NEWS AND VIEWS

RNA polymerase II transcription in living color

Angus I Lamond & Jason R Swedlow

Using a previously developed method, Singer and co-workers have now labeled RNA transcripts and quantitatively measured the synthesis of RNA in living cells. They find that only a small fraction of the RNA polymerases that bind the promoter actually produce messenger RNA. They also observed much faster elongation than previously reported, along with frequent long pauses.

Transcription by RNA polymerase II (Pol II) is fundamental to the mechanism of gene expression and a major target of biological regulatory mechanisms. It involves many separate steps, such as RNA Pol II molecules binding promoter sequences at the 5′ regions of genes, formation of initiation complexes, promoter escape, elongation and eventually termination and release of the RNA polymerase and the transcript from the DNA template. The structure of RNA Pol II and its interactions, both with promoter regions and with other transcription factors, have been studied in great detail. Much of this work has involved *in vitro* studies with purified components, which allow specific molecular interactions to be characterized in isolation. A variety of *in vivo* techniques have also been employed widely, including the use of reporter genes such as luciferase to measure changes in transcription activity and of chromatin immunoprecipitation (ChIP) assays to measure changes in the extent of protein binding to DNA under different conditions. In a study on page 796 of this issue, Darzacq *et al.*[1] have now taken analysis of the mechanism of gene transcription in eukaryotes to a new level by using advanced fluorescence-imaging techniques to quantitatively measure the kinetics of gene transcription by RNA Pol II in living mammalian cells[1]. The ultimate goal of this work is a quantitative model of gene transcription *in vivo*. The new

Angus I. Lamond and Jason R. Swedlow are in the Division of Gene Regulation and Expression, MSI/WTB Complex, University of Dundee, Dow Street, Dundee DD1 5EH, Scotland, UK. e-mail: angus@lifesci.dundee.ac.uk
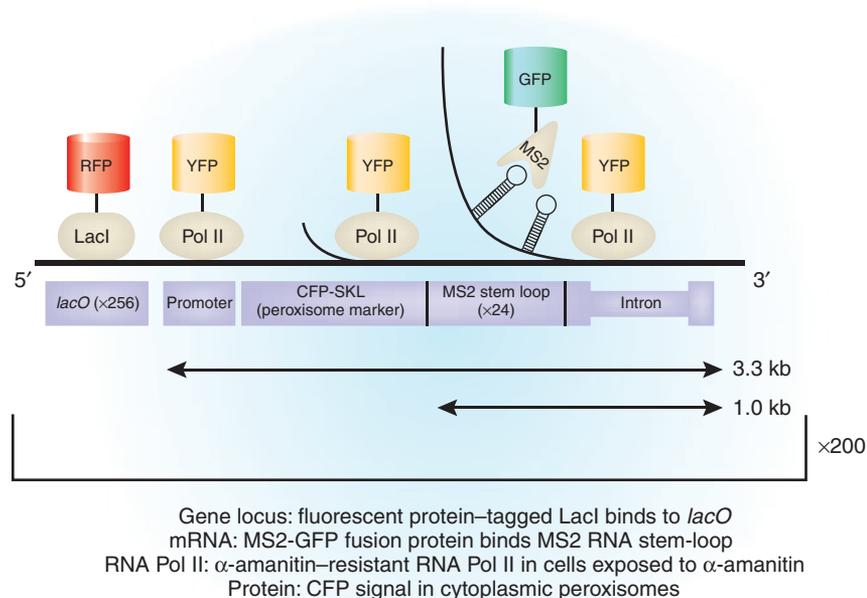

Gene locus: fluorescent protein–tagged LacI binds to *lacO*
mRNA: MS2-GFP fusion protein binds MS2 RNA stem-loop
RNA Pol II: α-amanitin–resistant RNA Pol II in cells exposed to α-amanitin
Protein: CFP signal in cytoplasmic peroxisomes

Figure 1 Cellular system for monitoring of Pol II transcription *in vivo*. Darzacq *et al.*[1] use a cell line bearing a tandemly repeated array containing 256 *lacO* cassettes to mark the gene locus, an inducible transcription-reporter fusion of the peroxisome marker SKL with CFP, and 20 MS2-binding sites[2]. This cell line is also transfected with sequences encoding red fluorescent protein (RFP)-LacI (or GFP-LacI), to fluorescently mark the gene array and YFP-MS2, which binds specifically to mRNA bearing the MS2 stem-loop insert.

study is an important step in this direction and provides quantitative estimates of the chemical and physical properties of RNA Pol II during transcription.

