

Eeny meeny miny moe, catch a transcript by the toe, or how to enumerate eukaryotic transcripts

Terence R. Strick^{1,3} and Nouria Hernandez^{2,3}

¹Institut Jacques Monod, CNRS UMR 7592, University of Paris-Diderot, 75205 Paris France; ²Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, CH-1015 Lausanne, Switzerland

In this issue of *Genes & Development*, Revyakin and colleagues (pp. 1691–1702) measure the relation between individual RNA polymerase II transcription events and transcription factor assembly by counting RNA transcripts retained on the template DNA using single-molecule fluorescence.

The marriage of biology, chemistry, mathematics, and physics has been a long and fruitful one. From the earliest days of optics and microscopy to the recent development of automated instrumentation with the power to resolve large numbers of individual molecules simultaneously, each discipline has found in the other a field in which to formulate and explore new ideas. Despite the ever-growing specialization required to perform at the cutting edge of each of these disciplines, research performed at their interfaces has served almost as a counterpoint to this specialization, underscoring the fundamental oneness of scientific knowledge. In this issue of *Genes & Development*, Revyakin et al. (2012) introduce the human RNA polymerase II (Pol II) transcriptional machinery to chemically modified glass coverslips sparsely coated with single DNA molecules and use advanced microscopy and imaging to count the number of times each separate DNA molecule is transcribed, thus laying the groundwork for quantitative characterization of eukaryotic transcription.

Gene regulation by the eukaryotic transcriptional machinery achieves specificity, in part, thanks to combinatorial libraries of proteins and promoter structure and sequence determinants. Six general transcription factors (GTFs; TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, and TFIIFH), which bind to the key core promoter elements, are required, in addition to Pol II itself, for “basal” levels of transcription; i.e., levels of transcription that can be detected in an *in vitro* transcription system. *In vitro*, their

order of assembly into the preinitiation complex (PIC) can be linear. Thus, on a TATA-box-containing promoter, TFIID is able to bind to the DNA on its own through binding of its TBP (TATA-box-binding protein) subunit to the TATA-box. This interaction is greatly stabilized by further binding of TFIIA and TFIIB. The binding of TFIIB then allows recruitment of Pol II and TFIIF onto the promoter, followed by binding of TFIIIE and TFIIFH (Orphanides and Reinberg 2002; Krishnamurthy and Hampsey 2009).

The first point made in this work is that the basal transcription reaction, dependent on the factors just mentioned, can be reconstituted on surface-tethered DNA templates, where the statistics of discrete transcription events can be obtained by single-molecule observation. Several important obstacles had to be overcome before components of the transcription machinery as well as its products could be enumerated at single-molecule resolution (Fig. 1). From a biochemical standpoint, just the six human GTFs involve upwards of 40 polypeptides and must be purified from either overexpression systems or human cell cultures. Furthermore, a surface chemistry compatible with the protein system is essential, as the high surface to volume ratio of reaction chambers makes some proteins prone to inactivation; for instance, by adsorption to the surface. The investigators describe a novel combination of standard silane-PEG (polyethylene glycol) functionalization, followed by treatment with polysiloxane to passivate glass surfaces. This modification prevents TFIID inactivation by standard silane-PEG surfaces, as shown by quantitatively comparing bulk with surface-based, enumerated transcription activities. DNA targets are fluorescently labeled at the surface-tethered end, allowing for their [x, y] localization to within tens of nanometers using low-background total internal reflection fluorescence (TIRF) microscopy and a sensitive electron-multiplying charge-coupled device (EMCCD) detector. The promoter used is a previously described, rationally designed “supercore” promoter (Juven-Gershon et al. 2006) with optimized TATA-box, initiator (Inr), motif ten element (MTE), and downstream promoter element (DPE), all of which favor efficient

[*Keywords*: single-molecule fluorescence, Pol II transcription, preinitiation complex, reinitiation, unstructured probes, surface passivation]

³Corresponding authors

E-mail terence.strick@ijm.univ-paris-diderot.fr

E-mail nouria.hernandez@unil.ch

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.199349.112>.

