Derailing the Locomotive: Transcription Termination*

Published, JBC Papers in Press, November 12, 2007, DOI 10.1074/jbc.R700032200

David S. Gilmour and Ruopeng Fan
From the Center for Gene Regulation, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802

Transcription can be divided into three stages: initiation, elongation, and termination. Initiation involves the binding of RNA polymerase (RNAP) to a gene’s promoter and the beginning of RNA synthesis. Elongation involves lengthening the nascent transcript by the addition of nucleotides to the 3’-end of the RNA located in the active site of RNAP. The last stage of transcription, termination, involves stopping RNA synthesis, release of the transcript, and dissociation of the RNAP from the DNA template.

Termination and Gene Regulation

Transcription termination is essential for subdividing and retrieving information from the genomes of all organisms. RNAPII transcribing from an upstream region can interfere with the function of a promoter in its path (1). Termination is also essential for recycling RNAP.

Termination has long been known to play a significant role in regulating transcription through genes in bacteria, and isolated cases hint at its potential importance in eukaryotes. In bacteria, attenuation and antitermination are related processes in which mutually exclusive patterns of RNA folding dictate whether the polymerase will continue transcription or prematurely terminate before completing transcription of an operon (2). It is less clear what role premature termination plays in regulating gene expression in eukaryotes. Measurement of the distribution of RNAPII throughout the genome in human cells identifies hundreds of genes in which the density of RNAPII at the 5’-end greatly exceeds the density of RNAPII at the 3’-end of the gene (3), but in many cases, RNAPII may be stalling as a result of its association with the negative elongation factors NELF and DSIF (4).

One case in eukaryotic cells in which premature termination is involved in repressing transcription is the HIV provirus. RNAPII is concentrated at the 5’-end of a transcriptionally repressed provirus (5). That this discontinuous distribution of RNAPII is a consequence at least in part of premature termination is indicated by the accumulation of short transcripts in the cytoplasm of the cell (6). Recently, we provided evidence that Pcf11, a protein previously implicated in termination at the ends of protein-encoding genes, could be responsible for premature termination on HIV (7). Activation of HIV transcription involves a protein called Tat. Tat can be thought of as an antiterminator. It functions by recruiting the kinase P-TEFb to the elongation complex (EC). Phosphorylation of RNAPII and other regulators of elongation renders RNAPII less prone to premature termination (4).

Corden and co-workers (8) recently uncovered a negative feedback mechanism that could be the first clear example of attenuation of an eukaryotic gene. Nrd1 participates in terminating RNAPII transcription at the ends of small nuclear RNA genes in yeast, and it binds specific sequences near the 3’-end of small nuclear RNA (9). Corden and co-workers (8) observed that mutations in the Nrd1 protein cause overexpression of NRD1. Further analysis revealed that there are numerous Nrd1-binding sites in the 5’-region of the NRD1 transcript. Mutation of these binding sites results in increased levels of NRD1 message even in the presence of normal Nrd1 protein.

Dismantling the Bacterial RNAP Elongation Complex

Once RNAP releases a nascent transcript, it cannot resume synthesis of the transcript. Hence, the EC must remain stably associated with the DNA template until it completes the transcript. This would seem diametric to its need to move along DNA. These requirements are met by the sliding clamp structure of the EC (10–12). RNAP has a claw-like structure with grooves that guide the spatial arrangement of the single- and double-stranded nucleic acid components of the EC.

Our current understanding of bacterial RNAP termination focuses on the nucleic acid components of the EC (13). This nucleic acid framework consists of a transcription bubble ~12 nucleotides in length (14). Within this bubble, eight nucleotides of the nascent transcript form a heteroduplex with the transcribed strand of the DNA. The eight-nucleotide heteroduplex is key to the stability of the EC (15). Shortening the heteroduplex by one nucleotide or weakening its stability by altering the nucleotide composition results in destabilization of the EC (16). The U:A base pair is the weakest of all possibilities, and U:A heteroduplexes are associated with many termination sites for bacterial RNAP, RNAPI, and RNAPIII.

