In Vivo Dynamics of Polymerase II Transcription

Darzacq, et al

Supplementary Information

Page

Supplementary Figure

Figure S1

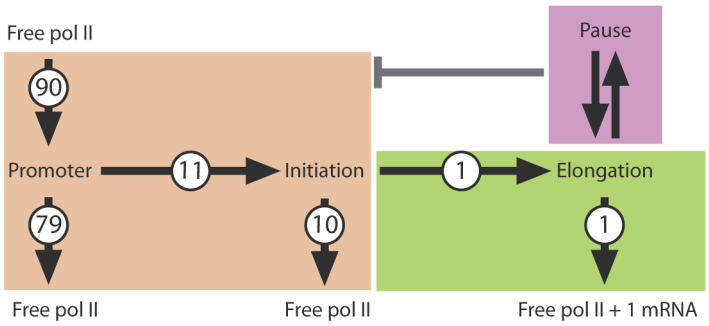

Figure S1 The assembly and processive steps of RNA polymerase II during transcription and in regulating transcription by pausing. Circled numbers indicate the relative flux of molecules between different steps. Initiation events (beige) are the sole factors dictating the amount of mRNAs produced per unit of time; subsequently each engaged polymerase will produce 1 RNA molecule. Elongation (green) represents the catalytic reaction and does not appear to be an obvious point of control. Pausing (violet) can feedback on initiation.

Supplementary Discussion

Testing of models

We modeled transcription from a gene array and compared the models derived from two different sets of experiments: first the enzyme (Pol II) and subsequently the product of the reaction (RNA). While these experiments measure transcription, each focuses on a specific aspect. For instance, the difference obtained for elongation times including the slow step, whether reading-out Pol II (517s) or the RNA (250s), is somewhat greater than simply scaling the transcription unit (3.3 kb vs 2.3kb, respectively). This may be due mainly to the low precision of the Pol II slowest component. Another possibility is that the slowest Pol II component is due to transcription beyond the cleavage or polyadenylation site. Alternatively, the 5' region of the gene, detected by YFP-Pol II may be more prone to pausing than the last 2.3 kb in the 3' region of the gene, detected by the MS2 reporter. These discrepancies will need to be addressed using molecular markers for specific molecular events.

Slow components in transcription, such as termination, or diffusion of mRNPs from the locus might account for the long times we attribute to polymerase pausing. However, all polymerases must terminate. If all polymerases were forced to go through a slow step, the kinetics of the transcription unit would be dominated by this slower process. There would then be only a slow recovery from FRAP detectable. We have modeled this option where increasing contributions of a slow component can be seen to progressively overwhelm any fast component (**Fig. 8c**). The data suggest that only 4.2% of the polymerases could account for this component, for an average of about four minutes. At the steady-state, this polymerase fraction comprises 26% of the total population of polymerases on the tandem array of 200 genes. Likewise, all mRNPs must diffuse away from their site of synthesis. If this were the slow component in the kinetics, it would show up in the models and likewise overwhelm the kinetics. The sensitivity of the assay to changes in the contribution to the slow component illustrates that the kinetics would change drastically from the observed data with only an increase from 4.2% to 10%. The model therefore cannot accommodate a universal slow step.

We considered that the slow step may be retention of a population of mRNPs at the

site. A subpopulation of post-transcriptional mRNPs at the locus would not be consistent with this model, since it would not take into account that the polymerases are retained in this kinetic component as well. Since it is likely that the polymerase slow step and the RNA slow step are associated with nascent RNAs, this also cannot be accommodated by a model where these processes are unlinked. It is likewise unlikely that polymerases transcribe at different speeds, one slow and one fast, because the camptothecin results indicate that the slow-down due to DNA torsion results in the same two components.

The analyses presented here were performed on an amplified gene array that provides the advantages of visualizing the transcriptional process on a specific gene due to the high signal-to-noise ratio of the measured fluorescence. While transcription has been shown to take place in assemblies of genes and polymerases termed 'transcription factories' one should be cautious when extrapolating from this array system to single genes or single factories. The results obtained from the array represent an ensemble of averaged measurements together with the assumption that all genes are equally active. However, the yield per gene would be different if a smaller number of different genes were transcribed. Still, the number of genes could not be small (e.g. 20) since the packing of the polymerases would get too close to allow our modeled low percentage of pausing. It is likely possible also that other genes are transcribed more efficiently that the construct described here. For instance, recent unpublished results using similar approaches (Edouard Bertrand, personal communication) suggest that viral genes may be much more efficient. Thus varieties of gene expression may be revealed and described more completely and mechanistically using these quantitative approaches.

