Title: A Novel Initiation Pathway in *Escherichia Coli* Transcription

Authors: Eitan Lerner[1‡], SangYoon Chung[1‡], Benjamin L. Allen[2], Shuang Wang[3], Jookyung J. Lee[4], Winson S. Lu[1], Wilson L. Grimaud[1], Antonino Ingargiola[1], Xavier Michalet[1], Yazan Alhadid[1], Sergei Borukhov[4], Terence Strick[3,] Dylan J. Taatjes[2], Shimon Weiss[1,5,6*]

Affiliations:

[1]Dept. of Chemistry & Biochemistry, University of California Los Angeles, Los Angeles, CA 90095.

[2]Dept. of Chemistry & Biochemistry, University of Colorado, Boulder.

[3]Institut Jacques Monod, Centre National de la Recherche Scientifique and University of Paris Diderot and Sorbonne Paris Cité, Paris, France.

[4] Rowan University School of Osteopathic Medicine, Stratford, NJ 08084, USA.

[5]California NanoSystems Institute, University of California Los Angeles, Los Angeles, CA 90095.

[6]Dept. of Physiology, University of California Los Angeles, Los Angeles, CA 90095.

[‡] Equal contributions

[*]Corresponding author: sweiss@chem.ucla.edu

Abstract:

Initiation is a rate-limiting step in transcription. We employed a series of approaches to examine the kinetics of RNA polymerase (RNAP) transcription initiation in greater detail. We show that, contrary to expectations, RNAP exit kinetics from later stages of initiation (e.g. from a 7-base transcript) was markedly slower than from earlier stages. Further examination implicated a previously unidentified intermediate in which RNAP adopted a long-lived backtracked state during initiation. In agreement, the RNAP-GreA endonuclease accelerated transcription kinetics from otherwise delayed initiation states and prevented RNAP backtracking. Our results indicate a previously uncharacterized RNAP initiation state that could be exploited for therapeutic purposes and may reflect a conserved intermediate among paused, initiating eukaryotic enzymes.

One Sentence Summary:

A previously undiscovered RNA polymerase initiation state is described, in which the nascent RNA backtracks and the enzyme pauses.

Main Text:

Transcription of genomic DNA requires formation of an RNA polymerase (RNAP)-promoter initially-transcribing complex (RP_{ITC}), in which RNAP unwinds DNA and generates a mechanically-stressed intermediate (*1*) through DNA scrunching (*2, 3*). The interaction of the σ70 RNAP subunit with the promoter creates a physical roadblock for the nascent RNA transcript (*4, 5*). For productive transcription to take place, this roadblock has to be removed (*4, 6*), otherwise RNAP enters repeated cycles of unsuccessful transcription attempts (*7, 8*). To quantitatively study the mechanism of transcription initiation by RNAP, we developed a single-round quenched kinetics assay (**Fig. 1**) to probe the kinetics of transcription by directly counting the number of transcripts produced per a given time. Using this assay, we examined the kinetics of *E. coli* RNAP transcription from distinct RP_{ITC} states generated via NTP starvation (**Figs. 1A** & **1B**). Because transcription initiation is much slower than elongation (i.e. initiation is rate-limiting) (*9*), the synthesis of relatively short, yet full-length RNA products (39- and 41-base transcripts) reflects the rate of transcription initiation (**Figs. 1C** & **1D**).

We anticipated that RNAP transcription kinetics from a late $RP_{ITC\leq i}$ state ('exit kinetics') would be similar, if not faster, than from an early $RP_{ITC\leq j}$ state (i>j). Strikingly, however, we observed that exit kinetics from the $RP_{ITC\leq 4}$, $RP_{ITC\leq 6}$, or $RP_{ITC\leq 7}$ states were slower than from $RP_{ITC=2}$ state (**Fig. 1D**). In fact, whereas exit kinetics from $RP_{ITC=2}$ was similar to that of RNAP already in the elongation state ($RD_{E=11}$), exit kinetics from $RP_{ITC\leq 7}$ was at least 3.5-fold slower (**Table S1**). These results suggest that, contrary to existing models, a previously undetected state exists for RNAP initiation complexes ($RP^*_{ITC\leq i}$, **Fig. 1E**). Importantly, this state is transient and overall RNAP activity remains unchanged, given that all "delayed" RNAP complexes ($RP_{ITC\leq 4}$, $RP_{ITC\leq 6}$, or $RP_{ITC\leq 7}$) eventually transition to elongation (**Fig. 1D**). RNAP activity was assessed as described in the Methods and Materials section.