The authors have exploited an elegant experimental system to derive rate constants for separate steps in the transcription process, on the basis of measurements made in single mammalian cells using fluorescence microscopy and analyzed by mathematical

modeling. The approach uses an inducible reporter system in which a tetracycline-regulated gene array is stably integrated into a human osteosarcoma cell line[2]. The array contains ~200 copies of a gene encoding a 3.3-kilobase (kb) mRNA with a single intron at the 3′ end (**Fig. 1**). Transcription of the array is induced by adding doxycycline to the cell culture medium. The 5′ end of the transcript encodes cyan fluorescent

protein (CFP) fused with a peroxisome-targeting peptide to facilitate detection of mRNA translation upon induction of expression. Detection of the nascent mRNA in living cells is possible because of the presence of 24 copies of the MS2-binding stem-loop, located in the 3′ untranslated region downstream of the open reading frame. Once transcribed, these are bound specifically by yellow fluorescent protein (YFP)-tagged MS2 coat protein. Importantly, the MS2 stem-loops are located ~1 kb downstream from the promoter. Thus, detection of RNA transcripts by YFP–MS2 coat protein binding reflects transcribing RNA Pol II molecules that have escaped from the promoter and traveled at least 1 kb along the gene. The site of transcription can also be conveniently colocalized with the gene array because of the presence of 256 *lac* operator (*lacO*) repeats upstream of each transgene, which are detected in living cells using a fluorescent protein–tagged *lac* repressor (LacI). By using MS2 coat protein and *lac* repressor fused with different chromatic variants of fluorescent protein tags, both the gene array and the mRNA products can be detected simultaneously in living cells. Finally, the authors were also able to detect the presence of RNA Pol II itself at the gene array locus by expressing a green fluorescent protein (GFP)-tagged, α-amanitin–resistant RNA Pol II subunit (GFP-RPB1αAm[r]) in cells exposed to α-amanitin.

A notable feature of this study is the use of careful mathematical modeling to derive values for rate constants and fractions of polymerase engaged in elongation and in promoter binding and initiation. The fluorescence data that were fit to the models were measured in single live cells, using both fluorescence recovery after photobleaching (FRAP) of either GFP-tagged RNA Pol II or YFP-tagged MS2 coat protein, and photoactivation of photoactivatable GFP-tagged MS2 coat protein. To interpret the data, it was important to evaluate whether the fluorescence measurements resulted mainly from binding events or rather were influenced significantly by diffusion of fluorescently tagged molecules into the region of the gene array. To test this, Darzacq *et al.*[1] analyzed the contribution of diffusion by measuring how the recovery time in FRAP experiments was affected by an increase in the diameter of the bleach spot[3]. They observed that the rate of fluorescence recovery did not change when the diffusional distance in the bleached volume was increased, and they therefore conclude that diffusion does not appreciably affect FRAP analysis in this experimental system.

On the basis of the resulting experimental data, the authors infer several novel and unexpected features of the transcription process *in vivo*. First, they conclude that only a surprisingly small fraction of RNA polymerases that bind the promoter, ~1%, actually go on to transcribe the gene and produce an mRNA. Their data argue against substantial local recycling of aborted polymerases back to the promoter, because this is inconsistent with the efficient recovery they observed upon photobleaching. They speculate that such apparently inefficient polymerase-promoter interactions could actually be useful in suppressing a potential background of nonspecific transcription events in the genome, because RNA synthesis would occur only at sites where initiation was attempted repeatedly.

