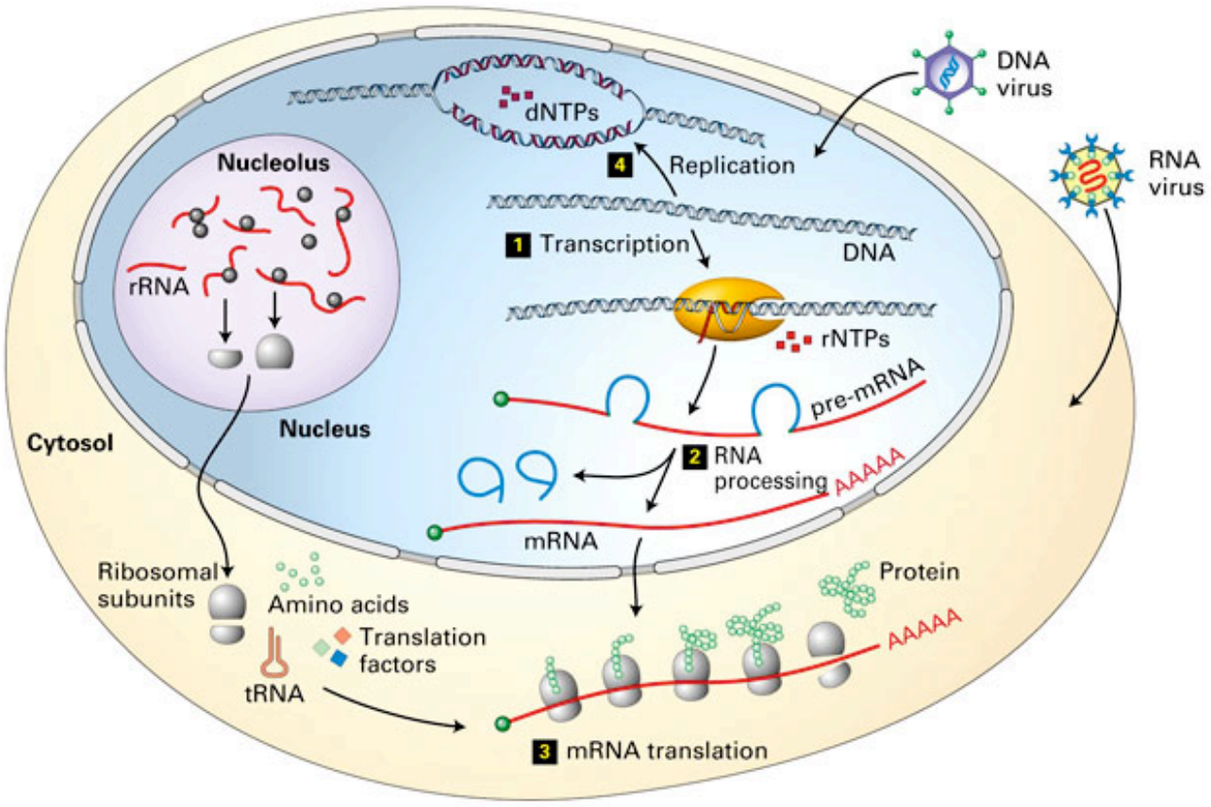
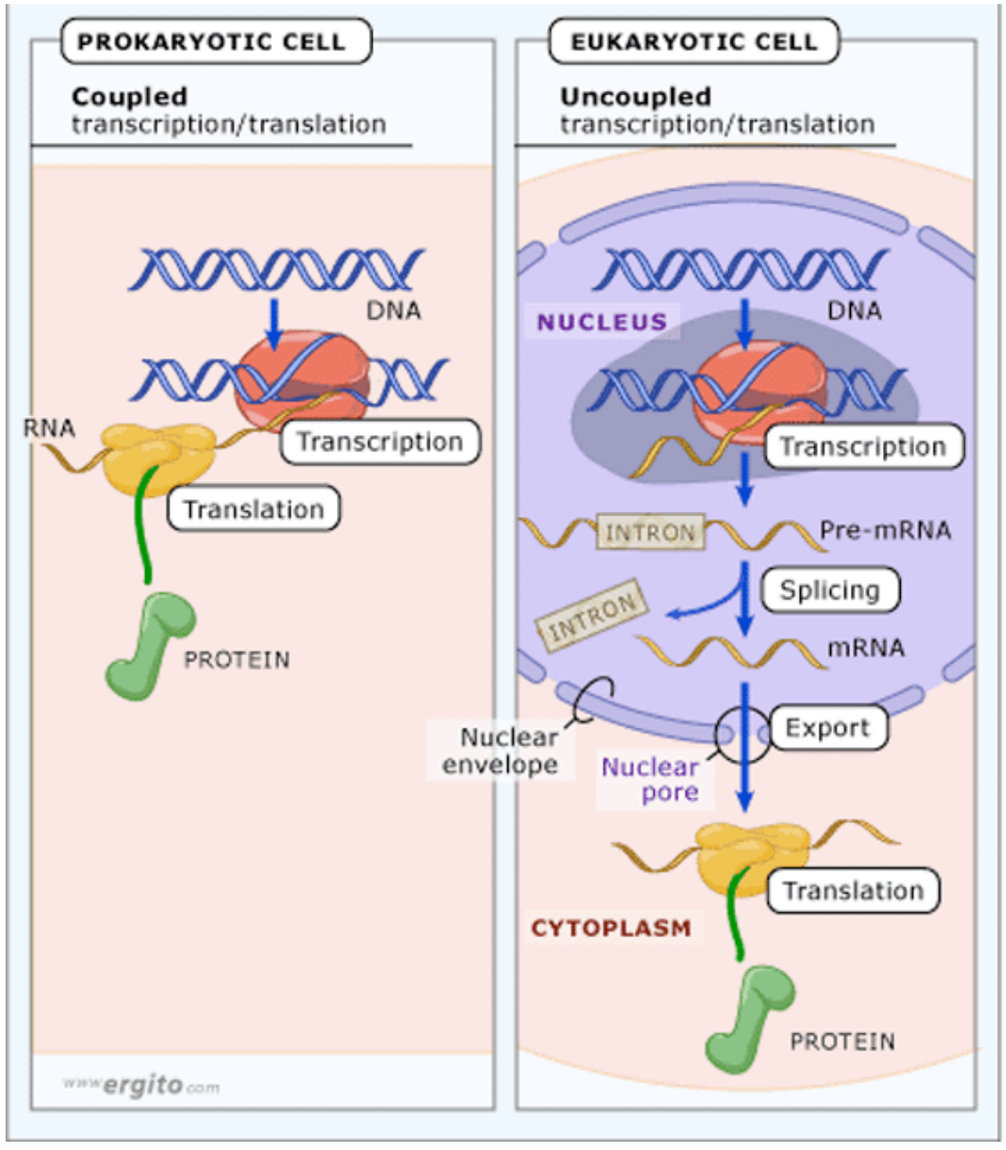
Direct Observation of Dynamics in Transcription at the Single Molecule Level