It is well-established that elongating RNAP enzymes can backtrack and pause (*10, 11*). In such circumstances, the nascent RNA 3'-end backtracks into the secondary channel, where it is subject to endonucleolytic cleavage by the GreA-RNAP complex (*10, 12, 13*). To test whether delayed exit kinetics for $RP_{ITC\leq 7}$ was due to backtracking during initiation, we assessed the effect of GreA using our single-round

quenched kinetics assay. As shown in **Figs. 2A** & 2**B**, 1 μM GreA accelerated the exit kinetics from $RP_{ITC\leq7}$ relative to the exit kinetics from $RP_{ITC=2}$. (~50% recovery from the $RP^*_{ITC\leq7}$ delayed state; see Methods and Materials and **Table S1**). These results are consistent with GreA-dependent release of RNAP from a backtracked and paused state during initiation.

To further test the effect of GreA during RNAP transcription initiation, we performed *in vitro* transcription assays in which ^{32}P-labeled RNAP transcripts were quantitated following poly-acrylamide gel electrophoresis (PAGE). This enabled identification of various abortive initiation products (band assignment in **Fig. S1A**) and thus allows determining whether GreA catalyzes cleavage of short transcripts during transcription initiation. As shown in **Figs. 2C** & 2**D**, the 7-base abortive RNA product was not generated in the presence of GreA. Because GreA stimulates cleavage only of "backtracked" RNA (i.e. with the 3'-end inserted into the secondary channel), these data, combined with our single-round kinetics data, confirmed that RNAP is capable of backtracking during transcription initiation. However, we cannot confirm whether only one base or more were cleaved by GreA in this assay (see Methods and Materials).

Although reduced NTP levels may occur *in vivo* (e.g. metabolic stress), the absence of select NTPs is unlikely. Therefore, we next asked whether RNAP backtracking during initiation would occur under more physiologically relevant conditions. We initially examined transcription with the quenched kinetics assay under NTP concentration imbalance (UTP and GTP >> ATP and CTP). Consistent with the results described above, we observed a delay in exit kinetics from the $RP_{ITC=2}$ state under conditions of NTP imbalance (compared to equimolar conditions; **Fig. S2**) at each of the two different promoter templates tested (**Fig. S3**).

The single-round quenched kinetics and transcription assays with ^{32}P-labeled UTP are ensemble experiments, which cannot reliably detect rare events. Indeed, with all NTPs in equimolar amounts (100 μM), we did not observe the 7-base abortive transcript, suggesting that when NTPs are abundant, RNAP rarely backtracks from $RP_{ITC\leq7}$ state (**Fig. S1B**). To detect potentially rare intermediates under more physiologically relevant conditions of equimolar NTPs, we implemented magnetic tweezers experiments with supercoiled promoter templates (**Fig. 3A**). This assay allowed us to track individual RNAP complexes

over time and simultaneously detect and identify distinct RP_{ITC} states, based upon well-established changes in DNA extension (*3*). In the absence of GreA, we observed short- and long-lived RP_{ITC} states (**Figs. 3B**, **3D**, **3F** & **3H**). The lifetimes spent in RP_{ITC} states are summarized in a histogram fitted with a double exponential in which 90% of events (n=216) were short-lived (τ=300±40 s SEM), and 10% were long-lived (τ=2600±700 s SEM; **Fig. 3H, blue**). Correlating these data with DNA bubble sizes (representing distinct RP_{ITC} states; **Figs. 3A** & **3F**) revealed that 90% of events were characterized by a 15 ± 3 bp transcription bubble (SD). The remaining 10% of events were characterized by a more homogeneous 10 ± 1 bp transcription bubble (SD). In agreement, correlative analyses (**Fig. 3D**) indicated RP_{ITC} states with a shorter bubble (<12 bases) can be long-lived (~2200 ± 350 s SEM, n=37;**Fig. 3D oval**) whereas RP_{ITC} states with a larger bubble (>12 bases) were generally shorter-lived (~1100 ± 110 s SEM, n=179).

