CHAPTER 9

Using Single-cell RNA Measurements to Decipher the Stochastic Kinetics of Transcription

MENGYU WANG[†a,b], JING ZHANG[†a,b] AND IDO GOLDING*[a,b]

[a]Department of Physics, Department of Microbiology and Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA; [b]Verna & Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA
*E-mail: igolding@illinois.edu

9.1 Premise: "Understanding" Transcription

The aim of the studies described in this chapter is to better understand prokaryotic transcription as it occurs *in vivo*. By "understand", what we typically mean is, form a satisfying narrative for an observed phenomenon. As reductionist scientists, this narrative typically consists of describing (or more strictly, predicting) complex observables in terms of simpler components and interactions. As physicists, we add the requirement that the narrative is formulated mathematically rather than verbally, and is able to describe and predict quantitative, rather than qualitative, observables.

[†]Equal contribution authors

Chemical Biology No. 19
RNA Polymerases as Molecular Motors: On the Road, 2nd Edition
Edited by Robert Landick, Terence Strick and Jue Wang

Published by the Royal Society of Chemistry, www.rsc.org

Even within the parameters defined above, what constitutes a satisfying narrative depends entirely on a subjective, personal choice of the desired scale, resolution and precision of the sought-after description. One may set as the goal the ability to predict the identity (identified, *e.g.* by their 5′ and 3′ ends) and copy-number of all the different RNA molecules in a given cell, at any moment in time, given the *a-priori* knowledge of the genome sequence and the cellular concentration of all the relevant molecules—RNA polymerase (RNAP), transcription factors, nucleotides, *etc.* Once the infeasibility of this level of detail becomes obvious (or does it?), the narrative's requirements may be relaxed in different ways. For example, one may aim to predict only the population-averaged, rather than the single-cell, values; or the relative change in levels between two different conditions (*e.g.* temperature) rather than the absolute numbers under a given condition.

It must be said that, using these definitions, we cannot claim to "understand transcription" in all but the coarsest resolution and precision. Not at all. At the same time, as in other areas of scientific investigation, the identification of scales at which we can and cannot provide a successful narrative helps us define the current limits to our knowledge, our rate of progress, and how far we still have to go. In this chapter, we attempt to do that. We report on some recent progress towards the formation of a quantitative narrative of transcription, and point to open questions, addressing which will lead to further refinements of this narrative.

9.2 Single-cell Measurements of RNA Copy-number Can be Used to Learn About the Stochastic Kinetics of Transcription

The approaches for using fluorescence-based, single-cell RNA measurements to study transcription kinetics build on earlier studies that used analogous methods to analyze single-cell *protein* data in bacteria.[1–3] The analysis comprises the following steps.[4–6] First, the copy number of RNA from a gene of interest is measured in many individual cells, under well-defined growth conditions. In chemically fixed cells, this can be reliably done using single-molecule fluorescence *in situ* hybridization (smFISH), in which a set of fluorescently-labeled oligonucleotides target the RNA of interest[7,8] (Figure 9.1A). Under the microscope, the labeled RNAs will appear as fluorescent foci ("spots"). The total signal intensity of all spots in a given cell can then be measured and, properly calibrated, provides an estimate of the copy-number of the RNA of interest. The measured values from a few hundreds or thousands of cells then yield the distribution of RNA copy-number within the population (Figure 9.1A).

The key to the investigative approach is the statistical concept of ergodicity. Ergodicity posits an equivalence between the above "snapshot" of RNA numbers from a population of cells at a given instant, and the stochastic trajectory of an individual cell over time—namely the kinetics of RNA copy

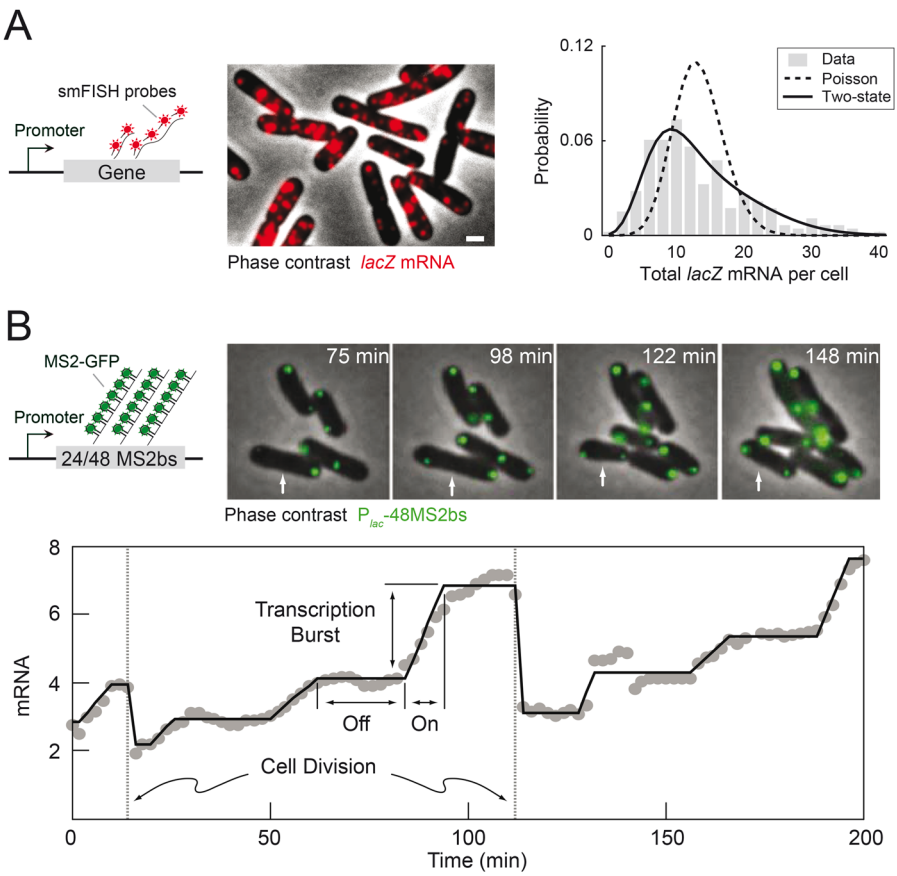

Figure 9.1 Probing the stochastic kinetics of transcription using single-cell RNA measurements. (A) The copy number of endogenous mRNA is measured in individual *E. coli* cells using single-molecule fluorescence *in situ* hybridization (smFISH). The resulting distribution is used to discriminate between alternative models for mRNA kinetics. The Poisson model, corresponding to the observed population average of ≈ 13 *lacZ* mRNA per cell, does not account properly for the observed variation in copy-number between individual cells. In contrast, the two-state model, which includes state-switching in the transcriptional state of the gene, successfully captures the observed variability. (B) In live cells, transcription is followed in real time *via* MS2–GFP labeling. The resulting time series shown below exhibits bursts of transcription, consistent with the broader-than-Poisson mRNA distribution found in (A). Panel B: Adapted from ref. 86 with permission from Taylor & Francis, Copyright 2012.

numbers, which we are seeking. The deduction of these single-cell properties is performed by writing down a theoretical model for the proposed stochastic kinetics of RNA. This model is then solved to yield the predicted distribution of RNA copy number in a population of cells. Comparing this prediction to the experimentally measured distribution allows testing the model's assumptions and estimating its kinetic parameters.

