Single-Molecule Approach to Molecular Biology in Living Bacterial Cells

X. Sunney Xie,[1] Paul J. Choi,[1] Gene-Wei Li,[2] Nam Ki Lee,[1] and Giuseppe Lia[1]

[1]Department of Chemistry and Chemical Biology and [2]Department of Physics, Harvard University, Cambridge, Massachusetts 02138;
email: xie@chemistry.harvard.edu

Annu. Rev. Biophys. 2008. 37:417–44

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

This article's doi:
10.1146/annurev.biophys.37.092607.174640

1936-122X/08/0609-0417$20.00

Key Words

living cells, stochastic gene expression, DNA-protein interactions, transcription, translation, replication

Abstract

Recent developments on fluorescent proteins and microscopy techniques have allowed the probing of single molecules in a living bacterial cell with high specificity, millisecond time resolution, and nanometer spatial precision. Recording movies and analyzing dynamics of individual macromolecules have brought new insights into the mechanisms of many processes in molecular biology, such as DNA-protein interactions, gene regulation, transcription, translation, and replication, among others. Here we review the key methods of single-molecule detection and highlight numerous examples to illustrate how these experiments are contributing to the quantitative understanding of the fundamental processes in a living cell.

Contents

INTRODUCTION

Much of our knowledge in molecular biology has been derived from experiments done in bulk with large numbers of lysed cells and even large numbers of specific DNA or protein molecules, usually in isolation from other macromolecules. However, if one could track a particular single macromolecule of interest in a living cell with millisecond time resolution and nanometer spatial precision while observing its biochemical reactions under native physiological conditions, many enduring questions in molecular biology could be answered. For example, how does a DNA binding protein find a specific target sequence buried in the genomic DNA? How does a particular gene get turned on and off, and how do transcription and translation processes occur in real time? How is chromosomal DNA replicated? How do DNA repair mechanisms restore the integrity of incorrectly synthesized or damaged DNA? Can quantitative information be extracted from these observations? Thanks to recent progress in fluorescent proteins (FPs), microscopy techniques, and single-molecule assays, we now can probe individual molecules in a single bacterial cell to address the above questions.

A living cell as a test tube is fundamentally different from the conventional test tubes of molecular biology and analytical chemistry in several respects. First, many macromolecules exist in low copy numbers in a living bacterial cell. As shown in **Table 1**, a particular gene has only one copy, or a few copies owing to DNA replication at the later stage of the cell cycle. A particular mRNA has only a few copies owing to the short cellular mRNA lifetime. Although the copy number for a particular protein varies from 1 to 10^4, some important proteins such as transcription factors and DNA polymerases are present at low copy numbers. This necessitates single-molecule sensitivity in single cells.

Second, cellular biochemical reactions of single molecules are inherently stochastic. Because of the low copy numbers of participating macromolecules, cellular processes, such as transcription, translation, gene regulation, and DNA repair, often exhibit stochastic reaction events. This means a particular time trace for one cell's behavior is not reproducible and cannot be synchronized with that of another cell, even though the statistical properties are reproducible. This is in contrast to conventional experiments in which a large ensemble of molecules or cells exhibits deterministic and reproducible temporal behavior. In recent years, stochastic gene expression in single cells

Table 1 Copy numbers in *Escherichia coli*

Molecular unit	Number	Reference(s)
Replication errors per genome	0.002	(23)
Double-strand breaks per genome	0.2	(66)
Replication forks per cell	1.5–6	(12)
Gene copies per cell	1–5	(12)
β-galactosidase tetramers per uninduced cell	1	(15)
F-plasmids per cell	1–3	(30)
Transposon copies per genome	1–15	(10)
lac repressor tetramers per cell	5	(33)
RNAPs per induced *lac* gene	5–20	(44)
DNA polymerase III per cell	10–20	(13)
lacZ mRNA per cell	10–30	(44)
Ribosomes per *lac* mRNA	20	(44)
DnaG primases per cell	50	(68)
Actively transcribing RNAPs per cell	200–2000	(12)
RecA molecules per cell	1000	(43)
Single-stranded DNA binding protein	1000–7000	(11, 75)
Total RNAPs per cell	1000–10000	(12)
Ribosomes per cell	7000–50000	(12)
β-galactosidase tetramers per induced cell	10000	(44)
Total nucleoid proteins (e.g., Fis, HU, H-NS) per cell	50000–200000	(4)
tRNA per cell	60000–400000	(12)

has been a subject of intense studies (28, 62, 65). The single-molecule sensitivity in a living cell allows for the study of stochasticity at the molecular level (15, 83). Meanwhile, within the past decade, in vitro single-molecule enzymology has yielded much understanding of the working of macromolecule machineries. Single-molecule measurements allow one to determine the distributions rather than simply the averages of bulk molecular properties. Furthermore, stochastic time traces of a single molecule reveal a wealth of dynamic information, such as reaction intermediates. In particular, numerous in vitro single-molecule assays have brought valuable mechanistic insights into DNA protein interactions (9, 59), transcription (20, 32, 53), translation, (50), replication (49, 54), and more. Similar to the in vitro experiments, statistical analyses for single-molecule stochastic time traces in living cells can provide mechanistic information otherwise hidden in ensemble experiments.

Third, cellular biochemical reactions often occur under nonequilibrium conditions. Although some reactions, such as a protein binding a specific DNA sequence, proceed as a dynamic equilibrium process, many cellular enzymatic reactions such as transcription, translation, and replication occur with a constant supply of free energy and substrates. Cellular single-molecule experiments of these processes usually occur under nonequilibrium steady-state conditions because the concentrations of the substrate and product molecules remain roughly constant, similar to the situation for in vitro single-molecule experiments of enzymes and molecular motors (41, 57, 64). In contrast, conventional ensemble kinetic studies of enzymes, for example, stop-flow experiments, occur under nonequilibrium non-steady-state conditions, in which the substrate and product concentrations vary with time in a deterministic fashion.

Finally, a large number of coupled reactions result in networks of complex interactions. Systems biology treats a cell as a dynamic interacting system (2). However, much of our knowledge in molecular biology has come from experiments on well-isolated molecules through a reductionistic approach. Live-cell experiments allow us to study single-molecule behaviors in the physiological context, which by definition reflect the holistic complexity.

All of the above points highlight the compelling need to conduct live-cell single-molecule experiments to understand the fundamental processes both at the molecular and cellular levels. The main challenge is the integration of specific probes and advanced microscopy in assay development for different dynamical processes. In this review, we discuss some of the key strategies in single-molecule live-cell imaging, highlight a few examples of some fundamental processes in bacterial molecular biology, and illustrate the wealth of new and quantitative information extractable from these experiments.

SINGLE-MOLECULE IMAGING IN LIVING BACTERIA

Fluorescence imaging is indispensable in probing living cells because of its high sensitivity, high time resolution, high spatial precision, noninvasiveness, and most importantly high specificity. The emergence of FPs as endogenous labels has not only revolutionized cell biology (16, 58, 71) but also enabled most of the single-molecule live-cell experiments discussed in this review. Meanwhile, advanced fluorescence microscopy techniques, especially the availability of sensitive charge-coupled device (CCD) cameras and high-quality optical filters, have facilitated the advances. **Table 2** lists the methods used for single or low-copy-number detection in living cells. Fluorescence correlation spectroscopy has been widely used to determine low concentrations and diffusion constants of FPs. Because of extensive review elsewhere (40), we do not cover it in this review. We discuss the other methods following a brief survey of FPs.

Fluorescent Proteins

FPs are the most popular fluorophores for single-molecule imaging in live cells, primarily because of the ability to genetically encode FP as a fusion with almost any protein or peptide. Although FP can be less bright and photostable than organic dyes and nanoparticles, specific labeling in living cells by the latter remains challenging, particularly for micron-sized bacteria that are not amenable to microinjection or introduction through endocytosis.

The particular choice of FP variant depends on the desired application (described in References 71 and 73). Because bacteria exhibit strong autofluorescence at short wavelength excitation, monomeric yellow and red FPs are preferred over blue and green variants for single-fluorophore detection, although the latter can be used for detection of tandem repeats. In addition to color, photostability and brightness are two major considerations. In general, yellow FPs are less photostable than their red counterparts, but they compensate for their lower stability with higher brightness. Photobleaching of FPs limits continuous monitoring of the same FP molecule to no more than several consecutive images. The length of observation time can be varied by adjusting the unilluminated time interval between consecutive images in time-lapse movies. Sometimes, one can purposely photobleach existing FP molecules inside the cell to count the subsequently born FP molecules (83). Repeated photoexcitation of a single FP molecule results in intensity fluctuations known as blinking (63), making it difficult to quantify fluorophore numbers from the intensity. However, under the excitation conditions of a single FP experiment, blinking usually occurs on a ~1-s timescale, which fortunately does not affect data collection with an exposure time of 0.1 s

Table 2 Methods for single-molecule detection in living bacteria cells

Technique	Description	Example
Fluorescence correlation spectroscopy (55)	Fluctuations from fluorophores moving through a small excitation volume yield diffusion constants and concentration.	• Monitoring mRNA concentration (48) • Monitoring protein diffusion (18, 26)
Detection of tandem repeats (7, 35)	Binding of multiple fluorophores to tandem arrays (~100) of mRNA or DNA binding sequences allows detection of a single target.	• Monitoring mRNA production (34) • Tracking chromosome movement (35, 46, 51) • Localization of RNAP (this review, 14, 52)
Single-fluorophore detection by localization (78, 83)	A fluorophore is detectable upon binding to DNA or cell membrane, whereas a diffusing fluorophore in the cytoplasm is overwhelmed by the cellular autofluorescence.	• Monitoring protein production (83) • Detection of replisome proteins (this review) • Tracking membrane diffusion (22) • Tracking cytoskeletal proteins (45)
Stroboscopic illumination (26, 81)	A short laser excitation pulse, during which little fluorophore diffusion occurs, allows detection of a rapidly moving single molecule despite the slow CCD frame rate.	• Detection of single cytoplasmic FP (81) • Tracking fast diffusion by a transcription factor (26)
Enzymatic amplification with microfluidics (15)	Enzymatic amplification of fluorescent product molecules confined to a small volume allows detection of a single enzyme molecule inside a live cell.	• Monitoring protein production (15)

nor time-lapse movies with a dwell time of minutes.

Another important consideration is the maturation time of the fluorophore. Upon folding of an FP, three key residues must be oxidized to form the fluorophore. Fluorophore maturation can be the rate-limiting step and determine the time resolution of a live-cell experiment measuring protein production. Furthermore, fast maturation ensures the replenishing of FPs after photobleaching. We find that the Venus variant (60) has a short maturation time of ~7 min in bacteria cells. There have been several reports of fast-maturing FP variants (16) but they are dimeric. Evolution and selection of FPs based on short maturation times will prove to be useful. In particular, fast-maturing variants of different colors would allow multiplexed observations in live cells.

Photoactivatable FPs, such as Kaede (3) and Dendra2 (17), allow the activation of specific FPs in space and time at the onset of a dynamical observation, similar to a pulse-chase experiment. Photoswitchable FPs such as Dronpa (38) allowed single-molecule nanometer-resolution fluorescence imaging in fixed cells (8, 25, 69) and hold promise for applications in live cells.

Does a protein fused to an FP have the same activity and structure as the wild-type protein? A case-by-case situation must be determined for each protein. For example, in experiments described below, FP-labeled T7 RNAP maintains polymerase activity when labeled on the N terminus. Likewise, two FP-labeled membrane proteins, Tsr and LcY, labeled on the C terminus still efficiently inserts into the membrane, with the FP remaining on the cytoplasmic side of the inner membrane. The FP-labeled *lac* repressor binds to DNA as a dimer with full wild-type activity, though the FP prevents the formation of a tetramer.