These data suggested that, in addition to the well-characterized RP_{ITC} state (**Fig. 3A**), a subset of RNAP complexes entered a distinct, long-lived state characterized by a smaller transcription bubble (denoted RP^*_{ITC} in **Fig. 3A**, **3B** & **3C**). We hypothesized that this long-lived initiation intermediate represented backtracked RNAP previously characterized in our quenched kinetics and gel-based transcription assays. If correct, the addition of GreA would be expected to markedly reduce the number of these long-lived events. In agreement, RP_{ITC} states (n=209) became uniformly short-lived (τ=350± 30 s SEM) in the presence of GreA, with transcription bubbles of larger sizes (16 ± 2 bases (SEM); **Figs. 3C**, **3E**, **3G** & **3H**). Experiments completed on a different promoter template showed the same trend (**Fig. S4**). In summary, without GreA, the bubble size in most events is small, while with GreA there is an increase in the number of events associated with larger bubble size as well as a net reduction in lifetime (**Fig. S4D & S4E, quadrants**). Collectively, these data further support the existence of a long-lived backtracked state during initiation and reveal a role for GreA in preventing RNAP from entering this state.

Our results demonstrate the existence of a previously uncharacterized state in which RNAP backtracks and pauses during transcription initiation, and revealing the key roles of GreA and NTP availability in regulating the flux in or out of this state. Because basic mechanisms of transcription by RNA

polymerases are broadly conserved (*14*) (e.g. scrunching (*2, 3, 15*), similar mechanistic role of σ70 and TFIIB (*16, 17*), trigger-loop function in catalysis (*18-22*)), it will be important to determine whether similar backtracked states are adopted during transcription initiation by eukaryotic RNA polymerases. Mammalian RNA polymerase II (Pol II) pauses during early stages of transcription, and this represents a common regulatory intermediate (*23*). Potentially, a mechanistic intermediate of paused mammalian Pol II enzymes may involve RNA backtracking; such backtracked intermediates may help explain why TFIIS, a eukaryotic ortholog of GreA, has been linked to transcription initiation and assembles with Pol II at the promoter (*24-26*). Finally, we emphasize that, because RNAP transcription initiation is rate-limiting and highly regulated *in vivo*, this previously unidentified backtracked/paused RNAP state may lead to potential new strategies for molecular therapeutics.

References and Notes

1. D. C. Straney, D. M. Crothers, A stressed intermediate in the formation of stably initiated RNA chains at the Escherichia coli lac UV5 promoter. *Journal of Molecular Biology* **193**, 267 (1987).

2. A. N. Kapanidis *et al.*, Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144 (2006).

3. A. Revyakin, C. Liu, R. H. Ebright, T. R. Strick, Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139 (2006).

4. D. Pupov, I. Kuzin, I. Bass, A. Kulbachinskiy, Distinct functions of the RNA polymerase sigma subunit region 3.2 in RNA priming and promoter escape. *Nucleic Acids Research* **42**, 4494 (2014).

5. Y. Zuo, T. A. Steitz, Crystal Structures of the E. coli Transcription Initiation Complexes with a Complete Bubble. *Molecular Cell* **58**, 534 (2015).

6. S. Samanta, C. T. Martin, Insights into the mechanism of initial transcription in Escherichia coli RNA polymerase. *The Journal of Biological Chemistry* **288**, 31993 (2013).

7. A. J. Carpousis, J. D. Gralla, Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the lac UV5 promoter. *Biochemistry* **19**, 3245 (1980).

8. L. M. Hsu, Promoter clearance and escape in prokaryotes. *Biochimica et Biophysica Acta* **1577**, 191 (2002).

9. E. Margeat *et al.*, Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. *Biophysical Journal* **90**, 1419 (2006).

10. S. Borukhov, J. Lee, O. Laptenko, Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Molecular Microbiology* **55**, 1315 (2005).

11. F. Toulme *et al.*, GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *The EMBO Journal* **19**, 6853 (2000).

12. O. Laptenko, J. Lee, I. Lomakin, S. Borukhov, Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *The EMBO Journal* **22**, 6322 (2003).

13. E. Stepanova *et al.*, Analysis of promoter targets for Escherichia coli transcription elongation factor GreA in vivo and in vitro. *Journal of Bacteriology* **189**, 8772 (2007).

14. F. Werner, D. Grohmann, Evolution of multisubunit RNA polymerases in the three domains of life. *Nature Reviews. Microbiology* **9**, 85 (2011).