In fact, the assumptions regarding RNA kinetics can also be tested more directly, by measuring these kinetics in individual live cells (Figure 9.1B). RNA detection and quantification in living cells is achieved using an RNA binding protein (most commonly, the phage MS2 coat protein), fused to a fluorescent protein. The RNA of interest is transcriptionally fused to multiple tandem copies of the coat protein's binding site. When the RNA is expressed, it serves as a target for the binding of multiple copies of a fluorescent form of the coat protein [*e.g.* MS2–green fluorescent protein (GFP)].[9,10] The end result, as in the case of smFISH above, is the appearance of fluorescent "spots", whose signal intensity can be quantified and converted to RNA copy-number.[4] Performing this measurement in a single cell over time provides the stochastic time series of RNA levels, and the characteristics of this time series can then be compared with the expectations from the proposed model.[4]

An early example of how the approach above can be used to illuminate transcription kinetics was the demonstration of transcription bursts in *Escherichia coli.*[4] The prevailing model prior to that study was that, for a fully induced ("constitutive") promoter, RNA production can be approximated as a Poisson process, namely, molecules are produced one at a time, with a constant probability over time.[2] Consequently, the RNA copy-number distribution in a population of cells at a steady state level (*i.e.* when RNA production and elimination are balanced) should follow the Poisson distribution. However, when the distribution was measured experimentally, it was found to be much broader than expected according to the model (Figure 9.1A). Other predictions of the Poisson assumption were also contradicted by the experimental data, for example, the fraction of cells having no RNA during the transition from promoter inactivity to full induction.[4]

Confronted with the failure of the Poisson hypothesis, Golding *et al.*[4] then resorted to an alternative kinetic scheme known as the two-state (or "telegraph") model. In this model, the promoter switches stochastically between two states. RNA production takes place, stochastically, only in one of these states (the "active state"), whereas the other promoter state is inactive. Promoter state transitions, RNA production in the active state, and RNA degradation, all have constant probabilities, *i.e.* are Poissonian. The two-state model, while being more elaborate, can still be solved to predict RNA copy-number statistics. Unlike the Poisson distribution, the predicted copy-number distribution from the two-state model is consistent with the experimentally measured histograms (Figure 9.1A). The two-state model is also able to reliably predict how the fraction of zero-RNA cells will diminish during gene induction.[4]

In light of its success in reproducing the population statistics, the authors next attempted to test whether the two-state model is consistent with the RNA kinetics observed at the single-cell level. A key prediction of the model is that, when the promoter spends most of its time in the inactive state, RNA production will be "bursty", characterized by short periods of activity where multiple RNAs are produced, interspersed by longer periods of inactivity. This prediction was borne out by following RNA production in individual

cells over time (Figure 9.1B). Moreover, the two-state model predicts that, in the bursty regime, the durations of active (RNA production) and inactive (no RNA production) periods will be exponentially distributed, reflecting the stochastic Poissonian switching between the active and inactive states. The experimentally measured durations were consistent with this prediction.[4]

In addition to providing strong evidence for the two-state hypothesis, the comparison of model predictions and experimental data yields an estimation of the relevant physiological parameters. In the example above, the transcriptional burst size (*i.e.* the average number of RNA molecules produced during each active period) can be evaluated, not only from the single-cell kinetics, but also from the RNA copy-number distribution. Specifically, it can be shown that the burst size is approximately equal to the ratio of population variance in RNA number per cell to the mean of this number, a ratio known as the "Fano factor" (note that measuring this ratio requires the ability to count the *absolute* number of molecules, rather than merely quantifying relative levels of expression[4]). In fact, measuring the RNA Fano factor has become the go-to method for estimating the transcription burst size. It was used, for example, to evaluate how burstiness varies when gene expression level changes, and to compare the burstiness of different promoters.[6,11] The effects of gene activity regulators (*e.g.* transcription factors[12] or DNA supercoiling[13]) on transcription kinetics could also be deciphered by examining the changes in RNA copy-number distributions. Similar analyses have been successfully applied for the analysis of bursty transcription in eukaryotes as well.[14–17]

9.3 Caveats

As with any method of scientific investigation, it is crucial to note the limitations of the approach described above. These reflect what can be thought of (in analogy to quantum physics) as the "uncertainty principle of single-cell biology", namely, the recognition that each measurement is also a *perturbation* to the system under study. This results in inherent, unavoidable limitations to the fidelity of the measurements performed.

In that context, we note that the MS2 method for live-cell RNA detection involves a significant modification of the endogenous system: The presence of an array of binding sites downstream of the promoter of interest, and the subsequent binding of MS2–GFP to the nascent RNA, are likely to perturb multiple aspects of the gene expression process, including transcription elongation, RNA translation, degradation, and spatial organization. Some of these expected perturbative effects have already been demonstrated, for example, the fact that RNA degradation in *E. coli*[4] and yeast[18] is inhibited by MS2–GFP binding. Furthermore, there are considerable limitations to the sensitivity, accuracy, and dynamic range of the MS2 method.[4] Thus, while live-cell RNA measurements provide a vivid impression of "the thing itself" by allowing real-time characterization of RNA kinetics, they rarely reflect the dynamics of the endogenous system in a precise manner.

As for measurements on fixed samples, these offer the advantage of being applied to the endogenous (genetically unmodified) system. In addition, smFISH typically provides better RNA counting performance than MS2.[6,7] This is due to the lower variability in the label-to-target stoichiometry, and the lower levels of background cellular fluorescence. Of course, it is not possible to track the temporal trajectory of individual cells which have been chemically fixed. Instead, as described above, kinetic information is deduced by examining "snapshots" of copy-number statistics under different conditions and comparing those snapshots to the prediction of a theoretical model for RNA kinetics. Our deduction in this case is thus only as good as the model we conceived of. Critically, the kinetic models used are rarely if ever unique, *i.e.*, more than one model can fit the data equally well.[19] In choosing among models, one typically applies Occam's razor and selects the simplest one, but it is doubtful whether this criterion, so powerful in physics, is foolproof when it comes to living systems.[20] When modeling cell-to-cell variability, we also commonly attribute to stochasticity (randomness) features that may actually reflect *deterministic* differences between individual cells, *e.g.*, cell cycle phase or gene copy number, resulting in mistaken conclusions regarding RNA kinetics.[21] We will return to this point later.

It is also important to remember that the models used for analyzing single-cell data are typically phenomenological. In other words, they are constructed, "top-down", to reproduce the experimental data, rather than "bottom-up", beginning from our knowledge of microscopic interactions. Consequently, the level of detail in the model is not determined by the actual complexity of the biological system, or by how well we know the particulars of the system. Rather, model complexity reflects the quality of available experimental data. Thus, for example, data describing single-molecule resolution counting of RNA in individual cells (using MS2 or smFISH) has been used to test and refine models that describe the discrete events of RNA production,[4–6] improving on earlier models that relied on relative measurements of protein fluorescence level.[1,2] However, the above single-cell RNA measurements were still limited to a resolution of one whole RNA. The corresponding models, accordingly, were unable to describe the kinetics of RNA elongation and degradation, which were approximated as instantaneous (occurring at infinite speed). Only further improvements in the resolution of RNA measurements, down to the sub-molecule level (partial transcripts), can allow for testable models that describe elongation and degradation kinetics. Such measurements, and the corresponding models, are now emerging.[22–26]

9.4 From Individual Cells to Individual Gene Copies

In addition to the general caveats discussed above, measurements that yield the total number of RNA copies in a single cell suffer specific limitations in terms of their ability to inform us about RNA kinetics. This is because whole-cell RNA measurements are blind to multiple features of RNA life history. One such feature is the identity of the individual gene copy from which the

RNA is transcribed. In a population of exponentially growing bacteria, individual cells will sample different cell-cycle phases, and thus exhibit a twofold range of gene copy numbers (between 1 and 2 during slow growth, but up to 8 or even 16 in fast growth medium[27]). Since whole-cell measurements yield no information on individual gene copies, the analysis of the resulting data relies on assumptions regarding the number of copies and the degree to which individual copies act independently or are, instead, correlated with each other.[6,12,28]

Whole-cell RNA measurements are also blind to the age of individual RNA molecules present in the cell. Note that a typical mRNA in *E. coli* spends a significant fraction of its minutes-long life still tethered to the template gene as an actively transcribed (nascent) message, before being released as a complete (mature) mRNA and degraded.[22] Nascent and mature mRNAs are subject to different kinetics and spatiotemporal dynamics, but those differences are all hidden from whole-cell measurements that cannot distinguish the two mRNA species.

In light of these limitations of the current single-cell measurements, it should not surprise us that fundamental questions regarding the spatiotemporal dynamics of mRNA in *E. coli* are still unresolved. First, temporally, there is still no consensus as to what determines the stochastic kinetics of mRNA production from the promoter (*i.e.* the transcriptional time series), and, in particular, what is the mechanism leading to transcription bursts. The leading hypotheses are that bursts reflect transcription-factor binding and unbinding at the promoter,[28] the coupling of transcription to DNA supercoiling,[13] or an as-yet unknown genome-wide process leading to non-gene-specific (universal) relation between gene expression level and the observed burst size.[6,29] Each of these theories can claim some support from whole-cell RNA data. Therefore, improved measurement capabilities may be required to resolve them.