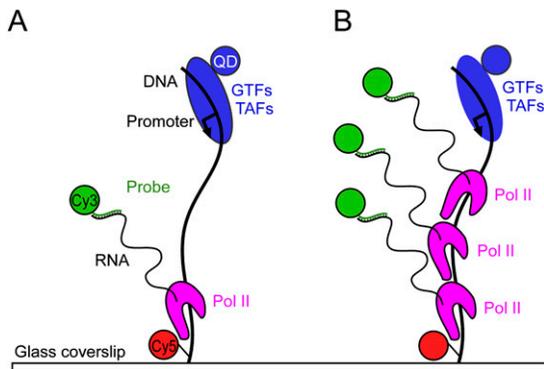


Figure 1. Multicolor, single-molecule observation of eukaryotic transcription. A DNA target bearing a promoter and labeled with a Cy5 fluorophore is detected and localized using single-molecule fluorescence imaging and localization. (A) If the DNA has undergone a single round of transcription, then only one Cy3-labeled probe complementary to the RNA transcript will be hybridized, colocalized with the DNA, and enumerated by single-step photobleaching. Template usage and initiation rates can be determined as a function of GTF or TAF presence by labeling the component of interest with a QDot (QD). (B) If the DNA has undergone several rounds of transcription, then that number can be determined by enumerating the hybridized, colocalized probes via detection of discrete photobleaching events.

TFIID binding. The supercore promoter on each DNA target directs transcription toward the surface, and the 5' end of the transcript contains an unstructured sequence that is probed for in situ with a complementary, fluorescently labeled probe. The transcription unit is long enough (on the order of a few hundred base pairs) that several elongating Pol II molecules can simultaneously be accommodated on the target DNA and captured at the surface, where they cannot terminate transcription.

A second EMCCD is dedicated to observing fluorescence of the labeled probe, and ultraresolution analysis of colocalization between the target DNA fluorescence and RNA probe fluorescence indicates whether or not an RNA is seen to colocalize with the DNA. Unbiased colocalization is used to assess probe specificity and efficiency, both of which are remarkably high (~100% and 80%, respectively). Counting RNA probe photobleaching steps provides a way of quantifying the number of transcripts per target DNA. Finally, a large imaging field of view must be obtained so as to observe enough DNA molecules (typically a few thousand in a 100- \times -100- μ m field of view) and maximize data collection. In this manner, Revyakin et al. (2012) were able to count the number of transcribed DNA molecules and the number of individual transcripts associated with each DNA molecule. The experiment consisted of localizing DNA templates in a field of view, introducing the Pol II transcription machinery, allowing it to work for several tens of minutes, hybridizing the RNA probe, and counting the number of probes per DNA.

Colocalization of DNA fluorescence and hybridization probe fluorescence indicates template utilization. As discussed further on, it was seen to be fully dependent on

promoter sequence and probe sequence and responds as expected to changes in DNA and protein determinants. The level of template utilization approached 40%, which was in good agreement with bulk biochemical results. Importantly, those molecules used are transcribed only once, on average, and the number of transcripts per DNA follows the Poisson distribution expected for this mean. This distribution is not affected by increasing Pol II, TFIID, or the other GTFs, indicating that the reaction is not rate-limited by the key reaction components. If transcription events are independent and uncorrelated, then indeed this Poisson distribution for the number of transcripts per used DNA can be expected. If transcription events are positively correlated—for instance, if TFIID stays bound for long periods and allows for repeated rounds of initiation—then one would expect the likelihood of the second transcription event to instead be greater than that of the first.

Revyakin et al. (2012) then used the properties of their assay to link binding of TFIID to transcription events. In these experiments, the DNA, TFIID, and RNA were probed for in succession. The DNA was labeled and localized and seen to photobleach in a single step as described before, and then the GTFs, Pol II, and ribonucleoside triphosphates (NTPs) were introduced into the reaction chamber. By using TFIID labeled with a quantum dot (QDot), long-term imaging of this component was accomplished. However, this comes at the cost of a photo-physical “blinking” behavior that these solid-state emitters can exhibit and that can cause their photon signal to fluctuate dramatically, leading to some uncertainty as to interpretation of the observed QDot-TFIID residence on a DNA template. Finally, the RNA was observed after the reaction by probe hybridization and was photobleached and enumerated as described earlier.

The arrival time for QDot-TFIID on a DNA template is slow and diffusion-limited in these experiments, and upon binding, the protein appears to reside on the DNA for extended periods of time, which is consistent with previous measurements in bulk reactions (Burley and Roeder 1996). However, after the GTF is seen to bind, the photon counts from QDot-TFIID undergo strong fluctuations. Although it is quite likely that this is due to QDot blinking behavior, it cannot yet be ruled out that the fluctuations correspond to transcription factor unbinding followed by rapid rebinding due to local concentration effects. Overall, this experiment indicates that in the absence of QDot-TFIID colocalization, <5% of DNA molecules show detectable transcription in this measurement set, whereas in the presence of QDot-TFIID colocalization, a full 25% of DNA molecules show detectable transcription. It is important to note that the 5% level of template utilization is also observed in the absence of key components of the reaction (e.g., TFIID or NTPs), meaning it likely corresponds to nonspecific probe-protein interactions and thus is an overestimate of template utilization. Finally, QDot-TFIID was observed to colocalize to a full 40% of DNA targets for which productive transcription was detected. Because TFIID is required for Pol II transcription, as verified by the