15. F. M. Fazal, C. A. Meng, K. Murakami, R. D. Kornberg, S. M. Block, Real-time observation of the initiation of RNA polymerase II transcription. *Nature* **525**, 274 (2015).

16. K. S. Murakami, S. A. Darst, Bacterial RNA polymerases: the wholo story. *Current Opinion in Structural Biology* **13**, 31 (2003).

17. S. Sainsbury, J. Niesser, P. Cramer, Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* **493**, 437 (2013).

18. L. A. Allison, M. Moyle, M. Shales, C. J. Ingles, Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**, 599 (1985).

19. G. Bar-Nahum *et al.*, A ratchet mechanism of transcription elongation and its control. *Cell* **120**, 183 (2005).

20. D. S. Hekmatpanah, R. A. Young, Mutations in a conserved region of RNA polymerase II influence the accuracy of mRNA start site selection. *Molecular and Cellular Biology* **11**, 5781 (1991).

21. V. Thuillier, I. Brun, A. Sentenac, M. Werner, Mutations in the alpha-amanitin conserved domain of the largest subunit of yeast RNA polymerase III affect pausing, RNA cleavage and transcriptional transitions. *The EMBO Journal* **15**, 618 (1996).

22. R. Weilbaecher, C. Hebron, G. Feng, R. Landick, Termination-altering amino acid substitutions in the beta' subunit of Escherichia coli RNA polymerase identify regions involved in RNA chain elongation. *Genes & Development* **8**, 2913 (1994).

23. K. Adelman, J. T. Lis, Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews. Genetics* **13**, 720 (2012).

24. B. Guglielmi, J. Soutourina, C. Esnault, M. Werner, TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16062 (2007).

25. B. Kim *et al.*, The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16068 (2007).

26. E. V. Stepanova, A. B. Shevelev, S. I. Borukhov, K. V. Severinov, [Mechanisms of action of RNA polymerase-binding transcription factors that do not bind to DNA]. *Biofizika* **54**, 773 (2009).

Acknowledgements:

Author Contributions: E.L. and S.C. developed and designed the quenched-kinetics experiments. E.L., S.C., W.S.L., W.L.G., and Y.A. performed the quenched-kinetics measurements, in consultation with A.I. and X.M. B.A. designed and performed gel-based experiments; S.W. and T.S. designed, performed and analyzed magnetic tweezers experiments; J.J.L. and S.B. prepared GreA protein. E.L., S.C., B.A., S.B., D.J.T., and S.W. analyzed the data; E.L., S.C., T.S., B.A., D.J.T., and S.W. wrote the paper.

We thank Prof. William Gelbart, Prof. Charles Knobler, Dr Cathy Yan Jin and Xiyu Yi for fruitful discussions and Maya Lerner for preparation of illustrations. This work was funded by the NIH (GM069709 to SW, GM095904 to XM and SW) and NSF (MCB-1244175 to SW and DJT).

Fig. 1: Quenched kinetics transcription results identify an initiation-related stalled state. (A) Representative promoter sequence used here to show how by changing the initially transcribed sequence (ITS, cyan), different NTP-starved states can be generated ($RP_{ITC=2}$, $RP_{ITC\leq4,\,6,\,7}$, $RD_{E=11}$). Other regions of the promoter include the promoter recognition sequence (PRS, pink) and the elongation sequence (yellow), including a probe target complementary sequence (red). All promoters measured are described in **Fig. S3**. (B) Schematic of RNAP run-off transcription starting from a particular NTP-starved state. Upon supplementing all NTPs, transcription kinetics starts and transcripts are quantified via hybridization to a FRET probe. (C) Example of quenched kinetics data generated from quantification of run-off transcripts. (D) Run-off kinetics from various NTP-starved states. Kinetics starting from late initiation states (e.g. $RP_{ITC\leq7}$, blue) are slower than from an earlier initiation state (e.g. $RP_{ITC=2}$, black). (E) A schematic of RNAP transcription initiation based upon classical models and the novel intermediate described here. We identify a slower initiation pathway (highlighted red) involving RNAP backtracking, in which backtracked RNA is ultimately cleaved or abortively released.