Imaging Tandem Repeats

For nucleic acids, such as chromosome loci or mRNA molecules, it is possible to detect sufficient numbers of FP molecules colocalized on a single nucleic acid binding site above the fluorescence background from freely

diffusing FPs and autofluorescence. This was first shown in bacteria by integrating 256 tandem *lac* operator sites into a specific location on a chromosome or plasmid (35). *lac* repressor FPs expressed inside the cell are colocalized at the tandem operators, allowing visualization of the specific site on a single chromosome or plasmid inside the cell. A similar method allows the detection of individual mRNA molecules in both eukaryotes (7) and bacteria (34). A gene with 96 tandem repeats of an RNA hairpin binding sequence was constructed to allow specific binding of many MS2-FP fusion protein molecules onto a single mRNA molecule.

Depending on how many FP are expressed in the cytoplasm, there is a minimum number of colocalized FPs necessary for detection. Considering that the fully induced expression level of the cytoplasmic FP may be several thousand per cell, localization at a diffraction-limited spot $\sim$0.04 μm^2 in a bacterial cell with a cross-sectional area of $\sim$1 μm^2 would theoretically require $\sim$100 colocalized FPs to give a signal larger than the background from free FP molecules. The perturbation of nucleic acid properties from these bulky repeated elements can be a disadvantage in some cases. For example, *lac* repressor arrays can block passage of the replisome and MS2 arrays inhibit mRNA degradation.

Single Fluorophore Detection by Localization

Detection of a macromolecule labeled with a single FP would be desirable, avoiding the perturbation from bulky tandem repeats. Single-molecule fluorescence detection at room temperature with various types of fluorescence microscopy has become a common practice. It is generally difficult to detect single fluorophores in a living cell because of strong cellular autofluorescence. Reducing the probe volume is needed to reject autofluorescence. Single molecules can be imaged with a confocal microscope with a focused laser beam and a point detector (80). However, commercial scanning confocal microscopes suffer from low signal collection and detector efficiency. Another way of reducing probe volume is total internal reflection (TIR) with a sensitive CCD camera, allowing detection of single FP molecules fused to membrane proteins (70). For bacteria, however, the low penetration depth of TIR only allows a small fraction of a bacterial cell to be uniformly illuminated.

Single FP molecules fused to membrane proteins (22, 83) or bound to DNA (26) in living bacteria cells can be visualized with an inverted microscope with wide-field laser excitation and CCD camera detection. The signal from a single FP is stronger than the autofluorescence from the thin monolayer of bacterial cells used but would be overwhelmed by thicker yeast or mammalian cells in the wide-field microscope.

Detection by localization is the key to detect single FP in these experiments (83). As illustrated in **Figure 1*a***, for a molecule rapidly diffusing through the cell cytoplasm (with a diffusion constant of $\sim$8 $\mu m^2\ s^{-1}$) (29), the signal is spread over the entire cell area during the 100 ms of a typical CCD camera exposure and is overwhelmed by the autofluorescent background. A membrane protein diffuses much slower (with a diffusion constant of $\sim 10^{-2}$ $\mu m^2\ s^{-1}$) (22), effectively confining the FP molecule to a diffraction-limited spot during the image exposure time. Similarly, a protein bound to a particular DNA locus is also largely stationary (with a diffusion constant of $\sim 10^{-5}$ $\mu m^2\ s^{-1}$) (27). The highly localized emission is above the autofluorescent background, allowing single-molecule detection as shown in **Figure 1*b***. The proof of detecting a single molecule is the quantized photobleaching of the fluorescence signal (**Figure 1*e***). Although the single FP spot size is diffraction limited to about half the wavelength of light, one can determine the position of the localized single FP to nanometer accuracy using centroid fitting of the fluorescence intensity, given enough detection photons (26, 74).

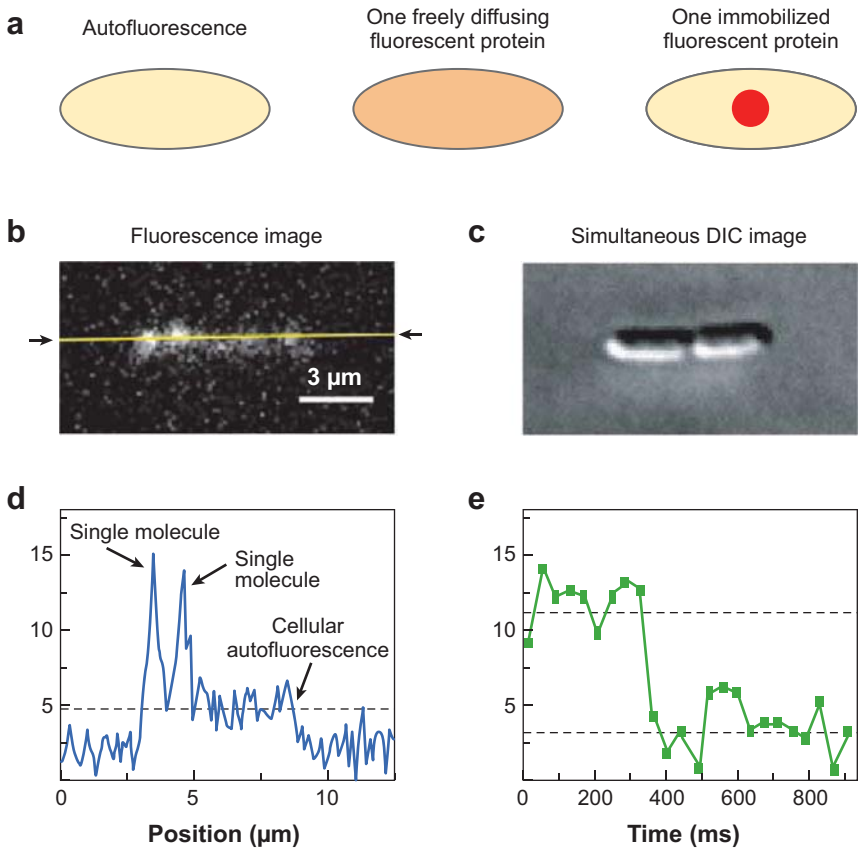

Figure 1

(*a*) When there are only a few freely diffusing FPs expressed inside the cell, their fluorescence is spread over the entire cytoplasm during a typical acquisition time of 50 to 100 ms, overwhelmed by the autofluorescence background. When an FP binds to DNA or the cell membrane, its diffusion significantly slowed, and its fluorescence can be detected as a distinct diffraction-limited spot over the background during typical acquisition times. (*b*) Fluorescence and (*c*) differential interference contrast (DIC) images of two *Escherichia coli* cells expressing Tsr-YFP. Two single-fusion protein molecules were detected as diffraction-limited fluorescent spots. (*d*) The fluorescence signal along the long axes of the two *E. coli* cells (the *yellow line* in panel *b*). (*e*) Fluorescence time trace of a single Tsr-Venus molecule in an *E. coli* cell. The abrupt photobleaching proves single-molecule detection. From Reference 83.

Single-FP detection by localization is easily accomplished using commercially available instrumentation and can be applied to a wide range of DNA binding and membrane proteins. A confocal microscope is not necessary because of the small size of the bacterial cells, as long as a thin layer of cells is used. The cells are fixed to a coverslip using an agarose pad sandwich or polylysine treatment of the glass. For this method to work, one also needs to have a low autofluorescence background and a low copy number of FPs in the cytoplasm. We use M9 minimal medium to reduce autofluorescence and keep the FP copy number inside a cell under 10.

STROBOSCOPIC EXCITATION

Extending the idea of detection by localization, even a single FP in the cytoplasm can now be detected. One would need a submillisecond exposure time to capture a

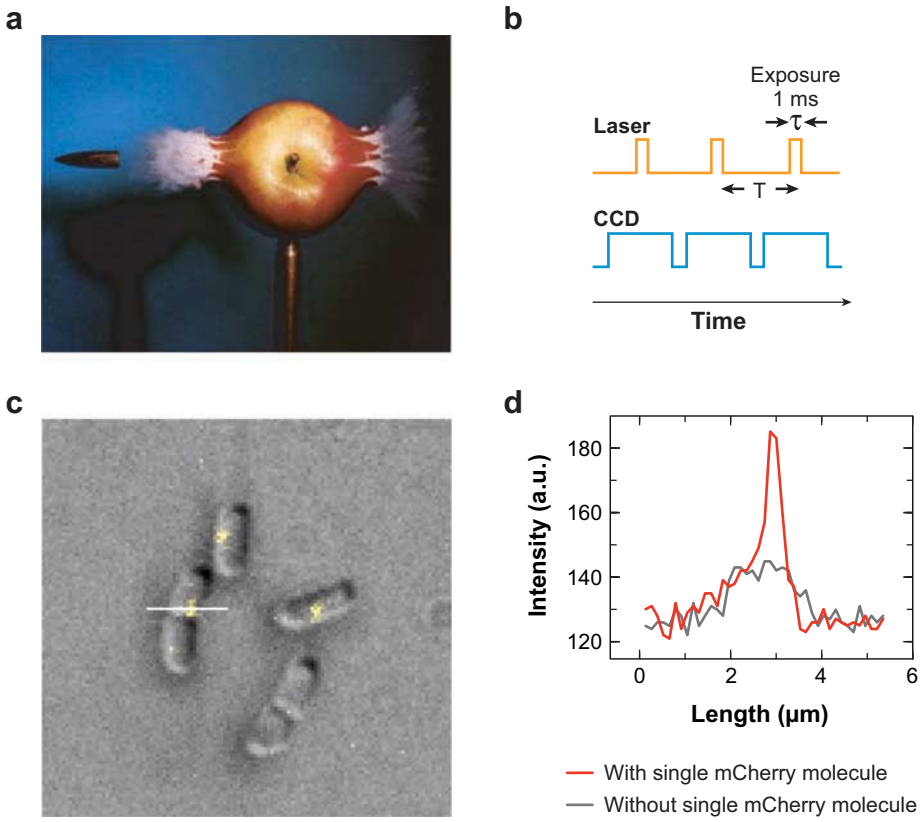

Figure 2

(*a*) Image of a bullet passing through an apple. From the Harold and Esther Edgerton Foundation. (*b*) Timing diagram for stroboscopic illumination. Each laser pulse is synchronized to a CCD frame, which lasts time *T*. (*c*) Fluorescence/DIC overlay of *Escherichia coli* cells, three of which contain single mCherry molecules in the cytoplasm. The fluorescence image was taken with a 300 μs, 50 kW cm^{-2} laser excitation. (*d*) Linescan of fluorescence signal (from the *white line* in panel *c*) for cell with (*red*) and without (*gray*) single mCherry molecule.

fast-diffusing FP, which is beyond the capability of a CCD camera. Our laboratory borrowed a method from strobe photography (81). **Figure 2*a*** shows a picture of a bullet going through an apple. Because the camera shutter was not faster than the bullet, the sharp picture was taken in a dark room with the shutter wide open, and the objects were exposed to an intense, short light pulse during which the bullet did not move far. Similarly, a short laser pulse can overcome the limitation of the slow shutter or frame rate of the camera. The shutter and the CCD can be left open for longer times (**Figure 2*b***). We applied an intense laser pulse for a short duration (~0.3 ms), during which a protein reporter does not diffuse beyond the diffraction limited spot. **Figure 2*c*** shows detection of the single red FPs mCherry in *Escherichia coli* cytoplasm with a high signal-to-noise ratio (**Figure 2*d***).