As for the spatial dimension, there is likewise still a debate as to where in the cell mRNA spends the duration of its lifetime. The textbook picture is that, upon completion, each transcript is immediately released from the template gene to diffuse freely in the cytoplasm, where it soon gets degraded.[30] However, single-cell experiments have provided conflicting evidence regarding the veracity of this picture, with results of one study indicating that mRNA never leaves the vicinity of the gene,[31] while results of later studies revealed mRNA distributed across the cytoplasm.[6,32] Messenger RNAs encoding inner-membrane proteins exhibit a tendency to localize to the membrane, but there is no consensus regarding the mechanism by which this localization is achieved.[32–34] Here, too, the current limitations on probing the life history of mRNA in a single cell stand in our way of resolving these disagreements.

9.5 Detecting Active Transcription from a Single Gene

To probe the spatiotemporal life history of mRNA in *E. coli* and address the open questions above, one would need to go beyond merely labeling and counting all mRNA from a given gene in the cell. Instead, it would be

necessary to measure the transcriptional activity of each individual copy of the gene, in order to examine how this activity depends on factors such as the presence of other copies of the same gene and the progression of the cell cycle. One would likewise need to discriminate nascent from mature mRNA, in order to characterize their different kinetics and spatial preferences.

In eukaryotic cells, active gene copies have been successfully identified on the basis of the presence of a strong RNA signal in the nucleus, corresponding to multiple nascent RNAs at the gene.[15,35–38] The presence of intronic RNA was likewise used to distinguish actively transcribed (pre-spliced) RNA from the mature species, where only exonic sequences can be detected.[39–41] However, the absence of a nucleus or introns prevents these methods from being directly applicable to bacteria.

As an alternative approach, inspired again by work in eukaryotes[42] as well as by a previous study in bacteria,[31] we decided to use the physical position of the gene locus as a fiducial marker for where active transcription takes place. We proposed the hypothesis that identifying and quantifying gene-proximal RNA would allow us to measure active transcription, whereas RNA further away would correspond to mature transcripts, which we would also measure in the same cell.[43]

We labeled the gene locus using the fluorescent repressor operator system (FROS) method,[44] in which an array of binding sites (here, 140 copies of *tetO*), engineered next to the gene of interest, is detected through the binding of the cognate transcription factor, fused to a fluorescent protein [here, tetracycline repressor–yellow fluorescent protein (TetR–YFP)] (Figure 9.2A). Other pairs of binding sites and proteins (*lacO*–LacI, *parS*–ParB) have been successfully used for the same purpose.[45,46] Consistent with previous reports, we find that the labeling scheme allows us to reliably measure the copy-number of the gene of interest in individual cells. By sorting the cells according to their length, we can track the gene copy number through the progression of the cell cycle (Figure 9.2A). Furthermore, by measuring the signal intensity of individual fluorescent foci (spots), we can detect unseparated sister copies (Figure 9.2A), which appear soon after gene replication.[44] This serves as a useful means of identifying the timing of gene replication, in both time-lapse movies of live cells (Figure 9.2B) and length-sorted fixed cells (see later).[43]

We next combined FROS gene tagging with the RNA labeling methods described above, namely MS2 in live cells and smFISH in fixed samples[43] (Figure 9.3). In the following paragraphs, we focus on our finding for the lactose promoter, P_{lac}. We have obtained similar results for P_{lac} and phage lambda P_R and P_{RM}. Applying the dual labeling approach to any promoter of interest is straightforward. In the analysis below, the live and fixed-cell methods for gene-plus-RNA detection are used to complement each other, as established previously for enumeration of RNA, to wit, precise measurements of an endogenous gene are made using smFISH, and interpreted using a theoretical model to yield kinetic schemes and parameter estimation. The live approach, less quantitative but more independent of model assumptions, is then used to validate the kinetic features inferred from fixed-cell data.

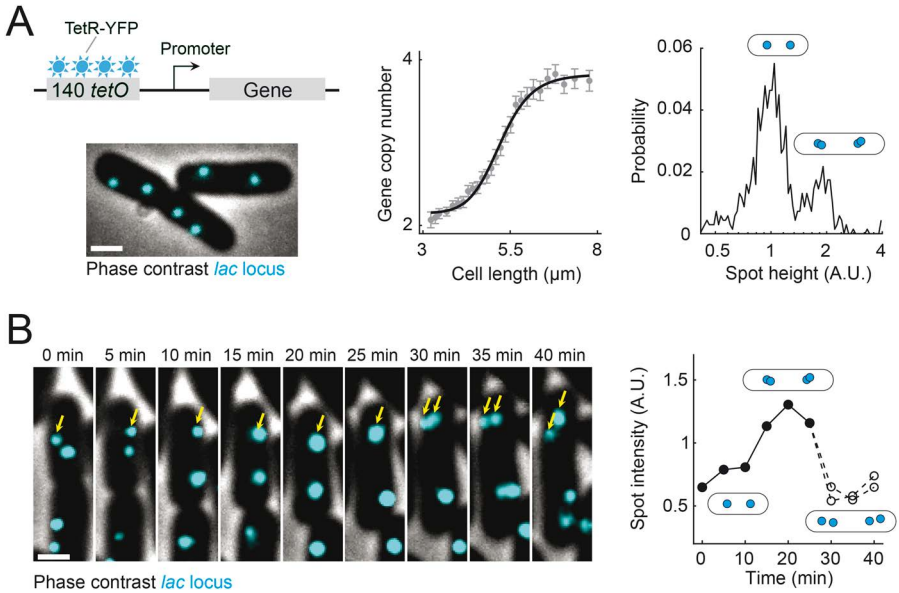

Figure 9.2 Detecting an individual DNA locus and the event of gene replication. (A) The locus of interest is labeled using the fluorescent repressor operator system (FROS). Sorting cells by length yields the locus copy number during the phase of the cell cycle. Measuring the fluorescent intensity of individual FROS spots allows the identification of newly replicated loci. (B) In time-lapse movies, gene replication is identified on the basis of the doubling of spot intensity followed by the spatial separation of the two sister copies.[43]

Our first test of the gene-plus-RNA labeling system is to ask whether it allows us to identify actively transcribed mRNA and distinguish it from the mature population. An indication that this is indeed the case is provided by examining the distance between each *lacZ* mRNA spot and the gene that encodes this mRNA (Figure 9.3A). The distribution of these distances revealed two distinct mRNA populations, one close to the gene and the other one further away. It is then plausible to propose that the gene-proximal population corresponds to nascent RNA, and this picture is supported by the following findings[43] (Figure 9.3A): (1) Under conditions of high expression, the gene-proximal signal is stronger than that from distal mRNA spots, consistent with the presence of multiple nascent mRNA molecules at the gene.[47] These nascent molecules are expected to appear as a single, diffraction-limited fluorescent spot, due to the limits of optical resolution. Previous smFISH-only experiments had already identified these putative transcription sites on the basis of their signal intensity.[7] The proximity of these strong RNA signals to the encoding gene, as seen here, solidifies this interpretation. (2) Two-color smFISH reveals that the gene-proximal mRNA is enriched for the 5′ region of the gene, as would be expected from the presence of partially transcribed