investigators, and because the QDot is tethered to TFIID via a potentially labile antibody–antigen interaction rather than a covalent interaction, it is likely that the remaining 60% of the DNA targets were bound to TFIID that was dissociated from its label and therefore simply not observed. Thus, to resume, the presence of QDot-TFIID makes it more than five times likelier for a given DNA molecule to support productive transcription than in the absence of QDot-TFIID.

In addition to reproducing factor-dependent basal transcription, Revyakin et al. (2012) showed that inclusion of the transcription factor Sp1 into the system leads to a twofold increase in transcription; i.e., to a low but clearly detectable level of transcription activation. Transcription activation encompasses a number of processes that rely on very different mechanisms. Perhaps the simplest and best-understood mechanism is the stabilization of the initiation complex by a transcription factor, an “activator,” that provides additional protein–DNA and/or protein–protein contacts. Many activators can stabilize the initiation complex by binding to a target DNA site and contacting members of the basal transcription machinery, including TBP-associated factors (TAFs) within TFIID. Sp1 is one such transcription factor and seems to activate transcription through a protein–protein contact with TAF4, at least in an *in vitro* transcription system. Thus, in previous work, Sp1-dependent transcription activation *in vitro* could be observed upon addition of TFIID or a partial TFIID complex containing TBP, TAF1 (TAF_{II}250), TAF2 (TAF_{II}150), and TAF4 (TAF_{II}110), but not upon addition of TBP alone or the partial TFIID complex lacking TAF4. This correlated with the ability of TAF4, but not TBP or other TAFs, to associate with Sp1 (Chen et al. 1994). The activation described by Revyakin et al. (2012) was observed only on a modified supercore promoter on which the initiator sequence had been weakened and did not result from an increased rate of reinitiation; in other words, the mean number of transcripts per DNA was unchanged. It is therefore highly likely that it reflects increased template utilization (in other words, an increased number of initiation complexes) and thus results from the transcription activation mechanism just described; i.e., stabilization of the initiation complex on the promoter.

In vivo, transcription activation is thought to result mostly from increased rates of reinitiation, which likely result at least in part from the persistent binding of some of the GTFs at the promoter after departure of the RNA polymerase. Previous studies have sought to determine what is left on the promoter after the first round of transcription, but in bulk biochemical assays, it is difficult to link such observations to reinitiation efficiency. Early studies combining assays on immobilized templates as well as competition assays showed that in a system consisting of purified GTFs including either TBP or TFIID and RNA Pol II, TBP or TFIID remained stably bound to the template after departure of the polymerase. The other GTFs were found to dissociate from the promoter: TFIIB was released upon addition of nucleotides, followed by TFIIE, TFIIIF, and TFIIH, all of which dissociated after transcription initiation; released TFIIIF was

capable of rebinding to stalled RNA polymerase (Zawel et al. 1995). As described above, Revyakin et al. (2012) found that in the most common pattern, TFIID binds relatively slowly to DNA and then remains bound for the duration of the experiment, consistent with the idea of TFIID binding being very stable. However, the jitter in the fluorescent signal prevented them from determining whether TFIID remained bound on the very promoters that had fired at least one RNA polymerase; i.e., that had produced an RNA molecule.

In any case, however, it is clear that more than just TFIID remaining bound to the promoter is needed to render reinitiation on the same template more efficient than the pioneer round of transcription initiation. In a yeast system, inclusion of the mediator and TFIIA in reactions performed on immobilized templates led to mediator itself, TFIID, TFIIA, TFIIE, TFIIH, and to a lesser extent, TFIIF remaining bound to the promoter after a single round of transcription, and this “scaffold” was more stable in the presence of a Gal4-VP16 activator bound to the promoter (Yudkovsky et al. 2000). Correlating with this larger stability of the scaffold, Gal4-VP16 increased transcription rates, suggesting an increased rate of reinitiation. The system described by Revyakin et al. (2012) should, in principle, allow the study of transcription activation by increased reinitiation, as it should be compatible with inclusion of additional factors in the transcription reactions, including various forms of the mammalian mediator and different activators. The effect of such additions on reinitiation could be monitored by counting the number of RNA molecules accumulated by each template.

