved only for RNAP detected between two refTags. As a result, we focused our analysis on RNAP bound to promoter Φ13, which lies between refTags one and three.

We localized the refTag and RNAP QDs by fitting a two-dimensional (2D) Gaussian to their fluorescence spots. The center of the Gaussian was extracted with nanometer accuracy and a table containing the *x* and *y* coordinates of the detected QDs was generated. Next, the transformation matrix was applied to bring RNAP and refTag locations to the same coordinate system. Using the known location of the refTags we calculate a stretching factor for each individual DNA molecule. This factor is used to convert the distance measured between the RNAP and the refTags into a genomically relevant value in base pairs. Figure 3-4 shows position histograms for RNAP detected on the promoter Φ13. To the left are histograms generated from the same data without using the refTags but instead relying on the DNA ends for mapping. To the right are histograms generated from data extracted using refTags. The width of the distribution, represented by the standard deviation of the fitted Gaussians, is a measure of the mapping precision while the difference between the peak of the distribution and the actual genomic promoter is the error. The width of the distribution is significantly reduced when using refTags and precision improved fivefold (from ca. 1.5 kb to ca. 310 bp when taking into account all the data and from ca. 1.2 kb to ca. 270 bp when mapping only strands longer than 80 pixels, corresponding to 70 % extension). This precision compares favorably to the precision of ChIP-

chip (chromatin imunoprecipitation followed by DNA-chip analysis) data and suggests that the use of refTags can overcome limitations on precision obtained when using the DNA strand end points. Moreover, it facilitates the mapping of arbitrary, long genomic fragments by uniquely identifying their genomic origin. By itself, SMILing DNA presents an attractive method for mapping structural variations in genomic DNA by visualizing the physical pattern created along long genomic stretches. Checking typical recognition sequences against the human genome shows that unique patterns composed of discrete labels can be generated. When combined with the visualization of an additional layer of information such as protein binding sites, optical mapping provides the missing contextual information lacking in bulk assays such as DNA arrays or sequencing where binding events originate in different genomes and single-cell genomic context is lost. By investigating such information over long distance scales on the single-molecule level, new information regarding the cooperative nature of certain binding proteins, as well as variations across individual chromosomes may be examined. Such analysis may give rise to subpopulations that are otherwise obscured by ensemble averaging. This analysis may be of particular relevance for diagnostic purposes where early detection of rare events may facilitate targeted and early medical intervention.

Methods:

Handling of long ds-DNA and protection from shearing. Optical mapping relies on the fact that the observed DNA is kept as long as possible thus maintaining its native contextual information. With increasing length, DNA becomes prone to shearing by mechanical stress induced during conventional processing such as pipetting and mixing. Here we adapt a technique

demonstrated by Cinque et al.[51] to protect genomic DNA from shearing by using spermine as a "DNA condensation agent" during the labeling process. We tested the utility of spermine for our application by conducting the following shearing assay:

a. Stock solution of 50 kb long bacteriophage DNA (Fermentas) was prepared in 10 mM Tris-HCl (pH 7.6) and 1 mM EDTA at 0.03 g/ l (~1 nM) and stained with YOYO-1 (Molecular Probes) for 15 min.
b. 4 M 4,9-diaza-1,12-dodecanediamine hydrochloride (Spermine tetrahydrochloride, Sigma Aldrich) solution was prepared in ultra-pure water (Millipore 18 M /cm) immediately before use.
c. Equal volumes of DNA and spermine solutions were gently mixed and the spermine-DNA solution as well as a control DNA solution of identical concentration was subject to repetitive pipetting and vortexing.
d. The DNA is released from spermine by adding 2 volumes of high-concentration (1 M, pH 6.5) MES buffer. Both DNA solutions were extended on modified glass surfaces before and after mechanical stress[5].

Figure 3-5 shows the effect of spermine on 50 kb Lambda phage DNA after vigorous pipetting.

Sequence-specific Methyltransferase-Induced Labeling (SMILing) of DNA. Position-controlled covalent one-step labeling of DNA was achieved by SMILing DNA[48,52–56]. This enzymatic method makes use of DNA methyltransferases (DNA MTases) in combination with synthetic cofactor analogues to deliver desired functional groups to DNA. The enzymes target specific DNA sequences and couple aziridine cofactors with their target nucleobase to yield covalently modified nucleobases within large DNA templates. In this report, a biotin moiety was

coupled with purified T7 phage DNA at the three 5'-ATCGAT-3' sequences using the adenine-specific DNA MTase M.BseCI (Figure 3-6).

The SMILing approach relies on the use of bacterial DNA MTases which are known to methylate specific DNA sequences using the ubiquitous cofactor *S*-adenosyl-L-methionine (AdoMet or SAM). Synthetic aziridine cofactors are used to couple desired functionalities in place of the methyl group with DNA. By replacing the amino acid side chain of AdoMet with an aziridine ring the enzymes are tricked to catalyse a nucleophilic aziridine ring opening reaction and attach the whole aziridine cofactor including a label to their target nucleobase in DNA. DNA MTases are a functionally, structurally and mechanistically well characterized class of enzymes which makes SMILing DNA a convenient one-step DNA labeling method with versatile DNA specificities[57].

Here, we used the DNA MTase from *Bacillus Stearothermophilus* (M.BseCI) and the synthetic analogues 6BAz to covalently incorporate biotin residues at the three target positions in genomic DNA from the T7 bacteriophage[58,59]. The reaction is performed in the presence of spermine in order to protect the DNA from shearing during labeling and purification. Naturally, M.BseCI catalyses the methylation of the exocyclic amino group of the second adenine within the 5'-ATCG**A**T-3' double-stranded (ds) DNA recognition sequence. T7 DNA, a genome of 39.9 kilobases, contains three distributed target positions for M.BseCI, two of them nearby each other at 22.9 kb and 27.1 kb and the third near to the end at 37.1 kb. The coupled 6BAz cofactor provides a linker length of approx. 2 nm between the target adenine in DNA and biotin which is expected to be negligible for positional resolution but can circumvent steric hindrance when binding streptavidin-modified quantum dots. Specifically, T7 phage DNA (10 µg, Bioron GmbH, Germany), M.BseCI (116 nM) and 6BAz (20 µM) were incubated in buffer (400 µL; 10 mM

Tris-HCl, 10 mM $MgCl_2$, 50 µM EDTA, 2 mM 2-mercaptoethanol, 2 µM spermine tetrahydrochloride, pH 7.4) at 37℃ for 7 h. Control reactions without M.BseCI or cofactor were performed in parallel. Afterwards the modified DNA was purified using the QIAGEN® Lambda Mini Kit (QIAGEN) and eluted with buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). DNA biotinylation was verified in a DNA protection assay using the restriction endonuclease R.ClaI. This enzyme cleaves unmodified 5'-ATCGAT-3' dsDNA sequences but is incapable of fragmenting the modified dsDNA. Modified DNA (100 ng) was treated with R.ClaI in 1 × Tango buffer (Fermentas) at 37 ℃ for 1 h and the fragmentation pattern was analyzed by agarose gel electrophoresis. The modified T7 DNA was fully protected against fragmentation by R.ClaI indicating quantitative biotinylation (Figure 3-7, lane 7).

Biotinylated RNAP: Expression and Purification. T7 RNA polymerase was amplified from pDL1917 and cloned into the vector pAN-4 (Avidity) containing an N-terminal BiotinAviTag peptide. The resulting plasmid, pNG301, containing the T7 RNAP with the N-terminal BiotinAviTag and an N-terminal hexa-histidine tag was transformed into AVB101 cells (Avidity). These cells overexpress a biotin ligase, which recognizes the peptide tag and introduces a single biotin molecule at the N-terminal. An overnight culture (10 mL) from a single colony was subcultured into 1 liter of LB media with 100 µg/mL of ampicillin and 25 µg/mL chloramphenicol at 37 ˚C and grown to an OD600 of 0.4-0.6. Cells were induced with 1 mM IPTG and 50 µM biotin. The cells were grown an additional 3 hours at 37˚C, and then harvested by centrifugation for 15 minutes at 5000 x g, 4 ˚C. The cell pellet was resuspended in 35 mL of nickel binding buffer (NBB: 50 mM sodium phosphate, 10 mM Tris-HCl, 0.5 M NaCl, 5mM imidazole, 4 mM 2-mercaptoethanol (2-ME), 5 % glycerol, pH 8.0), disrupted by sonication, and

the lysate was centrifuged for 30 min at 15,000 x rpm, 4 ˚C; all subsequent steps were maintained at 4 ˚C.

1 ml of Ni-NTA agarose beads (Qiagen) was equilibrated with 10 ml of NBB in a 20 ml disposable column (Econo-Pac, Bio-Rad), and the cleared lysate was passed through the column by gravity flow, followed by a 10 ml NBB (plus 10 mM imidazole) wash. Proteins were eluted with 3 column volume (CV) of NBB containing 20 mM, 40 mM and 150 mM imidazole, respectively, and fractions were collected at 1 CV each. The fractions were analyzed by SDS-PAGE, and peak fractions containing RNAP were pooled, and diluted with TGED buffer (20 mM Tris-HCl, 0.1 mM EDTA, 1 mM dithiothreitol (DTT), 5 % glycerol) to a final NaCl concentration of 0.1 M.

The resulting sample was loaded onto a Heparin-Sepharose column (GE Healthcare) pre-equilibrated with TGED with 0.1 M NaCl, using Aktä purifier (GE Healthcare). Protein was eluted with a gradient from TGED + 0.1 M NaCl to 1 M NaCl over 120 min at 0.25 mL/min. Fractions containing pure RNAP were pooled and dialyzed against storage buffer (10 mM Tris, pH 7.9, 50 % glycerol, 0.1 mM EDTA, 0.1 mM DTT, 0.05 M NaCl), then stored at -20 ˚C. The typical yield is about 5-10 mg of purified RNAP from a 1 liter culture.

Labeling of biotinylated DNA with quantum-dot (QD) refTags and RNAP-QD. Biotinylated DNA was labeled with QDs as sequence specific refTags. These tags were used to facilitate the precise mapping of QD-labeled RNAP bound to the same DNA. Experimental details for this reaction are as follows:

a. Binding was conducted in TB Buffer (50 mM Tris pH 7.9, 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 100 μg/ml BSA, 5 % glycerol).

b. QD labeling of T7 RNAP and T7 DNA: 1 μl of 4.5 nM biotinylated T7 DNA was mixed with 0.7 μl of 100 nM streptavidin (SA) conjugated 625 nm QDs from Invitrogen (Carlsbad, CA) in 1.7 μl TB buffer and incubated for 15 min at 25 ^{o}C. Similarly, 1.5 μl of 200 nM biotinylated T7 RNAP were mixed with 3 μl of 100 nM streptavidin (SA) conjugated 705 nm QDs from Invitrogen (Carlsbad, CA) and incubated for 15 min at 25 ^{o}C.

c. T7 RNAP binding reaction on T7 DNA: 4.5 μl of 705 nm QD-labeled T7 RNAP was added to 3.9 μl of TB buffer and incubated for 10 min at 30^{o}C. Then, 2.4 μl of 625 nm QD-labeled T7 DNA were mixed into the solution and incubated for 20 min at 37 ^{o}C to form RNAP 'open complex' (RPO). To eliminate nonspecific complexes that occur outside promoter sequences, this mixture was challenged with heparin Sepharose resin from GE Healthcare (Piscataway, NJ) as described elsewhere[24]. 0.4 μl of 1 mM G/A/UTP was added to initiate transcription and lock RNAP in a stable mode by feeding it an incomplete set of NTPs. Finally, 3.3 μl of 20 uM YOYO-1 dye from Invitrogen (Carlsbad, CA) was added to the mixture and incubated for 15 min at 25 ^{o}C to stain the DNA.

Surface preparation and deposition of DNA sample. No. 1 coverslips from Fisher Scientific (Pittsburgh, PA) were assembled in a rack and sonicated twice at 60 ^{o}C in 1 M KOH for 20 min. The coverslips were thoroughly washed by sonicating three times in deionized H_2O for 20 min at 60 ^{o}C. For complete drying and elimination of fluorescence background, the coverslips were baked at 450 ^{o}C for 3 h. 10 μl of 6 μg/ml poly-L-lysine from Sigma-Aldrich (St. Louis, MO) were sandwiched between two coverslips and dried in the dark for 24 h.

Before deposition of the labeled DNA sample, the sandwiched poly-L-lysine treated coverslips were separated, exposing the treated surface. An untreated coverslip was placed on top of a poly-L-lysine treated surface. 3 μl of QD-RNAP-DNA solution diluted 200 times in 100 mM Hepes, 100 mM βME, pH 7.0 buffer was deposited on the interface between the coverslips. Capillary flow spreads the drop between the coverslips away from the point of insertion. At micron distances, comparable to the DNA length, this results in deposition of linearly extended DNA molecules on the surface. 100 mM of βME was added to suppress QD blinking, as the βME has been experimentally proven to reduce QD blinking substantially at concentrations above 90 mM by creating a reductive environment[60].

Data acquisition. The use of QDs as fluorescence labels enabled us to use a simple imaging system with a single excitation band to record data in multiple colors. The extended DNA was imaged on an IX-71 Olympus microscope with a 60X 1.45 NA oil immersion Olympus objective. As the excitation source we used a Xenon lamp with a 40 nm band-pass centered at 470 nm and a 505 nm long-pass dichroic mirror. A filter wheel from Sutter Instruments was used to switch emission filters in order to image the three different fluorophors separately (535-50 band-pass for YOYO-1 stained DNA, 620-40 band-pass for 625 nm QDs on DNA, and 665 long-pass for 705 nm QD labeled RNAP).

