Mechanistic Interpretation of Promoter-Proximal Peaks and RNAPII Density Maps

Genome-wide RNA polymerase II (RNAPII) density profiles provide averaged snapshots of transcription but are difficult to interpret in the context of dynamic gene expression. We performed computational modeling to simulate RNAPII density profiles from individual transcription parameters and constructed simple mathematical models to explore general relationships between transcription parameters and density profiles.

The density of RNAPII on genes will depend mainly on three parameters: elongation rate (bases added per unit time), initiation frequency (number of start events per unit time that result in productive elongation), and processivity (fraction of polymerases remaining on the template after each catalytic event). Differences in elongation rate will affect the density via changes in the average spacing between polymerases (Figure 1A, top). It can be shown that this results in the density being inversely proportional to the elongation rate when the initiation frequency remains unchanged (see Data S1 available online for all mathematical derivations). A computational model, ChIPMOD, that simulates RNAPII density patterns based on the basic transcription parameters independently confirms this relationship, both for the total density across a gene and for density peaks that result from local pauses, defined here as regions of slow elongation (the reader is encouraged to test the relationships mentioned throughout the text by using the online version of the program (http://www.chipmod.org.uk; seeData S2 for details).

The initiation frequency is a key regulated step in the transcription cycle. In agreement with the intuitive expectation that a higher initiation frequency will result in a higher density of polymerases (Figure 1A, bottom), the average RNAPII density across a gene can be shown to be directly proportional to the average initiation frequency on that gene, assuming constant elongation rates.

Even though density patterns are affected by processivity, we assume here that the processivity of RNAPII is high, in line with the observation that chromatin immunoprecipitation sequencing (ChIP-seq) read densities are typically largely flat across most of the coding region of a gene. Nevertheless, as outlined below, imperfect processivity at the beginning of genes may contribute to specific density peaks observed at metazoan genes.

We note that, although the relationships between elongation rate, initiation frequency, and RNAPII density may appear intuitive, it can often be difficult to determine the impact of changes. This is in part because initiation frequency and elongation rate have similar but opposite effects on the density pattern, making it difficult to distinguish between the two. For example, Spt5 knockdown resulted in a significant increase of polymerase density across the body of several genes (Rahl et al., 2010). The authors concluded that this was due to increased release of RNAPII from the promoter. However, given that Spt5 is also an elongation factor, the higher RNAPII density across the coding region could arguably be equally well explained by a lower elongation rate.

It is worth emphasizing that an increased mRNA output can only be achieved by increasing the initiation frequency, but not by increasing the elongation rate: at an initiation frequency of ten polymerases per minute, for example, the mRNA synthesis rate can never exceed ten molecules per minute, regardless of the elongation rate. Indeed, only when the elongation rate is very low relative to the initiation frequency will it affect expression levels, as elongating polymerases may then saturate the beginning of a gene and prevent new rounds of initiation

(Figure 1B). This establishes a universal lower limit for the ratio of the elongation rate (v) to the initiation frequency (IF). Indeed, v/IF would have to be greater than 33 bp (the width of the transcribing elongation complex) to avoid clashes between newly recruited polymerases and polymerases leaving the promoter. And if v/IF = 33 bp, initiation is so rapid that the transcribing polymerases run immediately adjacent to one another. This, in turn, means that the average spacing between polymerases on a gene can be calculated from the initiation frequency and the elongation rate alone, and this can be plotted graphically (Figure 1C). The situation in which polymerases run immediately adjacent to one another can also be used to calculate the highest possible initiation frequency for any gene from the elongation rate alone ($IF_{MAX} = v$/33 bp). In human cells, where elongation rates are ~3.8 kb/min (Singh and Padgett, 2009), the maximal possible initiation rate (and mRNA output) will thus be ~115 min^{-1}. At the highly transcribed *hsp70* gene in *Drosophila* (elongation rate ~1.2 kb/min [O'Brien and Lis, 1993]), the maximal possible initiation rate will be ~36 min^{-1}, in agreement with the 20–40 min^{-1} determined experimentally (Lengyel and Graham, 1984). Together, these considerations make it clear that cause and effect in the early transcription cycle are not always unidirectional, as elongation can also restrict initiation.

Promoter-Proximal RNAPII Peaks

A defining feature of RNAPII ChIP and RNA-seq patterns is a peak of density found immediately downstream of the start site (Adelman and Lis, 2012). This promoter-proximal peak is often taken as an indication that gene expression is regulated after initiation through controlled release of an already initiated but halted polymerase. We used the computational model to examine how promoter-proximal peaks might arise and also explored whether their presence is sufficient to infer that the rate-limiting step in transcription occurs after initiation.