With such stroboscopic excitation, the time resolution of live-cell single-molecule detection can be submillisecond, no longer dictated by the shutter speed or frame rate of the CCD, but by the laser pulse width. In addition, the pulse width and dwell time can be varied to probe dynamical

properties such as residence times of weak binding and diffusion constants, as illustrated below. The drawback of the stroboscopic excitation is that single FP molecules are more photolabile under the high laser intensity.

Enzymatic Amplification with Microfluidics

Although this review focuses on experiments with the use of FPs in bacteria, single-molecule sensitivity can be realized with other reporter proteins, such as β-galactosidase (β-gal), a standard reporter for gene expression. A single molecule of β-gal can produce many copies of fluorescent product molecules by hydrolyzing a synthetic fluorogenic substrate, leading to amplification of the fluorescent signal, as first demonstrated in 1961 (67). However, the fluorescent product molecules are not retained in the cell but pumped out by efflux pumps on the cell membrane, which efficiently expel foreign organic molecules from cytoplasm. Because the fluorescent product molecules are pumped to the surrounding medium and rapidly diffuse away, Cai et al. (15) developed an assay to circumvent the loss of signal due to the efflux problem: A live cell is trapped in a polydimethylsiloxane (PDMS)-based microfluidic chamber in a small volume that confines and allows rapid mixing of the expelled fluorophores. The rate of fluorescence increase of the chamber reports the number of β-gal molecules within the live cell. Single-molecule sensitivity of this method has been demonstrated, and the technique has been applied to probe gene expression from *E. coli*.

The advantage of the method is that it is amenable to existing β-gal libraries across many species, well-suited for high-throughput measurements in microfluidic arrays, and less prone to photobleaching. A challenge for using β-gal to monitor gene expression in live cells is that the cell wall can act as a barrier for the influx of the enzymatic substrate.

DYNAMICS OF TRANSCRIPTION FACTORS

DNA protein interactions first arise from protein binding to DNA, which can be specific or nonspecific in nature. Specific binding means binding to a specific sequence of DNA, whereas nonspecific binding means transient, weaker binding to DNA regardless of the sequence context. Both specific and nonspecific binding are dynamic equilibrium phenomena with stochastic binding and unbinding events, which can now be probed by single-molecule experiment in a living cell.

Gene Regulation

A classic example of specific binding is that of the *lac* repressor, which controls gene expression of the *lac* operon by binding to specific operator sequences of DNA and has been a model system for understanding transcription factor-mediated gene regulation. Because the chromosomal DNA is largely stationary, we can take advantage of the idea of detection by localization. Elf et al. (26) reported the first real-time single-molecule observation of specific binding and unbinding of a *lac* repressor in live cells.

For single-molecule detection, the fact that the copy number of a *lac* repressor (five tetramers in the wild-type cell) is small is advantageous. In fact, we used a strain with even lower expression of *lac* repressor to improve the image contrast. There are three operators in the *lac* operon (O_1, O_2 and O_3), of which O_1 has the highest affinity. We replaced the O_3 with O_1, which further suppressed the expression level of the repressor (26). The *lac* repressor-Venus fusion protein binds to each O_1 operator as a dimer with full wild-type functionality.

Figure 3*a* shows that each *E. coli* cell under repressed condition has one fluorescent spot for each *lac* operon due to repressor-Venus fusion proteins specifically bound to the two O_1 operators. Some cells have two spots because DNA replication yields two *lac* operons late in their cell cycle. Upon addition of

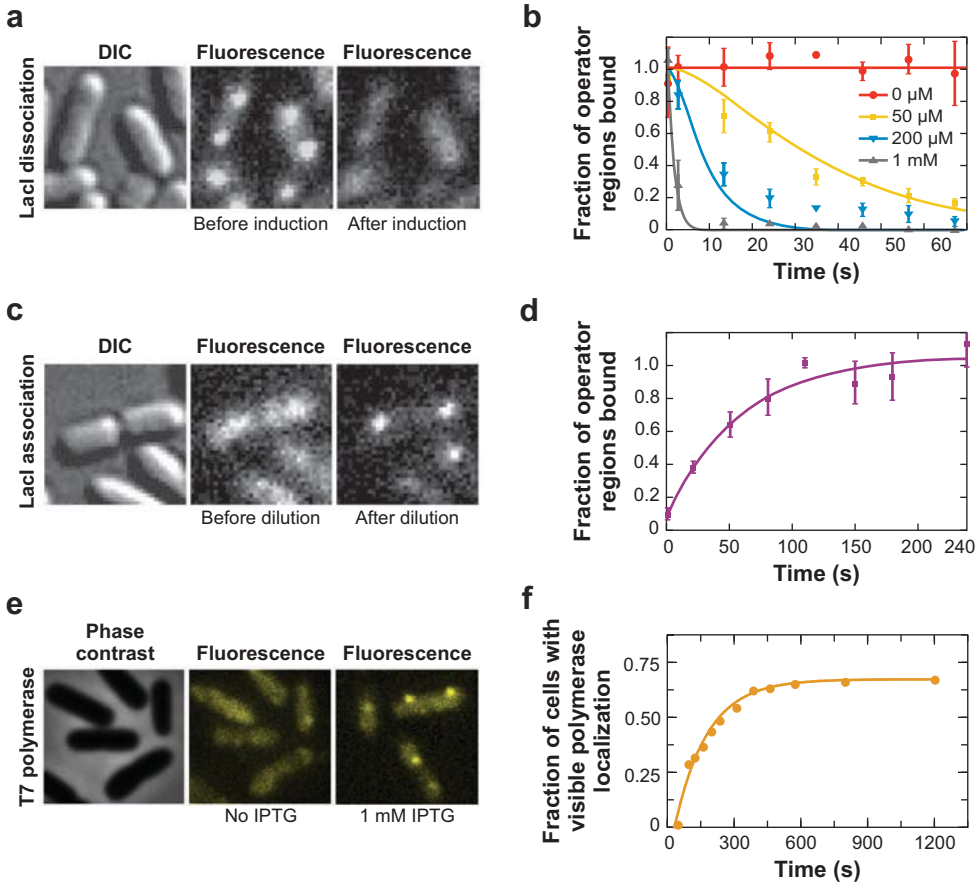

Figure 3

(*a*) *Escherichia coli* cells with Venus-labeled *lac* repressor before and 40 s after addition of IPTG to a final concentration of 1 mM. (*b*) Fraction of *lac* operons in a cell population bound to repressors is plotted as a function of time after induction by various concentrations of IPTG. (*c*) *E. coli* cells with Venus-labeled *lac* repressor before and 1 min after rapid dilution of IPTG from 100 to 2 mM. (*d*) Fraction of the operator regions that is bound to repressor as a function of time after the rapid dilution. The data are fitted with an exponentially distributed binding time, with a time constant of 60 s. From Reference 26. (*e*) *E. coli* cells expressing Venus-T7 RNAP fusions with T7 promoter replacing the *lac* promoter. Before adding IPTG, the *lac* repressor prevents transcription by the T7 RNAP. Following the addition of IPTG, multiple Venus-T7 RNAP molecules bound to the *lac* operon colocalize on the gene. (*f*) Fraction of cells with visible T7 RNAP localization as a function of time after adding IPTG.

isopropyl-D-1-thiogalactopyranoside (IPTG, an analog of the natural inducer *allo*-lactose) to the medium, the repressors dissociate, leading to the disappearance of the fluorescent spots. This loss of fluorescent spots signals the onset of gene expression.

On a single-molecule basis, the dissociation times are stochastic. **Figure 3*b*** shows the fraction of operons of a population of cells with at least one bound repressor as a function of time at different IPTG concentrations. This gives the rate of the repressor dissociation from the operon upon induction, which is the rate of IPTG binding to the repressor. The latter might be limited by the influx of IPTG through the cell membrane. This experiment demonstrates the measurement of live-cell kinetics of transcription factors.

Search for Target DNA Sequence

Much more information is derived from the reverse experiment involving the rebinding of the repressor to the operator upon removal of IPTG. **Figure 3*c*** shows the reappearance of the fluorescent spots after a sudden dilution of IPTG by a factor of 50, and **Figure 3*d*** shows the fraction of operators with a bound repressor as a function of time, yielding a time constant of ~60 s for the first repressor to rebind to an O_1 operator. Because in vitro data showed that the dissociation of IPTG from the repressor takes only ~1 s (24), the ~60 s should be the time for the first IPTG-free repressor to find the operator. Considering this is a bimolecular association and there are three dimeric repressors and two O_1 operators for each operon, it would take ~360 s for a single dimeric repressor in a single cell to find a single O_1 site. A faster search would require more repressor molecules inside the cell. This gives the first experimental measurement of how fast an individual DNA binding protein searches for its target sequence in a living cell.

Why does it take ~360 s for a single repressor to find the target? Von Hippel & Berg (76) studied this subject in the 1980s. The in vitro observation of unusually large bimolecular association rates for protein binding to DNA led to the model of facilitated diffusion. As illustrated in **Figure 4*a***, in searching for a target DNA sequence, a DNA binding protein first nonspecifically binds to DNA and undergoes Brownian diffusion along the DNA to scan for the target. However, inspection of the whole genome would be extremely time-consuming through such 1D diffusion alone. If the protein does not find the target in close vicinity, it dissociates from the DNA and undergoes much faster 3D diffusion through the cytoplasm to bind nonspecifically to a different DNA segment where the 1D search resumes. The combined 1D and 3D diffusion allows an efficient search throughout the whole genome. This model came from rigorous deduction based on indirect observations and has become the basis of many theoretical investigations (39, 72).

Considering the aforementioned single-molecule experiments, we were in a position to quantitatively examine the mechanism for sequence-specific DNA searching and recognition. First, in an in vitro experiment on a stretched DNA, we directly observed nonspecific binding and Brownian diffusion along DNA (9) and measured the 1D diffusion constant of the dimeric repressor-Venus to be $D_{1D} \sim 0.046\ \mu m^2\ s^{-1}$ (26). In general, the 1D diffusion constant of a DNA binding protein is independent of salt concentration, whereas the 1D sliding or residence time on DNA is dependent on the salt concentration.