Joshua W. Shaevitz
shaevitz@princeton.edu
www.princeton.edu/~shaevitz

Collaborators:
Steven Block (Stanford)
Elio Abbondanzieri (Stanford)
William Greenleaf (Stanford)
Kristina Herbert (Stanford)
Matthew Larson (Stanford)
Robert Landick (Univ Wisc)
and more ...

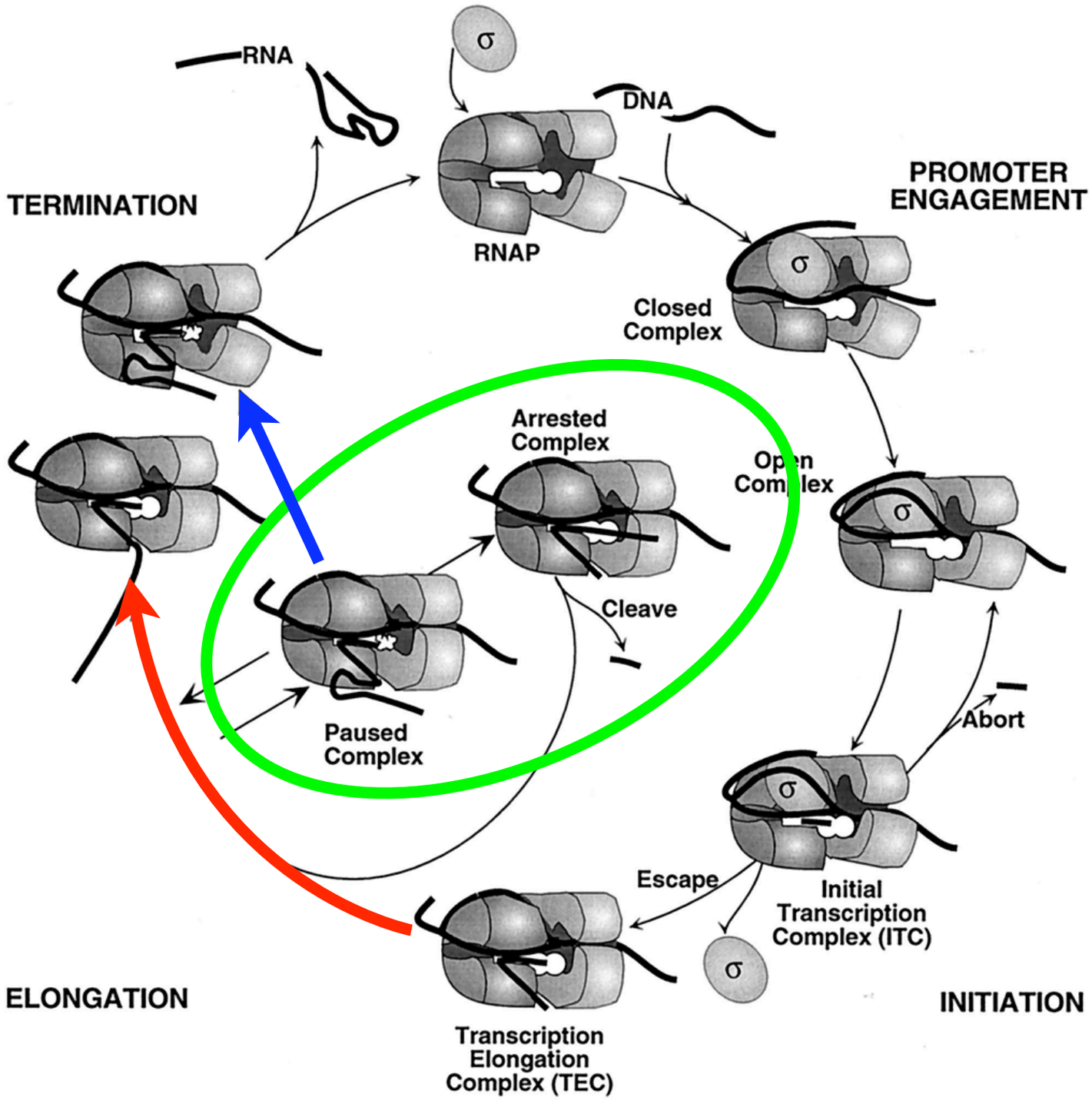
The central dogma of molecular biology





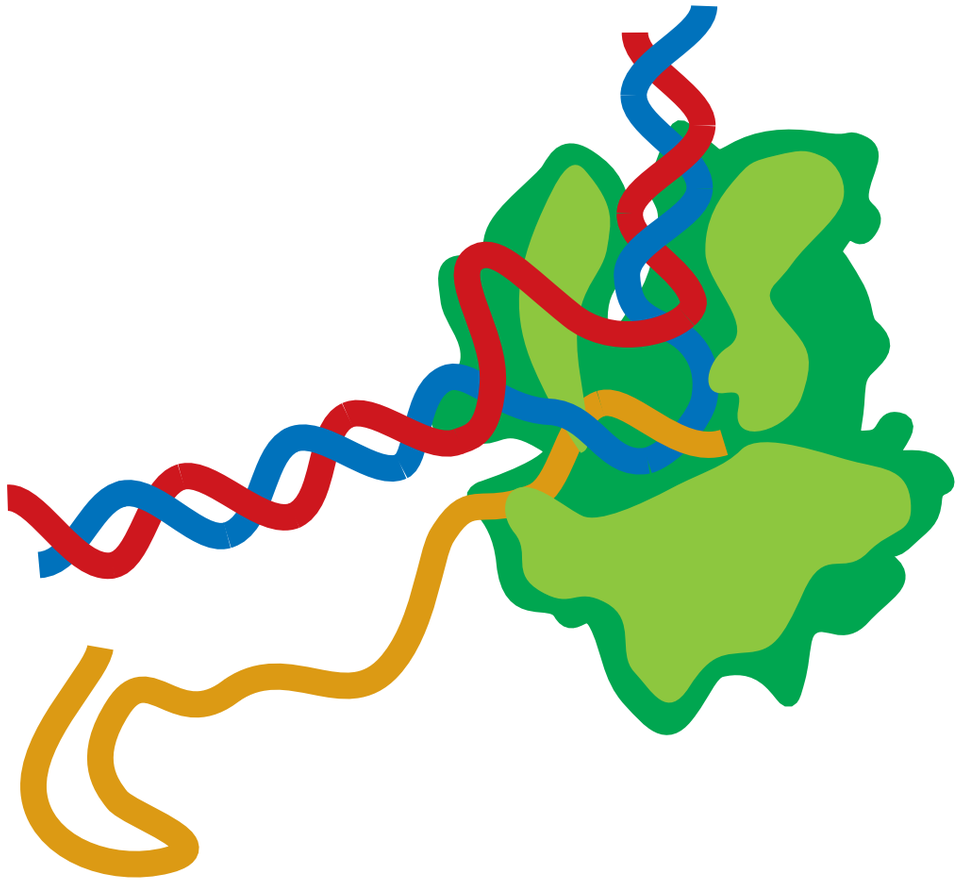
Regulation of transcription is the most common form of genetic control