Supplementary Methods

Microscopy:

Wide-field microscope for fixed-cell imaging. Images were acquired with an Olympus BX61 epifluorescence microscope with an internal focus motor and an UPlanApo 100X, 1.35 NA oil immersion objective using a 100 Watt mercury arc lamp for illumination (Olympus, Melville, NY). Digital images were acquired using a Photometrics CoolSNAP HQ camera (Photometrics, Tucson, AZ) as 3D images series (stacks) of 40 images taken with a Z step size of 0.1 □m using IPLab software (Windows v3, Scanalytics, Rockville, MD) and filter sets: 31044v2 (CFP), 41001 (FITC), 41007a (Cy3), 41008 (Cy5) and 41043 (RFP) (Chroma Technology, Rockingham, VT).

Confocal microscopes. FRAP and photoactivation was performed on a Leica TCS SP2 AOBS laser scanning confocal microscope equipped with a PlanApo 63X, 1.4 NA objective (Leica Microsystems Inc, Exton, PA). FRAP experiments with camptothecin were performed using Zeiss LSM 510 Meta laser scanning confocal microscope equipped with a Plan-Apochromat 63X, 1.4 NA oil objective (Jena, Germany). The spot size experiments were performed using Zeiss LSM 5 Live with DuoScan at Janelia Farm, HHMI, equipped with a Plan Apo oil 63X, 1.4 NA objective (Jena, Germany).

Wide-field microscope for live-cell imaging. The wide-field imaging of FRAP was performed on an Olympus IX-81 inverted microscope with a UApo 150x, 1.45 NA objective (Olympus, Center Valley, PA) and a YFP filter set (#41028, Chroma, Rockingham, VT). The system was equipped with a 300 mW Argon Ion 488 nm laser and a Mosaic Digital Diaphragm System (Photonic Instruments, St Charles, IL) used for bleaching. A Lambda DG-4 light source with 300W xenon lamp (Sutter Instruments, Novato, CA) was used for imaging recovery. The DG-4 illumination was attenuated by 33% using a 0.2 optical density metallic neutral density filter (Thor Labs, Newton, NJ). Images were acquired using a Cascade 512B EM-CCD camera (Photometrics, Tucson, AZ) driven by IPLab for Windows version 4 (BD Bioscience, Rockville, MD). All images were corrected for the bias drift of the EM-CCD camera. The cells were positioned in

the z-axis using a MS-2000-XYLE-PZ piezo-top stage (Applied Scientific Imaging, Eugene, OR). The acquisition of images, movement of the stage, shuttering of light and bleaching laser were synchronized using TTL digital logic circuitry built in the lab.

Image analysis. We wrote software in C++ to automatically identify and track transcription sites in the movies and also to report the mean intensity over time of user identified regions of interest. The transcription site was identified automatically by the software in each frame of the movie using the locus marker from the RFP channel. To segment the locus, the software filtered the data from the RFP channel using a 3x3 convolution kernel of ones and then applied a relative threshold equal to 99.5% of the maximum convolved image intensity. A binary mask was created from the thresholded data and the mask was dilated once, eroded twice and then dilated again. The largest contiguous object in the binary mask was the locus. The application of convolution filtering, a relative threshold and binary image processing allowed the software to identify the locus in each movie without user intervention. User defined options governed how the boundary of the locus was applied to the GFP channel to detect transcription. Either the locus boundary could be used without modification or the boundary could be enlarged by a fixed number of pixels or a circular region of fixed radius and centered on the geometric centroid of the locus boundary could be used or a circular region centered on the geometric centroid of the locus boundary with a radius sufficiently large to encompass the locus boundary could be used. The defined boundary was transferred to the GFP channel and the mean intensity and geometric centroid of the region encompassed by the boundary was reported for each time point of the movie. The software also reported the mean intensity and geometric centroid over time of user defined regions of interest in the movie. These regions were drawn outside the cell, in the cytoplasm and in the nucleus. The software provided an option to define automatically an intraneuclear background region of fixed thickness that circumscribed the transcription site boundary. For each time point, the background taken from a ROI outside of the cell was subtracted from all other measurements.