Channel registration for multi-color imaging. In order to measure precise distances between refTag QDs emitting at 625 nm and the desired RNAP QDs emitting at 705 nm we needed to overcome shifts between the two colors that arise due to filter changes and chromatic aberrations. In a first step a reference sample consisting of sparsely spin coated QDs emitting at 655 nm

which exhibited signals on both acquisition channels was recorded. For each channel the positions of the QDs where estimated by fitting their fluorescence point spread functions (PSF) with a Gaussian distribution and estimating its center. We use a program written in Igor Pro (kind gift of Peter Dedecker, Katholieke Universiteit Leuven, Belgium) to auto detect all spots in the field of view and extract their center coordinates. Each detected QD had two center coordinates (one in channel one, the other one in channel two). The registration was done by applying a local weighted mean fitting algorithm (Matlab 2009b 'cp2tform') to minimize the mismatch between the QD coordinates from both channels. The algorithm results in a registration matrix that was subsequently used for the recorded experimental data. The histograms presented in Figure 3-8 show the effect of the channel registration procedure. The upper left shows the distribution of distances between fluorescent spots recorded in the two color channels but arising from a single QD emitter. In the lower panel are the same data after applying the transformation matrix. The average distance between the channels has shrunk from over 150 nm to less than 30 nm. Similar results were obtained for a test dataset recorded on a different day (right), indicating that the registration procedure is robust.

Manual refTag mapping. We first mapped the positions of the 3 refTags in order to verify that they are located in the expected positions and easily distinguished one from the other in case that not all three refTags are present on a given DNA molecule. In order to determine the genomic base-pair position of refTags we measure the distance (in pixels) between the refTag signal maxima and the far end of the DNA. This value is divided by the measured full length of the DNA to give a normalized position value (ranging from 0 to 1). Multiplying this value by the known genomic length of 39936 bp, results in a genomic refTag position for each molecule.

RefTags were mapped manually using a Matlab program written for this purpose. The spectrally separated images of the stretched DNA and of the QD fluorescent spots (625 nm QD channel and YOYO-1 stained DNA channel) were overlaid to visualize QDs bound on the DNA refTags. Using the Matlab function Improfile, a line was manually drawn along each non-bundled, long DNA strand (> 60 pixels) that showed at least one QD along its contour. The Improfile function projects the intensity values of each pixel along the drawn line on a pixel vs. intensity plot. The DNA lengths were measured by subtracting the two y-intercepts on the DNA channel that represent DNA ends. The refTag-QD position was recorded by finding the peak of the pixel vs. intensity plot in the QD channel and measuring its distance to one end of the DNA in order to assign a genomic position. Figure 3-9 shows the impact of multiple QDs on the stretching factor. We analyze the length distribution of genomes baring 1, 2 and 3 QDs and find that stretching factor increases (< 5 %) as the number of bound QDs increase. This may be an effect of the increased drag force acting on the DNA due to the large hydrodynamic radius of the QDs. Next, we plot histograms of the three refTag locations extracted from the total measured population (Figure 3-9a) and compare to analysis of DNA strands containing 2 (Figure 3-9b) and 3 (Figure 3-9c) refTags. In all three cases the 3 peaks are fairly separated and allow assignment of the refTags based on their visualized location on the genome. This distinction is important for the mapping procedure since once a RefTag is identified, its exact sequence location is used (rather than its measured location) as a reference for mapping other unknown detected signals along the DNA.

RefTag localization based RNAP mapping. RefTag and RNAP fluorescent spots may be accurately localized by 2D Gassian fitting. By measuring the distance between detected RNAP

signals and the known refTags, we were able to map RNAP binding positions on the studied genome. Images were analyzed using a custom Matlab program that extracts the position coordinates of fluorescent spots from the different color channels. The spectrally separated images (625 nm RefTag-QD channel, 705 nm RNAP-QD channel, and YOYO-1 stained DNA channel) were overlaid to visualize QDs bound to DNA. Examples of such overlay images are given in Figure 3-11. QDs on each straight, non-bundled and long DNA strand (> 60 pixels) were 2D Gaussian fitted and the X and Y centers of the Gaussian fits were recorded orthogonally for each channel (625 nm and 705 nm). Since the refTag positions are known, the stretching factor and orientation of the genome is calculated based on the distances between the refTag QDs (625 nm). In order to obtain the genomic binding position of RNAP we measured the distance between RNAP and one of the refTags and then use the stretching factor to convert this physical distance to the base-pair location of RNAP on the genome.

Future directions and possible applications. In this work we have shown that the use of refTags enhances the mapping precision of DNA binding proteins such as RNAP and brings this technology one step closer towards mapping of biologically relevant information on complex genomes. The generation of a unique pattern of refTags, generated via sequence-specific recognition sites of various DNA methytransferases will allow the assignment of both orientation and genomic identity of arbitrary DNA fragments such as expected from large genomes that are inevitably fragmented during extraction and processing.

An immediate extension of this experiment that does not require further development is the mapping of nucleosomes that are reconstituted *in-vitro* on selected DNA into what is referred to as reconstituted chromatin[61]. Such experiments readily offer information regarding the sequence preferential assembly of nucleosomes and may be further enhanced by differential color labeling

of histones with various post-translational modifications. Figure 3-12 shows preliminary results of reconstituted chromatin on lambda phage DNA. The atomic force microscope (AFM) image of the sample confirms nucleosome assembly on the DNA seen as white bumps along the DNA contour (left panel). Despite the high protein content the DNA readily stretches and deposits on the surface of a glass coverslip using the same experimental procedures described above (right panel). Another immediate application for refTags is genotyping. Being a single molecule technique, it should enable for example the detection of rare bacterial strains within a sample. Similar single molecule techniques such as optical restriction mapping and direct linear analysis (DLA) have already proven their potential for this application and have been commercialized[62,63]. The ability to visualize low resolution representations of genomic DNA over hundreds of kb may be used in concert with high throughput sequencing to reduce the needed sampling coverage by aiding sequence assembly. Moreover, by comparing these long range patterns with a consensus sequence database the method may serve to detect structural variations such as inversions, deletion, duplications and other copy number variations with potential in both basic science and clinical settings.

Figure 3-1. Schematics of the experiment. (a) Schematic representation of QD-labeled RNAP bound to sequence-specifically labeled T7 bacteriophage DNA. The DNA-methyltransferase M.BseCl recognizes specific rare sequences in the T7 genome (5'-ATCGAT-3') and conjugates an aziridine cofactor 6BAz with an attached biotin (yellow triangle). This biotin group can be used for site-specific labeling with QDs. (b) RNAP-bound T7 bacteriophage genome flow-stretched on a poly-lysine surface. (c) Image of flow-stretched YOYO-1 stained T7 bacteriophage DNA (white) with QD-labeled RNAP (green) and M.BseCl refTags labeled with spectrally distinct QDs (red). (d) Conceptual representation of a genome-wide map of promoter and DNA methyltransferase sites. Histogram of M.BseCl positions represented in red and RNAP positions represented in green.

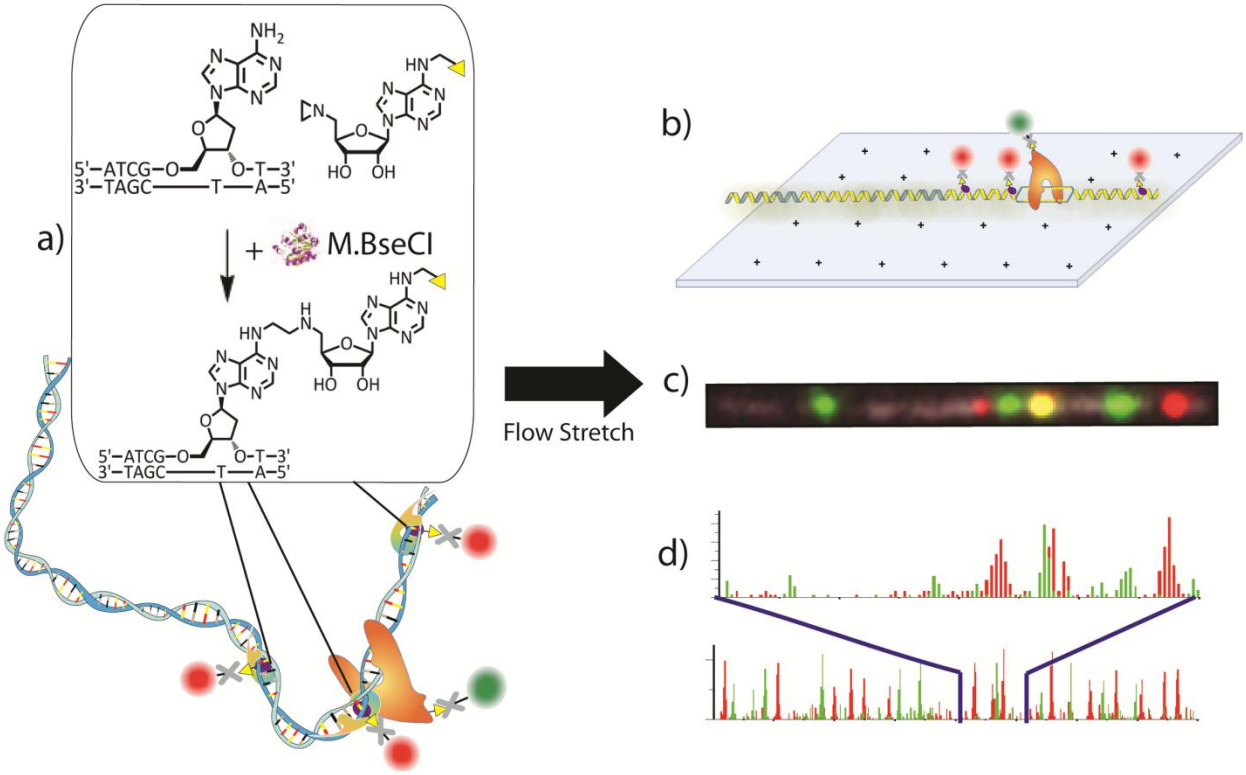

a)
M.BseCI
Flow Stretch
b)
c)
d)

Figure 3-2. MTase labeling specificity. (a) Histogram obtained from localizing refTag - QDs on stretched T7 bacteriophage genome. Red vertical lines represent expected position of recognition sequences. (b) Selected images of extended YOYO-1 stained T7 genomes (white) labeled with QDs on M.BseCl DNA-MTase refTag sites (red) (scale bar 3 μm).

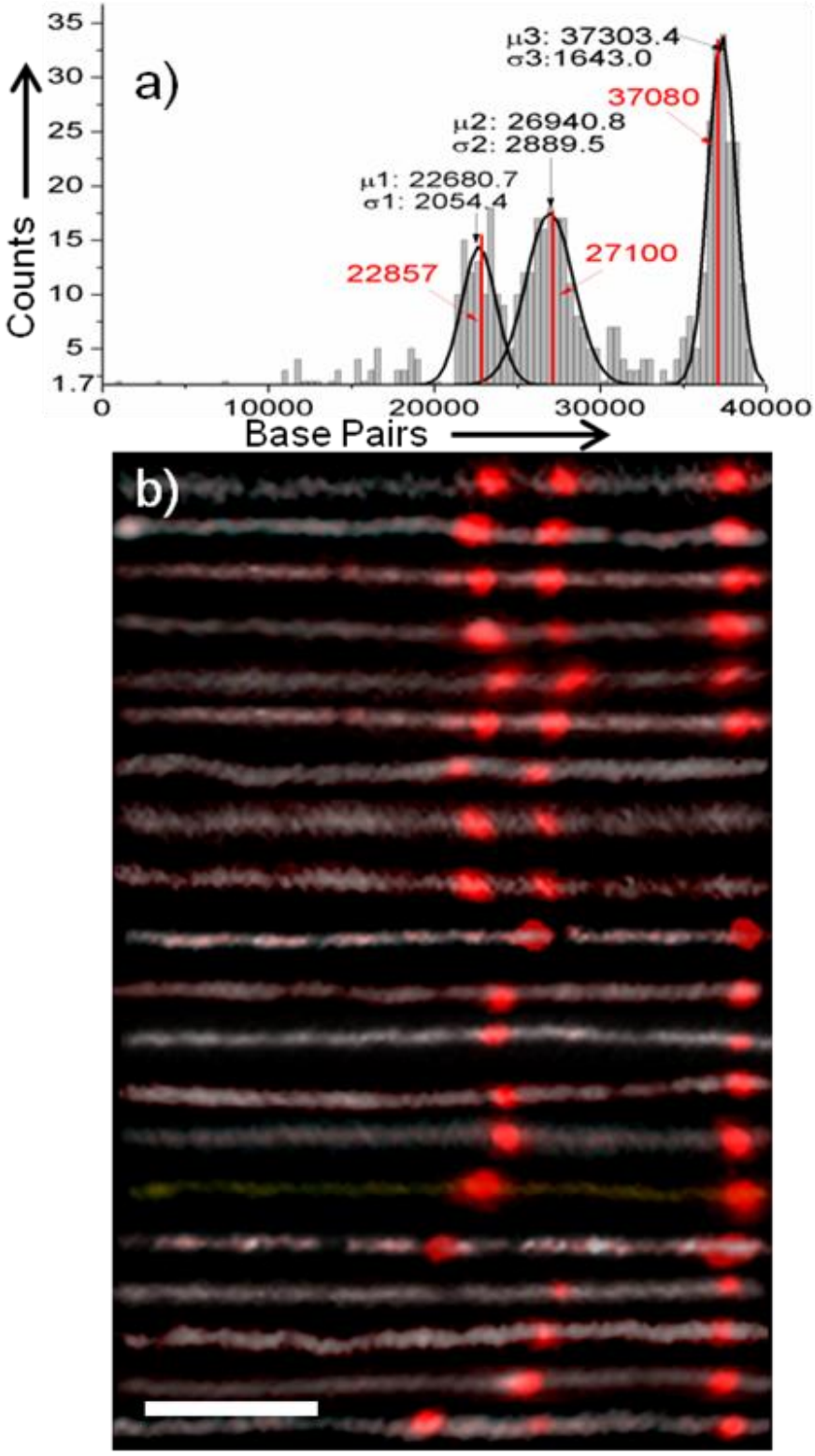

a)
μ3: 37303.4
σ3:1643.0
37080
μ2: 26940.8
σ2: 2889.5
μ1: 22680.7
σ1: 2054.4
22857
27100
Counts
Base Pairs
35
30
25
20
15
10
5
1.7
0
10000
20000
30000
40000
b)

Figure 3-3. Sequence-specific Methyltransferase-Induced Labeling of DNA (SMILing DNA) creates a distinct "bar code" that reports the preferential binding sites of RNAP on T7 bacteriophage DNA. Selected images of QD-labeled RNAP (green) bound to promoters on stretched YOYO-1 stained T7 bacteriophage DNA (white or blue) with refTag sites labeled with spectrally-distinct QDs (red). Overlapping red and green signals are shown in yellow (scale bar 3 μm).