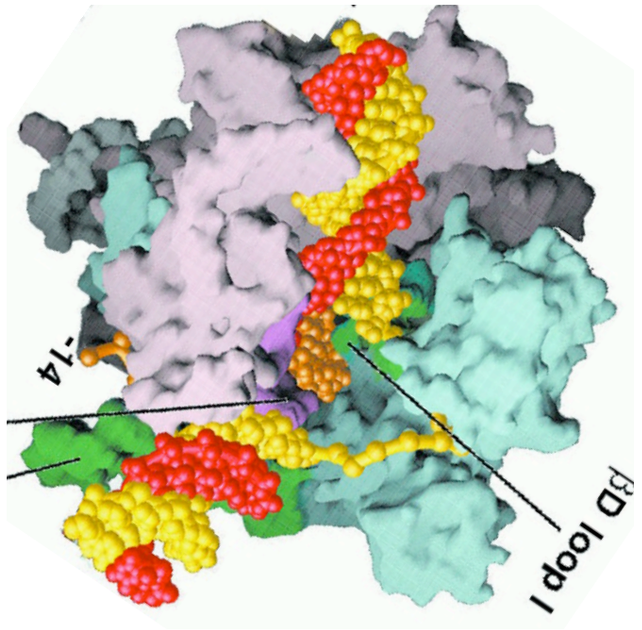
Transcription is a complex process



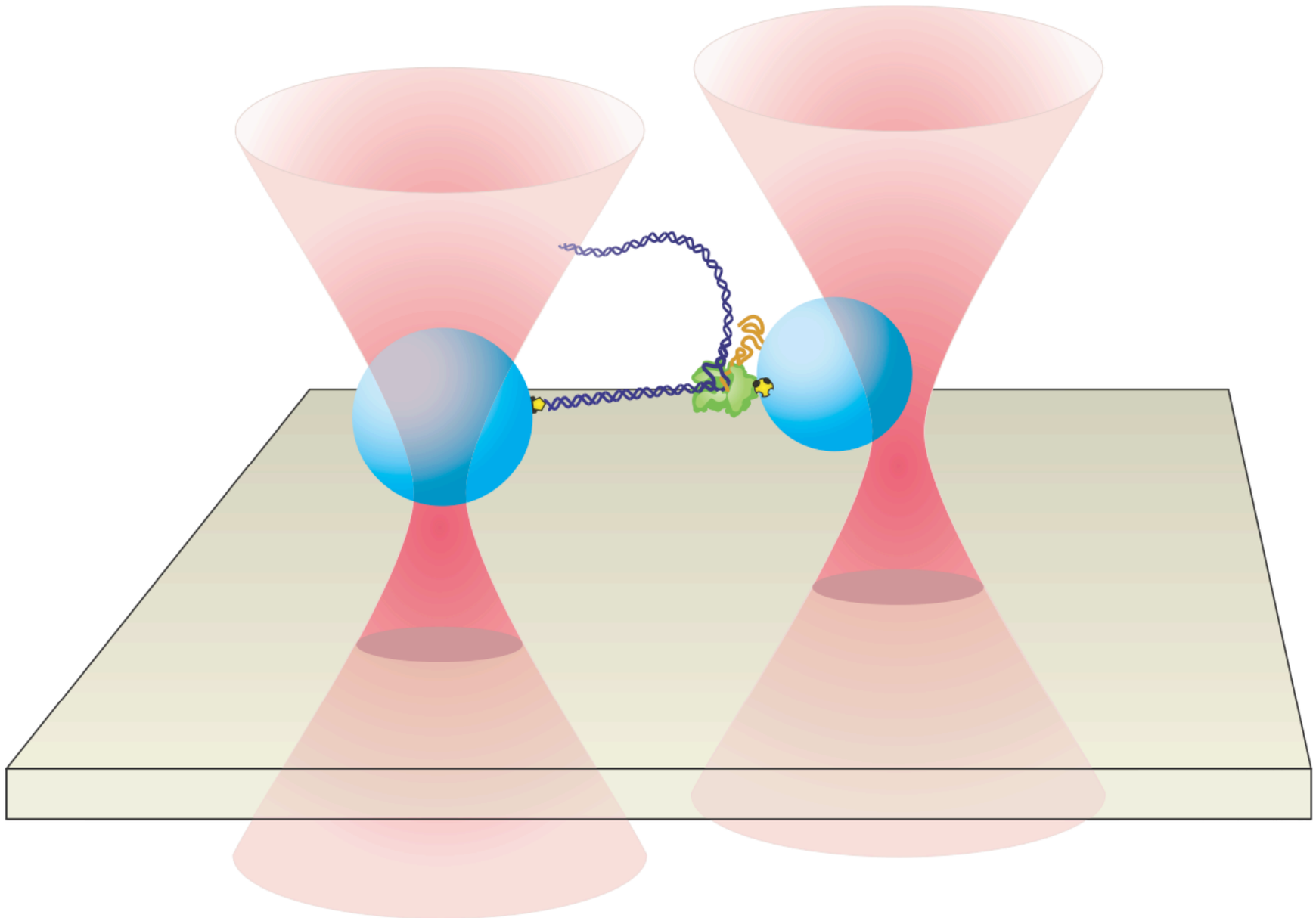
Mooney et al. *J. Bact* (1998)