T(t) and I(t) were measured for each time point as the mean intensity of the nucleus and the mean intensity in the transcription site (position defined by the software tracking

algorithm), respectively. One image was collected prior to bleaching/photoactivation and these initial conditions are referred to as Ti = nuclear intensity and Ii = intensity in ROI before bleaching. Ic(t) is the corrected intensity of the bleached ROI at time t For FRAP experiments:

$$Ic_{(t)} = \frac{I_{(t)} \cdot T_i}{I_i \cdot T_{(t)}}$$

For photoactivation experiments:

$$Ic_{(t)} = I_{(t)} - Ii$$

Calculation for Ic(t) in FRAP experiments corrects for imaging photoattenuation while calculation of Ic(t) in photoactivation experiments does not. This is because under the imaging conditions we used for both experiments, we found that GFP signals were decaying (about 20%) during the course of the experiment (e.g. 80 frames). paGFP was imaged in conditions where the intensity of a not fully activated cell would remain nearly constant (variations were inferior to experimental error) over 80 frames. We believe this is not the consequence of an absence of bleaching due to the scanning laser but rather an equilibrium in between some leak activation counterbalancing photobleaching because when we tuned the laser to higher values, we observed an imaging dependent increase of the signal rather than a decay (not shown).

Transcription site ‘x’ and ‘y’-axis planar motion during the imaging of the fluorescence recovery after photobleaching (~10 min) was compensated for by tracking the signal. When imaging two-dimensional time series using a confocal instrument, we excluded data sets in which the transcription locus drifted in the z-axis, since we could not compensate for this drift in the analysis. To verify the robustness our results, we repeated the polymerase II experiments using a wide-field microscope equipped with technology that permitted faster temporal resolution sampling than a confocal, and that also had rapid three-dimensional imaging capabilities. In this fashion, the total fluorescence at the transcription locus was captured regardless of the XYZ movements. Both confocal and wide-field imaging platforms yielded the same results

Single RNA quantification. In order to quantify the number of mRNA molecules produced at the sites of transcription, single molecule mRNA identification and

quantification was performed as previously described[1,2] using a probe which has only one binding site on the transcript and that therefore can serve for quantification. We used the wide-field microscope for fixed cell imaging described below. First, 3D stacks of cells labeled with the probe at the transcription site were acquired. Second, a point spread function (PSF) was obtained for each experiment. These 3D images were then deconvolved using Exhaustive Photon Reassignment (EPR, Scanalytics) which uses a quantitative, constrained-iterative algorithm[3] with an acquired PSF. The total fluorescence intensity (TFI) at each of the deconvolved sites of active transcription was measured with a script written for IPLab. The number of mRNAs at the transcription sites was determined by calculating the total fluorescent intensity (TFI) of individual RNA molecules, and dividing that value by the TFI per probe as follows: TFI per probe was calculated for each probe by collecting images from serial probe dilutions. 5 □l of each probe dilution (ranging from 4 ng/□l to 4 $x10^{-4}$ ng/□l) were placed between a coverslip and a slide, onto which 170 nm blue fluorescent beads had previously been dried. Using the beads as markers, the distance between the coverslip and slide was measured (in microns) and the center plane of the dilution was located. A range of interest (256 X 256 pixels) that excluded beads was identified, and a single image was captured at the center plane, using exposure times identical to that used to capture cell images. This procedure was repeated three times, each at a different location on the coverslip, for each dilution. The TFI per probe was then obtained by plotting the integrated fluorescence in the total imaged volume against the known number of molecules in that volume. The slope of the resulting curve represents the TFI per one fluorescent probe molecule and is further used in the calculations. To compensate for the deconvolution, this value was then divided by the number of planes in the PSF used by EPR.

Statistical Analysis. To determine model order, that is the number of resolvable model state variables, data from RNA polymerase II FRAPs were fitted to sums of exponentials. Fitting was carried out by the method of generalized least squares using the SAAM II software (http://depts.washington.edu/saam2/). Three criteria were adopted to decide the model order: 1) the residuals must appear randomly distributed around

zero, 2) the Akaike information criterion (AIC)[4] for a more complex model must be less than the AIC of a simpler model to justify the increased complexity, and 3) ideally, the Bayesian Information Criterion (BIC)[5], which penalizes additional parameters even more severely than the AIC, should also be smaller for the more complex of two candidate models. This test was also applied to the MS2 FRAP and photoactivation data.

Supplementary References

1. Femino, A.M., Fay, F.S., Fogarty, K. & Singer, R.H. Visualization of single RNA transcripts in situ. *Science* **280**, 585-90 (1998).
2. Shav-Tal, Y. et al. Dynamics of single mRNPs in nuclei of living cells. *Science* **304**, 1797-800 (2004).
3. Carrington, W.A. & Lisin, D. Cluster computing for digital microscopy. *Microsc Res Tech* **64**, 204-13 (2004).
4. Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716-723 (1974).
5. Schwarz, G. Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464 (1978).