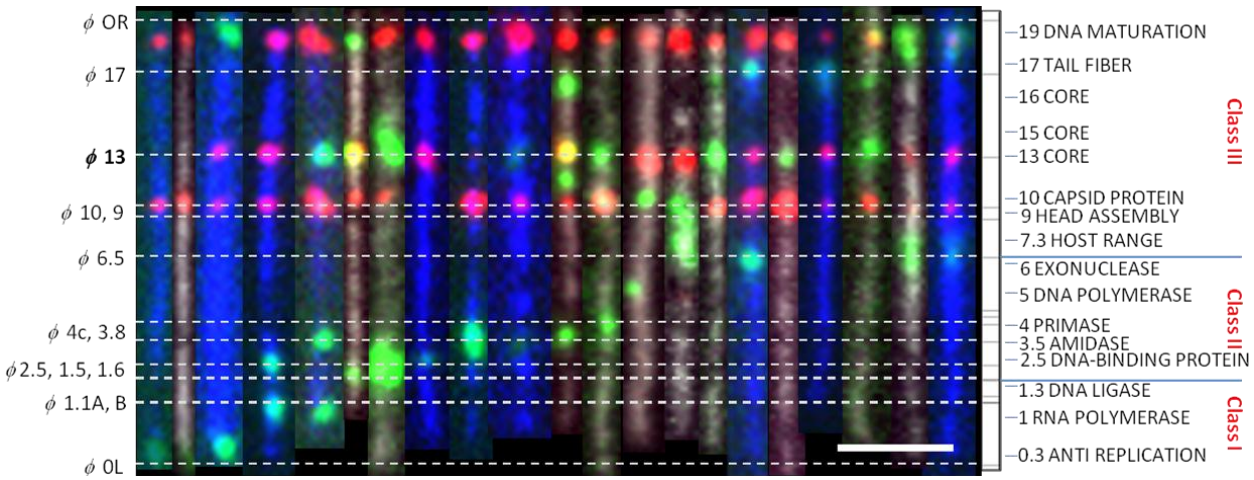

ϕ OR
ϕ 17
ϕ 13
ϕ 10, 9
ϕ 6.5
ϕ 4c, 3.8
ϕ2.5, 1.5, 1.6
ϕ 1.1A, B
ϕ 0L
19 DNA MATURATION
17 TAIL FIBER
16 CORE
15 CORE
13 CORE
10 CAPSID PROTEIN
9 HEAD ASSEMBLY
7.3 HOST RANGE
6 EXONUCLEASE
5 DNA POLYMERASE
4 PRIMASE
3.5 AMIDASE
2.5 DNA-BINDING PROTEIN
1.3 DNA LIGASE
1 RNA POLYMERASE
0.3 ANTI REPLICATION
Class III
Class II
Class I

Figure 3-4. Histograms for localized RNAP on T7 bacteriophage using distance measurment to the DNA ends (left) versus localization by refTags (right). (a & b) Histograms of all analyzed strands with gaussian fit, (c & d) histograms based on selection of long strands (>80 pixels) with gaussian fit. Dotted lines represent actual position of *Φ13* promoter. Histograms using refTags yield a 5-fold increase in accuracy as evidenced by sharp reductions in the width of promoter localization distributions.

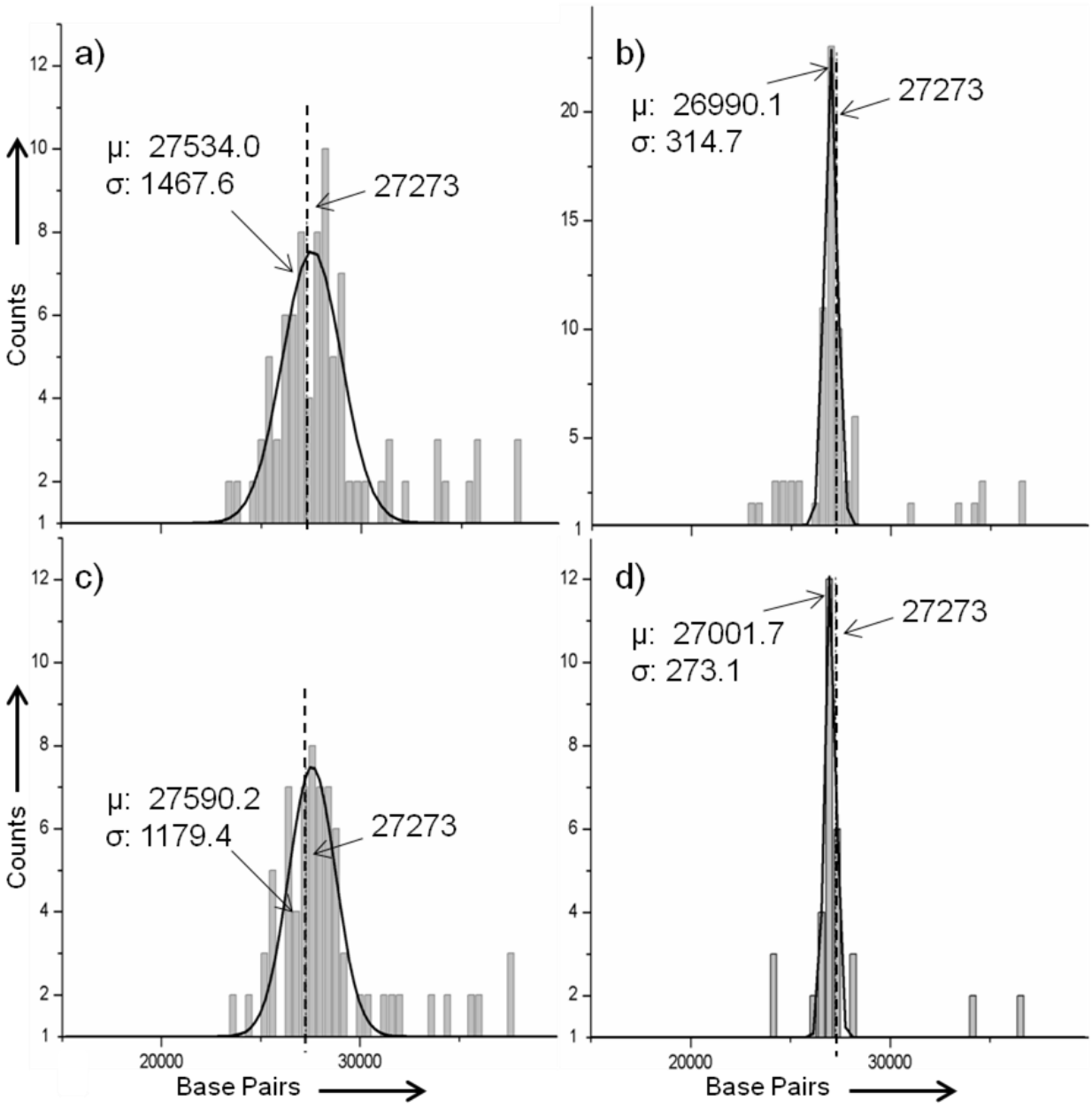

a)
μ: 27534.0
σ: 1467.6
27273
b)
μ: 26990.1
σ: 314.7
27273
c)
μ: 27590.2
σ: 1179.4
27273
d)
μ: 27001.7
σ: 273.1
27273
Counts
Counts
Base Pairs
Base Pairs
20000
30000
20000
30000

Figure 3-5. Capability of Spermine to protect DNA from shearing. A) Lambda DNA in buffer solution (Initial solution). B) DNA molecules after mechanical stresses without any treatment. c) DNA molecules compacted by Spermine tetrahydrochloride (SP^{4+}). D) DNA molecules stretched after applying mechanical stress during C) (scale bar 10 μm).

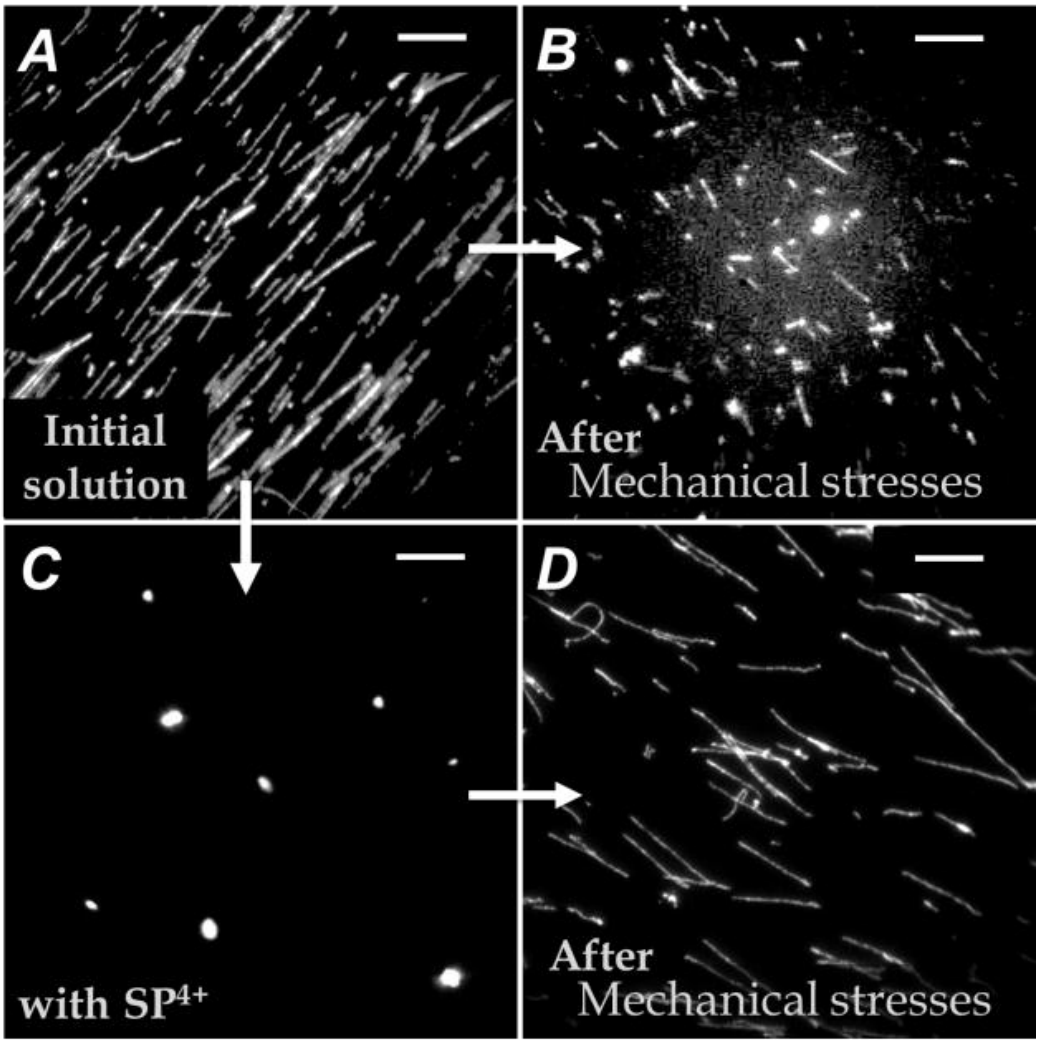

A
B
Initial
solution
After
Mechanical stresses
C
D
with SP4+
After
Mechanical stresses

Figure 3-6. Sequence-specific Methyltransferase-Induced Labeling of DNA (SMILing DNA) for the incorporation of functional groups in DNA. Reaction scheme for DNA biotinylation with the adenine-specific DNA MTase M.BseCI and the aziridine cofactor 6BAz.

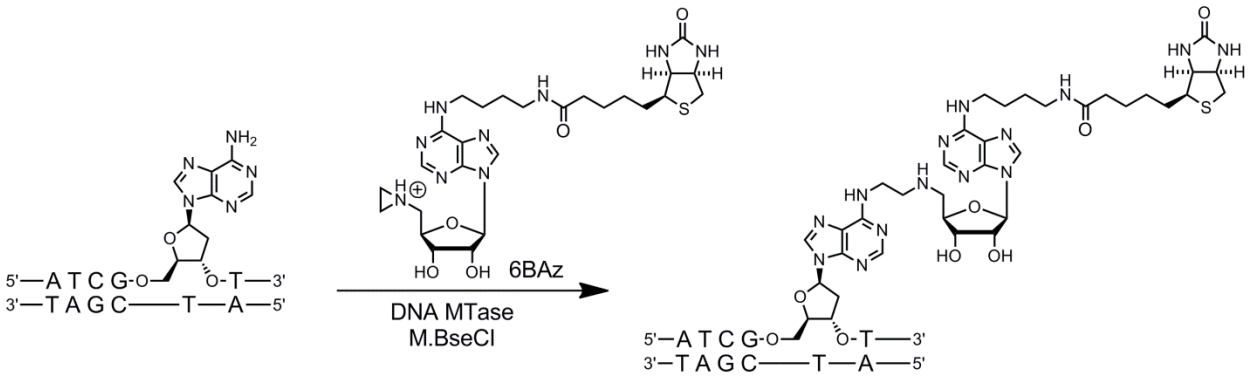

5'—A T C G-O
O-T—3'
3'—T A G C——T—A—5'
6BAz
DNA MTase
M.BseCl

Figure 3-7. Sequence-specific biotinylation of T7 phage DNA using 6BAz and M.BseCI. Labeling was verified by protection against R.ClaI fragmentation and analyzed by agarose gel electrophoresis. Lane 1: T7 DNA incubated with M.BseCI and afterwards with R.ClaI; lane 2: T7 DNA incubated with M.BseCI; lane 3: T7 DNA incubated with 6BAz and afterwards with R.ClaI; lane 4: T7 incubated with 6BAz; lane 5: T7 DNA incubated alone and afterwards with R.ClaI; lane 6: T7 DNA incubated alone; lane 7: T7 DNA incubated with M.BseCI and 6BAz and afterwards with R.ClaI; lane 8: T7 DNA incubated with M.BseCI and 6BAz.