In living bacteria, residence times should be shorter than in vitro because of the high cellular salt concentration, whereas the 1D diffusion constant should be similar to the in vitro value. Because DNA is coiled inside the cell, the 1D diffusion constant is difficult to probe. However, the nonspecific binding can be observed by stroboscopic illumination. As shown in **Figure 4*b***, 10-ms exposure times resulted in sharp images under the induced condition, suggesting that the residence time of nonspecific binding on DNA is less than 10 ms, which was previously unknown. Furthermore, by varying the time interval between two laser excitation pulses, we measured the apparent diffusion constant $D_{eff} \sim 0.40\ \mu m^2\ s^{-1}$ (**Figure 4*c***) inside the cell. On the other hand, the cytoplasmic 3D diffusion constant of the repressor-Venus fusion protein with its DNA binding domain removed is measured by fluorescence correlation spectroscopy to be $D_{3D} \sim 3\ \mu m^2\ s^{-1}$ (**Figure 4*d***). Realizing $D_{eff} = D_{3D}(1 - F) + FD_{1D}/3$, where F is the fraction of time the repressor is nonspecifically bound to DNA, and that the second term is negligible, it follows that $F \sim 90\%$. This suggests that in searching for the operator, a *lac* repressor spends ~90% of its time nonspecifically bound to and diffusing along DNA. This is consistent with the von Hippel et al. (82) estimate (>90%) based

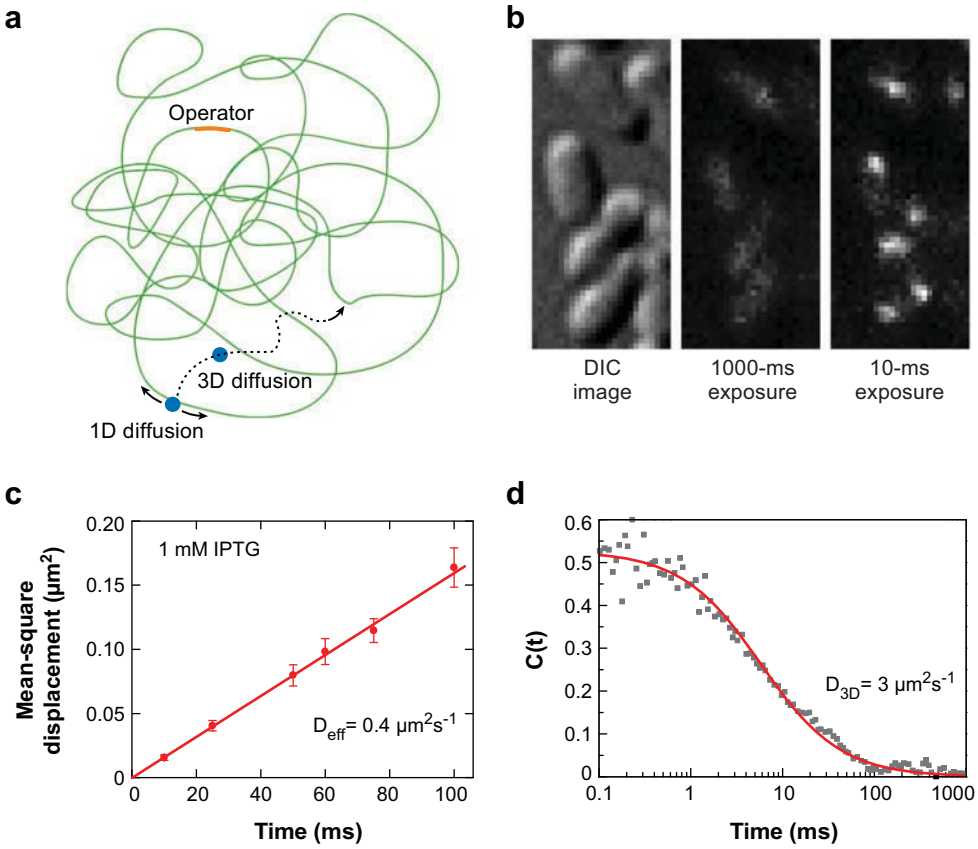

Figure 4

(*a*) In searching for a target DNA sequence, a DNA binding protein first nonspecifically binds to DNA and undergoes 1D diffusion along a short segment of DNA before dissociating from DNA, diffusing in 3D through the cytoplasm, and rebinding to different DNA segment. (*b*) Two fluorescence images with different exposure times and the corresponding DIC image of IPTG-induced *Escherichia coli* cells. At 1000 ms, individual *lac* repressor-Venus molecules appear as diffuse fluorescence background. At 10 ms, they are clearly visible as nearly diffraction-limited spots. (*c*) Mean-square displacement for nonspecifically bound transcription factors for different time intervals. The red line shows a linear fit of the mean-square displacement. The fitting agrees well with normal diffusion in the imaging plane, $<\Delta x^2> = 4D_{eff}\Delta t$, with $D_{eff} = 0.4\ \text{mm}^2\ \text{s}^{-1}$. (*d*) The autocorrelation function for cytoplasmic Venus from fluorescence correlation spectroscopy of individual cells is well-fit by a one-component diffusion model to give $D_{3D} \sim 3\ \mu\text{m}^2\ \text{s}^{-1}$. From Reference 26.

on the measurement of tetrameric repressor concentrations in minicells, which are derived from *E. coli* cells but contain no DNA.

During a cellular residence time of ~5 ms, the repressor scans ~85 base pairs according to the measured 1D diffusion constant. As stated in **Table 1**, the occupancy of the *E. coli* genome by DNA binding proteins is dominated by RNP and nucleoid proteins such as HU and Fis, though the binding of the latter is relatively weak. Therefore, at least 80% of the genome is accessible (42). The single-repressor/single-target search time of ~360 s is close to the back-of-the-envelope estimate for repeated unsuccessful 1D searches in the ~5-ms durations separated by almost instantaneous and random hopping among the ~85-bp segments in ~80% of the

4.8×10^6 bp genome (79). Again, the search time can be shortened if the copy number of the repressor is increased. This gives a benchmark for the discussion of genome searches by other DNA binding proteins, such as DNA repair enzymes.

TRANSCRIPTION

Dissociation of the repressor triggers transcription, the first step of gene expression. Transcription initiation, elongation, and termination have been studied in detail by in vitro ensemble and single-molecule experiments (20, 32, 47). However, the copy number of most mRNA transcripts per bacterial cell is low because of their limited lifetime, preventing single-molecule sensitivity for single-cell analysis by conventional methods such as DNA arrays and Northern blot analyses, let alone these methods cannot monitor living cells. Here we describe live-cell experiments used to study these processes at the single-molecule level.

Probing RNAP Activity

The study of transcription through the labeling of RNA polymerase (RNAP) with FPs has been demonstrated first at high growth rates, at which RNAP-FP fusions localize into tight foci in *Bacillus subtilis* (52) and *E. coli* (14). These foci exist only during active transcription and likely correspond to sites of stable rRNA or tRNA transcription. Because native RNAP promoters are ubiquitous in bacterial genomes, it is only possible to distinguish sites of extremely active transcription with colocalization of many FP-labeled polymerase molecules, similar to the array of tandem repeats.

The activity of a single RNAP can be observed from the production of single mRNA transcripts. Single-molecule sensitivity is required to count the mRNA even under fully induced conditions for the *lac* operon because the short mRNA lifetime leads to a low copy number, as described in **Table 1**. Although powerful PCR and in situ hybridization methods exist for sensitive measurements of mRNA in lysed or fixed cells, to date, there has been only one method for detection of individual mRNA molecules in living bacteria cells (34), adapted from work in eukaryotes (7). The target transcript includes on its 3′ UTR 96 tandem repeats of the sequence corresponding to a hairpin motif that binds to the MS2 viral coat protein. Expression of an MS2-FP fusion protein results in colocalization of 50 to 100 FP molecules for detection of the transcript. Using this technique, Golding et al. (34) observed the pulsed production of individual mRNA in real time in living *E. coli* cells (**Figure 5**).

Transcriptional Pulsing

Under fully induced conditions, a promoter might be expected to continually produce new transcripts. However, the real-time visualization of mRNA production revealed that transcription occurred in bursts of multiple transcripts separated by windows of inactivity (34). This finding implied that transcription events were correlated, and following initiation of the first transcript, subsequent events were more likely to occur. An OFF-ON model for the promoter toggles, with the promoter shifting between active and inactive states, certainly explains the data. However, it is unclear what mechanism would lead to this proposed toggling. Bacterial DNA lacks the extensive structural rearrangement as in eukaryotic chromatin remodeling, although they do contain several histone-like proteins, such as HU, Fis, and H-NS.

Instead of monitoring mRNA, we are developing another assay to probe transcriptional bursting. Bursting may be evident by directly imaging RNAP numbers at a specific gene. We use an FP-labeled T7 RNAP in bacteria containing a strong T7 promoter integrated into the chromosome to replace the *lac* promoter. The *lac* repressor still blocks transcription at this artificial promoter, but only T7 RNAP, not native *E. coli* RNAP, binds

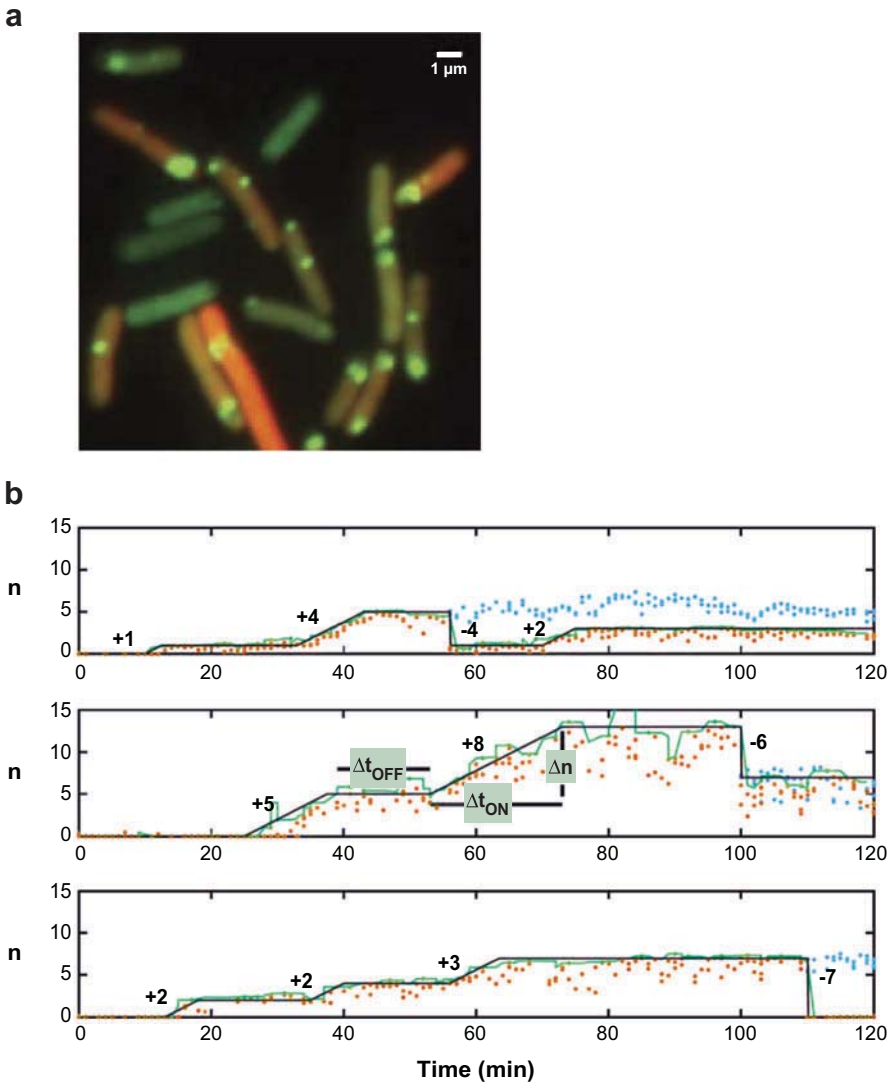

Figure 5

(*a*) Fluorescence image of single mRNA labeled on the 3′ UTR with 96 tandem repeats of an MS2-binding hairpin that bind multiple FP-MS2 coat protein fusion proteins. The tagged mRNA also expresses red FP. The image shows an overlay of mRNA expression (green localization) and protein expression (orange intensity). (*b*) mRNA numbers for three cells as a function of time. Red spots correspond to raw data, the green line corresponds to smoothed data, and the black line corresponds to a visual aid. Cyan spots correspond to mRNA in a sister cell following cell division. The mRNA production shows period of inactivity, Δt_{OFF}, and periods of mRNA production, Δt_{ON}. From Reference 34.

to this promoter. Because there are no endogenous T7 promoters in the *E. coli* genome, localized T7 RNAP molecules signal loading and transcription at the engineered promoter. In this way, it is possible to monitor transcription activity of this gene in real time. **Figure 3*e*** shows the localization of actively transcribing T7 RNAPs at the T7 promoter upon dissociation of repressor from the operator with IPTG, and **Figure 3*f*** shows

changes in T7 loading over time for a cell population. Analysis of such data is underway to reveal the working of transcriptional machineries in a live bacterial cell.