RNA polymerase (RNAP) carries out transcription





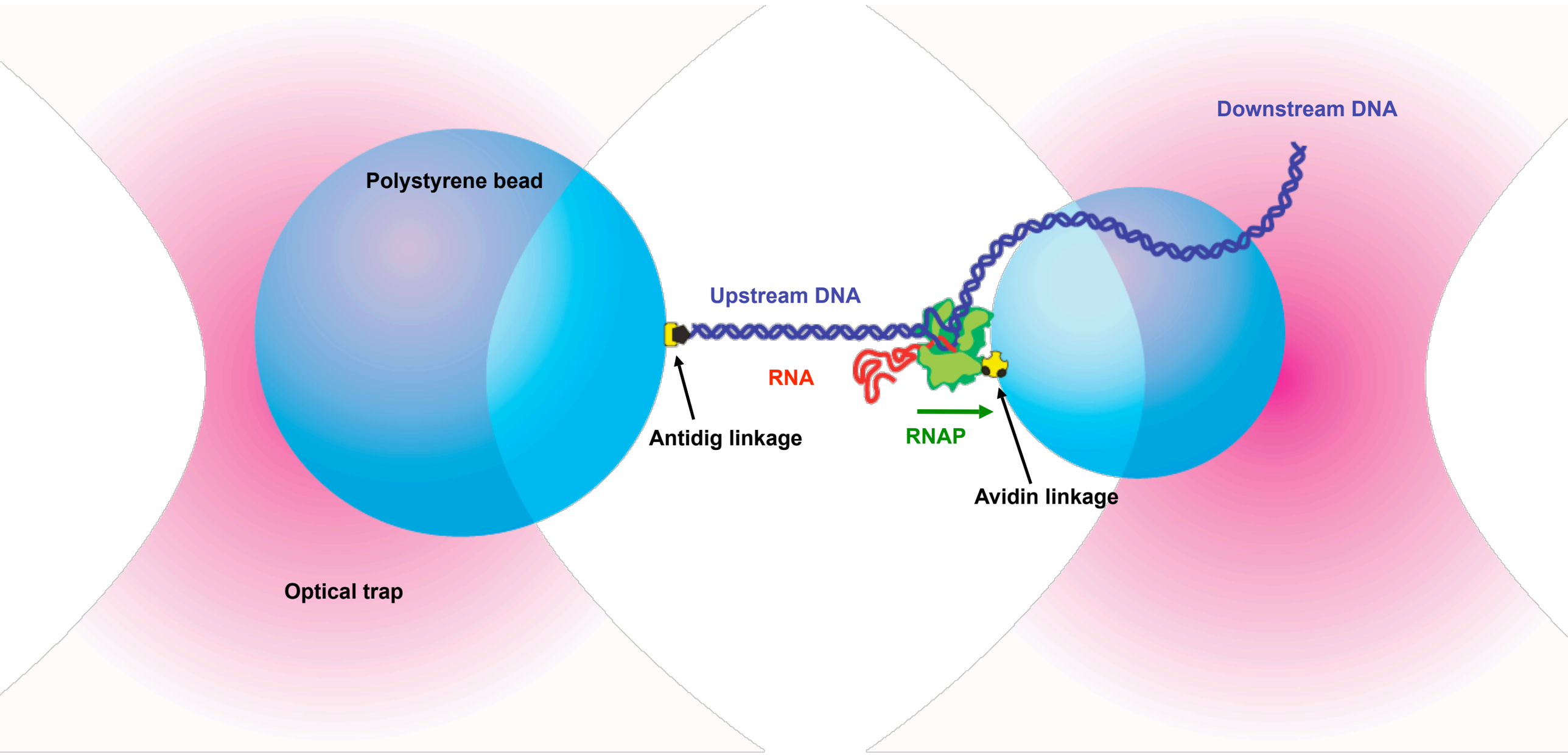
Levitated experiment avoids noise and drift