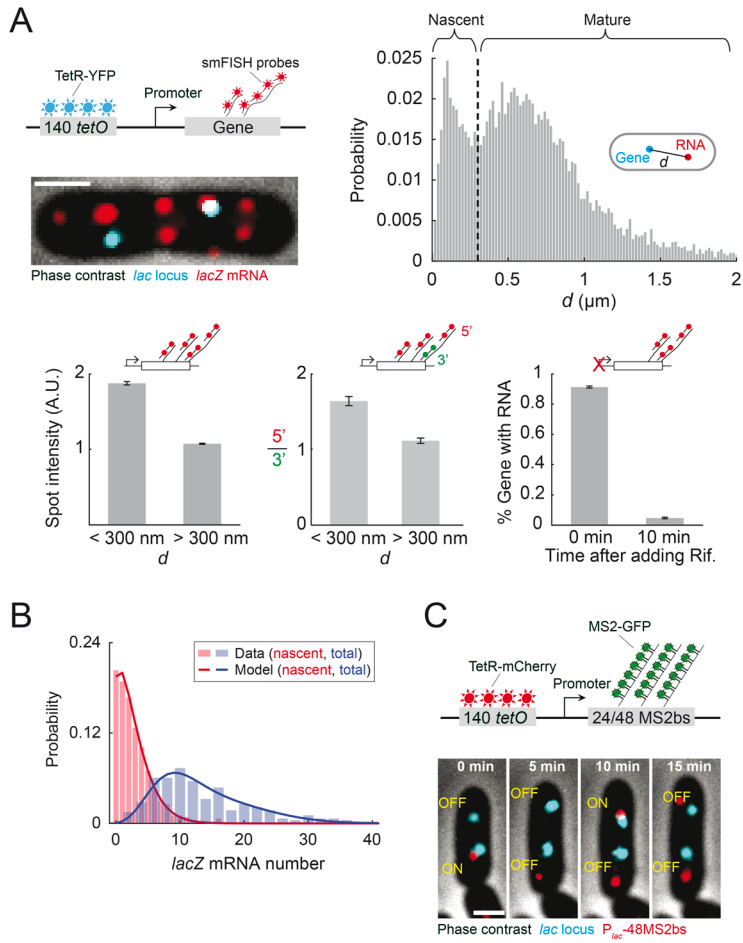

Figure 9.3 Detecting active transcription at a single gene copy. (A) Simultaneous labeling of endogenous mRNA and the gene that encodes it in fixed cells. The distribution of RNA-to-gene distances indicates the existence of gene-proximal and gene-distal mRNA populations. Bottom: Evidence supporting the identification of gene-proximal mRNA as actively transcribed (nascent) molecules: the gene-proximal RNA signal is stronger, is enriched for the 5′ region of the gene, and is eliminated following rifampicin treatment. (B) Nascent (per gene copy) and total (per cell) mRNA are measured to yield the copy-number distributions, which are then compared with the prediction by a theoretical model for mRNA kinetics. (C) In live cells, the transcriptional state (ON/OFF) of each gene copy can be followed over time.[43]

(incomplete) mRNA.[47] (3) On the other hand, when transcription initiation is inhibited using rifampicin, the gene-proximal mRNA signal disappears within a few minutes, consistent with the completion and release of the transcripts already initiated.[22] (4) The gene-proximal population is depleted altogether when the labeled locus and mRNA correspond to different genes.[43]

The evidence thus indicates that using an RNA-to-gene distance criterion allows us to classify mRNA molecules in the cell as nascent or mature. When applying this procedure, we computationally correct for the probability of random co-localization of gene and mRNA, which would lead to overcounting of nascent mRNA.[43] In fixed samples, we next quantify the nascent mRNA signal (following the same approach used for smFISH-only data[7]) to obtain the copy-number distribution across individual cells (Figure 9.3B). In live cells, we use the distance criterion to establish whether a given gene copy is transcriptionally active or not, and then follow the activity of individual genes over time (Figure 9.3C).

9.6 Analyzing Nascent mRNA to Reveal the Kinetics of Initiation, Elongation, and Degradation

We next wish to leverage these measurement capabilities to evaluate physiological parameters and test mechanistic hypotheses regarding some of the processes acting on mRNA. To do so, we again take our cue from physics and engineering, where a canonical way to probe the function of a system is to sharply perturb it from its steady state, and then follow the system's dynamics as it gradually approaches a new steady state (or returns to the pre-perturbation state).[48] The ability to test assumptions about the function of the system under study comes about, again, by forming a mathematical abstraction of these assumptions and comparing model predictions with the experimentally measured behavior. Beyond its application in human-made systems, the perturbative approach has also been successful in the interrogation of living ones, notably, characterizing chemotactic adaptation in *E. coli.*[49,50]

Here we perturb the cell by inducing (turning on) P_{lac} using the gratuitous inducer isopropyl-β-D-thiogalactoside (IPTG), which drives unbinding of the Lac repressor.[51] This procedure is commonly used to probe transcription kinetics[4,52] (the inverse scenario, turning the gene off, can also be used to obtain kinetic insights[51]). P_{lac} induction using IPTG provides a generous dynamic range of more than 100-fold (from a mean expression level of less than 0.1 to more than 10 *lacZ* mRNA per cell) provided that cyclic adenosine monophosphate (cAMP) metabolism is genetically abrogated and cAMP receptor protein (CRP) is instead activated by providing cAMP exogenously.[6,53]

Figure 9.4A depicts the results of a typical induction experiment. At different times after adding IPTG, samples were taken and the cells fixed, labeled, and imaged. The analysis described earlier in this chapter was used to measure the amounts of nascent, mature, and total *lacZ* mRNA per cell, which were then population-averaged. Even before performing detailed analysis, a few features can be observed in the data. The amount of total *lacZ* mRNA per cell shows the expected continuous increase, eventually saturating (after a few mRNA lifetimes) at a steady-state level that reflects the balance of mRNA production

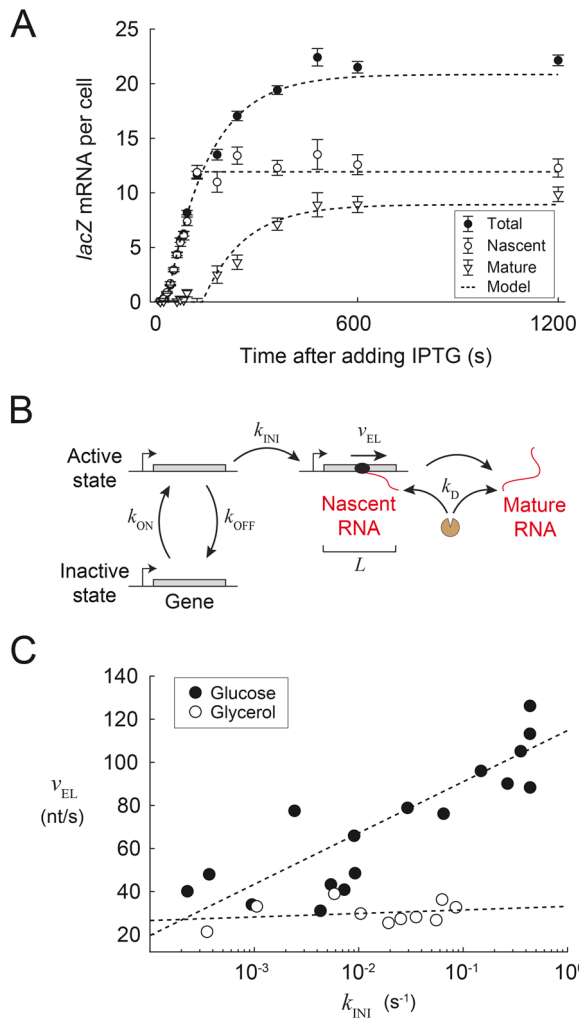

Figure 9.4 Analyzing nascent mRNA reveals the stochastic kinetics of transcript initiation, elongation, and degradation. (A) The levels of total, nascent, and mature *lacZ* mRNA following induction using IPTG, in cells grown in glucose. (B) A stochastic model for mRNA kinetics. The model is able to reproduce the experimental data in panel A. (C) The estimated mRNA elongation speed as a function of the transcription initiation rate (corresponding to different *lacZ* expression levels), for cells grown in glucose and glycerol.[43]

and degradation.[4] Mature mRNA, in contrast, only appears after a delay, corresponding to the time it takes to complete and release the first *lacZYA* transcript. At that same moment, nascent *lacZ* mRNA reaches its steady-state level, reflecting the balance of transcription initiation and release.[25]

To render these interpretations more rigorous and to estimate the underlying kinetic parameters, we formulated a mathematical model for the transcription

process[43] (Figure 9.4B). In the model, transcription initiation (at rate k_{INI}) is followed by mRNA elongation at a constant speed v_{EL}. Upon completion of the full transcript (of length L), nascent mRNA is released from the gene to become a mature one. Both nascent and mature mRNA are degraded, with degradation initiated at rate k_D and proceeding, from the 5′ end of the gene, at a speed that is also equal to $v_{EL.}$[22]

Here again, as in previous efforts discussed above, the level of detail included in the model is dictated by the resolution of the experimental data. Thus, for example, the processes of elongation and degradation, which we modeled as instantaneous in earlier work,[4–6] can now be considered more realistically thanks to the higher-resolution data. Our formalism, which explicitly describes each position (0–L) on the gene, follows the model used by Chen *et al.*[22] to interpret RNA-Seq data, but extends the original model to differentiate between nascent and mature species,[25] a distinction which the authors' experiments (and therefore, their model) did not allow. At the same time, the strong simplifying assumptions made in our model, for example, that the speeds of elongation and degradation are equal to each other and constant along the gene, also reflect the fact that our data is not sufficiently resolved to constrain more detailed assumptions.