TRANSLATION

Following transcription, the second step in gene expression involves translation of the mRNA by ribosomes. Similar to transcription, much about translation has been learned from in vitro experiments, including recent single-molecule experiments (50). Conventional methods for characterizing protein expression such as Western blot analyses, mass spectrometry, and flow cytometry lack the sensitivity necessary to detect the low levels of many proteins in individual cells. In 2006, our group demonstrated two approaches to detect low-level protein expression via localization of a membrane-FP fusion and using a microfluidic enzymatic assay. These techniques allowed us to make the first observations on the real-time basal expression of a gene under repressed conditions in a living cell.

Yu et al. (83) reported a time-lapse movie of dividing *E. coli* cells randomly producing single-protein molecules from a repressed promoter (**Figure 6*a***). This was done using the method of detection by localization, in which an FP gene was fused to a gene for the membrane protein, Tsr, in place of the *lacZ* gene (83). **Figure 6*b*** shows the stochastic production of proteins with translational bursts, visible only during repressed conditions. Under these repressed condition, the repressor is tightly bound to its operator sites. Nevertheless, its infrequent dissociation still results in a transcription event that generates a few copies of protein molecules. This movie was taken with a 100-ms exposure during which the FP molecules are not photobleached, followed by a 1-s laser exposure during which the FP molecules generated in the previous image are photobleached to avoid the accumulation of FPs. Under these conditions, the cell division time was not affected by the laser illumination, suggesting minimal phototoxicity.

Although most studies have been limited to protein expression at high levels because of low sensitivity, it is important to study the repressed condition for two reasons. First, many important regulatory proteins, such as transcription factors, have low protein copy numbers. Second, the analysis of the stochastic time trajectories at a low expression level allows us to obtain quantitative information about the size and shape of translational bursts and the cell cycle dependence of the bursts.

Translational Burst Size

What do the translational burst size and shape tell us? First, we discuss temporal burst width. We have proven that each burst arises from only one mRNA under the repressed conditions (83). According to the translational rate and the close spacing between subsequent ribosomes on mRNA observed in electron microscopy images (36), the protein molecules should appear with a time spread of less than 1 min. However, the bursts have a larger characteristic width of 7 min, given by the time constant of the exponential autocorrelation function of **Figure 7*a***. This ~7 min time constant is consistent with the live-cell maturation time of Venus measured by an ensemble assay. Faster-maturing FPs may allow the study of protein folding or translocation, if they are the rate-limiting step for signal detection.

Second, we examine the burst size. **Figure 7*b*** shows that the average burst size, b, equals four protein molecules per mRNA, and the distribution of burst sizes follows an exponential distribution. Although such an exponential distribution has been theoretically predicted (6, 56), this was the first experimental observation. The origin of the exponential distribution is related to the short cellular lifetime of mRNA because of the presence of RNase E, a ribonuclease, which degrades mRNA upon binding to the ribosome binding site. **Figure 7*c*** shows the

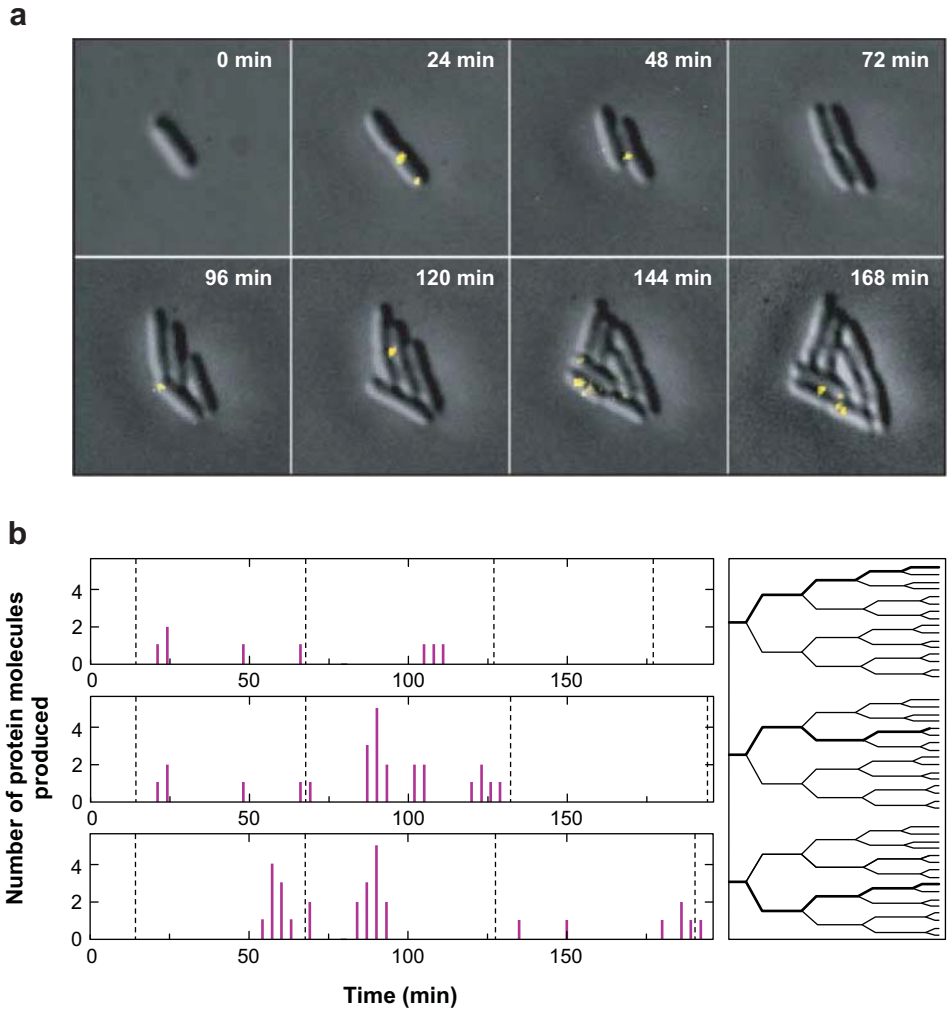

Figure 6

(*a*) Time-lapse movie of fluorescence images (*yellow*) overlaid with simultaneous DIC images (*gray*) of *Escherichia coli* cells expressing Tsr-Venus fusion proteins under the repressed condition. In the experiment, images are collected every 3 min with a 100-ms exposure immediately followed by a 1-s exposure for photobleaching to prevent accumulation of FPs. Representative images are shown. (*b*) Time traces of the expression of Tsr-Venus protein molecules (*left*) along three particular cell lineages (*right*) extracted from time-lapse fluorescence movies. The vertical axis is the number of protein molecules newly synthesized during the last three minutes. The dotted lines mark the cell division times. The time traces show that protein production occurs in random bursts, within which variable numbers of protein molecules are generated. From Reference 83.

ensemble-averaged mRNA lifetime measurement, with a 1.5 min exponential decay of mRNA over time, which means, on a single-molecule basis, the probability density of cellular mRNA lifetime also follows an exponential distribution with a 1.5 min time constant. The longer a particular mRNA evades degradation, the more protein can be translated. As a result, the number of proteins translated per mRNA follows an exponential distribution.

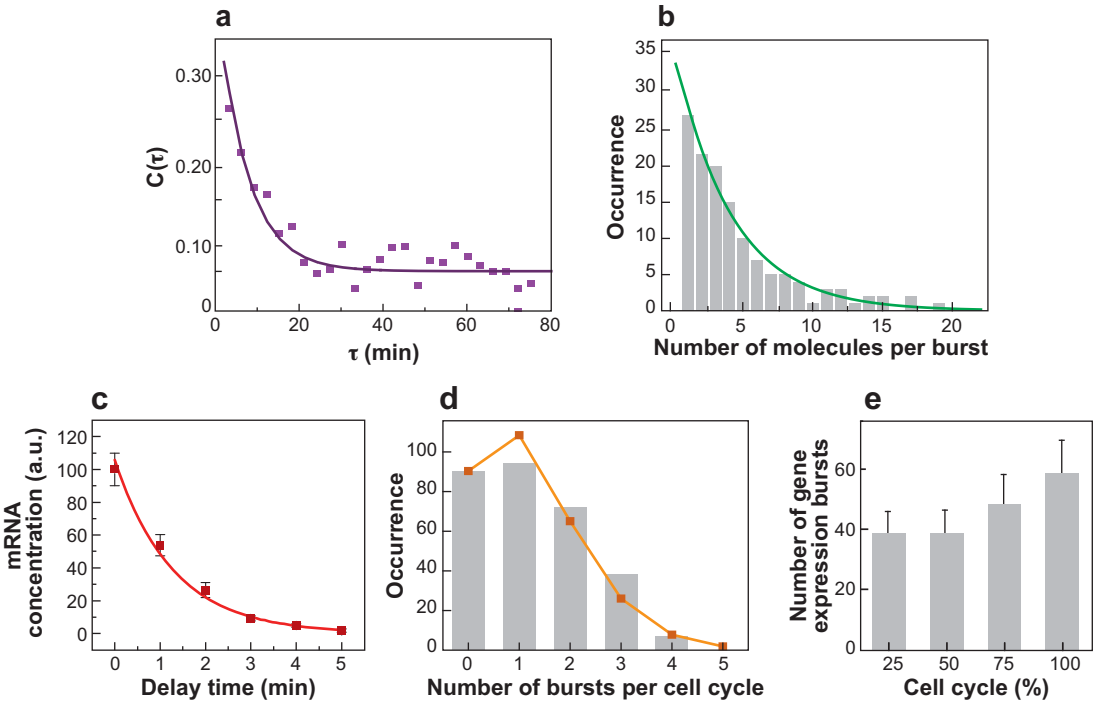

Figure 7

(*a*) Autocorrelation function of over 30 protein production time traces of cell lineages as shown in **Figure 6*b***. The decay constant of 7 min is attributed to fluorophore maturation. (*b*) Distribution of the number of FPs in each gene expression burst, which follows an exponential distribution (*solid line*), giving a probability of ribosome binding of $p = 0.8$ and an average of $b = 4$ molecules per burst. (*c*) Degradation of *tsr-venus* mRNA. After a short period of induction, the mRNA production is inhibited by an antibiotic and the amount of *tsr-venus* mRNA is measured using real-time PCR. The single exponential decay of the *tsr-venus* mRNA population (*solid curve*) yields an mRNA degradation time constant of 1.5 ± 0.2 min. (*d*) Histogram (*gray bars*) of the number of expression events per cell cycle. The data fit well to a Poisson distribution (*solid line*), with an average of 1.2 gene expression bursts per cell cycle. (*e*) Cell cycle dependence of the gene expression rate for *lac* promoter under repressed conditions. The division cycle of each cell was divided into four time windows of equal length. The result shows that more gene expression bursts are observed at the later stage of the cell cycle. From Reference 83.

More quantitatively, assigning relative probabilities for ribosome and RNase E binding to the ribosome binding site of p and $1 - p$, respectively, the probability, $P(n)$, of producing n proteins from one mRNA is simply the probability of n consecutive ribosome bindings followed by one RNase E binding, or $P(n) = p^n (1 - p)$. This is an exponential distribution for discrete integers. Fitting the distribution data in **Figure 7*b*** with the model yielded $p = 0.8$.