A series of elegant experiments has provided evidence for one termination mechanism known as the forward translocation model (17). This model posits that an external force pushes RNAP and the nucleic acid framework forward without accompanying addition of nucleotides to the 3’-end of the nascent transcript. This results in shortening of the heteroduplex through the concerted formation of base pairs between strands of DNA on the upstream side of the transcription bubble and the disruption of base pairs both at the downstream side of the bubble and at the part of the heteroduplex most distal to the 3’-end of the nascent transcript (Fig. 1, A and B). Shortening the heteroduplex destabilizes the EC. Recent work indicates that the forward translocation model is equally applicable to the single subunit T7 phage RNAP (18).

Three sources of external force that push the polymerase forward causing termination of RNAP have been identified in bacteria. These are the RNA hairpin that constitutes part of

*This minireview will be reprinted in the 2008 Minireview Compendium, which will be available in January, 2009.

1 To whom correspondence should be addressed. Tel.: 814-863-8905; Fax: 814-863-7024; E-mail: dsg11@psu.edu.

2 The abbreviations used are: RNAP, RNA polymerase; HIV, human immunodeficiency virus; EC, elongation complex; CTD, C-terminal domain.
MINIREVIEW: Transcription Termination

Intrinsic terminators and the translocation activities of the proteins Rho and Mfd. Intrinsic terminators consist of a DNA sequence that encodes a hairpin structure in RNA, followed by a U-rich stretch of nucleotides (Fig. 1C). They constitute about half of the terminators in Escherichia coli. Yarnell and Roberts (19) observed that an oligonucleotide could substitute for one strand of the hairpin, indicating that the hairpin structure per se is not responsible for termination. This was followed by experiments showing that blocking the forward translocation of the RNAP inhibits oligonucleotide-mediated transcript release (20). The importance of the U'A-rich heteroduplex in the forward translocation model is that the weakness of the base pair facilitates unwinding of the heteroduplex as the RNAP is pushed forward by the incipient hairpin.

Rho-dependent terminators constitute the other half of the terminators found in E. coli (21). Rho binds RNA and couples ATP hydrolysis with translocation in a 5' to 3' direction. Upon reaching RNAP, Rho can push the RNAP forward (Fig. 1E). Mfd is involved in transcription-coupled repair, and it couples ATP hydrolysis with translocation along DNA (22). Translocation of Mfd along the DNA allows it to push RNAP forward (Fig. 1D). In the presence of a DNA lesion that blocks RNAP per se, it is likely that the applicability of the prokaryotic paradigms to RNAP in eukaryotes might be similar. The eight-nucleotide heteroduplex appears to be a major determinant in the stability of both the bacterial RNAP and RNAPII ECs (15). A DNA translocase called TTF2 appears analogous to Mfd1 in being able to dissociate RNAP from DNA. TTF2 has been implicated in DNA repair and in clearing ECs from DNA during condensation of mitotic chromosomes (24, 25). Also, an intrinsic terminator from E. coli was observed to disrupt a yeast RNAPII EC by a mechanism that depends on hairpin formation (16). Notably, RNA hairpins have not been implicated in termination for any naturally occurring DNA in eukaryotes, suggesting that the applicability of the prokaryotic paradigms to eukaryotic genes is limited.

RNAPIII terminates on its own at T-rich sequences, but unlike intrinsic termination in bacteria, these T-rich sequences are not preceded by hairpin-forming sequences (26). Instead, sequences downstream from the terminating T-rich stretch affect termination efficiency. Structural data identify two RNA-P III-specific subunits, C53 and C37, that are located toward the front of the EC (27). Removal of these two subunits alters the location of termination, although dissociation still occurs on T-rich sequences (28). Add-back experiments indicate that C53 and C37 slow the rate of elongation, so they could function to cause RNAPIII to pause long enough at particular sites so that
the instability of the U:A heteroduplex leads to release of the transcript and dissociation of RNAPIII.

Termination by yeast RNAPI at the terminator located at the end of the ribosomal gene can be reconstituted in vitro with a protein called Reb1p (29). Reb1p associates with a region downstream from a T-rich sequence and causes RNAPI to halt, so the EC has a U:A-rich heteroduplex (Fig. 1F). The Lac repressor can substitute for Reb1p in vitro to block elongation and cause termination. This suggests that causing RNAPI to pause with a U:A-rich heteroduplex could be sufficient for termination (30). Because Reb1p and the Lac repressor are thought to halt the forward progress of RNAPI, dissociation may not involve the forward translocation mechanism.