Our model can be solved to obtain mathematical expressions for the population-averaged levels of nascent and mature mRNA (more specifically, the expected smFISH or MS2–GFP signals for each species) as a function of time after adding the inducer.[43] As seen in Figure 9.4A, our model indeed captures the essential features of mRNA kinetics during P_{lac} induction. Moreover, the resulting estimates for the rates of *lacZ* mRNA elongation (approximately 42 nucleotides s^{-1}) and degradation (approximately 0.008 s^{-1}) are consistent with measurements using total mRNA only, and with values reported in the literature.[6,22,32,54]

Despite being highly simplified, our model, when applied to the experimental data, illuminates a number of non-trivial biological features. First, by evaluating elongation speed at different induction levels, we found that this speed is correlated with the rate of transcription initiation (Figure 9.4C). This confirms a previous report, which attributed this correlation to RNAP–ribosome interactions.[54,55] Second, the assumptions made in the model (*i.e.* the existence of co-transcriptional degradation[22] and the equality between rate of elongation and degradation[22]) are strengthened by the failure to fit our experimental data using alternative models that do not include these assumptions.[43]

9.7 Evidence for Partial Retention of Mature mRNA at the Gene

We next performed similar measurements for cells grown in glycerol, a slow growth medium (Figure 9.5A). Surprisingly, we found that our model failed to predict the level of gene-proximal (presumably, nascent) *lacZ* mRNA, both

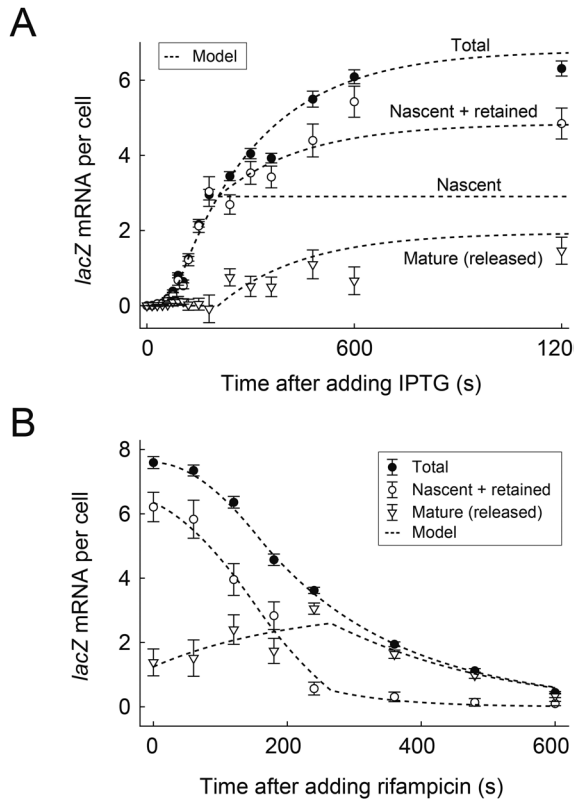

Figure 9.5 Partial retention of mature mRNA at the gene. (A) The levels of total, nascent, and mature *lacZ* mRNA following induction using IPTG, in cells grown in glycerol. The experimental data is reproduced by a model that includes the retention of approximately 55% of mature *lacZ* mRNA in the vicinity of the gene. (B) The levels of total, nascent, and mature *lacZ* mRNA following the addition of rifampicin. The transient increase in cytoplasmic (released) mRNA indicates that the mature mRNA retention depends on active transcription.[43]

during P_{lac} induction and at the steady-state level of expression. Specifically, the fraction of cellular mRNA that was localized to the encoding gene was higher than expected on the basis of the time it takes to complete one transcript. This observation was reminiscent of an earlier report of strong DNA–RNA colocalization at the *lac* locus under similar growth conditions.[31] In contrast to these results however, where the authors did not detect any non-gene-proximal mRNA, our data revealed both proximal and cytoplasmic populations, albeit not at the theoretically expected ratio.

To interpret these results, we added a feature to our theoretical model, where upon completion of transcript elongation a fraction of mature mRNAs remained in the vicinity of the gene rather than being released to diffuse in the cytoplasm. The hypothesis of mature mRNA retention is supported by

a few pieces of data.[43] First, the gene-proximal *lacZ* smFISH signal showed a lower 5′:3′ ratio than that expected for nascent mRNA, consistent with the presence of complete transcripts among the proximal mRNA. Second, we found that rapidly centrifuging the cells (4500 *g* for 5 min) restored the gene–proximal *lacZ* levels, and the 5′:3′ ratios, to the expected values for nascent (actively transcribed) mRNA, indicating that the mature population retained near the gene is not physically tethered to the DNA, in contrast to the actively transcribed molecules.[43] Solving the modified model yielded good agreement with the experimental data, during both induction (Figure 9.5A) and steady-state expression. The theoretical fit provided an estimate for the fraction of mature *lacZ* mRNA that is retained at the gene, approximately 55%.

At this stage, we can only speculate as to the mechanism of mature mRNA retention. It is well documented that multiple features of the bacterial cell are modulated by growth conditions, including the spatial organization of the nucleoid, RNAPs, and ribosomes,[56] as well as the fluidity of the bacterial cytoplasm.[57] mRNA retention may be a consequence of any of those effects. Interestingly, following mRNA kinetics after inhibiting transcription initiation using rifampicin revealed a transient accumulation of cytoplasmic *lacZ* mRNA prior to the expected decline in levels (Figure 9.5B). This suggests to us that mature mRNA retention depends on active transcription. This dependence, in turn, highlights yet again the tight spatiotemporal coupling between the different genomic processes of transcription, translation, mRNA release (and, as we see below, gene replication), a coupling which poses a challenge to identifying the mechanism for any single phenomenon.

9.8 A Stochastic Kinetic Model Captures Single-cell Statistics

The analysis above used a deterministic model for mRNA kinetics, and was applied to the population-averaged experimental values. All stochastic effects, and the resulting differences between individual cells, were ignored. As in the case of whole-cell RNA measurements discussed earlier, we can next formulate a stochastic version of the mRNA model (Figure 9.4B), and use it to capture the measured copy-number statistics of nascent mRNA. Specifically, in this new model, we describe the events of transcription initiation, mRNA release, and degradation, as stochastic. In addition, promoter state now follows the two-state kinetics introduced earlier.[4] Inclusion of this latter feature is further supported by the observed statistics of gene "on" and "off" durations when using the live MS2 FROS reporter.[43] We solved the stochastic model of mRNA kinetics and fitted it to the measured distributions of nascent and total *lacZ* mRNA, during induction[43] and at steady-state levels of expression (Figure 9.3B above). This fitting procedure yielded estimates for the model parameters. In particular, it revealed that the probability of

switching from inactive to active promoter state, k_{ON}, is the main feature modulated as P_{lac} expression is increased. This result is supported by the direct observation of promoter kinetics in live cells,[43] and reminiscent of similar reports in other organisms.[16,36,41,58]

9.9 Sister Gene Copies Can be Highly Correlated in Their Transcriptional Activity

Two or more sister copies of a gene are often present in the same cell. Under slow-growth conditions, this will only happen during the latter part of the cell cycle, after the gene has replicated. At shorter doubling times, the simultaneous presence of multiple replication forks results in the presence of two or more copies of each gene even in newborn cells. Consequently, a locus may be present at anywhere from 1 to 16 copies per cell, depending on the genome position, bacterial growth rate, and cell-cycle phase.[27]

Whereas each copy of the gene is regulated locally, through the binding of transcription factors and RNAP, early studies of stochastic gene expression evaluated the hypothesis that fluctuations in the cytoplasmic concentration of these actors would result in correlations between the activities of individual gene copies.[59] These correlations were conceptualized using the term "extrinsic noise" (to contrast from "intrinsic noise", which describes the uncorrelated fluctuations of individual gene copies),[59] and were probed using a two-color reporter, comprising two spectrally-distinct fluorescent proteins, driven by identical promoters, and placed symmetrically on the left and right sides of the chromosome from the origin of replication.[1] Measuring the relative signals of the two proteins provided a window into possible copy–copy correlations. However, despite the ingenuity of the engineered system, the two reporter genes, unlike true sister copies, do not share the same genomic context. It is thus not obvious that the measurements reflect the behavior of endogenous sister copies in the replicating chromosome.