Translational Burst Frequency

Figure 7*d* shows the distribution of the number of translational bursts per cell cycle under repressed conditions, which follows a Poisson distribution with an average burst frequency of $a = 1.2$ per cell cycle. Knowing that each burst corresponds to one mRNA, this means that transcription is a Poisson process under the repressed condition. This is because infrequent and purely random dissociations of the repressor from its operator

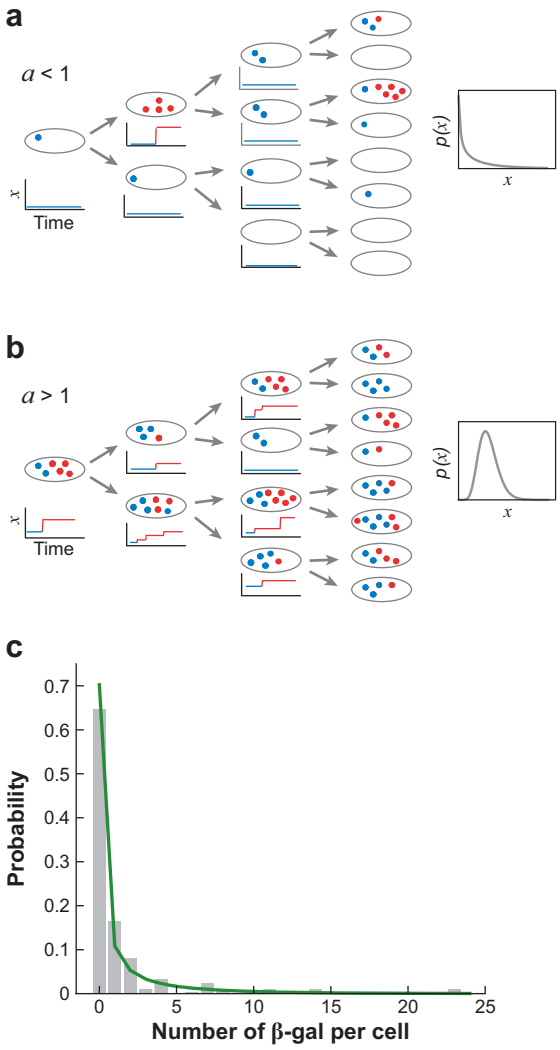

Figure 8

Protein molecules are produced in bursts (*red*), mixed with existing molecules (*blue*), and randomly partitioned upon cell division, leading to a steady-state distribution, $p(x)$, of protein copy number, x, in the cell population. (*a*) With the burst frequency $a < 1$, $p(x)$ peaks at $x = 0$ and a large fraction of cells do not contain a single copy of the protein regardless of the burst size, b. (*b*) With the burst frequency $a > 1$, most cells express the protein at some level. Both types of distribution can be described by the Γ distribution with two adjustable parameters, a and b. (*c*) Copy number distribution of β-galactosidase (β-gal) molecules in single *Escherichia coli* cells under the repressed condition measured by enzyme amplification in microfluidics. The solid line shows a Γ distribution with $a = 0.16$ bursts per cell cycle and $b = 7.8$ β-gal molecules per burst. These numbers agree with the real-time measurements of a and b using the same microfluidics method. From Reference 15.

were the rate-limiting step for transcription initiation.

We studied the cell cycle dependence of burst frequency. At least two processes may alter the frequency of translational bursts during the cell cycle. First, replication of the chromosome results in two or more gene copies at later stages of the cell cycle, which increases the protein production rate. Second, a collision between the replication machinery and the repressor as the replication fork moves through the gene may result in the dissociation of the repressor from its operator and hence a burst of proteins (37). **Figure 7*e*** shows the cell cycle dependence of protein burst frequencies by analyzing movies such as the one in **Figure 6*a***. Later stages of the cell cycle have an increased probability of protein bursts because of replication of the gene. However, the probability of bursts is still well distributed across the entire cell cycle, which suggests that basal level of expression does not result solely from the collision between the replication machinery and repressor. Had that been the case, the bursts would only occur at a specific point during the cell cycle. This question is easy to address with the single-molecule experiment.

Understanding Stochasticity in Protein Expression

Stochasticity in protein expression includes both the temporal fluctuations in a single cell and the variation of protein copy numbers across a cell population at a given moment in time. Both kinds of fluctuations can be important in determining cell phenotype. The single-molecule studies of translation provided the first direct measurements of the burst size, b, and the burst frequency, a. These two kinetic parameters uniquely define the dynamic fluctuations in protein expression of a single cell over time. Assuming every cell is identical, these same parameters should be sufficient to describe the steady-state variation in protein expression across a population, as illustrated in **Figure 8**.

Friedman et al. (31) developed a model connecting these two different fluctuations based on burst size and frequency of protein production. They assumed that (*a*) all cells are identical, (*b*) transcription is a Poisson process, (*c*) the number of proteins per mRNA follows an exponential distribution, as discussed above, and (*d*) partitioning of molecules to daughter cells leads to an effective dilution of copy numbers. Then, burst frequency a and burst size b predict the probability $p(x)$ of observing a cell in a population with x molecules to be $p(x) = b^{-a}x^{a-1}e^{-x/b}/\Gamma(a)$, known as the Γ distribution. **Figure 8*a*,*b*** depict the situation for $a < 1$ and $a > 1$, respectively. This is a simple but important result, allowing the reconciliation of the real-time protein number fluctuations in a single cell with the steady-state distribution of a cell population.

The protein copy number distribution in a cell population has been experimentally shown be well approximated by a Γ distribution, which can be fit well with the parameters a and b determined from single-cell real-time time traces. The Γ distribution provides not only a better fit of experimental data but also a better mechanistic justification than other phenomenological fits previously used, such as normal or log-normal distributions. **Figure 8*c*** shows the distribution for β-gal expression under the repressed condition using the enzymatic amplification assay (15). The Γ distribution is not specific to repressed conditions but can be generalized to induced conditions (31). In practice, a and b can be obtained from the variance and mean of a steady-state distribution. Because most protein expression profiling with FP libraries were conducted with flow cytometry or fluorescence microscopy on a population of cells at a particular time, being able to deduce a and b, the two fundamental parameters of gene expression, from these experiments is useful.

MEMBRANE PROTEINS

Many proteins involved in cell sensing or signaling reside on the membrane. Because of their small diffusion constants, these membrane proteins can be subject to detection by localization, provided FP labeling does not perturb their structure or function. Single-molecule experiments can probe how such membrane proteins assemble and function to regulate cell behavior.

Assembly of Membrane Proteins

The translation studies discussed above relied on the use of membrane-localized FPs. Tsr is a membrane receptor existing in oligomeric forms. Although the above experiments use only monomeric Tsr-FP encoded in the *lac* operon for detection by localization, one could use similar experiments to study the assembly of the oligomeric protein. In fact, we observed under repressed conditions that newly produced Tsr-FP monomers were randomly distributed on the membrane. We know that the Tsr-FPs eventually migrate to one of the cell poles, as indicated by the image taken under induced conditions (**Figure 9*a***). The nascent Tsr-FPs appeared to mature and assemble initially on random membrane positions before migrating to the poles on a much longer timescale. Although fast photobleaching under continuous illumination did not allow us to determine the timescale of oligomerization, a time-lapse movie with adjustable delays between image acquisitions would track the assembly dynamics. In addition to oligomerization, the genesis of membrane proteins, including protein translation, folding, and insertion on the membrane, could also be well suited for studies using detection by localization.

Membrane Transporters

We have also recently applied detection by localization to study lactose permease, a membrane transporter for lactose encoded by the *lac* operon. Expression of the permease is controlled by the *lac* repressor responding to the level of inducer in the media. Under partial induction, genetically identical cells in a

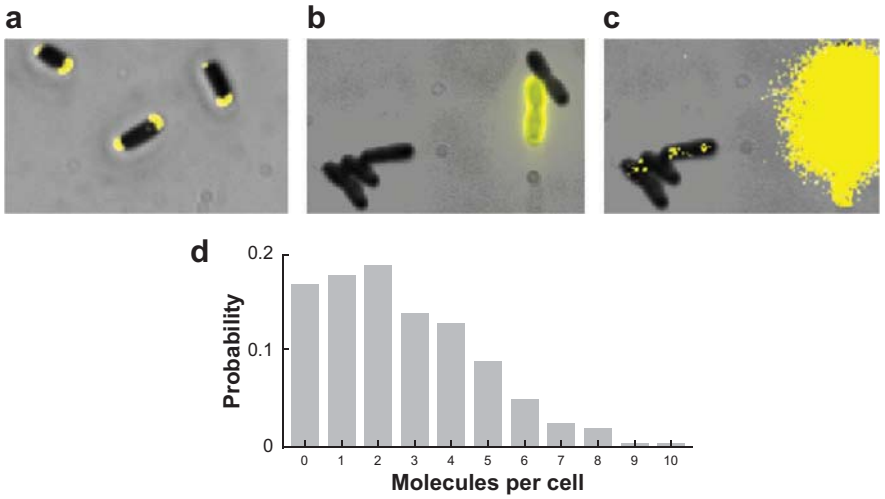

Figure 9

(*a*) Overlay of fluorescence on phase contrast image of induced Tsr-FP localized at the cell poles. (*b*) Overlay of fluorescence on phase contrast image of membrane-bound lactose permease-FP fusion proteins under a partially induced condition. Two induced cells with a high level of expression show FPs over the entire membrane. The other uninduced cells show little fluorescence when their signals are scaled to the induced cells. (*c*) The same image with single-molecule sensitivity after rescaling the signal of uninduced cells. Individual permease-FP molecules are now visible in the uninduced cells. Induced cells' signals saturate with this scaling. (*d*) Distribution of permease molecules per cell in the uninduced fraction of cells.

population can exhibit two different phenotypes, induced and uninduced, with extremely different expression levels of the permease (**Figure 9*b***). This variation in lactose metabolism can be beneficial for the survival of the entire population in an uncertain or changing environment. The transition from uninduced to induced state occurs when a cell randomly expresses a large enough burst of permease molecules, leading to a high influx of inducer that affects the *lac* repressor.

Induced cells have strong fluorescence along the cell membrane, whereas uninduced cells have minimal signal not detectable by conventional means. In the literature, it was proposed that one permease molecule was enough to drive the induction process (61). By directly counting the number of molecules in the uninduced population with single-molecule sensitivity, we found that there were 0 to 10 permease molecules in the uninduced fraction. This proves that one molecule is not enough to drive induction. Had one permease molecule been enough to drive the induction process, uninduced cells would only have had zero molecules. **Figure 9*d*** shows the distribution with burst frequency $a = 1.5$ and burst size $b = 1.7$ under a partially induced condition, corresponding to the $a > 1$ situation of **Figure 8*a***. Our measurement sets a lower bound of 10 on the threshold number of permease molecules needed to drive full induction (P.J. Choi, L. Cai, K. Frieda & X.S. Xie, manuscript in preparation). This gives an example of the importance of low-copy-number proteins involved in cell sensing or signaling in determining the phenotype and function of the cell.