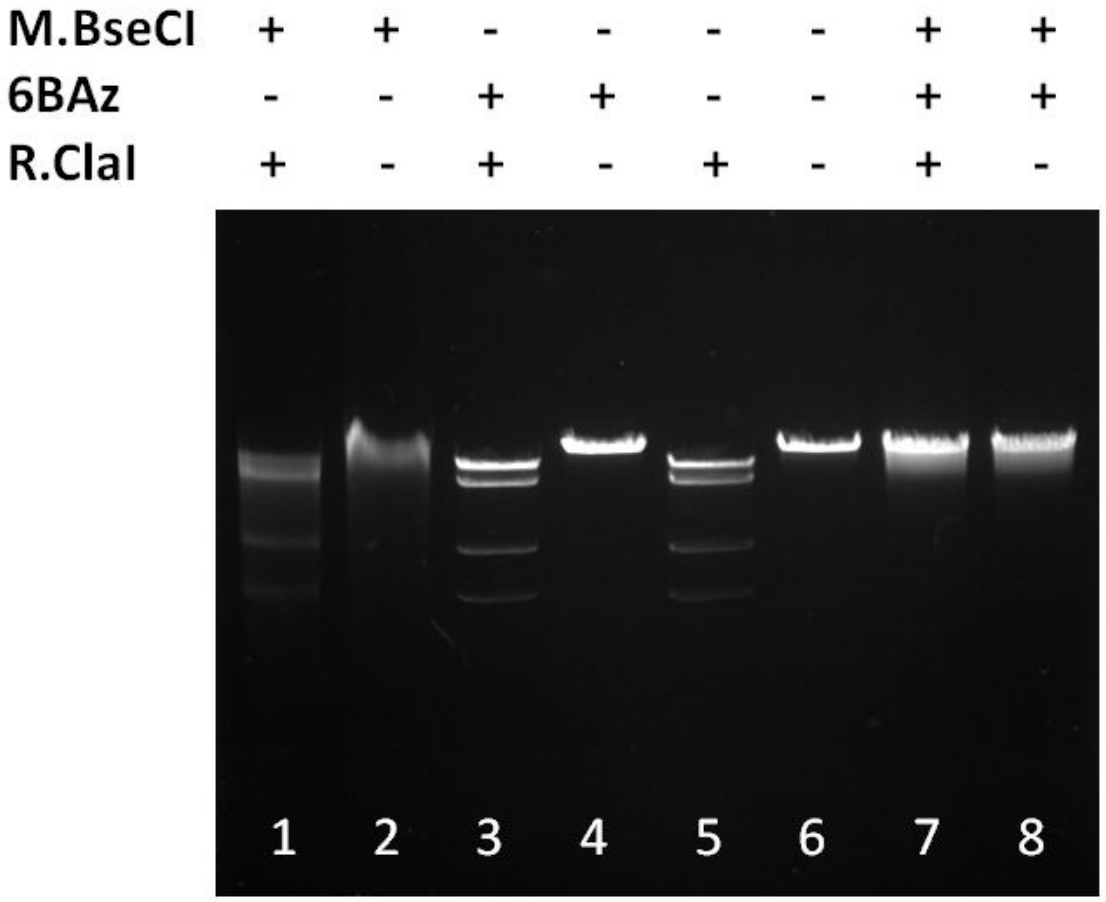

M.BseCI + + - - - - + +
6BAz - - + + - - + +
R.ClaI + - + - + - + -
1 2 3 4 5 6 7 8

Figure 3-8. Channel registration and shift correction for the image used for generating the conversion matrix (left) and a test image acquired separately (right). The top histograms show the pixel difference between the locations of a QD extracted from the different color channels. The average distance is about 1.25 pixels which correspond to about 150 nm. After applying the correction matrix, the average difference between channels is reduced to about 0.2 pixels corresponding to about 30 nm

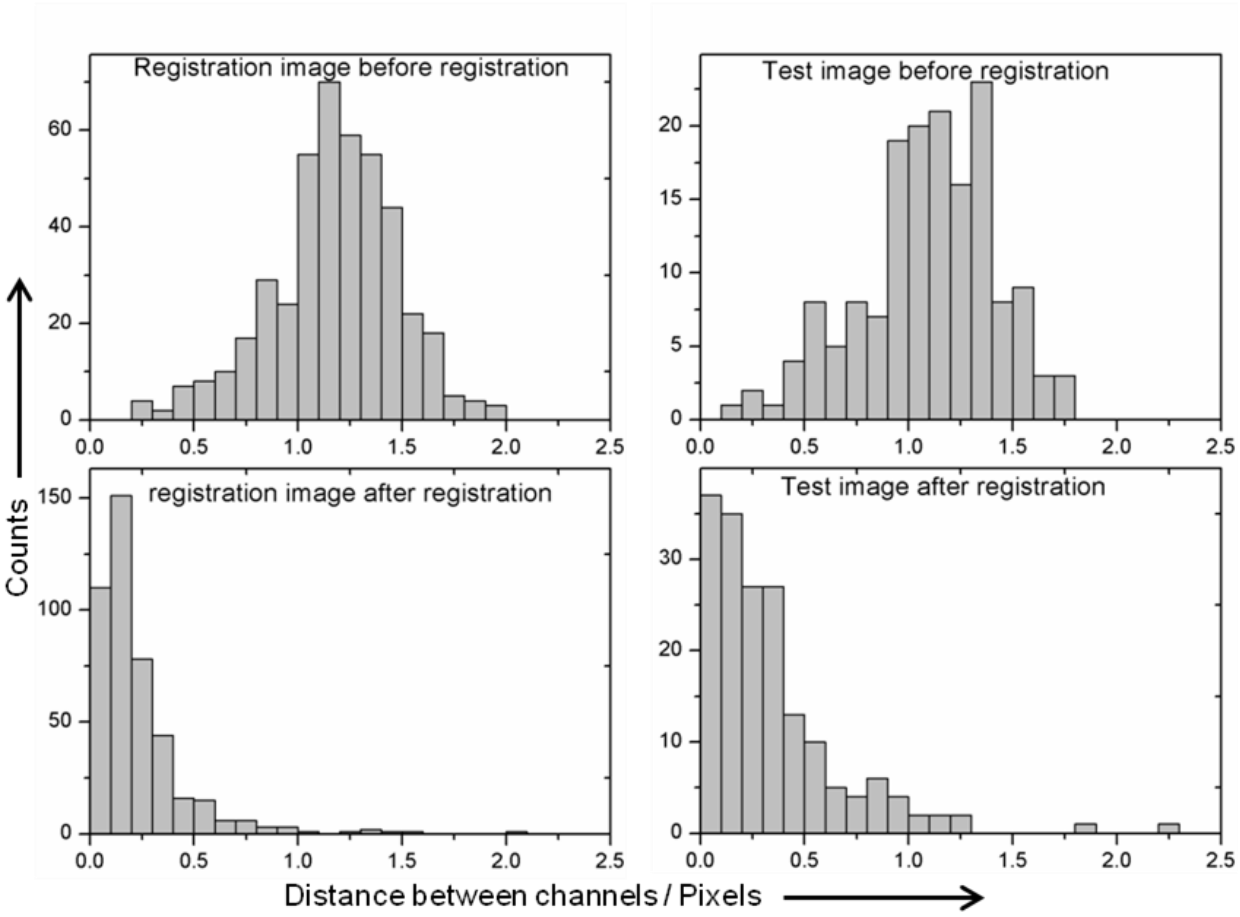

Registration image before registration
Test image before registration
registration image after registration
Test image after registration
Counts
Distance between channels / Pixels

Figure 3-9. Length distribution of extended T7 genomes with 1 (black), 2 (red) and 3 (blue) bound QDs. Stretching factor slightly increases with addition of QDs.

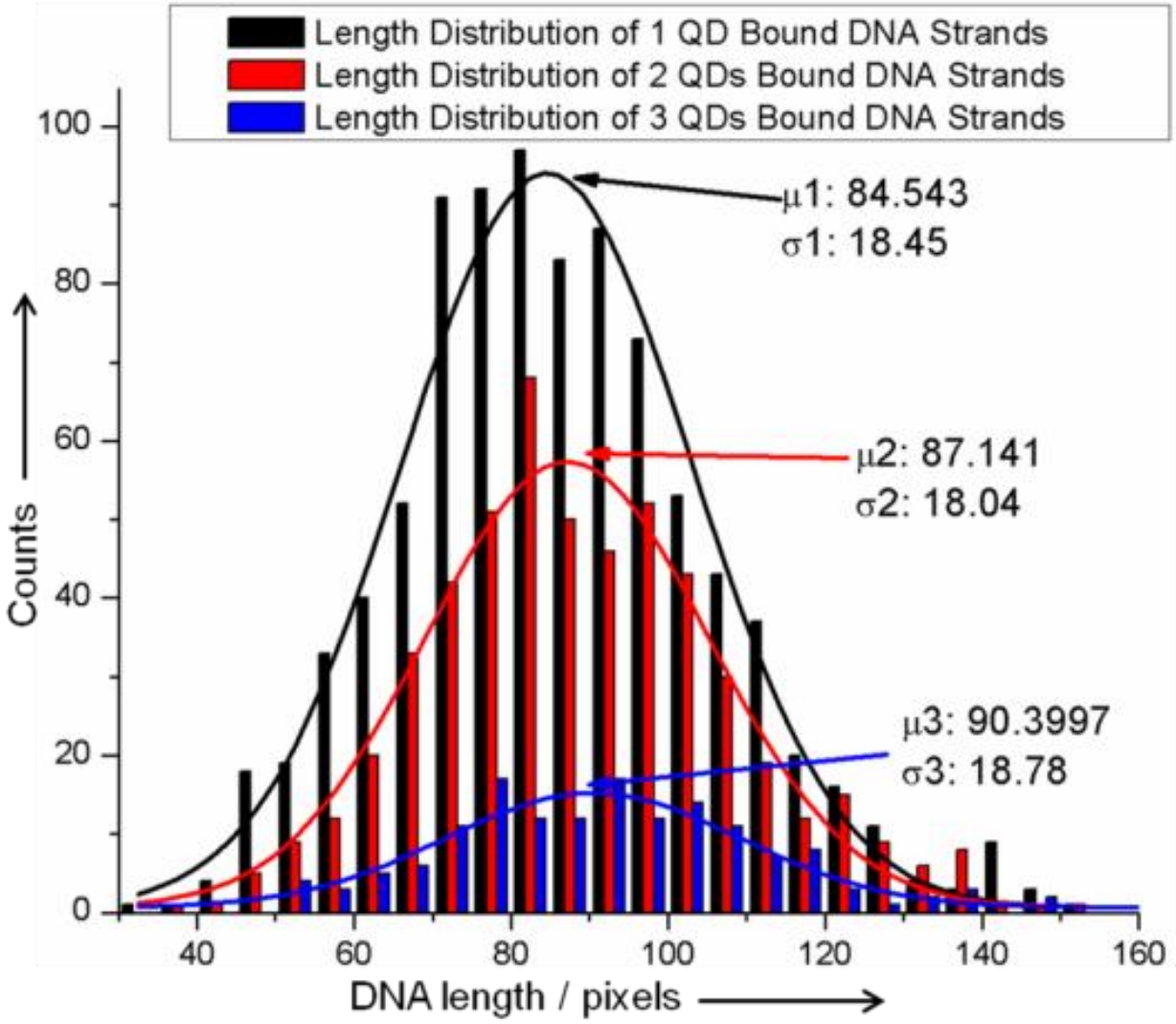

Length Distribution of 1 QD Bound DNA Strands
Length Distribution of 2 QDs Bound DNA Strands
Length Distribution of 3 QDs Bound DNA Strands
μ1: 84.543
σ1: 18.45
μ2: 87.141
σ2: 18.04
μ3: 90.3997
σ3: 18.78
Counts
DNA length / pixels
0
20
40
60
80
100
40
60
80
100
120
140
160

Figure 3-10. Histograms of refTag locations measured manually for a) all observed DNA molecules (N=603 points (329 strands)), b) DNA molecules with 2 or more bound QDs (N=478 (204 strands)) and c) DNA molecules with 3 bound QDs (N=190 (60 strands)).

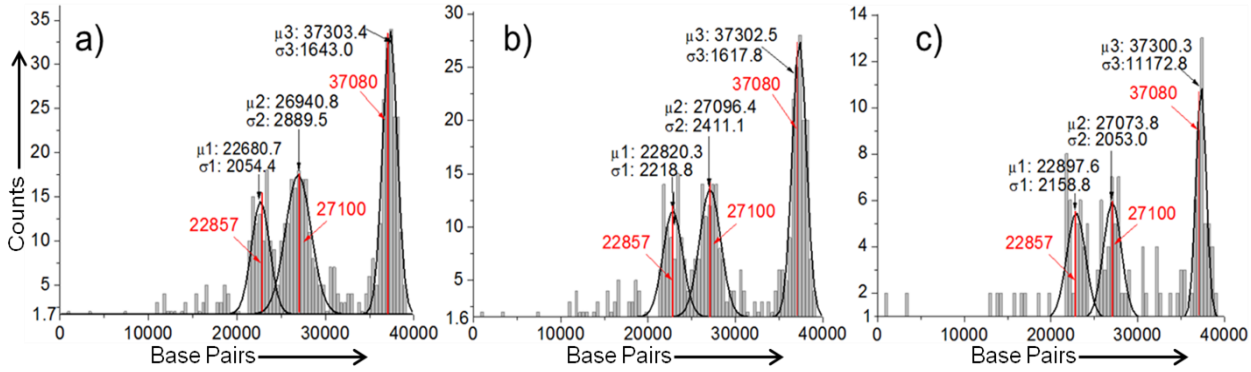

a)
μ3: 37303.4
σ3:1643.0
37080
μ2: 26940.8
σ2: 2889.5
μ1: 22680.7
σ1: 2054.4
22857
27100
Counts
Base Pairs
b)
μ3: 37302.5
σ3:1617.8
37080
μ2: 27096.4
σ2: 2411.1
μ1: 22820.3
σ1: 2218.8
22857
27100
Base Pairs
c)
μ3: 37300.3
σ3:11172.8
37080
μ2: 27073.8
σ2: 2053.0
μ1: 22897.6
σ1: 2158.8
22857
27100
Base Pairs

Figure 3-11. Cropped color overlay images of T7 genomes (white) carrying both refTag QDs (red) and RNAP QDs (green) (scale bar 3 μm).