Second, they find that RNA polymerases transcribe (elongate) more rapidly than previously thought, measuring rates of ~70 bases per second, in contrast to the previous estimates of ~30 bases per second. A likely explanation for this difference is that the data of Darzacq *et al.*[1] specifically measure the maximum velocity of the elongating RNA polymerases, rather than measuring the average velocity during the total time from the induction of transcription to the subsequent accumulation of mRNA products. The latter value is expected to be an underestimate of the maximal rate of elongation, because Darzacq *et al.*[1] also propose that RNA polymerases sporadically pause for extended periods during their engagement with the gene. They suggest that at any single time point a small fraction of RNA polymerases pause in this way. At steady state, they estimate ~25% of the RNA polymerases associated with the array are paused. Thus, the array of 200 gene copies has on average 200–400 RNA Pol II molecules bound when the genes are induced, with approximately three-quarters of the RNA polymerase molecules actively elongating. Interestingly, previous studies on RNA Pol I, which transcribes the endogenous ribosomal DNA repeats in the nucleolus, have also measured a high rate of transcription, of over 90 bases per second, although there is no evidence that RNA Pol I molecules pause for extended periods[4].

The pausing of RNA polymerases transcribing the array that is reported by Darzacq *et al.*[1] is distinct from previous examples of polymerases pausing close to the promoter, as revealed by ChIP assays, because in the new study, detection of RNA transcripts requires the presence of the MS2 protein–binding sites that are located ~1 kb downstream from the

promoter. Therefore, the observations must represent pausing of RNA polymerases that have already actively transcribed at least 1 kb of DNA. The authors suggest that both the elongation and pausing activities of RNA polymerase are the most likely sources of their MS2 binding data, which show two separable components, both of which are lost upon treatment with the elongation inhibitor DRB. In contrast, treatment with the topoisomerase inhibitor camptothecin reduces the velocity of the polymerase by inhibiting DNA unwinding but does not prevent pausing. Camptothecin thus produces an increase in the paused polymerase component.

Detection of such stochastic pausing events, involving a small number of RNA polymerase molecules that may be distributed randomly across the gene, is made possible only by the combination of single-cell fluorescence time-lapse measurements and modeling used in this study. It would not be detected by other methods commonly used to analyze polymerase-DNA interactions or transcription factor–DNA interactions, such as ChIP assays, which typically sample average values from large numbers of cells in a population rather than time-resolved events in a single cell. Moreover, these new data raise the possibility that in certain cases where ChIP assays or other footprinting techniques reveal strong DNA occupancy at a gene locus, this may reflect pausing events, or possibly even abortive initiation, rather than efficient transcription activity. This highlights the importance of directly measuring *in vivo* the efficiency with which mature, completed RNA transcripts are generated from a gene under different conditions, rather than focusing only on how long the corresponding gene is bound by RNA polymerases and other factors.

No method, however, is without limitations, and novel approaches often require further tests and applications to validate their conclusions fully. At this stage, therefore, there are reasons to consider some of the results presented by Darzacq *et al.*[1] with caution. For example, very few endogenous genes in mammalian cells are expressed from large, tandemly repeated arrays. The highly repetitive nature of the gene array locus used by Darzacq *et al.*[1], which was required to provide a sufficient signal-to-noise ratio for reliable fluorescence measurements, could well have local effects on chromatin structure and transcriptional supercoiling that in turn affect the transcription reporter used in the study. Thus, the question of whether local chromatin effects at the array locus are affecting the measured properties of RNA Pol II must be kept in mind. It may be that the fraction of RNA Pol II molecules engaging

in productive transcription is considerably larger at endogenous genes than the value of ~1% measured in this study. Nonetheless, the results of Darzacq et al.[1] are striking and suggest that a substantially larger number of RNA Pol II molecules abort transcription than previously suspected. It should also be noted that the mathematical approach used by Darzacq et al.[1] requires assumptions to be made in fitting the data to a specific model. The authors make a very plausible argument in favor of the assumptions inherent in their model, which appears to be the simplest model that is consistent both with their experimental observations and with the known biological properties of transcription. Nonetheless, it remains to be determined whether the model will continue to explain future data when other genes are analyzed using similar methods.