Termination by RNAPII and RNAPIII appears to be more complicated in animal cells. For RNAPI, a Reb1p homolog called TTF-I causes RNAPI to pause in a T-rich region (31). A second protein called PTRF must be added in vitro to achieve transcript release. PTRF interacts with TTF-I and RNAPII, but the mechanism by which it dissociates the EC is not known. For RNAPIII, the La protein has been implicated in transcript release, although this point remains controversial (32, 33). Notably, in contrast to Rho- and Mfd-mediated termination, the aforementioned dissociation reactions for mammalian RNAPI and RNAPIII do not require nucleotide hydrolysis (31, 33). A bioinformatics analysis determined that RNAPIII terminators in mammals most frequently contain 4-base stretches of Ts, whereas those in yeast most frequently contain 6- or 7-base stretches of Ts (26). The longer U:A stretch might render the yeast EC less stable than the human EC, so transcript release is factor-independent.

Termination by RNAPII has been the most enigmatic. The stability of the RNAPII EC is quite remarkable: RNAPII remains engaged on the dystrophin gene for ~16 h to complete transcription of this giant gene (34). It was observed that purified RNAPII will terminate at T-rich sequences that have a propensity to bend DNA, although this was inefficient at physiological salt concentrations (35). No one has mapped a precise site where RNAPII disengages from the template in vivo (36). Instead, termination seems to be a stochastic event whose probability increases markedly at pause sites once RNAPII has transcribed past the polyadenylation signal located at the end of the gene. Analysis of yeast mutants has established a connection between polyadenylation and termination (37). The polyadenylation signal in the nascent transcript can cause the association of several protein complexes involved in processing of the 3'-end of the mRNA (38), suggesting that recruitment of these factors leads to termination. Alternatively, these factors associate with the gene upstream from the polyadenylation signal and become active after the polyadenylation signal is transcribed (39).

Two RNAPII termination models have dominated the field (36, 40). Both models explain how transcription of the polyadenylation signal increases the probability of termination in the region downstream from this signal. The torpedo model posits that cotranscriptional cleavage of the nascent transcript at the polyadenylation signal provides an entry site for a 5’ to 3’ exonuclease called Rat1 in yeast and Xrn2 in animals. Disruption of this protein in yeast or animal cells results in RNAPII transcript release beyond regions where it normally terminates (41, 42).

The allosteric model posits that the polyadenylation site causes a change in the EC, which increases its tendency to terminate. The change in the EC could involve release of a factor that normally inhibits termination or the recruitment of a factor that causes termination. Recent work provides support for the latter. The protein Pcf11 was found to cause RNAPII to dissociate from DNA by associating with the CTD of RNAPII and the nascent transcript (Fig. 1G) (43, 44). The CTD of RNAPII is the C-terminal domain of the largest subunit, and it is involved in transcription initiation, elongation, and termination and also RNA processing (45). Because Pcf11 dismantling activity requires an intact CTD, it provides a possible reason why the CTD is required for termination in animal cells (46). Pcf11 is concentrated at the 3’-ends of genes, where it might be recruited by the polyadenylation signal in the nascent transcript (38, 44, 47).

It is unclear if Rat1 directly causes RNAPII to dissociate from the DNA because extensive degradation of the nascent transcript in an RNAPII EC by a variety of RNases does not dismantle the complex (48). In contrast, Pcf11 alone dismantles the EC by a mechanism that is dependent on the CTD but independent of nucleotide hydrolysis (43, 44). Several lines of evidence support a model in which Pcf11 forms a bridge between the CTD of RNAPII and the nascent transcript (43, 44). Because the action of Pcf11 depends on RNAPII pausing, we suggest that connecting the nascent transcript to the CTD might allow force that destabilizes the EC to be exerted on the RNA.