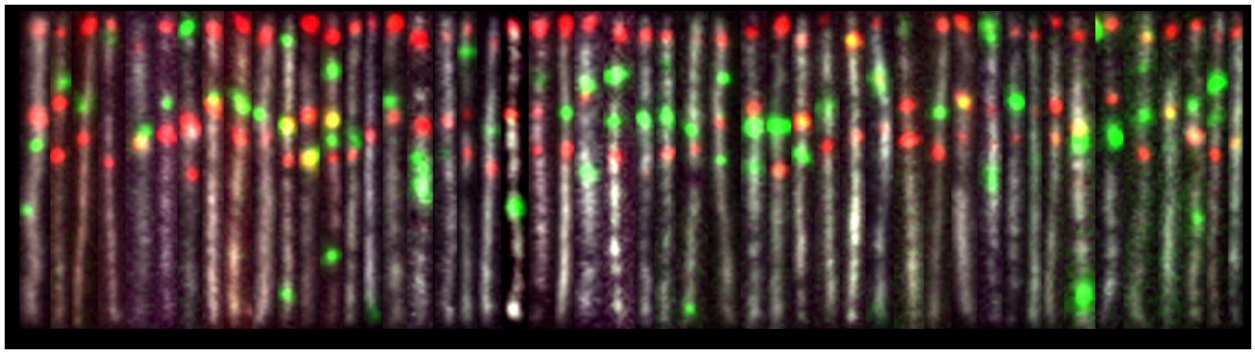
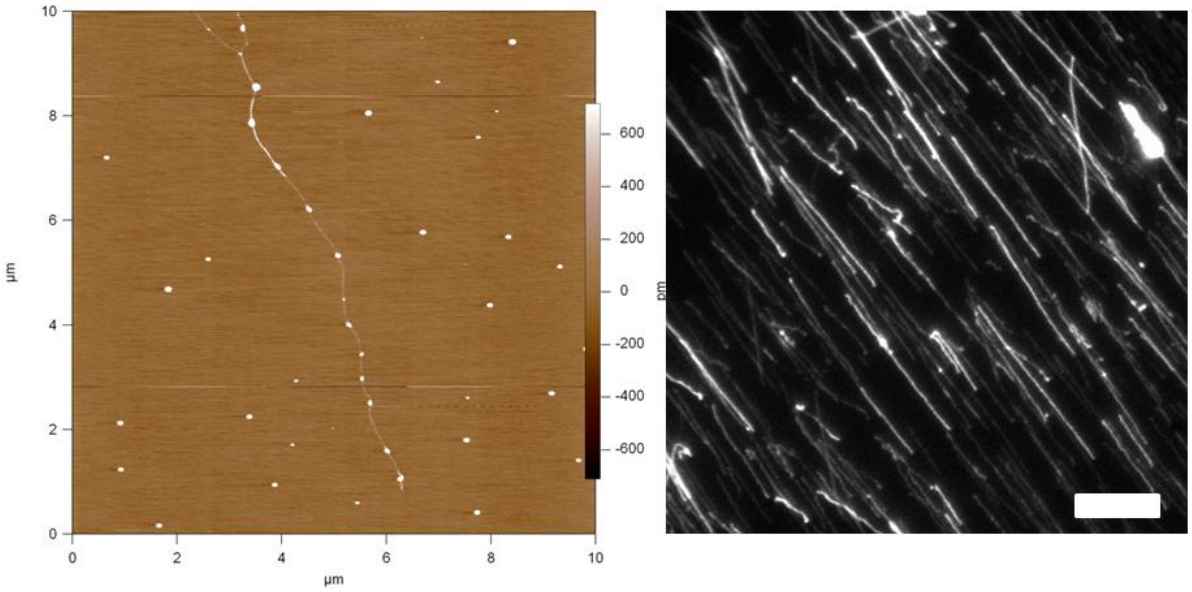


Figure 3-12. Reconstituted chromatin on lambda phage DNA. Left: atomic force microscope (AFM) image of the reconstituted sample showing multiple nucleosomes assembled along the viral genome. Right: fluorescence image of the reconstituted chromatin sampled stained with the intercalating dye YOYO-1 and stretched on a glass coverslip before imaging (scale bar 10 µm).


10
8
6
4
2
0
µm
0
2
4
6
8
10
µm
600
400
200
0
-200
-400
-600
pm

Appendix:

Handling and Characterization of Transcription

Introduction and Background

Transcription is a multi-component system that involves multiple proteins, nucleic acids, and chemicals. Prokaryotic transcription system is much simpler than that of the eukaryotes. However, even the less intricate bacterial *e. coli* transcription machinery and the reconstruction of its functional state *in vitro* can be complicated. This is due to the difficulties of screening the individual components and chemicals that comprises an operational RNAP complex. As no outcome in biology is binary, the comparison should be more-active versus less-active, rather than active versus non-active. The problem is the lack of a reliable high throughput method to compare the activities of different RNAP complex samples. The batch-to-batch variation of the activity and various protein purification methods only adds to the complication.

RNAP transitions into different states as during its transcription cycle:

$$core + \sigma^{70} + DNA \leftrightarrow Holo + DNA \leftrightarrow RPc \rightarrow RPo \rightarrow RPitc \rightarrow RDe$$

where core refers to the RNAP core that comprises of subunits α^2, β, and β'; σ^{70} is the cofactor for the RNAP that helps it recognize and bind to its promoter sequence on the template DNA; Holo is formed from the core binding to the sigma factor; RPc is the closed complex where RNAP Holo enzyme is bound to its promoter; RPo is the open complex where the transcription bubble is melted open; RPitc is the initial transcribing complex where the RNAP is has not yet escaped its promoter, transcribing only short abortive transcripts; and finally RDe is the elongation complex where the RNAP has escaped its promoter after several abortive initiation processes and transcribes the mRNA throughout the gene. Comparison of the RNAP complex has to be done at different stages of the enzymatic state, and can be done through different methodologies. The three processes studied were RPc formation, the transition from RPo to RPitc, and its progression into RDe.

I. **RNAP binding assay.** RNAP complex formation has been assayed using two different methods; native gel assay and ALEX.

a. **Native gel assay.** The most direct way of assaying RNAP's recognition of its promoter is a native gel assay. The signal studied is the mobility of the stained promoter DNA on a 5% polyacrylamide gel. The mobility of the template DNA at a fixed electric potential is reduced upon RNAP binding. The binding efficiency and the number of proteins bound can be estimated by referencing the mobility of the DNA-only control lane (lane 1 & 2, figure A-1 (a)). RNAP core has electrostatic affinity towards the negatively charged DNA, which results in non-specific binding of RNAP to the promoter DNA (lanes 3, 5, 7, 12, figure A-1 (a)). Heparin is a chemical that has a higher affinity towards RNAP than non-specific sequences, but is weaker than the interaction of promoter sequences and the RNAP. The non-specifically bound enzymes can be removed by the addition of heparin in the mixture (figure A-1 (a)). Different RNAP cores have been compared and titrated for most efficient RPo complex formation (figure A-1 (a)).

b. **Single molecule binding assay using ALEX.** ALEX binding assay detects the coincidence of dye labeled DNA with a different color dye labeled RNAP while they diffuse through the confocal excitation volume. By alternating the excitation source faster than the diffusion time scale, each dye can directly be excited during their diffusion to check for the presence of each component. Subpopulations with the S value between 0.25-0.75 confirm that the diffusing molecule is a complex rather than independent components (please refer to the methods section of chapter 2 for details of ALEX[21,24]). The degree of complex formation can be

obtained by the quantitative comparison of S = 0.25-0.75 subpopulations with the rest of the subpopulations in the ES histogram. Native gel assay is an ensemble assay that has a lower limit of ~10 nM of DNA for fluorescence staining and band detection. For equilibrium single molecule assays on diffusing molecules, the concentration has to be lowered by a few orders of magnitude to ensure the occupation of one molecule, at the most, inside the excitation volume. Different concentrations may lead to contradicting results between the ensemble gel assay and the ALEX assay (figure A-1 (a) versus (b) (left & middle)). This discrepancy is a result of the dissociation constant K_d allowing complex formation only at higher concentrations. K_d is a variable dependent to the chemical environment (figure A-2). Single molecule binding assay is a necessity to all single molecule complex studies as the interaction between the individual components of the complex may be too weak at single molecules concentrations.

II. RNAP activity assay. RNAP activity has been checked at two stages; before and after the dissociation from its promoter, abortive initiation assay and elongation assay, respectively.

a. Abortive initiation assays. Abortive initiation assay checks for the enzyme's ability to transcribe abortive transcripts by either the detection of the abortive short transcripts or the distance change of the RNAP with respect to a fixed position on the DNA**.**

i. FDAI (fluorescence detected abortive initiation). FDAI uses the fluorescence signal from dyes labeled on NTPs (γ-AmNS UTP) de-

quenched as the dyes are cleaved upon their uptake by the RNAP for mRNA synthesis. The DNA template construct used (figure A-3 (a)) has TTAAC as their first five transcript sequences. By only adding ApA or ATP and γ-AmNS UTP without the GTP in the reaction mixture, the RNAP complex is unable to transcribe beyond the first four sequences, and thus being forced into abortive initiation. The fluorescence signal will increase with time as more γ-AmNS UTP are used, freeing more de-quenched dyes in the solution. Comparing the slope between samples in the fluorescence intensity versus time graph will provide a relative activity of the RNAP complex with respect to the control that has no template DNA in the mixture (figure A-3 (b)).

ii. **LE FRET assay (leading edge FRET).** RNAP activity can be assayed by the distance change between two labeled points of the RPitc complex induced by the abortive initiation. The DNA construct has an acceptor dye on the non-template strand 15 nucleotide downstream of the transcription start site and the σ^{70} has a donor dye at the 366 amino acid position (figure A-4). The sequences of the DNA template has been engineered such that the RNAP can be stalled at different positions on the DNA by depriving a certain NTP in the reaction mixture: ATP only results in RPitc+2, and A/U/GTP results in RPitc+7, where RPitc+2 refers to RNAP initial transcribing complex stalled 2 nucleotide sequences downstream from the transcription start site. As the DNA is reeled into the RNAP during abortive initiation, the reduced distance between the labeled positions will

cause the FRET efficiency to increase (figure A-5). The addition of the full set of NTPs, A/U/G/CTP, will cause the RNAP to run off from its template and the subpopulations with S values between 0.25 and 0.75 will decrease as the RNAP complex will dissociate into free DNA and free RNAP. At the single molecule concentration (~100pM), RNAP will rarely rebind to DNA for another round of abortive initiation as the concentration is below the dissociation constant.

III. **Elongation assays.** Elongation assay is done by the detection of the mRNA transcript through fluorescence readouts or through the detection of the formation of the mRNA/DNA duplex by the hybridization of the mRNA transcript to its complementary probe DNA.

i. **FDE (fluorescence detected elongation).** The concept is identical to FDAI, however, with all the NTPs added in the mixture. Rates of abortive transcription and elongation can be compared by evaluating the slope of the fluorescence versus time graph between the two transcription states (figure A-3). The activity of the RNAP complex with different σ^{70} (wt σ^{70} versus σ^{70} labeled with a dye on amino acid sequence 366) was compared (figure A-6).

ii. **smTranscription assay (Methods and concept has been detailed in chapter 2).** While FDE is an ensemble measurement, single molecule transcription assay (smTranscription assay) is a method that can report percentage of "active" RNAP in a given batch. FDE requires high concentration of RNAP complex for sufficient fluorescence signal, thus

allowing multiple rounds of transcription. smTranscription assay works in the single molecule concentration where RNAP will rarely rebind to the DNA after having "run off". Additionally, the DNA probe will only hybridize to transcripts resulting from RDe transcription since the DNA construct has a 20dA "track" after 15 sequences downstream of the transcription start site (figure A-7). Different batches of RNAP have been assayed for activity comparison (table A-1).

Methods

Native gel. On a 5% polyacrylamide gel 20 ng of DNA samples with and without RNAP complex (see chapter 2 methods for RNAP complex formation reaction and buffers) with final volume of 10ul were loaded and run for 4 - 6 hours at 100 mV. After the run the gel was stained with Sybr green I by incubating and rocking in Sybr green I solution for 30 minutes. The gel was washed with water then imaged under gel imager with appropriate excitation (488 nm) and filters.

FDAI & FDE. RNAP Holo complex was formed by mixing 7.4 µl of TB buffer (see chapter 2 methods for TB), 5.6 µl 1.11 µM core, and 2.1 µl 6 µM σ^{70} and incubating the mixture for 20 minutes at 30 C. The RNAP reaction mix was prepared by adding and incubating for 10 minutes at 37 C 43.75 µl TB, 17.25 µl H_2O, 10 µl 1mM G/A/CTP (or ApA), 1 µl 5mM γ-AmNS UTP, and 15 µl of previously prepared RPo. DNA reaction mix was prepared separately by adding 7.5 µl TB, 3.6 µl DNA, and 4 µl H_2O followed by 37 C incubation for 10 minutes. The RNAP reaction mix and DNA reaction mix were added together in a cuvette and placed in a fluorometer at 37 C. The fluorescence signal was gathered at 500 nm by exciting the mixture at 360 nm.