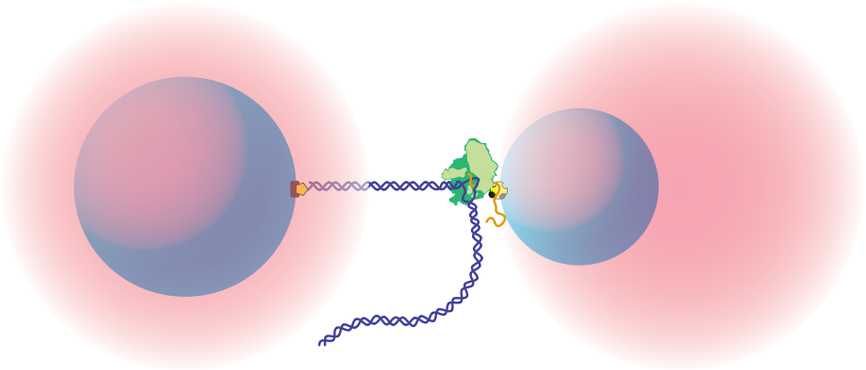
Massively ~~Parallel~~ Serial Enzymology!
(but very precise ...)

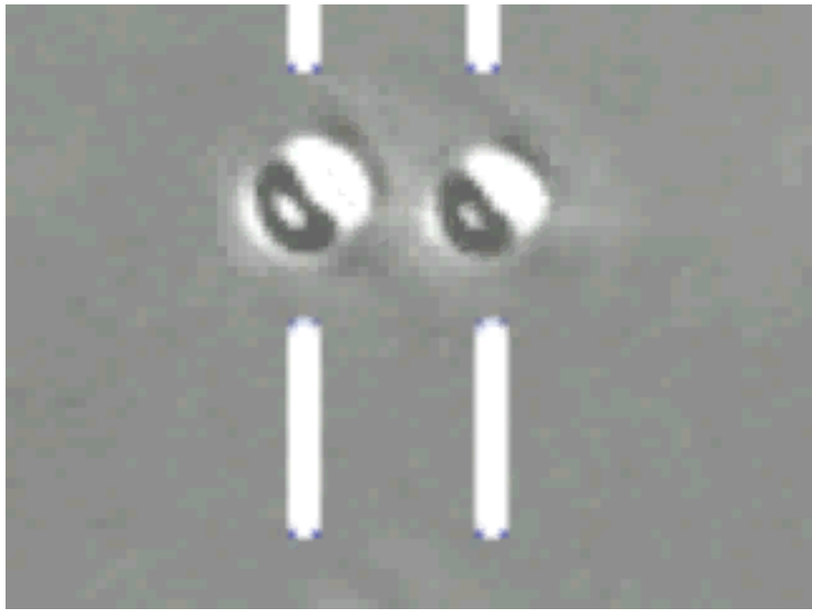
Shaevitz et al. *Nature* (2003)