Our novel ability to examine transcription at a single gene copy now allows us to directly measure the correlation between sister copies. When doing so, we observed a dramatic difference in the behavior of P_{lac} between cells grown in glucose and glycerol. Focusing on induction conditions where the fraction of active gene copies is approximately ½, and examining only those cells having two P_{lac} copies, we found that, in glucose, the number of active copies per cell closely followed a binomial distribution, as would be expected if gene copies are independent in their activity (Figure 9.6). However, this was not the case in glycerol. There, most two-copy cells had both P_{lac} copies in the same transcriptional state (either active or inactive), with only a small minority of cells showing mixed activity of the two copies (Figure 9.6). Consistent with these observations, the measured correlation in activity was low in glucose and high in glycerol. A similar trend was seen when comparing the measured correlation in nascent *lacZ* levels, as well as when measuring the temporal cross-correlation between P_{lac} copies in live cells.[43]

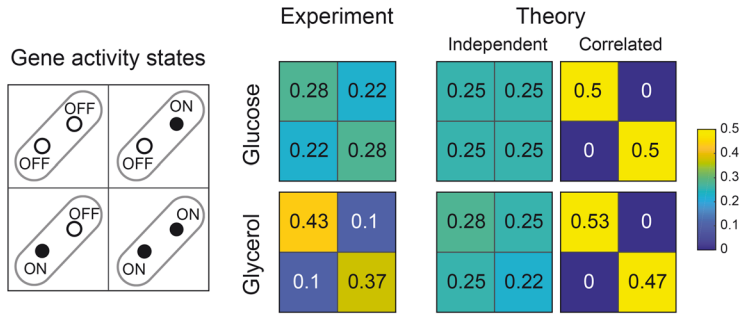

Figure 9.6 The correlation in transcriptional activity of two sister gene copies. The activity state (ON/OFF) of each P_{lac} copy was measured in cells having two copies of the locus. In glucose (top row), the distribution of number of active copies is consistent with each copy acting independently. In glycerol (bottom row), the two sister copies are highly correlated.[43]

Expanding our measurements beyond these two samples revealed a complex dependence on the experimental parameters.[43] Some of these dependencies indicated that attributing sister-copy correlations to fluctuations in the cytoplasmic concentration of an upstream regulator may be misplaced, or at least incomplete. For one, we found that a different promoter, phage lambda P_R, exhibited a similar trend of higher correlation in a slower growth medium. The behavior of P_R, which in the absence of the lambda repressor is considered "constitutive",[60] indicates that a gene-specific transcription factor is not the regulator whose fluctuations create correlations. This still leaves open the possibility of fluctuations in a genome-wide regulator, such as RNAP[61] or guanosine tetraphosphate (ppGpp).[62] However, we also found that sister-copy correlation strongly depended on the genomic locus of the promoter.[43] This appears to go against any cytoplasmic regulator as the single driver of correlations, and points, perhaps, to a possible role for temporal changes in the local chromosomal environment,[63] which would create a common regulatory effect on sister copies in two chromosomes.

9.10 Transcription from a Repressed Promoter is Coupled to the Event of Gene Replication

The way in which promoter activity in *E. coli* depends on the cell cycle remains an open question. *A priori*, cell-cycle progression involves multiple changes that could plausibly affect transcription. Some of these changes are, to a good approximation, continuous, *e.g.* cell volume, genome size, the concentrations of RNAP and other regulators, whereas others are discrete, *e.g.* cell division and the replication of the gene of interest and the genes encoding its upstream regulators. The possible consequences of these cell-cycle events for bacterial gene expression have received recent theoretical attention.[64–66] but experimental progress has lagged.

Of particular interest, the event of gene replication has long been speculated to affect that gene's transcription.[67] Early studies, using populations of synchronized bacteria, detected an increase in expression (beyond the obvious doubling of dosage) at a time corresponding to gene replication,[68,69] although no similar effect has been reported in the years since using newer experimental approaches. Multiple mechanisms that could couple gene replication and its transcription have been put forward.[67] These include changes in supercoiling of DNA surrounding the replication fork; movement of the replicated gene to the surface of the nucleoid, increasing its accessibility to RNAP; and transient de-repression of the new gene copy (or of both old and new copies). Regarding the latter, the idea that a tightly-bound transcription factor would only unbind DNA during passage of the replication fork is supported by single-cell measurements of P_{lac} repression by LacI[70] and of the binding kinetics of catalytically inactive Cas9 (dCas9).[71]

To probe the possible coupling between promoter activity and gene replication, we needed to identify the replication event within the cell cycle. In chemically-fixed cells, we first used length sorting as a proxy for cell-cycle progression.[72] Gene replication was then detected through the increase in FROS signal intensity, corresponding to the appearance of replicated, but not yet separated, sister copies (see Figure 9.2; note that the increase in the number of spots, indicating separation of the cohesive sister copies, is a lagging, and thus inefficient, indicator of replication[44]). Using this approach, we detected gene replication events at the expected approximately twofold length intervals,[72] and in a manner consistent with the expected dependence on distance from the origin of replication.[43]

We next proceeded to examine how the degree of active transcription, as indicated by the amount of nascent mRNA, changes during the cell cycle. The null hypothesis would be that the activity of each gene copy is constant throughout the cell cycle, and, therefore, that the level of nascent mRNA per cell closely follows gene dosage. Measuring the activity of a P_R reporter (in cells lacking the transcription factors to repress it), revealed just that: A doubling in the amount of nascent RNA is observed at the cell-cycle phase corresponding to gene replication (Figure 9.7A).

Very different behavior was revealed in P_{lac}, under conditions where the promoter is strongly repressed. Nascent mRNA levels increased transiently around the estimated time of gene replication (Figure 9.7A). This was the case for both the endogenous promoter and for reporters placed at other genomic loci.[43] A similar pulse of activity could be detected near the event of gene replication in live cells (Figure 9.7B). In fact, even in wild-type, genetically unmodified cells (strain MG1655), a higher propensity for transcription was observed in cells whose length corresponded to the cell-cycle phase of gene replication.[43]

Thus, our data provides evidence for an instance in which transcription from a repressed promoter is coupled to the replication of the gene. The data, however, does not offer an obvious means to distinguish between the