More to come from single-molecule observation methods

In the past few years, methods using fluorescently labeled molecules have proven increasingly valuable for studying complex multicomponent assemblies at single-molecule resolution (note that similar approaches are also being adopted that do not have single-molecule sensitivity but are still extremely valuable) (Bonham et al. 2009). Named CoSMoS (for colocalization single-molecule spectroscopy), the single-molecule approaches have led to striking successes, including but not limited to real-time, single-molecule DNA sequencing (Eid et al. 2009); the direct observation of charged tRNAs binding to, and depleted tRNAs releasing from, the ribosome during polypeptide synthesis (Uemura et al. 2010); observation of bacterial transcriptional activation by the hexameric AAA⁺ protein NtrC (Friedman and Gelles 2012); and direct observation of spliceosome assembly and kinetics (Hoskins et al. 2011). Just as remarkably, the latter experiments have demonstrated that it is possible not only to study complexes reconstituted from purified components *in vitro*, but to make a large step toward *in vivo* conditions by working with complete cell extract in which the proteins of interest have been replaced by versions bearing fusion tags (e.g., Snap, Clip, and Holo tags). Several protein components can therefore be labeled simultaneously in

extract with different fluorophores, providing an attractive alternative to working with purified protein components in a reductionist approach. Despite the simultaneous presence of multiple components, discrete binding events can still be observed, and the kinetics of assembly can be quantified.

Indeed, although single-molecule enumeration of a complex reaction at its beginning and end is extremely informative, one of the unique strengths of surface-based single-molecule observation is its intrinsic ability to simultaneously monitor multiple components interacting in multistep reactions in real-time. In conventional biochemical assays, diffusion and unsynchronized molecular interactions can make the measurement of multistep reaction kinetics particularly challenging. Often, reaction conditions may need to be tuned to optimize each individual step at which the reaction is to be synchronized, leading to a multiplicity of experimental conditions and complicating global reconstruction of reaction kinetics. In addition, although order-of-addition experiments can allow one to tease apart the causal relationships between components of such reactions, they may themselves represent perturbations of the system.

In the appropriate single-molecule assay, many of these problems are eliminated, and a single experimental design can provide several observables. By simultaneously visualizing multiple components acting on an individual, spatially localized target, there is no need to synchronize a reaction, as the beginning and end of each binding event can, in principle, be directly observed and timed. Thus, microscopic on and off rates can both be measured in the same assay without changing experimental conditions. Observing components coming and going from their target will obviously also provide information on the order of assembly of a complex, and by using photobleaching-based enumeration as described earlier, one may even count how many instances of a given component are involved in a complex. Ultimately, one expects that the experimental data can even be used to derive a fuller kinetic description of multipathway systems in unified reaction conditions. Furthermore, by providing a complete and continuous timeline of the reaction, even transient and rare events can be detected and accounted for.

If applied to the question of PIC assembly, these methods could shed light on the order in which the basal transcription factors join various types of core promoters in a crude extract. In particular, one could determine whether TFIIA always joins the PIC and whether it does so before or after TFIIB or both. Similarly, the order of binding of a transcriptional activator relative to basal factors could be determined, leading to better insight into how the PIC is stabilized, as illustrated by Sp1 in the experiments of Revyakin et al (2012).

As interactions between distant regulatory elements and PICs presumably require DNA wrapping or looping to bring sites into contact with one another, it is clear that the mechanical properties of DNA and DNA wrapped into nucleosomes (i.e., chromatin) can also come into play as an additional "DNA determinant" for the regulation of gene expression. In fact, transcription reinitiation

seems to be dependent on the formation of a loop bringing the transcription start site in close proximity to the polyadenylation site in a manner dependent on TFIIB (Singh and Hampsey 2007; Perkins et al. 2008; Lainé et al. 2009).

In principle, such DNA looping interactions can be observed and controlled in single-molecule DNA nanomanipulation experiments using optical tweezers or magnetic trapping. More generally, these nanomanipulation approaches can provide for a direct, real-time readout of protein activity and have been used, for instance, to study promoter opening and escape (Revyakin et al. 2004, 2006), transcription elongation (Abbondanzieri et al. 2005), collisions with nucleosomes (Jin et al. 2010), and transcription termination (Larson et al. 2008) by *Escherichia coli* RNA polymerase. Optical trapping experiments have also recently provided insight into the mechanisms of Pol II elongation (Larson et al. 2012), stalling (Galburt et al. 2007), and interactions with nucleosomes (Hodges et al. 2009). Combining a nanomanipulation method with the single-molecule fluorescence-based detection of component presence as described by Revyakin et al. (2012) could help address questions of timing of complex activity as it relates to complex assembly. The main obstacle here is the required throughput: As eukaryotic transcription initiation is not a deterministic process, one must be able to watch large numbers of DNA molecules in order to detect with statistical significance the 20%–40% fraction that initiates transcription. Magnetic trapping, which can be scaled to work on numerous molecules in parallel, may provide a means for combining high-throughput single-molecule visualization and manipulation. However, even on their own, these novel high-throughput visualization methods hold enormous promise for detailed understanding of multicomponent reaction kinetics and show what is to come when the field of single molecules goes beyond single molecules.