Recent reports raise questions about Rat1 and Pcf11. An important discrepancy concerning the mutant allele pcf11-13 has arisen. This mutation was originally reported to be defective in CTD binding in vitro and termination of transcription from a plasmid in yeast in vivo (49). However, a more recent study found that the pcf11-13 mutation does not exhibit defective termination on two chromosomal genes (50). In addition, recent data suggest that association of Pcf11 and Rat1 with the 3’-ends of genes in yeast is interdependent (51). This interdependency is likely to confound the interpretation of results from genetic studies.

Small nucleolar RNAs transcribed by RNAPII are not polyadenylated. Analysis of yeast mutants revealed that termination on the corresponding genes involves Pcf11 but not Rat1 (50). Termination at these genes also involves two additional proteins: Nrd1 and Nab3 (9). Nrd1 binds specific RNA sequences in the terminator and the CTD of RNAPII. Cordon and coworkers (8) recently speculated that Nrd1 might dismantle an EC by bridging the CTD to the nascent transcript in a mechanism that could be analogous to Pcf11.

Transcriptional pausing is the nexus where the decision between continued elongation and termination is made. Any given sequence of DNA will have multiple sites where the sequence alone causes RNAP to pause. Certain pause sites may be reinforced by sequence-specific DNA-binding protein (52, 53) or position of nucleosomes. Recent evidence shows that one of the polyadenylation factors, CPSF, associates with the body of RNAPII and induces pausing upon its binding to the AAUAAA hexamer of the polyadenylation signal (54). Because pausing of RNAPII can be induced by multiple mechanisms, this could
explain the stochastic nature of the location where RNAPII terminates.