Fig. 2: GreA rescues the initiation-related delay and cleaves backtracked RNA in initiating RNAP. (A) Run-off transcription kinetics is slower when starting from $RP_{ITC\leq7}$ (blue) than from $RP_{ITC=2}$ (black). (B) With 1 μM GreA, the delay in transcription initiation is reduced. (C & D) Gel-based abortive initiation kinetics: without GreA, NTP-starved $RP_{ITC\leq7}$ produced abortive transcripts of up to 7 bases long, whereas this product is not produced with 1 μM GreA, suggesting 1-2 bases of 3'-backtracked RNA is cleaved by GreA during initiation.

Fig. 3: Backtracking in initiation correlates with RNAP pausing in the presence of equimolar NTPs. (A) Schematics of the magnetic tweezer transcription assay (see Materials and Methods). Representative bead extension trajectories shown for single-molecule transcription experiments without (B) or with (C) 1 μM GreA unwinding levels (grey lines) shown, indicating different bubble sizes imposed by different RNAP states (below). Yellow lines highlight typical lifetimes in each state. Unwinding levels and RP_{ITC} and RP^*_{ITC} lifetimes are summarized in scatter plots without (D) or with (E) 1 μM GreA; their 1D projections are shown in (F), (G), and (H). Ellipse in (D) highlights events with long lifetimes and reduced unwinding (see also in (F)), that are absent with GreA (E, G). 20-50 DNA templates used for each condition, with 5-10 transcription pulses per template.