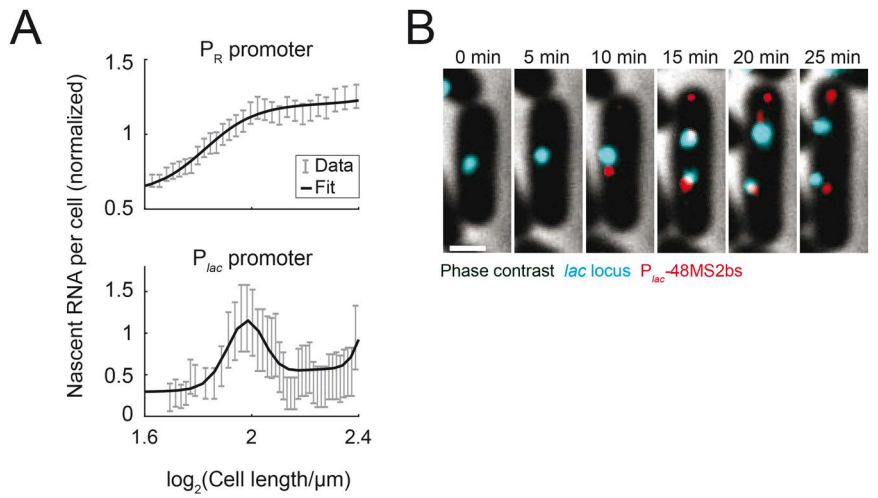

Figure 9.7 Promoter activity is coupled to gene replication. (A) In length-sorted fixed cells, nascent RNA from the P_R promoter (top) exhibits the expected doubling at the cell-cycle phase corresponding to gene replication. In contrast, a repressed copy of P_{lac} exhibits a pulse of transcription around gene replication. (B) In time-lapse movies, transcription from a repressed P_{lac} copy is observed close to the time of gene replication.[43]

proposed mechanisms for this coupling. Our current resolution does not allow us to determine the relative timing between replication and the increase in transcription, an observable that could yield useful information regarding the mechanistic questions posed. In terms of the magnitude of the effect, however, the phenomenology is quite clear, with the relative amplitude of the replication-induced pulse monotonically decreasing as gene expression level increases, until the effect becomes undetectable, as in highly expressed P_{lac} and P_R.[43]

9.11 Summary

In this chapter, we described our recent progress in using single-cell RNA measurements, interpreted using stochastic models of the underlying kinetics, to elucidate the spatiotemporal life history of mRNA in *E. coli.* As is evident from the discussion above, fundamental questions regarding those dynamics remain open.

In the temporal dimension, it is still unclear what drives the stochastic kinetics of mRNA production from the promoter. A large body of data is consistent with the picture of two-state kinetics, but the molecular underpinning of the active and inactive states is still debated.[13,28,29] Improvements in the temporal resolution of live-cell RNA measurements could help illuminate this question, and may reveal more elaborate kinetic schemes.[73] Importantly, it is also unclear how much of the observed population heterogeneity in mRNA

levels actually reflects the stochastic aspects of transcription, as opposed to other drivers of cell-to-cell differences. As described earlier, we demonstrated two such drivers, namely the coupling of transcription to gene replication and the correlation between sister copies. In the absence of proper analysis, the cellular heterogeneity in mRNA numbers that stems from these effects may be erroneously attributed to stochastic transcription.[21]

Spatially, the life history of mRNA is just as mysterious, part of our greater ignorance of how the flow of genetic information in *E. coli* is organized within the cell.[74] For example, beyond the partitioning of the cell between nucleoid and cytoplasm, is the cytoplasm itself well mixed, or do key gene-expression players localize to improve function? RNAP "transcription factories", localized to ribosomal RNA operons, have been well-documented,[75,76] but their functional significance is still unclear.[77] The *lac* repressor, too, has been reported to exhibit a heterogeneous spatial distribution in the cell.[78] Considering the exciting recent observations of RNAP and transcription-factor "hubs" in eukaryotes,[79–81] this promises to be a fruitful direction for future research. Subsequent to transcription, mRNA translation has been shown to be modulated by the different spatial preferences of DNA, RNAP, and ribosomes in the cell,[56,82,83] and would plausibly be affected by the partial retention of mature mRNA discussed above. Finally, mRNA degradation has been suggested to take place mostly at the cell membrane,[32] but direct evidence for this is still lacking. The spatial dimension of the Central Dogma thus still needs to be experimentally elucidated. The theoretical modeling of these aspects, too, is only at an early stage.[84,85]

Acknowledgements

Work in the Golding lab is supported by grants from the National Institutes of Health (R35 GM140709), the National Science Foundation (PHY 1147498, PHY 1430124 and PHY 1427654), the Welch Foundation (Q-1759) and the John S. Dunn Foundation (Collaborative Research Award). We gratefully acknowledge the computing resources provided by the CIBR Center of Baylor College of Medicine.

References

1. M. B. Elowitz, A. J. Levine, E. D. Siggia and P. S. Swain, *Science*, 2002, **297**, 1183–1186.
2. E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman and A. van Oudenaarden, *Nat. Genet.*, 2002, **31**, 69–73.
3. N. Friedman, L. Cai and X. S. Xie, *Phys. Rev. Lett.*, 2006, **97**, 168302.
4. I. Golding, J. Paulsson, S. M. Zawilski and E. C. Cox, *Cell*, 2005, **123**, 1025–1036.
5. C. Zong, L. H. So, L. A. Sepúlveda, S. O. Skinner and I. Golding, *Mol. Syst. Biol.*, 2010, **6**, 440.

6. L. H. So, A. Ghosh, C. Zong, L. A. Sepulveda, R. Segev and I. Golding, *Nat. Genet.*, 2011, **43**, 554–560.
7. S. O. Skinner, L. A. Sepúlveda, H. Xu and I. Golding, *Nat. Protoc.*, 2013, **8**, 1100–1113.
8. A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden and S. Tyagi, *Nat. Methods*, 2008, **5**, 877–879.
9. I. Golding and E. C. Cox, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11310–11315.
10. E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer and R. M. Long, *Mol. Cell*, 1998, **2**, 437–445.
11. Y. Taniguchi, P. J. Choi, G. W. Li, H. Chen, M. Babu, J. Hearn, A. Emili and X. S. Xie, *Science*, 2010, **329**, 533–538.
12. L. A. Sepúlveda, H. Xu, J. Zhang, M. Wang and I. Golding, *Science*, 2016, **351**, 1218–1222.
13. S. Chong, C. Chen, H. Ge and X. S. Xie, *Cell*, 2014, **158**, 314–326.
14. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas and S. Tyagi, *PLoS Biol.*, 2006, **4**, e309.
15. D. Zenklusen, D. R. Larson and R. H. Singer, *Nat. Struct. Mol. Biol.*, 2008, **15**, 1263–1271.
16. A. Senecal, B. Munsky, F. Proux, N. Ly, F. E. Braye, C. Zimmer, F. Mueller and X. Darzacq, *Cell Rep.*, 2014, **8**, 75–83.
17. K. Bahar Halpern, S. Tanami, S. Landen, M. Chapal, L. Szlak, A. Hutzler, A. Nizhberg and S. Itzkovitz, *Mol. Cell*, 2015, **58**, 147–156.
18. E. Tutucci, M. Vera, J. Biswas, J. Garcia, R. Parker and R. H. Singer, *Nat. Methods*, 2018, **15**, 81–89.
19. G. Neuert, B. Munsky, R. Z. Tan, L. Teytelman, M. Khammash and A. van Oudenaarden, *Science*, 2013, **339**, 584–587.
20. I. Golding, *Annu. Rev. Virol.*, 2016, **3**, 1–20.
21. I. Golding, *Curr. Opin. Microbiol.*, 2017, **43**, 9–13.
22. H. Chen, K. Shiroguchi, H. Ge and X. S. Xie, *Mol. Syst. Biol.*, 2015, **11**, 808.
23. S. Choubey, J. Kondev and A. Sanchez, *PLoS Comput. Biol.*, 2015, **11**, e1004345.
24. S. Iyer, B. R. Park and M. Kim, *Nucleic Acids Res.*, 2016, **44**, e142.
25. H. Xu, S. O. Skinner, A. M. Sokac and I. Golding, *Phys. Rev. Lett.*, 2016, **117**, 128101.
26. S. Kim and C. Jacobs-Wagner, *Biophys. J.*, 2018, **114**, 1718–1729.
27. H. J. Nielsen, B. Youngren, F. G. Hansen and S. Austin, *J. Bacteriol.*, 2007, **189**, 8660–8666.
28. D. L. Jones, R. C. Brewster and R. Phillips, *Science*, 2014, **346**, 1533–1536.
29. A. Sanchez and I. Golding, *Science*, 2013, **342**, 1188–1193.
30. B. Alberts, *Molecular Biology of the Cell*, Garland Science, New York, 4th edn, 2002.
31. P. M. Llopis, A. F. Jackson, O. Sliusarenko, I. Surovtsev, J. Heinritz, T. Emonet and C. Jacobs-Wagner, *Nature*, 2010, **466**, 77–81.
32. J. R. Moffitt, S. Pandey, A. N. Boettiger, S. Wang and X. Zhuang, *Elife*, 2016, **5**, e13065.