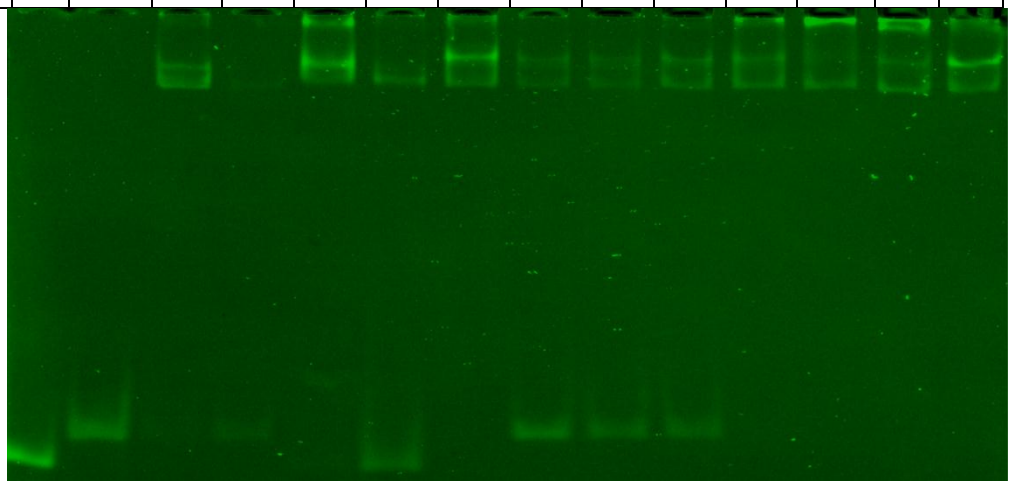
LE FRET. We incubated a mix of 1.6 µl of commercially purified 1.111 µM RNAP core (Epicentre), 0.4 µl 366-Cy3B-σ^{70}, and 0.6 µl of 1 µM template DNA (labeled with Atto647 at +15 downstream of the transcription start site) at 37 °C in 26 µl of 50 mM Tris HCl, 100 mM KCl, 10 mM MgCl2, 1 mM DTT, 100 µg ml−1 BSA and 5% glycerol (pH 8.0) for 15 min to form the RNAP open complex (RPO). This mixture was challenged with heparin sepharose resin as described previously to eliminate nonspecific complexes. The resulting concentration of RPo was 20nM and was diluted in KG7 to 1nM. Right before taking a measurement 1 µl RPo was added to 1 µl 1mM NTP (UTP for RPitc4, U/GTP for RPtic7, and U/A/G/C for RDe) and 8 µl KG7. The ALEX measurement was done at 37 C for 10 minutes with excitation powers 160 µW 532 nm and 80 µW 635 nm (see chapter 2 for details of ALEX).

smTranscription. (Details in chapter 2)

Figure A-1. RNAP binding assay. (a) Native gel assay with reactions done in buffer 1. The band signals are DNA stained with SyBr green I. DNA 1 is control 49 bp oligomer without the promoter; DNA2 is the promoter DNA, core1 and 2 are cores prepared by different people in the lab at different time points. (b) (left & middle) ALEX binding assay done on samples used in (a). (b) (right) ALEX binding assay done on a sample with different buffer.

a)

DNA1	+	-	-	-	+	+	-	-	-	-	-	-	-	-
DNA2	-	+	+	+	-	-	+	+	+	+	+	+	+	+
Core1	-	-	+	+	+	+	2µl	2µl	3µl	4µl	6µl	8µl	-	-
Core2	-	-	-	-	-	-	-	-	-	-	-	-	+	+
366-Cy3B-σ^{70}	-	-	-	-	+	+	+	+	+	+	+	+	+	+
Heparin	-	-	-	+	-	+	-	+	+	+	+	+	-	+



b)

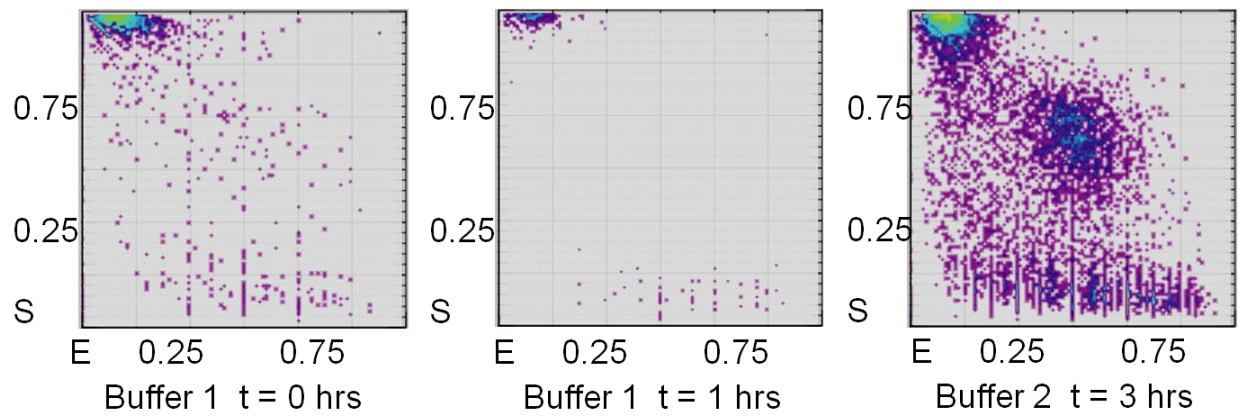


Figure A-2. Single molecule RNAP binding assay using ALEX. ALEX measurements on RPo complexes diluted in (a) $KG7_1$ with hepes (RPI), (b) TB, (c) T50, and (d) $KG7_2$ (Fisher). Samples were measured for 10 minutes within one hour after dilution. Buffer recipes are described in chapter 2.

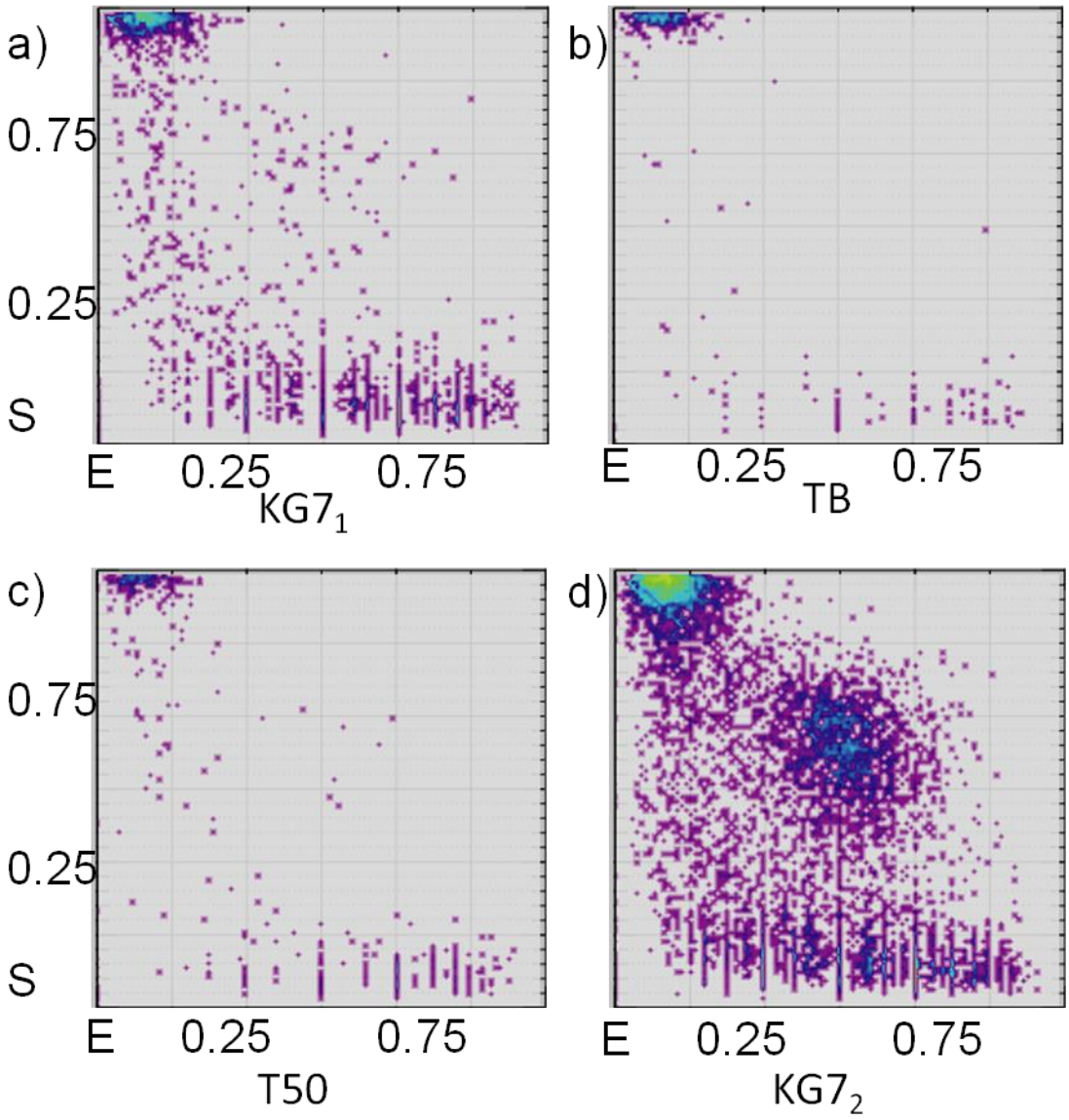
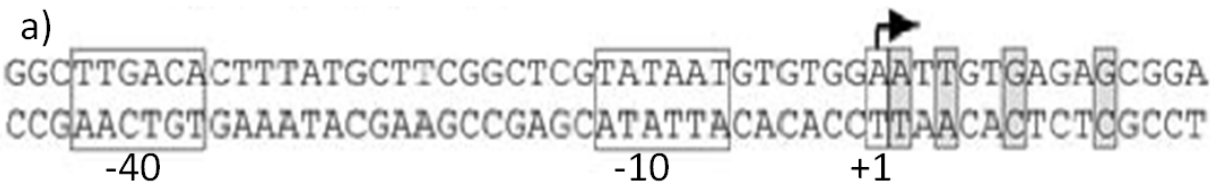

a)
0.75
0.25
S
E
0.25
0.75
KG7$_1$
b)
E
0.25
0.75
TB
c)
0.75
0.25
S
E
0.25
0.75
T50
d)
E
0.25
0.75
KG7$_2$

Figure A-3. FDAI. (a) DNA promoter construct with the template DNA as the bottom strand. Transcription sequences engineered for RPitc stall sites on the DNA. Adding incomplete set of NTPs will prevent the RNAP from dissociating from the promoter to transition into elongation complex. (b) Fluorescence signals increased as dyes are de-quenched and released into the solution as transcription proceeds. Transcription rates are determined by the comparison of the slopes. Abortive cycle rates can be compared to elongation rates (green and brown versus yellow). A control sample has no DNA template in the mixture (red).

a)


GGCTTGACACTTTATGCTTCGGCTCGTATAATGTGTGGAATTGTGAGAGCGGA
CCGAACTGTGAAATACGAAGCCGAGCATATTACACACCTTAACACTCTCGCCT
-40
-10
+1

b)

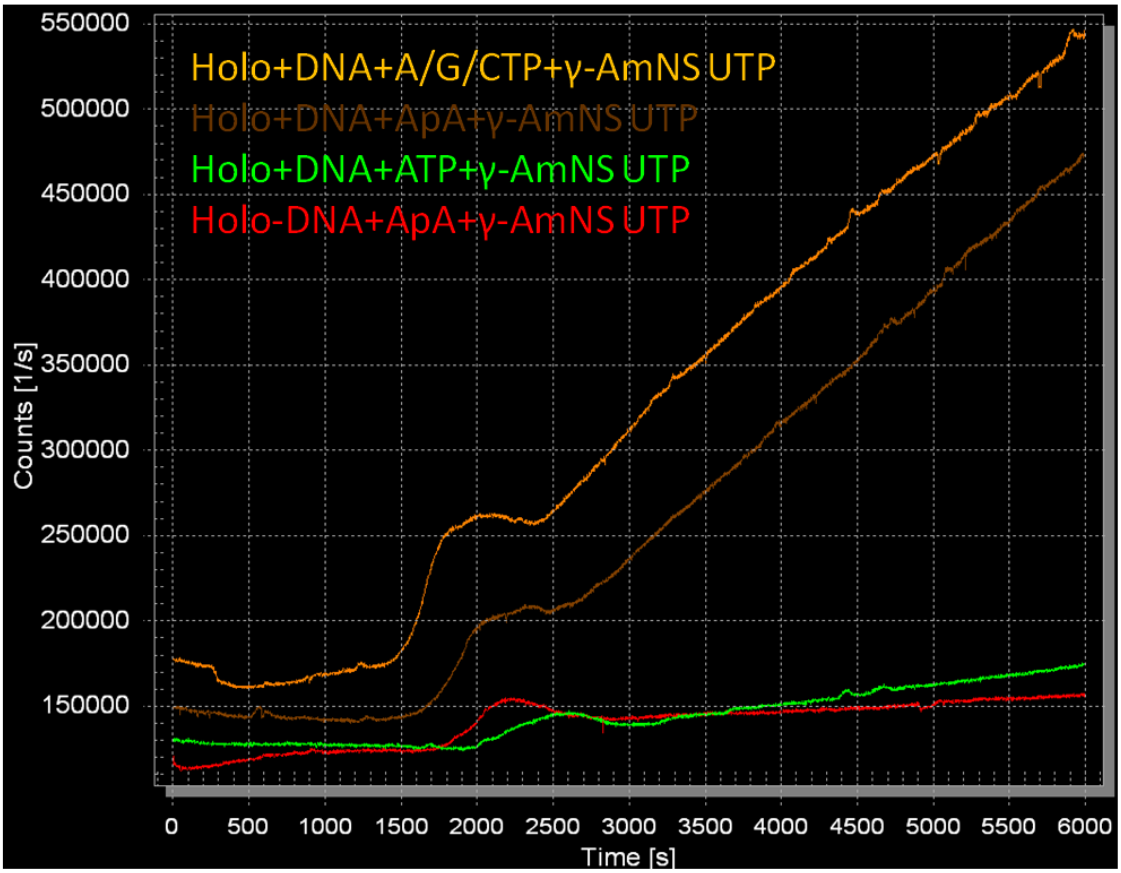

Holo+DNA+A/G/CTP+γ-AmNS UTP
Holo+DNA+ApA+γ-AmNS UTP
Holo+DNA+ATP+γ-AmNS UTP
Holo-DNA+ApA+γ-AmNS UTP
Counts [1/s]
550000
500000
450000
400000
350000
300000
250000
200000
150000
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
Time [s]

Figure A-4. Labeled DNA template construct and RNAP complex. LacCONS promoter is labeled with Atto647 acceptor dye at +15 position on the non-template strand. The DNA template engineered to stall RNAP at different positions as a function of NTPs added (ATP-RPitc2, A/UTP-RPitc4, ApA/U/GTP-RPitc7, A/U/G/CTP-RDe). The σ^{70} is labeled with Cy3B donor dye at 366 amino acid sequence that will show increase in FRET efficiency as the RPo transitions into RPitc4/RPitc7.