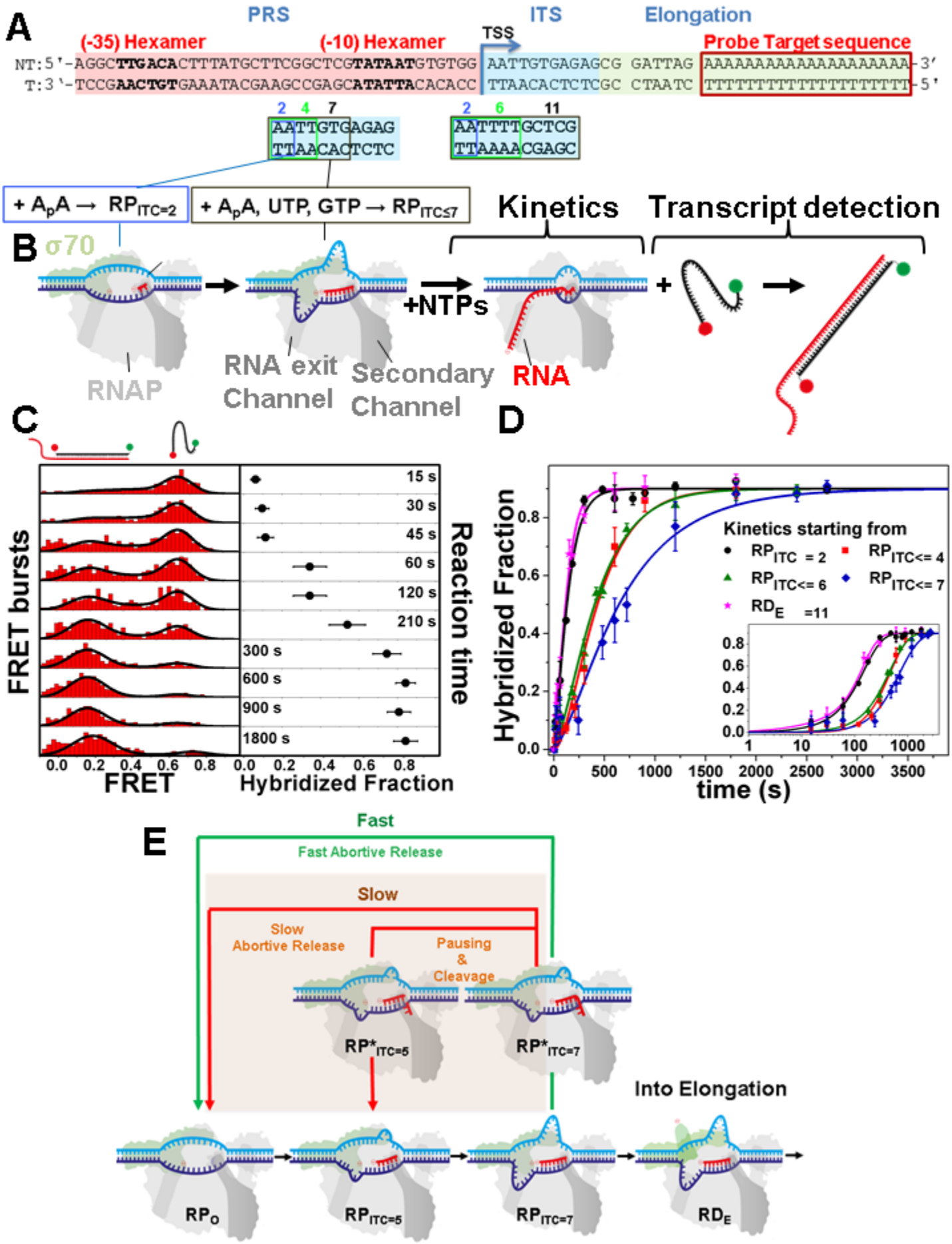

Fig. 1 Lerner, Chung et al.

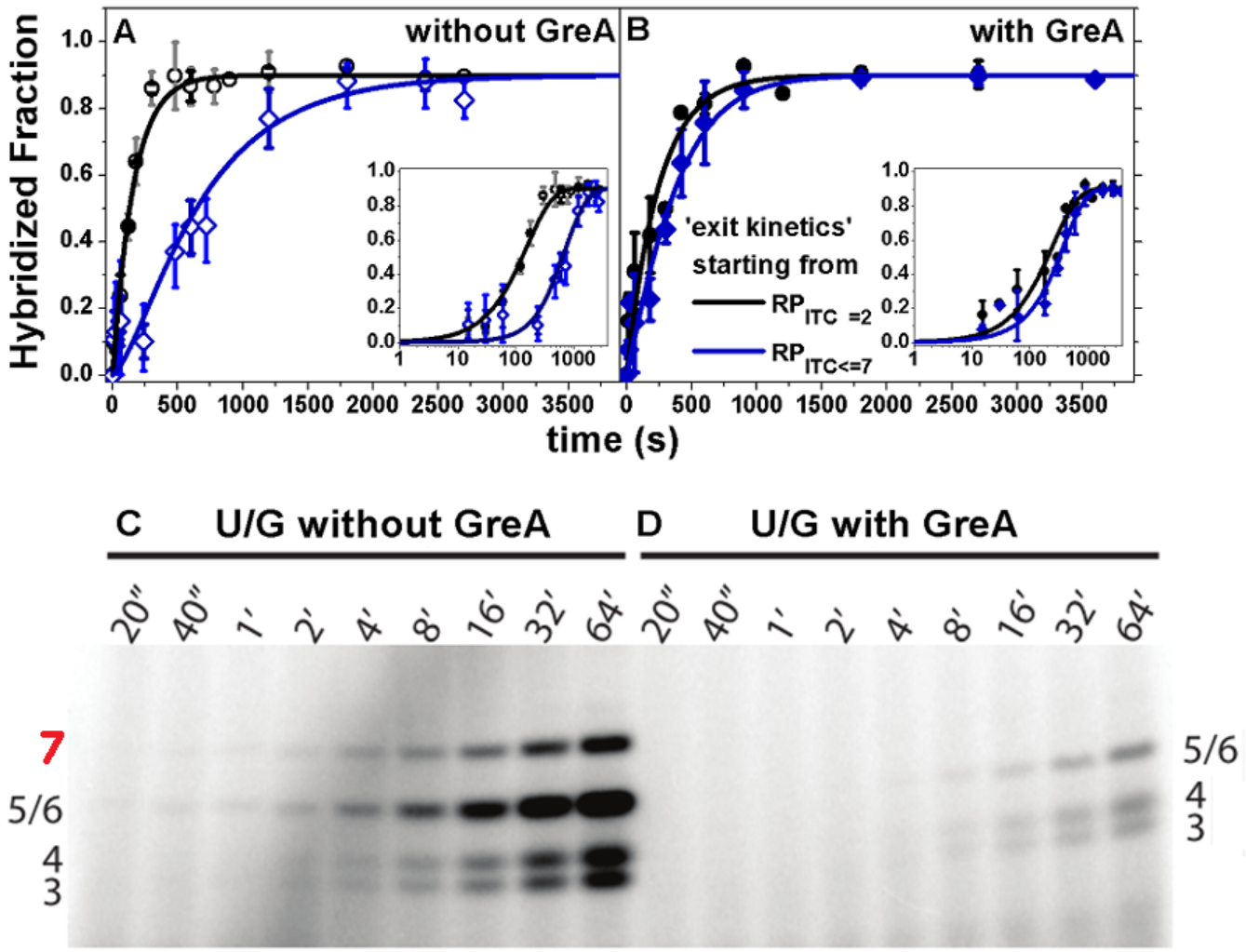

Fig. 2 Lerner, Chung et al.