Three mechanisms could conceivably explain the promoter-proximal peak: arrest, slow elongation (pausing), and imperfect processivity (Figure 1D). Strikingly, when simulated computationally, all three models resulted in similar peaks,
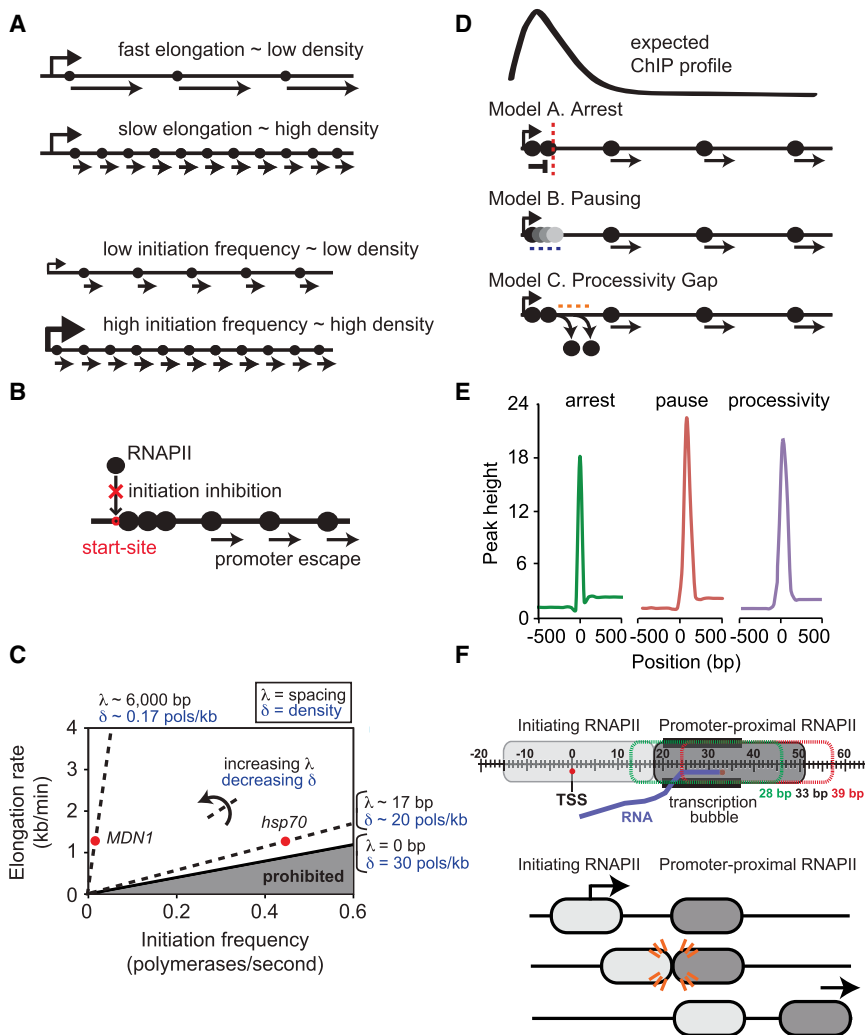
Figure 1. Effects of Transcription Dynamics on RNAPII Density and Promoter-Proximal Peaks

(A) (Top) Slower elongation results in higher density. (Bottom) Higher initiation frequency results in higher density.

(B) With slow elongation or high initiation frequency, transcribing RNAPII fails to advance far enough to liberate the start site, resulting in inhibition of new initiation.

(C) Graphic representation of the relationship between elongation rate, initiation frequency, density, and spacing between polymerases. Prohibited zone represents an area in which initiation frequency and elongation rate are incompatible. Data for *MDN1* and *hsp70* were taken from Larson et al. (2011), O'Brien and Lis (1991), and Lengyel and Graham (1984).

(D) Models for promoter-proximal density peaks. Arrest, complete halt of elongation for defined amount of time; pausing, slow early transcription; processivity gap, low processivity early in gene.

(E) Computer simulation illustrates the feasibility of obtaining promoter-proximal peaks by all three mechanisms.

(F) (Top) Alignment of promoter-proximal and initiating RNAPII. Initiating polymerase with its active site (red dot) at the start site, using the footprint of 33 bp obtained from crystal structures. Three promoter-proximal elongation complexes for RNAs of 28, 33, and 39 nucleotides are overlaid in green, black, and red, respectively. Transcription bubble, filled black bars. The peak of the RNA-seq distribution (position +33; Nechaev et al., 2010) corresponds to a snug fit between the initiating and promoter-proximal RNAPII. (Bottom) Initiation can be rate limiting despite the presence of a paused polymerase (see also Animations 1 and 2 at http://www.chipmod.org.uk/animation.html).

which closely resembled those observed in vivo when combined with antisense transcription (Figure 1E; also try online version of program at http://www.chipmod.org.uk). By themselves, density patterns are therefore insufficient to distinguish between arrest, slow elongation, and imperfect processivity as the mechanism for generation of promoter-proximal peaks. Of the three models, slow elongation in the promoter-proximal area is, in our opinion, particularly attractive, as it conforms with current knowledge about transcript elongation and can be explained by a simple, slow transition into the fully competent elongation complex through, for example, slow transit through the first nucleosome, a need for recruitment of elongation factors or CTD kinases, and maturation of the core RNAPII elongation complex. Still, it is conceivable that promoter-proximal peaks arise by any one of the three mechanisms or by combinations thereof and that different mechanisms govern different genes.