33. K. Nevo-Dinur, A. Nussbaum-Shochat, S. Ben-Yehuda and O. Amster-Choder, *Science*, 2011, **331**, 1081–1084.
34. S. Kannaiah, J. Livny and O. Amster-Choder, *Mol. Cell*, 2019, **76**, 574–589.
35. D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao and R. H. Singer, *Science*, 2011, **332**, 475–478.
36. H. Xu, L. A. Sepúlveda, L. Figard, A. M. Sokac and I. Golding, *Nat. Methods*, 2015, **12**, 739–742.
37. S. C. Little, M. Tikhonov and T. Gregor, *Cell*, 2013, **154**, 789–800.
38. K. Tantale, F. Mueller, A. Kozulic-Pirher, A. Lesne, J. M. Victor, M. C. Robert, S. Capozi, R. Chouaib, V. Backer, J. Mateos-Langerak, X. Darzacq, C. Zimmer, E. Basyuk and E. Bertrand, *Nat. Commun.*, 2016, **7**, 12248.
39. D. Y. Vargas, K. Shah, M. Batish, M. Levandoski, S. Sinha, S. A. Marras, P. Schedl and S. Tyagi, *Cell*, 2011, **147**, 1054–1065.
40. C. H. Hansen and A. van Oudenaarden, *Nat. Methods*, 2013, **10**, 869–871.
41. S. O. Skinner, H. Xu, S. Nagarkar-Jaiswal, P. R. Freire, T. P. Zwaka and I. Golding, *eLife*, 2016, **5**, e12175.
42. X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair and R. H. Singer, *Nat. Struct. Mol. Biol.*, 2007, **14**, 796–806.
43. M. Wang, J. Zhang, H. Xu and I. Golding, *Nat. Microbiol.*, 2019, **4**, 2118–2127.
44. M. C. Joshi, A. Bourniquel, J. Fisher, B. T. Ho, D. Magnan, N. Kleckner and D. Bates, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 2765–2770.
45. Y. Li, K. Sergueev and S. Austin, *Mol. Microbiol.*, 2002, **46**, 985–996.
46. I. F. Lau, S. R. Filipe, B. Soballe, O. A. Okstad, F. X. Barre and D. J. Sherratt, *Mol. Microbiol.*, 2003, **49**, 731–743.
47. S. L. French and O. L. Miller Jr, *J. Bacteriol.*, 1989, **171**, 4207–4216.
48. S. H. Strogatz, *Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering*, Addison-Wesley Pub., Reading, Massachusetts, 1994.
49. J. E. Segall, S. M. Block and H. C. Berg, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 8987–8991.
50. T. L. Min, P. J. Mears, I. Golding and Y. R. Chemla, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 9869–9874.
51. J. Elf, G. W. Li and X. S. Xie, *Science*, 2007, **316**, 1191–1194.
52. S. Kim, B. Beltran, I. Irnov and C. Jacobs-Wagner, *Cell*, 2019, **179**, 106–119 e116.
53. T. Kuhlman, Z. Zhang, M. H. Saier Jr and T. Hwa, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 6043–6048.
54. S. Proshkin, A. R. Rahmouni, A. Mironov and E. Nudler, *Science*, 2010, **328**, 504–508.
55. V. Epshtein and E. Nudler, *Science*, 2003, **300**, 801–805.
56. S. Mohapatra and J. C. Weisshaar, *Mol. Microbiol.*, 2018, **110**, 262–282.
57. B. R. Parry, I. V. Surovtsev, M. T. Cabeen, C. S. O'Hern, E. R. Dufresne and C. Jacobs-Wagner, *Cell*, 2014, **156**, 183–194.
58. O. Padovan-Merhar, G. P. Nair, A. G. Biaesch, A. Mayer, S. Scarfone, S. W. Foley, A. R. Wu, L. S. Churchman, A. Singh and A. Raj, *Mol. Cell*, 2015, **58**, 339–352.

59. P. S. Swain, M. B. Elowitz and E. D. Siggia, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12795–12800.
60. N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain and M. B. Elowitz, *Science*, 2005, **307**, 1962–1965.
61. S. Yang, S. Kim, Y. Rim Lim, C. Kim, H. J. An, J. H. Kim, J. Sung and N. K. Lee, *Nat. Commun.*, 2014, **5**, 4761.
62. V. Hauryliuk, G. C. Atkinson, K. S. Murakami, T. Tenson and K. Gerdes, *Nat. Rev. Microbiol.*, 2015, **13**, 298–309.
63. J. K. Fisher, A. Bourniquel, G. Witz, B. Weiner, M. Prentiss and N. Kleckner, *Cell*, 2013, **153**, 882–895.
64. J. R. Peterson, J. A. Cole, J. Fei, T. Ha and Z. A. Luthey-Schulten, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 15886–15891.
65. J. Lin and A. Amir, *Nat. Commun.*, 2018, **9**, 4496.
66. C. A. Vargas-Garcia, K. R. Ghusinga and A. Singh, *Curr. Opin. Syst. Biol.*, 2018, **8**, 109–116.
67. P. Guptasarma, *Bioessays*, 1995, **17**, 987–997.
68. P. Hanawalt and R. Wax, *Science*, 1964, **145**, 1061–1063.
69. J. M. Mitchison, *Science*, 1969, **165**, 657–663.
70. P. Hammar, M. Wallden, D. Fange, F. Persson, O. Baltekin, G. Ullman, P. Leroy and J. Elf, *Nat. Genet.*, 2014, **46**, 405–408.
71. D. L. Jones, P. Leroy, C. Unoson, D. Fange, V. Curic, M. J. Lawson and J. Elf, *Science*, 2017, **357**, 1420–1424.
72. M. Wallden, D. Fange, E. G. Lundius, O. Baltekin and J. Elf, *Cell*, 2016, **166**, 729–739.
73. C. Scholes, A. H. DePace and A. Sanchez, *Cell Syst.*, 2017, **4**, 97–108 e109.
74. M. Campos and C. Jacobs-Wagner, *Curr. Opin. Microbiol.*, 2013, **16**, 171–176.
75. T. Gaal, B. P. Bratton, P. Sanchez-Vazquez, A. Sliwicki, K. Sliwicki, A. Vegel, R. Pannu and R. L. Gourse, *Genes Dev.*, 2016, **30**, 2272–2285.
76. X. Weng, C. H. Bohrer, K. Bettridge, A. C. Lagda, C. Cagliero, D. J. Jin and J. Xiao, *bioRxiv*, 2018, 320481.
77. X. Weng, C. H. Bohrer, K. Bettridge, A. C. Lagda, C. Cagliero, D. J. Jin and J. Xiao, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 20115–20123.
78. T. E. Kuhlman and E. C. Cox, *Mol. Syst. Biol.*, 2012, **8**, 610.
79. M. Mir, A. Reimer, J. E. Haines, X.-Y. Li, M. Stadler, H. Garcia, M. B. Eisen and X. Darzacq, *Genes Dev.*, 2017, **31**, 1784–1794.
80. S. Chong, C. Dugast-Darzacq, Z. Liu, P. Dong, G. M. Dailey, C. Cattoglio, A. Heckert, S. Banala, L. Lavis and X. Darzacq, *Science*, 2018, eaar2555.
81. W.-K. Cho, J.-H. Spille, M. Hecht, C. Lee, C. Li, V. Grube and I. I. Cisse, *Science*, 2018, eaar4199.
82. A. Sanamrad, F. Persson, E. G. Lundius, D. Fange, A. H. Gynnå and J. Elf, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 11413–11418.
83. M. Stracy, C. Lesterlin, F. Garza de Leon, S. Uphoff, P. Zawadzki and A. N. Kapanidis, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, E4390–E4399.

84. M. Castellana, S. Hsin-Jung Li and N. S. Wingreen, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 9286–9291.
85. T. M. Earnest, J. A. Cole, J. R. Peterson, M. J. Hallock, T. E. Kuhlman and Z. Luthey-Schulten, *Biopolymers*, 2016, **105**, 735–751.
86. R. Phillips, H. Garcia, J. Theriot and J. Kondev, *Physical Biology of the Cell*, Taylor & Francis Group, 2nd edn, 2012.