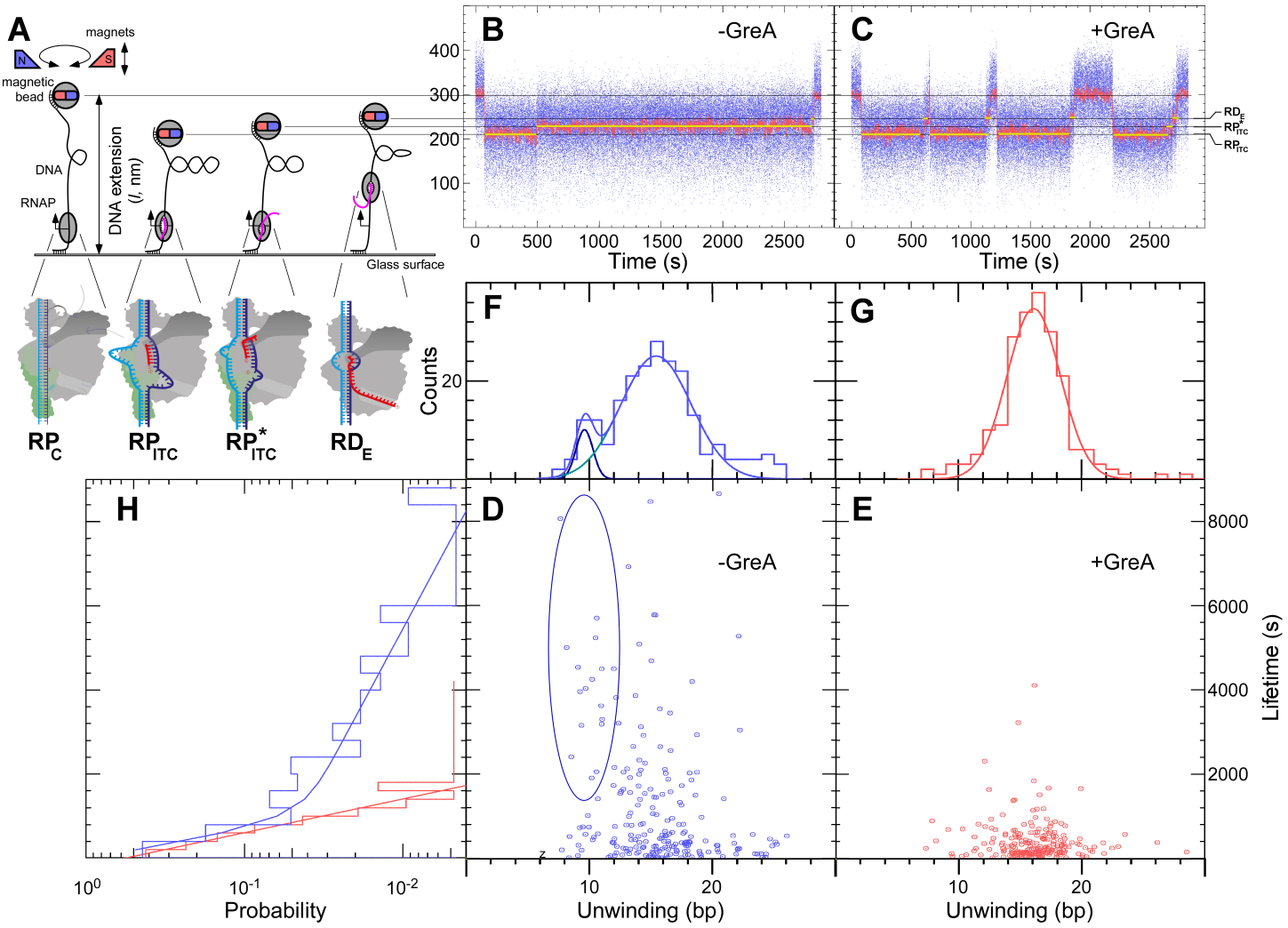

Fig. 3 Lerner, Chung et al.