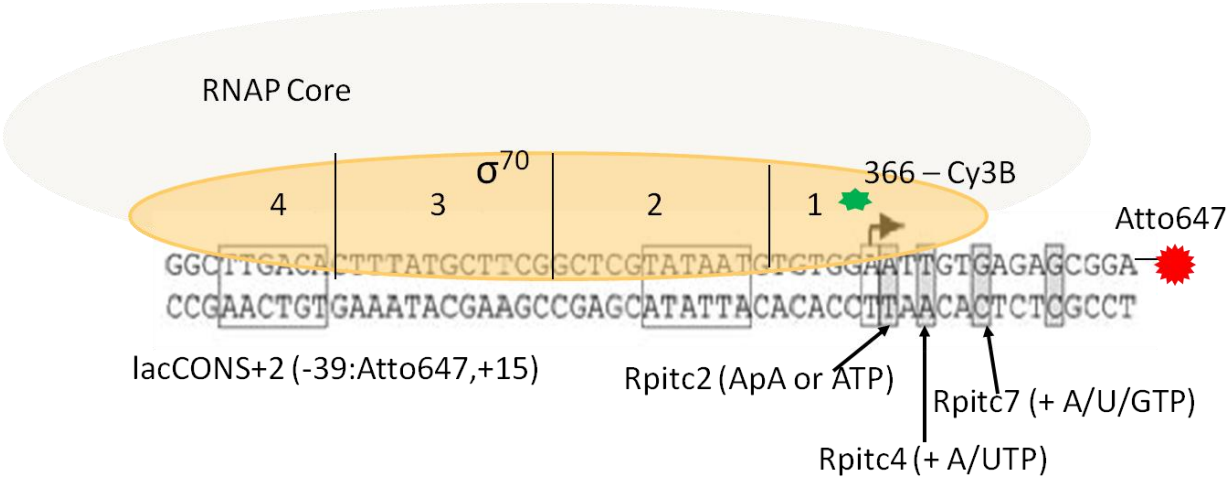

RNAP Core
σ70
4
3
2
1
366 – Cy3B
Atto647
lacCONS+2 (-39:Atto647,+15)
Rpitc2 (ApA or ATP)
Rpitc4 (+ A/UTP)
Rpitc7 (+ A/U/GTP)

Figure A-5. LE ALEX assays. FRET efficiency (E*) between donor dye on 366 position of the σ^{70} and acceptor dye on +15 position on the non-template strand of the promoter DNA as a function of NTPs added. As the RPitc2 transitions into +4/+7 the FRET efficiency increases, suggesting decrease in distance between the FRET dye pair. When all the NTPs have been added the complex dissociates, leading to the diminish of the subpopulation of S values 0.25-0.75.

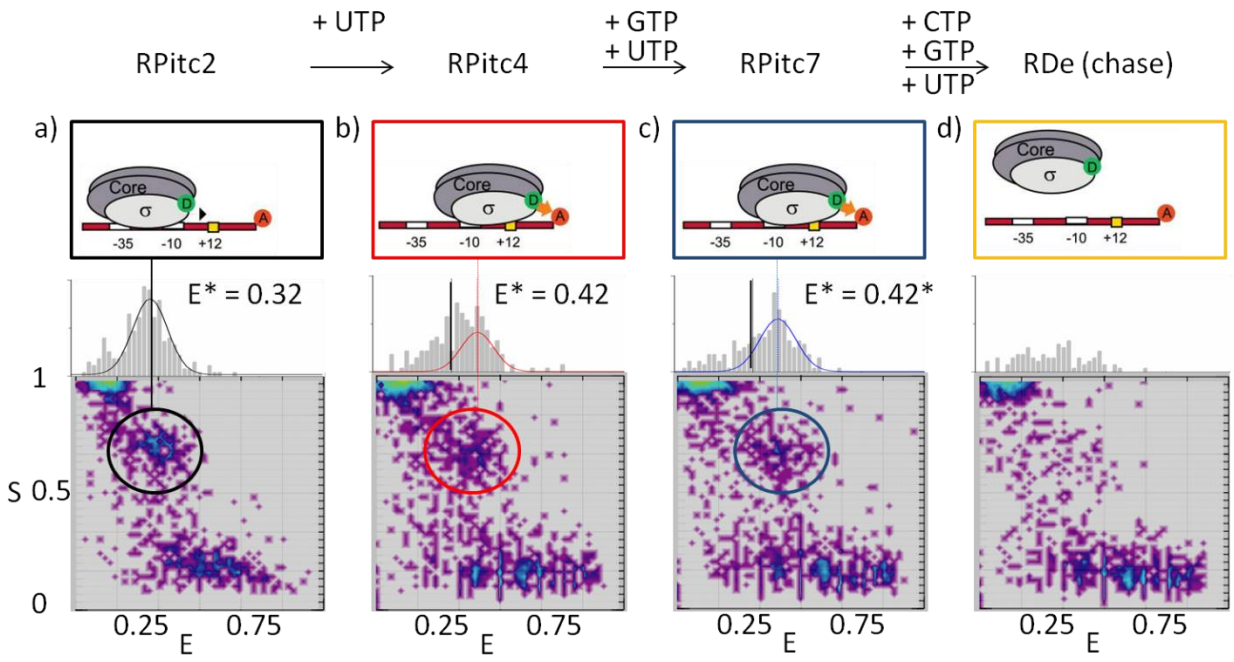

RPitc2
+ UTP
RPitc4
+ GTP
+ UTP
RPitc7
+ CTP
+ GTP
+ UTP
RDe (chase)
a)
b)
c)
d)
Core
σ
D
A
-35
-10
+12
E* = 0.32
E* = 0.42
E* = 0.42*
1
S 0.5
0
0.25
0.75
E

Figure A-6. FDE. Transcription rates at elongation mode of different RNAP complexes are compared; RNAP complex with wtσ^{70} (red), 366-cy3Bσ^{70} (green), and control with no σ^{70} (yellow).

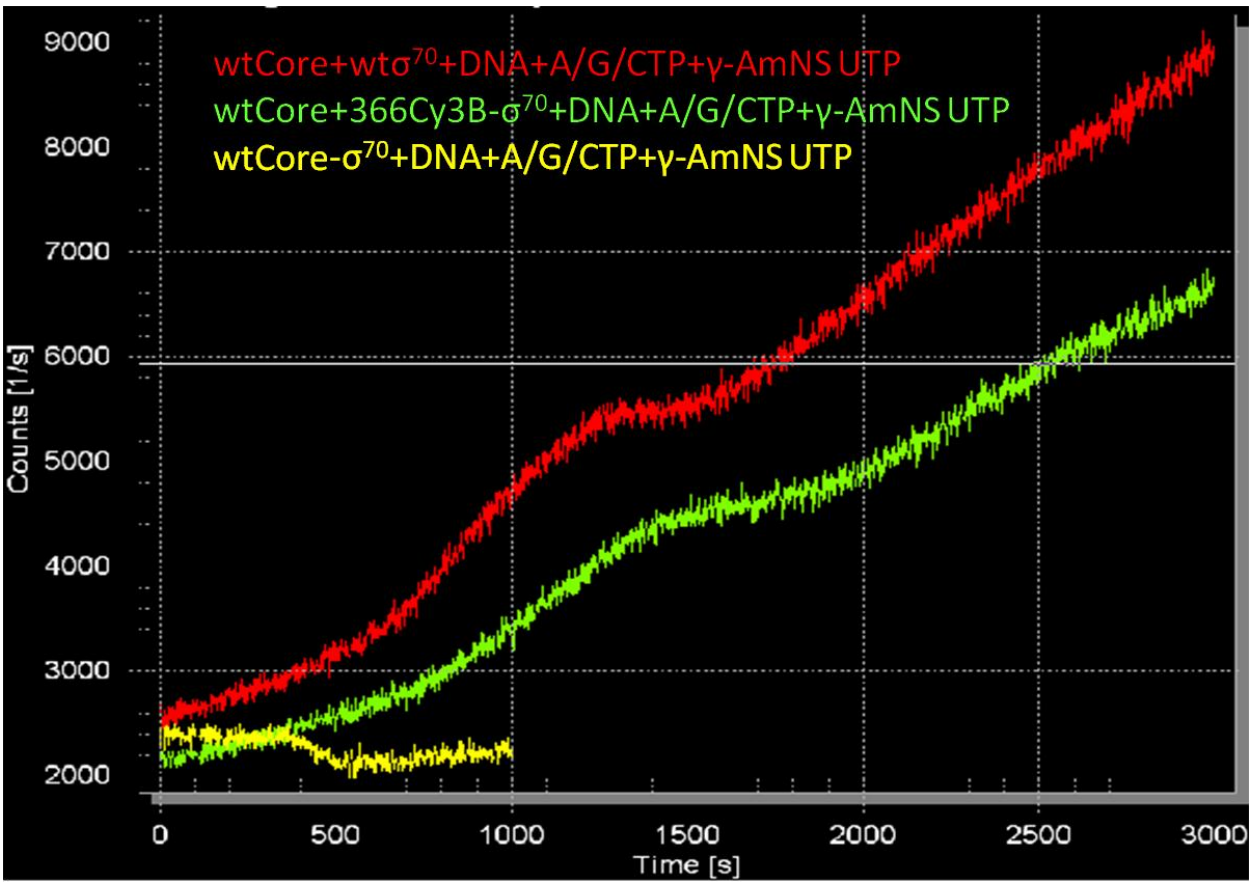
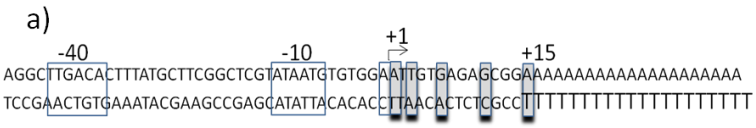
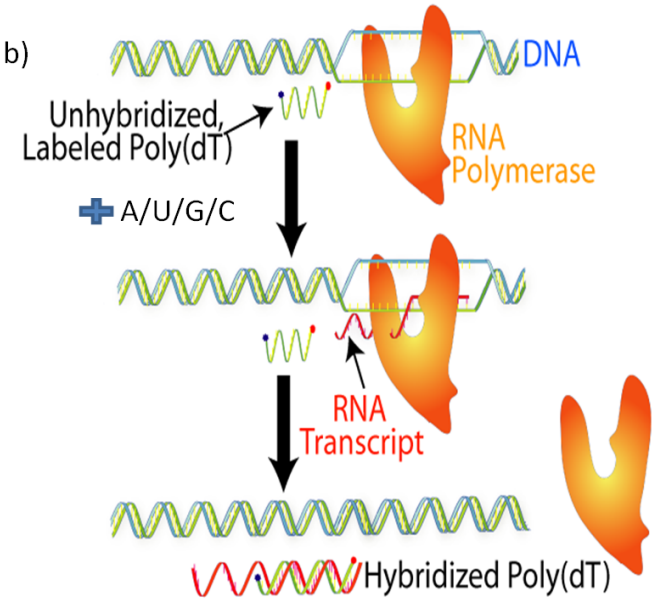

wtCore+wtσ70+DNA+A/G/CTP+γ-AmNS UTP
wtCore+366Cy3B-σ70+DNA+A/G/CTP+γ-AmNS UTP
wtCore-σ70+DNA+A/G/CTP+γ-AmNS UTP
Counts [1/s]
9000
8000
7000
6000
5000
4000
3000
2000
0
500
1000
1500
2000
2500
3000
Time [s]

Figure A-7. smTranscription assay. (a) LacCONS promoter DNA was engineered to have 20 consecutive A's after +15 position of the non-template strand. The mRNA transcribed by elongation hybridizes to the ssDNA probe that is a 20dT oligomer labeled with a donor dye and an acceptor dye at the ends. The mRNA/DNA duplex will shift the FRET between the two FRET dye pair from high FRET to low FRET. (b) The schematic of the assay. (c) RNAP activity, not transcription rate, is reported by the ratio between the high FRET population and low FRET population. As more RNAP complexes are added in the mixture the high FRET peak diminishes while the low FRET peak emerges.


a)
-40
-10
+1
+15
AGGCTTGACACTTTATGCTTCGGCTCGTATAATGTGTGGAATTGTGAGAGCGGAAAAAAAAAAAAAAAAAAAAA
TCCGAACTGTGAAATACGAAGCCGAGCATATTACACACCTTAACACTCTCGCCTTTTTTTTTTTTTTTTTTTTT


b)
DNA
Unhybridized,
Labeled Poly(dT)
RNA
Polymerase
A/U/G/C
RNA
Transcript
Hybridized Poly(dT)

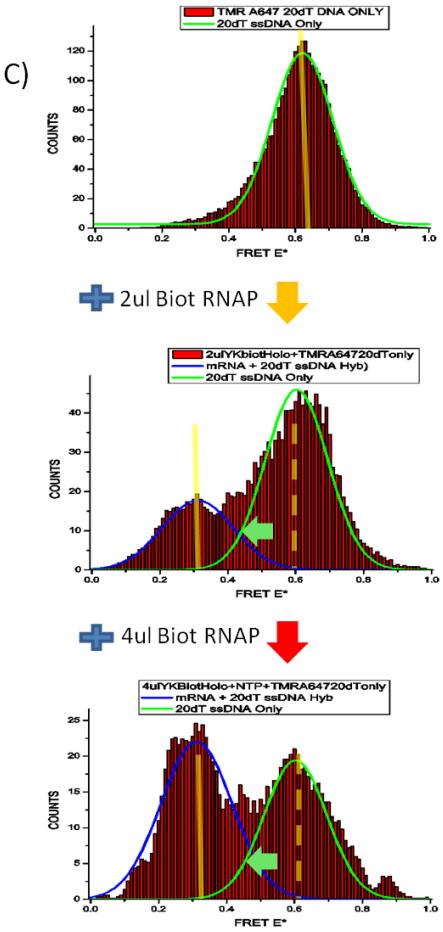

C)
TMR A647 20dT DNA ONLY
20dT ssDNA Only
COUNTS
FRET E*
2ul Biot RNAP
2ulYKbiotHolo+TMRA64720dTonly
mRNA + 20dT ssDNA Hyb)
20dT ssDNA Only
4ul Biot RNAP
4ulYKBiotHolo+NTP+TMRA64720dTonly
mRNA + 20dT ssDNA Hyb
20dT ssDNA Only

Table A-1. RNAP activity comparison. The activities of different batches of RNAP were compared using smTranscription assay. Refer to chapter 2 for activity calculation and details of the method.