References

- Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, Block SM. 2005. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**: 460–465.
- Bonham AJ, Neumann T, Tirrell M, Reich NO. 2009. Tracking transcription factor complexes on DNA using total internal reflectance fluorescence protein binding microarrays. *Nucleic Acids Res* **37**: e94. doi: 10.1093/nar/gkp424.
- Burley SK, Roeder RG. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem* **65**: 769–799.
- Chen JL, Attardi LD, Verrijzer CP, Yokomori K, Tjian R. 1994. Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell* **79**: 93–105.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* **323**: 133–138.
- Friedman LJ, Gelles J. 2012. Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation. *Cell* **148**: 679–689.

- Galbur EA, Grill SW, Wiedmann A, Lubkowska L, Choy J, Nogales E, Kashlev M, Bustamante C. 2007. Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **446**: 820–823.
- Hodges C, Bintu L, Lubkowska L, Kashlev M, Bustamante C. 2009. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* **325**: 626–628.
- Hoskins AA, Friedman LJ, Gallagher SS, Crawford DJ, Anderson EG, Wombacher R, Ramirez N, Cornish VW, Gelles J, Moore MJ. 2011. Ordered and dynamic assembly of single spliceosomes. *Science* **331**: 1289–1295.
- Jin J, Bai L, Johnson DS, Fulbright RM, Kireeva ML, Kashlev M, Wang MD. 2010. Synergistic action of RNA polymerases in overcoming the nucleosomal barrier. *Nat Struct Mol Biol* **17**: 745–752.
- Juven-Gershon T, Cheng S, Kadonaga JT. 2006. Rational design of a super core promoter that enhances gene expression. *Nat Methods* **3**: 917–922.
- Krishnamurthy S, Hampsey M. 2009. Eukaryotic transcription initiation. *Curr Biol* **19**: R153–R156. doi: 10.1016/j.cub.2008.11.052.
- Lainé J-P, Singh BN, Krishnamurthy S, Hampsey M. 2009. A physiological role for gene loops in yeast. *Genes Dev* **23**: 2604–2609.
- Larson MH, Greenleaf WJ, Landick R, Block SM. 2008. Applied force reveals mechanistic and energetic details of transcription termination. *Cell* **132**: 971–982.
- Larson MH, Zhou J, Kaplan CD, Palangat M, Kornberg RD, Landick R, Block SM. 2012. Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc Natl Acad Sci* **109**: 6555–6560.
- Orphanides G, Reinberg D. 2002. A unified theory of gene expression. *Cell* **108**: 439–451.
- Perkins KJ, Lusic M, Mitar I, Giacca M, Proudfoot NJ. 2008. Transcription-dependent gene looping of the HIV-1 provirus is dictated by recognition of pre-mRNA processing signals. *Mol Cell* **29**: 56–68.
- Revyakin A, Ebricht RH, Strick TR. 2004. Promoter unwinding and promoter clearance by RNA polymerase: Detection by single-molecule DNA nanomanipulation. *Proc Natl Acad Sci* **101**: 4776–4780.
- Revyakin A, Liu C-Y, Ebricht RH, Strick TR. 2006. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**: 1139–1143.
- Revyakin A, Zhang Z, Coleman RA, Li Y, Inouye C, Lucas JK, Park S-R, Chu S, Tjian R. 2012. Transcription initiation by human RNA polymerase II visualized at single-molecule resolution. *Genes Dev* (this issue). doi: 10.1101/gad.194936.112.
- Singh BN, Hampsey M. 2007. A transcription-independent role for TFIIB in gene looping. *Mol Cell* **27**: 806–816.
- Uemura S, Aitken CE, Korlach J, Flusberg BA, Turner SW, Puglisi JD. 2010. Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* **464**: 1012–1017.
- Yudkovsky N, Ranish JA, Hahn S. 2000. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**: 225–229.
- Zawel L, Kumar KP, Reinberg D. 1995. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev* **9**: 1479–1490.