REFERENCES

1. Greger, I. H., Aranda, A., and Proudfoot, N. (2000) *Proc. Natl. Acad. Sci. U. S. A.* 97, 8415–8420
2. Gollnick, P., and Babitzke, P. (2002) *Biochim. Biophys. Acta* 1577, 240–250
3. Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., Wu, Y., Green, R. D., and Ren, B. (2005) *Nature* 436, 876 – 880
4. Peterlin, B. M., and Price, D. H. (2006) *Mol. Cell* 23, 297–305
5. Zhang, Z., Klett, A., Gilmour, D. S., and Henderson, A. J. (2007) *J. Biol. Chem.* 282, 16981–16988
6. Cullen, B. R. (1991) *FASEB J.* 5, 2361–2368
7. Zhang, Z., Klett, A., Henderson, A. J., and Gilmour, D. S. (2007) *Genes Dev.* 21, 1609–1614
8. Arigo, J. T., Carroll, K. L., Ames, J. M., and Corden, J. L. (2006) *Mol. Cell* 21, 641–651
9. Steinmetz, E. J., Conrad, N. K., Brow, D. A., and Corden, J. L. (2001) *Nature* 413, 327–331
10. Landick, R. (1997) *Cell* 88, 741–744
11. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. (2001) *Science* 292, 1876–1882
12. Kozheneva, N., Mustaev, A., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. A. (2000) *Science* 289, 619–625
13. von Hippel, P. H. (1998) *Science* 281, 660–665
14. Tadigotla, V. R., Ö Maiošlídigh, D., Sengupta, A. M., Epshtein, V., Ebright, R. H., Nudler, E., and Ruchkin, A. E. (2006) *Proc. Natl. Acad. Sci. U. S. A.* 103, 4439–4444
15. Kireeva, M. L., Komissarova, N., Vaugh, D. S., and Kashlev, M. (2000) *J. Biol. Chem.* 275, 6530–6536
16. Komissarova, N., Becker, J., Solter, S., Kireeva, M., and Kashlev, M. (2002) *Mol. Cell* 10, 1115–1162
17. Park, J. S., and Roberts, J. W. (2006) *Proc. Natl. Acad. Sci. U. S. A.* 103, 4870–4875
18. Zhou, Y., Navaroli, D. M., Eunamhe, M. S., and Martin, C. T. (2007) *Proc. Natl. Acad. Sci. U. S. A.* 104, 10352–10357
19. Yarnell, W. S., and Roberts, J. W. (1999) *Science* 284, 611–615
20. Santangelo, T. J., and Roberts, J. W. (2004) *Mol. Cell* 14, 117–126
21. Ciampi, M. S. (2006) *Microbiology (Read.)* 152, 2515–2528
22. Deaconescu, A. M., Savery, N., and Darst, S. A. (2007) *Curr. Opin. Struct. Biol.* 17, 96–102
23. Touloupkine, I., and Landick, R. (2003) *Mol. Cell* 12, 1125–1136
24. Jiang, Y., Liu, M., Spencer, C. A., and Price, D. H. (2004) *Mol. Cell* 14, 375–385
25. Hara, R., Selby, C. P., Liu, M., Price, D. H., and Sancar, A. (1999) *J. Biol. Chem.* 274, 24779–24786
26. Braglia, P., Percudani, R., and Dieci, G. (2005) *J. Biol. Chem.* 280, 19551–19562
27. Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C., Steuerwald, U., Ruirog, R. W., Sentenac, A., Muller, C. W., and Schoehn, G. (2007) *Mol. Cell* 25, 813–823
28. Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. (2006) *EMBO J.* 25, 118–128
29. Lang, W. H., and Reeder, R. H. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 9781–9785
30. Jeong, S. W., Lang, W. H., and Reeder, R. H. (1995) *Mol. Cell. Biol.* 15, 5929–5936
31. Jansa, P., and Grummt, I. (1999) *Mol. Gen. Genet.* 262, 508–514
32. Maraia, R. J., and Intine, R. V. (2001) *Mol. Cell. Biol.* 21, 367–379
33. Maraia, R. J., Kenan, D. J., and Keene, J. D. (1994) *Mol. Cell. Biol.* 14, 2147–2158
34. Tennyson, C. N., Klamut, H. J., and Morton, R. G. (1995) *Nat. Genet.* 9, 184–190
35. Kerppola, T. K., and Kane, C. M. (1990) *Biochemistry* 29, 269–278
36. Rosonina, E., Kaneko, S., and Manley, J. L. (2006) *Genes Dev.* 20, 1050–1056
37. Birse, C. E., Minvielle-Sebastia, L., Lee, B. A., Keller, W., and Proudfoot, N. J. (1998) *Science* 280, 298–301
38. Kim, M., Krogan, N. J., Greenblatt, J. F., and Buratowski, S. (2004) *EMBO J.* 23, 354–364
39. Venkataraman, K., Brown, K. M., and Gilmartin, G. M. (2005) *Genes Dev.* 19, 1315–1327
40. Luo, W., and Bentley, D. (2004) *Cell* 119, 911–914
41. West, S., Gromak, N., and Proudfoot, N. J. (2004) *Nature* 432, 522–525
42. Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., and Buratowski, S. (2004) *Nature* 432, 517–522
43. Zhang, Z., Fu, J., and Gilmour, D. S. (2005) *Genes Dev.* 19, 1572–1580
44. Zhang, Z., and Gilmour, D. S. (2006) *Mol. Cell* 21, 65–74
45. Buratowski, S. (2003) *Nat. Struct. Biol.* 10, 679–680
46. McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M., and Bentley, D. L. (1997) *Nature* 385, 357–361
47. Gross, S., and Moore, C. (2001) *Proc. Natl. Acad. Sci. U. S. A.* 98, 6080–6085
48. Wu, W., Wind, M., and Reines, D. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 6935–6940
49. Sadowski, M., Dichtl, B., Hubner, W., and Keller, W. (2003) *EMBO J.* 22, 2167–2177
50. Kim, M., Vasiljeva, L., Rando, O. J., Zhelkovsky, A., Moore, C., and Buratowski, S. (2006) *Mol. Cell* 24, 723–734
51. Luo, W., Johnson, A. W., and Bentley, D. L. (2006) *Genes Dev.* 20, 954–965
52. Connelly, S., and Manley, J. L. (1989) *Mol. Cell. Biol.* 9, 5254–5259
53. Ashfield, R., Patel, A. J., Bossone, S. A., Brown, H., Campbell, R. D., Marcu, K. B., and Proudfoot, N. J. (1994) *EMBO J.* 13, 5656–5667
54. Nag, A., Narisinh, K., and Martinson, H. G. (2007) *Nat. Struct. Mol. Biol.* 14, 662–669