It is commonly assumed that the presence of a density peak downstream of the start site implies that the rate-limiting step of transcription occurs after initiation. We explored this issue further by combining knowledge from the crystal structure of RNAPII with the lengths of early transcripts obtained from genome-wide fine-mapping of promoter-proximal polymerases (Nechaev et al., 2010). Interestingly, by drawing footprints of the initiating and paused polymerases to scale, it became apparent that the two polymerases are either immediately adjacent to each other or overlapping and interfering with one another (Figure 1F, top). This striking proximity would be expected to affect the mechanism of gene regulation, even if the precise nature of the interaction remains unknown. Although it is possible that the promoter-proximal polymerase restricts binding of the initiating polymerase, it is equally possible that the converse is true and that release of the promoter-proximal polymerase is triggered by the arrival of a new polymerase. Crucially, it follows from this that the rate-limiting step of transcription could still be initiation, even with postinitiation arrest or pausing (illustrated in Animations 1 and 2 at http://www.chipmod.org.uk/animation.html). This would hold true even if the paused polymerase is located further downstream from the start site (Figure 1F, bottom). At least two biochemical mechanisms could be envisioned for this model: initiation factors might continue to retain the promoter-proximal polymerase until the next polymerase is recruited to release it, or the initiating

polymerase might help to "nudge" the promoter-proximal polymerase out of an early stage of slow transcription. Support for both mechanisms has been obtained experimentally (Takahashi et al., 2011; Saeki and Svejstrup, 2009). Genetic and biochemical evidence over the last 30 years has shown that initiation is the principal rate-limiting step in transcription through processes such as chromatin remodeling and recruitment of the general transcription machinery. The considerations and models presented here could help to reconcile that knowledge with the widespread occurrence of promoter-proximal RNAPII density peaks.

It is very important to emphasize that we do not argue that transcription is never regulated at the postinitiation level—only that the mere presence of a promoter-proximal peak is insufficient to draw conclusions about mechanism or about the stage at which transcription is regulated. If we nevertheless accept the possibility that promoter-proximal RNAPII does not generally signify a rate-limiting, regulated step in transcription, what could be its alternative function? One possibility is that it is simply an accidental consequence of slow transcription or low processivity close to the promoter. Moreover, given that metazoan genes are often devoid of strong promoter sequence elements such as a TATA box, it is an intriguing possibility that promoter-proximal retention of polymerases might have evolved not as a regulatory feature but as an effective marker for the beginning of genes. In this view, promoter-proximal RNAPII acts as the demarcation of a "landing pad" for initiating RNAPII, providing additional interaction points beyond those established with the general transcription factors at the core promoter.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two data files and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2013.07.032.

WEB RESOURCES

The URLs for data presented herein are as follows:

ChIPMOD Computer Simulation, http://www.chipmod.org.uk
Animations, http://www.chipmod.org.uk/animation.html

Andreas H. Ehrensberger,[1,*]
Gavin P. Kelly,[2]
and Jesper Q. Svejstrup[1,*]
[1]Mechanisms of Transcription Laboratory, Clare Hall Laboratories, Cancer Research UK London Research Institute, South Mimms EN6 3LD, UK
[2]Bioinformatics & Biostatistics Group, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK
*Correspondence: andreas.ehrensberger@cancer.org.uk (A.H.E.), jesper.svejstrup@cancer.org.uk (J.Q.S.)
http://dx.doi.org/10.1016/j.cell.2013.07.032

REFERENCES

Adelman, K., and Lis, J.T. (2012). Nat. Rev. Genet. *13*, 720–731.

Larson, D.R., Zenklusen, D., Wu, B., Chao, J.A., and Singer, R.H. (2011). Science *332*, 475–478.

Lengyel, J.A., and Graham, M.L. (1984). Nucleic Acids Res. *12*, 5719–5735.

Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Science *327*, 335–338.

O'Brien, T., and Lis, J.T. (1991). Mol. Cell. Biol. *11*, 5285–5290.

O'Brien, T., and Lis, J.T. (1993). Mol. Cell. Biol. *13*, 3456–3463.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). Cell *141*, 432–445.

Saeki, H., and Svejstrup, J.Q. (2009). Mol. Cell *35*, 191–205.

Singh, J., and Padgett, R.A. (2009). Nat. Struct. Mol. Biol. *16*, 1128–1133.

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., et al. (2011). Cell *146*, 92–104.