Complex	dsDNS Ratio *R*	[ssDNA Probe]	[RPo]	ACTIVITY (%)
4ul YK Biotin Holo+705QD	0.488111	100 pM	540 pM	9.03%
4ul YK Biotin Holo	0.562892	100 pM	600 pM	9.38%
2ul YK Biotin Holo	0.304573	100 pM	300 pM	10.15%
2ul EPICENTRE Holo	0.643339	100 pM	200 pM	32.17%
1ul EPICENTRE Holo	0.569694	70 pM	100 pM	39.879%
3ul YK Biotin Core + $_{wt}\sigma^{70}$	0.388515	100 pM	800 pM	4.85%
1 ul YK Biotin Core + $_{wt}\sigma^{70}$	0.158598	100 pM	400pM	3.96%

References

1. Meyer, K.D., Lin, S.-C., Bernecky, C., Gao, Y. & Taatjes, D.J. p53 activates transcription by directing structural shifts in Mediator. *Nature structural & molecular biology* **17**, 753-60 (2010).

2. Herbert, K.M. *et al.* E. coli NusG Inhibits Backtracking and Accelerates Pause-Free Transcription by Promoting Forward Translocation of RNA Polymerase. *Journal of molecular biology* **399**, 30-17 (2010).

3. Melin, J. & Quake, S.R. Microfluidic large-scale integration: the evolution of design rules for biological automation. *Annual review of biophysics and biomolecular structure* **36**, 213-31 (2007).

4. Kim, S. *et al.* High-throughput single-molecule optofluidic analysis. *Nat Meth* **advance on**, (2011).

5. Ebenstein, Y. *et al.* Lighting Up Individual DNA Binding Proteins with Quantum Dots. *Nano Letters* **9**, 1598-1603 (2009).

6. Ebenstein, Y., Gassman, N., Kim, S. & Weiss, S. Combining atomic force and fluorescence microscopy for analysis of quantum-dot labeled protein–DNA complexes. *Journal of Molecular Recognition* **22**, 397-402 (2009).

7. Pljevaljcić, G., Schmidt, F. & Weinhold, E. Sequence-specific methyltransferase-induced labeling of DNA (SMILing DNA). *Chembiochem : a European journal of chemical biology* **5**, 265-9 (2004).

8. Kim, S. *et al.* Enzymatically Incorporated Genomic Tags for Optical Mapping of DNA-Binding Proteins. *Angewandte Chemie International Edition* n/a-n/a (2012).doi:10.1002/anie.201107714

9. Maxwell, K.L. *et al.* Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein science : a publication of the Protein Society* **14**, 602-16 (2005).

10. Ha, T. Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences* **93**, 6264-6268 (1996).

11. Melin, J. & Quake, S.R. Microfluidic large-scale integration: the evolution of design rules for biological automation. *Annual review of biophysics and biomolecular structure* **36**, 213-31 (2007).

12. Squires, T. & Quake, S. Microfluidics: Fluid physics at the nanoliter scale. *Reviews of Modern Physics* **77**, 977-1026 (2005).

13. Hamadani, K.M. & Weiss, S. Nonequilibrium Single Molecule Protein Folding in a Coaxial Mixer. *Biophys. J.* **95**, 352-365 (2008).

14. Hertzog, D.E. Femtomole Mixer for Microsecond Kinetic Studies of Protein Folding. *Anal. Chem.* **76**, 7169-7178 (2004).

15. Lipman, E.A., Schuler, B., Bakajin, O. & Eaton, W.A. Single-Molecule Measurement of Protein Folding Kinetics . *Science* **301** , 1233-1235 (2003).

16. Pfeil, S.H., Wickersham, C.E., Hoffmann, A. & Lipman, E.A. A microfluidic mixing system for single-molecule measurements. *The Review of scientific instruments* **80**, 055105 (2009).

17. Vandelinder, V., Ferreon, A.C.M., Gambin, Y., Deniz, A.A. & Groisman, A. High-Resolution Temperature−Concentration Diagram of α-Synuclein Conformation Obtained from a Single Förster Resonance Energy Transfer Image in a Microfluidic Device. *Analytical Chemistry* **81**, 6929-6935 (2009).

18. Lemke, E.A. *et al.* Microfluidic Device for Single-Molecule Experiments with Enhanced Photostability. *Journal of the American Chemical Society* **131**, 13610-13612 (2009).

19. Hansen, C.L., Sommer, M.O.A. & Quake, S.R. Systematic investigation of protein phase behavior with a microfluidic formulator . *Proceedings of the National Academy of Sciences of the United States of America* **101** , 14431-14436 (2004).

20. Ridgeway, W.K., Seitaridou, E., Phillips, R. & Williamson, J.R. RNA–protein binding kinetics in an automated microfluidic reactor . *Nucleic Acids Research* **37** , e142-e142 (2009).

21. Kapanidis, A.N. *et al.* Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules . *Proceedings of the National Academy of Sciences of the United States of America* **101** , 8936-8941 (2004).

22. Gralla, J.D. & Huo, Y.-X. Remodeling and activation of Escherichia coli RNA polymerase by osmolytes. *Biochemistry* **47**, 13189-96 (2008).

23. Colyer, R.A. *et al.* High-throughput multispot single-molecule spectroscopy. *Proceedings - Society of Photo-Optical Instrumentation Engineers* **7571**, 75710G-75710G11 (2010).

24. Kapanidis, A.N. *et al.* Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science (New York, N.Y.)* **314**, 1144-7 (2006).

25. Unger, M.A., Chou, H.-P., Thorsen, T., Scherer, A. & Quake, S.R. Monolithic Microfabricated Valves and Pumps by Multilayer Soft Lithography . *Science* **288** , 113-116 (2000).

26. Zimm, B.H., Roe, G.M. & Epstein, L.F. Solution of a Characteristic Value Problem from the Theory of Chain Molecules. *The Journal of Chemical Physics* **24**, 279 (1956).

27. Allawi, H.T. & SantaLucia, J. Thermodynamics and NMR of Internal G·T Mismatches in DNA. *Biochemistry* **36**, 10581-10594 (1997).

28. Owczarzy, R. *et al.* Effects of Sodium Ions on DNA Duplex Oligomers: Improved Predictions of Melting Temperatures. *Biochemistry* **43**, 3537-3554 (2004).

29. Sugimoto, N. *et al.* Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes. *Biochemistry* **34**, 11211-11216 (1995).

30. Erickson, D. Modeling of DNA hybridization kinetics for spatially resolved biochips. *Analytical Biochemistry* **317**, 186-200 (2003).

31. Majumdar, D.S. *et al.* Single-molecule FRET reveals sugar-induced conformational dynamics in LacY . *Proceedings of the National Academy of Sciences* **104** , 12640-12645 (2007).

32. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nature biotechnology* **26**, 1135-45 (2008).

33. Neely, R.K., Deen, J. & Hofkens, J. Optical mapping of DNA: single-molecule-based methods for mapping genomes. *Biopolymers* **95**, 298-311 (2011).

34. Meng, X., Benson, K., Chada, K., Huff, E.J. & Schwartz, D.C. Optical mapping of lambda bacteriophage clones using restriction endonucleases. *Nature genetics* **9**, 432-8 (1995).

35. Michalet, X. Dynamic Molecular Combing: Stretching the Whole Human Genome for High-Resolution Studies. *Science* **277**, 1518-1523 (1997).

36. Chan, E.Y. *et al.* DNA mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags. *Genome research* **14**, 1137-46 (2004).

37. Xiao, M. *et al.* Rapid DNA mapping by fluorescent single molecule detection. *Nucleic acids research* **35**, e16 (2007).

38. Das, S.K. *et al.* Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. *Nucleic acids research* **38**, e177 (2010).

39. Reisner, W. *et al.* Single-molecule denaturation mapping of DNA in nanofluidic channels. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 13294-9 (2010).

40. Zohar, H., Hetherington, C.L., Bustamante, C. & Muller, S.J. Peptide nucleic acids as tools for single-molecule sequence detection and manipulation. *Nano letters* **10**, 4697-701 (2010).

41. Neely, R.K. *et al.* DNA fluorocode: A single molecule, optical map of DNA with nanometre resolution. *Chemical Science* **1**, 453 (2010).

42. Narzisi, G. & Mishra, B. Comparing de novo genome assembly: the long and short of it. *PloS one* **6**, e19175 (2011).

43. Narzisi, G. & Mishra, B. Scoring-and-unfolding trimmed tree assembler: concepts, constructs and comparisons. *Bioinformatics (Oxford, England)* **27**, 153-60 (2011).

44. Aston, C., Mishra, B. & Schwartz, D.C. Optical mapping and its potential for large-scale sequencing projects. *Trends in Biotechnology* **17**, 297-302 (1999).

45. Caburet, S. *et al.* Human ribosomal RNA gene arrays display a broad range of palindromic structures. *Genome research* **15**, 1079-85 (2005).

46. Kidd, J.M. *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56-64 (2008).

47. Zohar, H. & Muller, S.J. Labeling DNA for single-molecule experiments: methods of labeling internal specific sequences on double-stranded DNA. *Nanoscale* **3**, 3027-39 (2011).

48. Pljevaljčić, G., Schmidt, F. & Weinhold, E. Sequence-specific Methyltransferase-Induced Labeling of DNA (SMILing DNA). *ChemBioChem* **5**, 265-269 (2004).

49. Wilkinson, S. *et al.* Molecular scale architecture: engineered three- and four-way junctions. *Bioconjugate chemistry* **19**, 470-5 (2008).

50. Braun, G. *et al.* Enzyme-directed positioning of nanoparticles on large DNA templates. *Bioconjugate chemistry* **19**, 476-9 (2008).

51. Cinque, L., Ghomchi, Y., Chen, Y., Bensimon, A. & Baigl, D. Protection of human genomic DNA from mechanical stress by reversible folding transition. *Chembiochem : a European journal of chemical biology* **11**, 340-3 (2010).

52. Pignot, M., Siethoff, C. & Linscheid, M. Coupling of a nucleoside with DNA by a methyltransferase. *Angewandte Chemie* **37**, 2888-2891 (1998).

53. Pljevaljcic, G., Pignot, M. & Weinhold, E. Design of a new fluorescent cofactor for DNA methyltransferases and sequence-specific labeling of DNA. *Journal of the American Chemical Society* **125**, 3486-92 (2003).

54. Pljevaljcić, G., Schmidt, F., Scheidig, A.J., Lurz, R. & Weinhold, E. Quantitative labeling of long plasmid DNA with nanometer precision. *Chembiochem : a European journal of chemical biology* **8**, 1516-9 (2007).

55. Pljevaljcic, G., Schmidt, F., Peschlow, A. & Weinhold, E. Sequence-specific DNA labeling using methyltransferases. *Methods in molecular biology (Clifton, N.J.)* **283**, 145-61 (2004).

56. Schmidt, F.H.-G. *et al.* Sequence-specific Methyltransferase-Induced Labelling (SMILing) of plasmid DNA for studying cell transfection. *Bioorganic & medicinal chemistry* **16**, 40-8 (2008).

57. Klimasauskas, S. & Weinhold, E. A new tool for biotechnology: AdoMet-dependent methyltransferases. *Trends in biotechnology* **25**, 99-104 (2007).

58. Wilkinson, S. *et al.* Molecular scale architecture: engineered three- and four-way junctions. *Bioconjugate chemistry* **19**, 470-5 (2008).

59. Braun, G. *et al.* Enzyme-directed positioning of nanoparticles on large DNA templates. *Bioconjugate chemistry* **19**, 476-9 (2008).

60. Hohng, S. & Ha, T. Near-complete suppression of quantum dot blinking in ambient conditions. *Journal of the American Chemical Society* **126**, 1324-5 (2004).

61. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362-6 (2009).

62. Samad, A., Huff, E.F., Cai, W. & Schwartz, D.C. Optical mapping: a novel, single-molecule approach to genomic analysis. *Genome Research* **5**, 1-4 (1995).

63. Protozanova, E. *et al.* Fast high-resolution mapping of long fragments of genomic DNA based on single-molecule detection. *Analytical biochemistry* **402**, 83-90 (2010).