REPLICATION

DNA replication occurs with high efficiency and fidelity at replisomes (**Figure 10*a***), where several proteins participate in the highly choreographed DNA synthesis at the replication fork. The *E. coli* replisome

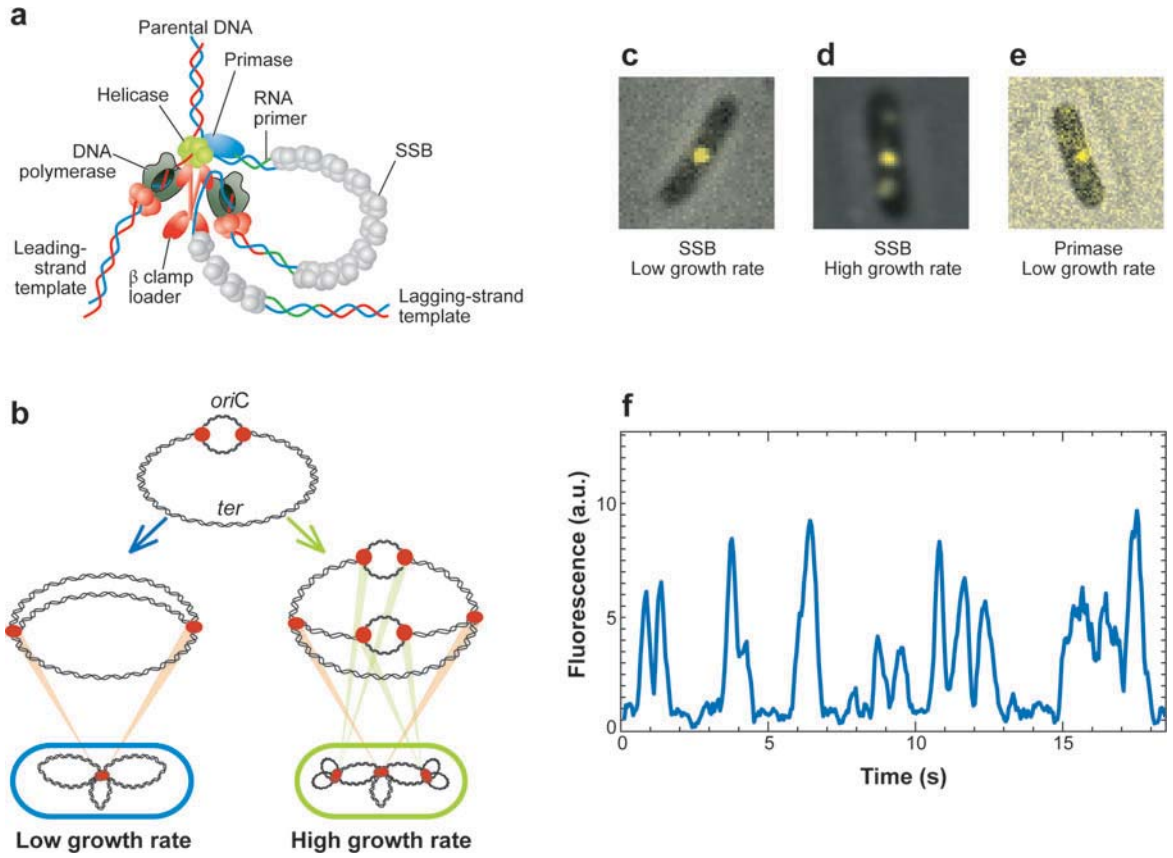

Figure 10

(*a*) Trombone model for the replication fork with the components of *Escherichia coli* replisome. Adapted from Reference 1. (*b*) Sketch of *E. coli* genome replication showing the replisome locations in *E. coli* cells at low and high growth rates. (*c*) Overlay of phase contrast and fluorescence images showing the location of Venus-labeled single-stranded DNA binding (SSB) proteins bound to the replisome at the cell center at a low growth rate. (*d*) At a high growth rate, Venus-labeled SSB proteins also appeared at one-quarter and three-quarter lengths of the cell. (*e*) Overlay of phase contrast and fluorescence images showing the binding of a single Venus-labeled primase molecule bound to the replisome in the center of an *E. coli* cell. (*f*) Fluorescence time trace of an *E. coli* cell center showing the events of transient binding and unbinding of individual Venus-labeled primase molecules at the replisome, with each cycle possibly corresponding to the formation of an Okazaki fragment.

consists of a DNA helicase that unwinds the double-stranded DNA, DNA polymerases that synthesize DNA on each of the two template strands, and single-stranded DNA binding (SSB) protein that protects single-stranded (ss)DNA prior to synthesis. The fact that both the leading- and lagging-strand DNA polymerase synthesize DNA from the 5′ to 3′ direction presents a topological difficulty, which has been resolved by the trombone model, in which a replication loop (**Figure 10*a***) is formed between the helicase and lagging-strand DNA polymerase (1). This model accounts for the continuous synthesis of DNA in the leading strand and discontinuous synthesis in the lagging strand. The replication loops are formed repetitively to produce short fragments of DNA called Okazaki fragments. Each Okazaki fragment is initiated with an RNA primer, which is synthesized

by the primase upon transient binding to the helicase.

The crystal structures of the major players of the replisome have been solved. Although most mechanic studies of the replisome have been carried out in vitro, recent single-molecule real-time observations on reconstructed replisomes have revealed significant mechanistic insights into DNA replication (49). Here we describe cellular observations of the location of the replisome and assembly of its components, as well as the study of the dynamics of DNA replication in individual living *E. coli* cells.

Replisome Localization

In *E. coli*, the genomic DNA is replicated bidirectionally by two replication forks starting from the origin of replication, termed *oriC* (**Figure 10*b***). The two replication forks move around the chromosome until they encounter the termination of replication sequence, termed *ter*. The two replication forks in bacterial cells are located near the cell center (5, 21, 77). In addition to the two replisome at the cell center, newly formed replisomes localized nearby at one-quarter and three-quarter lengths of the cell under high growth rates. Also, GFP-labeled DNA polymerase molecules can be imaged at the replisome location (51).

To demonstrate the assembly of protein components on the replisome, we first examined a Venus-labeled SSB protein. Because the cellular copy number of an SSB protein is several thousand (**Table 1**), labeling all copies of SSB proteins with Venus would prevent "detection by localization" of single SSB proteins bound to the replisome. We instead use a low copy plasmid to limit the copy number of labeled SSB proteins. **Figure 10*c*** shows the localization of a small number of Venus-labeled SSB proteins bound to ssDNA at the sites of the two replisomes under conditions of slower growth. It is evident that, ssDNA appears only at the replication fork. Under conditions of faster growth, we obtained additional spots at positions one-fourth and three-fourths along the cell length, consistent with the reported replisome locations in *E. coli* (5) and *B. subtilis* (51). The approach of partial labeling of SSB proteins for single-molecule detection by immobilization can be applied to other abundant proteins in a cell.

Dynamic of Lagging-Strand Synthesis

To probe the highly coordinated actions of lagging-strand DNA synthesis, we studied the primase, which has a low endogenous copy number (**Table 1**) and is subject to detection by localization using an FP-primase fusion. **Figure 10*e*** shows one *E. coli* cell having a single YFP-labeled primase bound to the replisome at the center at a particular moment. **Figure 10*f*** shows the fluorescence time trace at the center of a cell. The fluorescence signal is indicative of individual primase molecules transiently bound to the replisome, again based on detection by localization. Previously, in vitro biochemical studies have suggested that the interaction between the primase and helicase is transient in *E. coli* and that the lengths of Okazaki fragments have a broad distribution of 0.5 to 2 kb (19, 84). Assuming a DNA synthesis rate of ~600 nt s^{-1}, each Okazaki fragment requires 1–3 s to be produced, according to a recent single-molecule in vitro experiment (49). On the basis of the average time separation between two fluorescence bursts being ~2 s in **Figure 10*f***, we tentatively assign each fluorescence burst to the onset of each Okazaki fragment. This could well be the first direct observation of Okazaki fragments in real time in a living cell.

OUTLOOK

Opening up a molecular biology textbook, one sees the fundamental processes in molecular biology described by cartoons on a single-molecule basis even though our knowledge has been derived primarily from in vitro

studies. Thanks to the advances in FPs and single-molecule imaging, many of these processes can now be monitored in individual bacterial cells at the single-molecule level in real time. The wealth of information derived from these studies is new and highly quantitative, even for the well-characterized systems discussed here, yielding fresh insights into the intracellular workings of these processes.

We have discussed the single-molecule live-cell assay developments with the aim of providing a toolbox for bacterial molecular biology. Meanwhile, techniques will continue to evolve: Super-resolution imaging in live cells is emerging and new fluorescent proteins with short maturation times and high photostability will be developed. Similar single-molecule 3D imaging in live eukaryotic cells is being made possible by two-photon fluorescence microscopy (G.W. Li & X.S. Xie, manuscript in preparation).

The methods described in this article can be widely applicable and accessible. There are countless compelling problems that the single-molecule approach can address. We hope future single-molecule live-cell experiments will lead to new knowledge in molecular and cell biology for years to come.

SUMMARY POINTS

1. Many important cellular processes, such as transcription, translation, and replication, occur with low copy numbers of macromolecules and hence require single-molecule sensitivity to probe their dynamics.
2. The low copy numbers of macromolecules result in the stochastic behavior of biochemical reactions and molecular motions, which cannot be synchronized among a population of molecules or cells.
3. Detection by localization allows a single FP to be imaged upon binding on DNA or attaching to the cell membrane. Stroboscopic excitation allows detection of a single FP nonspecifically bound to DNA or freely diffusing in the cytoplasm with submillisecond time resolution. Tandem repeats of fluorophores can be used to visualize single DNA loci and mRNA molecules.
4. A transcription factor searches for a target sequence on the genome by repeated nonspecific binding on and 1D diffusion along different DNA segments, with a residence time of less than 5 ms, separated by much faster diffusion through the cytoplasm between two segments in less than 0.5 ms.
5. Under repressed conditions, mRNA molecules are generated in a Poisson process owing to random dissociation of the repressor from the operator. Under induced conditions, however, multiple mRNA molecules are generated in pulses of transcriptional activity.
6. Protein production occurs in bursts with one mRNA generating a few copies of protein molecules. The number of protein molecules produced per mRNA follows an exponential distribution.
7. Stochasticity in gene expression is manifested both in the random events of transcription or translation in time in a single cell, and in the variation of copy numbers of protein or mRNA per cell in a population of cells at a particular time. Both measurements give the same values for the transcription frequency and translational burst size.
8. Primary replication forks are located at the cell center. The assembly of protein complexes onto the replisome can be realized by detection by localization.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We are grateful to the former members of our group, Paul Blainey, Long Cai, Johan Elf, Nir Friedman, Ji Yu, Jie Xiao, and Kirsten Frieda, for their contributions to the work summarized in this article. We thank Jeremy Hearn for technical assistance. This work is supported primarily by the National Institute of Health Director's Pioneer Award program and in part by the Department of Energy, Office of Biological and Environmental Research, Genome-to-Life program.

LITERATURE CITED

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. 2002. *The Molecular Biology of the Cell*. New York: Garland. 1463 pp.
2. Alon U. 2006. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. New York: Chapman & Hall/CRC Press. 320 pp.
3. Ando R, Hama H, Yamamato-Hino M, Mizuno H, Miyawaki A. 2002. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc. Natl. Acad. Sci. USA* 99:12651–56
4. Azam TA, Iwata A, Nishimura A, Ueda S, Ishihama A. 1999. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.* 181:6361–70
5. Bates D, Kleckner N. 2005. Chromosome and replisome dynamics in *E. coli*: Loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell* 121:899–911
6. Berg O. 1978. A model for the statistical fluctuations of protein numbers in a microbial population. *J. Theor. Biol.* 71:587–603
7. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. 1998. Localization of ASH1 mRNA particles in living yeast. *Mol. Cell.* 2:437–45
8. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, et al. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* 313:1642–45
9. Blainey PC, van Oijen AM, Banerjee A, Verdine GL, Xie XS. 2006. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. USA* 103:5752–57
10. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–62
11. Bobst AM, Perrino FW, Meyer RR, Rein DC. 1985. Variability in the nucleic acid binding site size and the amount of single-stranded DNA-binding protein in *Escherichia coli*. *FEBS Lett*. 181:133–37
12. Bremer H, Dennis PP. 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In Escherichia coli *and* Salmonella*: Cellular and Molecular Biology*, ed. FC Neidhardt, R Curtiss, JL Ingraham, ECC Lin, KB Low, et al., pp. 1553–69. Washington, DC: ASM
13. Burgers PMJ, Kornberg A. 1983. The cycling of *Escherichia coli* DNA polymerase III holoenzyme in replication. *J. Biol. Chem.* 258:7669–75