An important goal for the future is therefore to repeat methods of *in vivo* analysis similar to those reported by Darzacq et al.[1] with other gene loci, to determine how general their findings prove to be. Ideally, this will eventually be done using either much smaller gene arrays or, if possible, single-copy genes, preferably with endogenous DNA sequences and chromatin structure. We look forward also to future technical developments that will facilitate the imaging of single RNA Pol II molecules in living cells. One possibility is that fluorescent speckle microscopy, a live-cell technique that allows visualization of single fluorescent molecules, will enable analysis of DNA transcription by single polymerases[5]. This would avoid the need to derive molecular models from analysis of populations of molecules. Single-molecule data directly reflect the heterogeneity in populations and thus would provide information about the distribution of paused and transcribing RNA Pol II molecules. Combining single-molecule analysis with either RNA interference–based knockdown or genetic depletion of specific factors could allow the analysis of both paused and transcribing states, thus providing a more detailed understanding of the properties of RNA Pol II during initiation and elongation.

Despite the limitations discussed above, this study represents a crucial milestone on the way to building a quantitative understanding of the mechanism of transcription in single live cells. Regardless of how general the results presented ultimately prove to be for most endogenous cellular genes, the work of Darzacq et al.[1] demonstrates what is already possible in the rapidly emerging field of live-cell fluorescence imaging and points the way to what should become possible in future. It will be important now to determine whether RNA Pol II's initiation efficiency and pausing frequency are modulated as a regulatory mechanism during transcriptional activation. Once such mechanistic understanding in live cells matures, we will begin to understand properly the systems biology of DNA transcription *in vivo*.

1. Darzacq, X. et al. Nat. Struct. Mol. Biol. **14**, 796–806 (2007).
2. Janicki, S.M. et al. Cell **116**, 683–698 (2004).
3. Sprague, B.L. & McNally, J.G. Trends Cell Biol. **15**, 84–91 (2005).
4. Dundr, M. et al. Science **298**, 1623–1626 (2002).
5. Danuser, G. & Waterman-Storer, C.M. Annu. Rev. Biophys. Biomol. Struct. **35**, 361–387 (2006).

When it comes to couple(r)s, do opposites attract?

Sharsti Sandall & Arshad Desai

A recent study using electron microscopy provides a detailed view of the oligomerization of a protein complex on the surface of a microtubule polymer. The findings point to a new type of interaction that may be well suited to couple the movement of cargo to dynamic cytoskeletal polymers.

The budding yeast Dam1 complex, comprised of ten subunits, localizes to centromeres and is centrally important for chromosome–spindle microtubule interactions[1]. Inhibition of this complex leads to extensive chromosome mis-segregation and cell lethality because replicated chromosomes cannot properly connect to opposite poles of the mitotic spindle. Wang et al.[2] have used electron microscopy to investigate the binding

Sharsti Sandall and Arshad Desai are at the Ludwig Institute for Cancer Research, Department of Cellular & Molecular Medicine, University of California, San Diego, CMM East, Room 3052, 9500 Gilman Dr., La Jolla, California 92093-0653, USA.
e-mail: abdesai@ucsd.edu

of the Dam1 complex to microtubules. Single-particle analysis of the free complex and helical reconstruction of polymer-bound oligomeric complexes reveal a striking conformational change between the monomeric and oligomeric forms. More significantly, the structure of oligomeric Dam1 complex on the microtubule highlights a new mode of association between accessory proteins and cytoskeletal polymers that has broad implications for the functions of the cytoskeleton.

The work of Wang et al.[2] was made possible by the biochemical reconstitution of the Dam1 complex, accomplished by coexpression of all ten subunits in bacteria[3]. This technical breakthrough has laid the groundwork for *in vitro* biophysical as well as structural studies of Dam1 complex interaction with microtubules[3–8]. In prior work, negative-stain electron microscopy has revealed multiple ten-subunit Dam1 complexes associated on the surface of a microtubule, forming both ring-shaped and spiral structures[3,6]. Assembly of the oligomeric structures is markedly promoted by the microtubule surface, although the free complex does oligomerize at high concentrations. Wang et al.[2] extended these earlier studies by first obtaining a 28-Å-resolution view of the free Dam1 complex using single-particle analysis. The free complex exists mostly as dimers of the ten-subunit monomer; each of the monomers has a projecting arm that extends out of a long base (see schematic in **Fig. 1a**). Reconstructions of mutant-containing complexes reveal that the arm is comprised of