14. Cabrera JE, Jin DJ. 2003. The distribution of RNAP in *Escherichia coli* is dynamic and sensitive to environmental cues. *Mol. Microbiol.* 50:1493–505
15. Cai L, Friedman N, Xie XS. 2006. Stochastic protein expression in individual cells at the single molecule level. *Nature* 440:358–62
16. Chudakov DM, Lukyanov S, Lukyanov KA. 2005. Fluorescent proteins as a toolkit for in vivo imaging. *Trends Biotechnol.* 23:605–13
17. Chudakov DM, Lukyanov S, Lukyanov K. 2007. Tracking intracellular protein movements using photoswitchable fluorescent proteins PS-CFP2 and Dendra2. *Nat. Protoc.* 2:1–9
18. Cluzel P, Surette M, Leibler S. 2000. An ultrasensitive bacterial motor revealed by monitoring signal proteins in single cells. *Science* 287:1652–55
19. Corn JE, Berger JM. 2006. Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions. *Nucleic Acids Res.* 34:4082–88
20. Dalal R, Larson MH, Neuman KC, Gelles J, Landick R, Block SM. 2006. Pulling on the nascent RNA during transcription does not alter kinetics of elongation or ubiquitous pausing. *Mol. Cell* 23:231–39
21. Danilova O, Reyes-Lamothe R, Pinskaya M, Sherratt D, Possoz C. 2007. MukB colocalizes with the oriC region and is required for organization of the two *Escherichia coli* chromosome arms into separate cell halves. *Mol. Microbiol.* 65:1485–92
22. Deich J, Judd EM, McAdams HH, Moerner WE. 2004. Visualization of the movement of single histidine kinase molecules in live Caulobacter cells. *Proc. Natl. Acad. Sci. USA* 101:15921–26
23. Drake JW. 1991. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. USA* 88:7160–64
24. Dunaway M, Olson JS, Rosenberg JM, Kallai OB, Dickerson RE, Matthews KS. 1980. Kinetic studies of inducer binding to the *lac* repressor-operator complex. *J. Biol. Chem.* 255:10115–19
25. Egner A, Geisler C, von Middendorff C, Bock H, Wenzel D, et al. 2007. Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters. *Biophys. J.* 93:3285–90
26. Elf J, Li G, Xie XS. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316:1191–94
27. Elmore S, Müller M, Vischer N, Odijk T, Woldringh CL. 2005. Single-particle tracking of *oriC*-GFP fluorescent spots during chromosome segregation in *Escherichia coli*. *J. Struct. Biol.* 151:275–87
28. Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002. Stochastic gene expression in a single cell. *Science* 297:1183–86
29. Elowitz MB, Surette MG, Wolf P-E, Stock JB, Liebler S. 1999. Protein motility in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* 181:197–203
30. Frame R, Bishop JO. 1971. The number of sex-factors per chromosome in *Escherichia coli*. *Biochem. J.* 121:93–103
31. Friedman N, Cai L, Xie XS. 2006. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* 97:168302
32. Galburt EA, Grill SW, Wiedmann A, Lubkowska L, Choy J, et al. 2007. Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* 446:820–23
33. Gilbert W, Müller-Hill B. 1966. Isolation of the *lac* repressor. *Proc. Natl. Acad. Sci. USA* 56:1891–98
34. Golding I, Paulsson J, Zawilski SM, Cox EC. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell* 123:1026–36

35. Gordon GS, Sitnikov D, Webb CD, Teleman A, Straight A, et al. 1997. Chromosome and low copy plasmid segregation in *E. coli*: visual evidence for distinct mechanisms. *Cell* 90:1113–21
36. Gotta SL, Miller OL Jr, French SL. 1991. rRNA transcription rate in *Escherichia coli*. *J. Bacteriol.* 173:6647–49
37. Guptasarma P. 1995. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*? *Bioessays* 17:987–97
38. Habuchi S, Ando R, Dedecker P, Verheijen W, Mizuno H, et al. 2005. Reversible single-molecule photoswitching in the GFP-like fluorescent protein Dronpa. *Proc. Natl. Acad. Sci. USA* 102:9511–16
39. Halford SE, Marko JF. 2004. How do site-specific DNA-binding proteins find their targets? *Nucleic Acid. Res.* 32:3040–52
40. Hess ST, Huang S, Heikal AA, Webb WW. 2002. Biological and chemical applications of fluorescence correlation spectroscopy: a review. *Biochemistry* 41:697–705
41. Hill TL. 2005. *Free Energy Transduction and Biochemical Cycle Kinetics*. Mineola, NY: Dover. 119 pp.
42. Johnson RC, Johnson LM, Schmidt J, Gardner JF. 2005. Major nucleoid proteins in the structure and function of the *Escherichia coli* chromosome. In *The Bacterial Chromosome*, ed. P Higgins, pp. 65–132. Washington, DC: ASM
43. Karu AE, Belk ED. 1982. Induction of *E. coli recA* protein via *recBC* and alternate pathways: quantitation by enzyme-linked immunosorbent assay (ELISA). *Mol. Gen. Genet.* 185:275–82
44. Kennell D, Riezman H. 1977. Transcription and translation initiation frequencies of the *Escherichia coli lac* operon. *J. Mol. Biol.* 114:1–21
45. Kim SY, Gitai Z, Kinkhabwala A, Shapiro L, Moerner WE. 2006. Single molecules of the bacterial actin MreB undergo directed treadmilling motion in *Caulobacter crescentus*. *Proc. Natl. Acad. Sci. USA* 103:10929–34
46. Lau IF, Filipe SR, Søballe B, Økstad OA, Barre FX, Sherratt DJ. 2003. Spatial and temporal organization of replicating *Escherichia coli* chromosomes. *Mol. Microbiol.* 49:731–43
47. Lawson CL, Swigon D, Murakami KS, Darst SA, Berman HM, Ebright RH. 2004. Catabolite activator protein: DNA binding and transcription activation. *Curr. Opin. Struct. Biol.* 14(1):10–20
48. Le TT, Harlepp S, Guet CC, Dittmar K, Emonet T, et al. 2005. Real-time RNA profiling within a single bacterium. *Proc. Natl. Acad. Sci. USA* 102:9160–64
49. Lee JB, Hite RK, Hamdan SM, Xie XS, Richardson CC, van Oijen AM. 2006. DNA primase acts as a molecular brake in DNA replication. *Nature* 439:621–24
50. Lee TH, Blanchard SC, Kim HD, Puglisi JD, Chu S. 2007. The role of fluctuations in tRNA selection by the ribosome. *Proc. Natl. Acad. Sci. USA* 104:13661–65
51. Lemon KP, Grossman AD. 1998. Localization of bacterial DNA polymerase: evidence for a factory model of replication. *Science* 282:1516–19
52. Lewis PJ, Thaker SD, Errington J. 2000. Compartmentation of transcription and translation in *Bacillus subtilis*. *EMBO J.* 19:710–18
53. Lia G, Bensimon D, Croquette V, Allemand JF, Dunlap D, et al. 2003. Supercoiling and denaturation in Gal repressor/heat unstable nucleoid protein (HU)-mediated DNA looping. *Proc. Natl. Acad. Sci. USA* 100:11373–77
54. Luo G, Wang M, Konisberg WH, Xie XS. 2007. Single-molecule and ensemble fluorescence assays for a functionally important conformational change in T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* 104:12610–15

55. Magde D, Elson E, Webb WW. 1972. Thermodynamic fluctuations in a reacting system—measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.* 29:705–8
56. McAdams HH, Arkin A. 1997. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA* 94:814–19
57. Min W, Jiang L, Yu J, Kou SC, Qian H, Xie XS. 2005. Nonequilibrium steady state of a nanometric biochemical system: determining the thermodynamic driving force from single enzyme turnover time traces. *Nano Lett.* 5:2373–78
58. Miyawaki A. 2005. Innovations in the imaging of brain functions using fluorescent proteins. *Neuron* 2:189–99
59. Myong S, Bruno MM, Pyle AM, Ha T. 2007. Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. *Science* 317:513–16
60. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A. 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* 20:87–90
61. Novick A, Weiner M. 1957. Enzyme induction as an all-or-none phenomenon. *Proc. Natl. Acad. Sci. USA* 43:553–66
62. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. 2002. Regulation of noise in the expression of a single gene. *Nature* 31:69–73
63. Peterman EJG, Brasselet S, Moerner WE. 1999. The fluorescence dynamics of single molecules of green fluorescent protein. *J. Phys. Chem. A* 103:10553–60
64. Qian H. 2006. Phosphorylation energy hypothesis: open chemical systems and their biological functions. *Annu. Rev. Phys. Chem.* 58:113–42
65. Raser JM, O'Shea EK. 2005. Noise in gene expression: origins, consequences, and control. *Science* 309:2010–13
66. Robbins-Manke JL, Zdraveski ZZ, Marinus M, Essigmann JM. 2005. Analysis of global gene expression and double-strand-break formation in DNA adenine methyltransferase- and mismatch repair-deficient *Escherichia coli*. *J. Bacteriol.* 187:7027–37
67. Rotman B. 1961. Measurement of activity of single molecules of β-D-galactosidase. *Proc. Natl. Acad. Sci. USA* 47:1981–91
68. Rowen L, Kornberg A. 1978. Primase, the dnaG protein of *Escherichia coli*. An enzyme which starts DNA chains. *J. Biol. Chem.* 253:758–64
69. Rust MJ, Bates M, Zhuang X. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* 3:793–95
70. Sako Y, Minoghchi S, Yanagida T. 2000. Single-molecule imaging of EGFR signalling on the surface of living cells. *Nat. Cell Biol.* 2:168–72
71. Shaner NC, Steinbach PA, Tsien RY. 2005. A guide to choosing fluorescent proteins. *Nat. Methods* 2:905–9
72. Slutsky M, Mirny LA. 2004. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* 87:4021–35
73. Smith C. 2007. Keeping tabs on fluorescent proteins. *Nat. Methods* 4:755–61
74. Thompson RE, Larson DR, Webb WW. 2002. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* 82(5):2775–83
75. Villani G, Pierre A, Salles B. 1984. Quantification of SSB protein in *E. coli* and its variation during RECA protein induction. *Biochimie* 66:471–76
76. Von Hippel PH, Berg OG. 1989. Facilitated target location in biological systems. *J. Biol. Chem.* 164:675–78
77. Wang X, Liu X, Possoz C, Sherratt DJ. 2006. The two *Escherichia coli* chromosome arms locate to separate cell halves. *Genes Dev.* 20:1727–31

78. Watanabe N, Mitchison TJ. 2002. Single-molecule speckle analysis of actin filament turnover in lamellipodia. *Science* 295:1083–86
79. Winter RB, Berg OG, von Hippel PH. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli lac* repressor-operator interaction: kinetic measurements and conclusions. *Biochemistry* 20:6961–77
80. Xie XS, Trautman JK. 1998. Single-molecule optical studies at room temperature. *Annu. Rev. Phys. Chem.* 49:441–80
81. Xie XS, Yu J, Yang WY. 2006. Living cells as test tubes. *Science* 312:228–30
82. Ying KH, Revzin A, Butler AP, O'Connor P, Noble DW, von Hippel P. 1977. Non-specific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound *Escherichia coli lac* repressor in vivo. *Proc. Natl. Acad. Sci. USA* 74:4228–32
83. Yu J, Xiao J, Ren X, Lao K, Xie XS. 2006. Probing gene expression in cells, one molecule at a time. *Science* 311:1600–3
84. Zechner E, Wu C, Marians K. 1992. Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. II. Frequency of primer synthesis and efficiency of primer utilization control Okazaki fragment size. *J. Biol. Chem.* 267:4045–53