Transcription assay in an optical trap

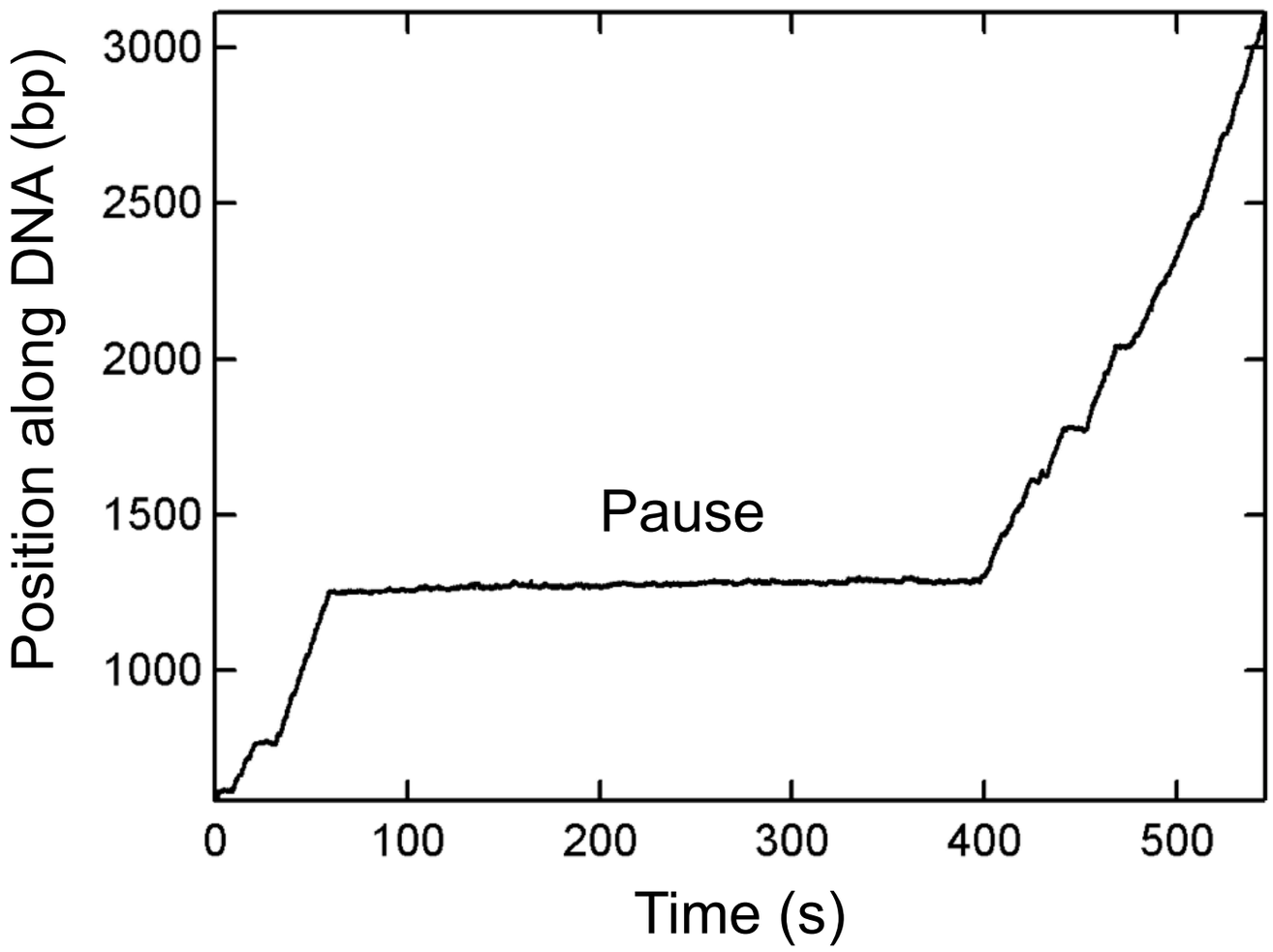


RNAP motion in the optical trap

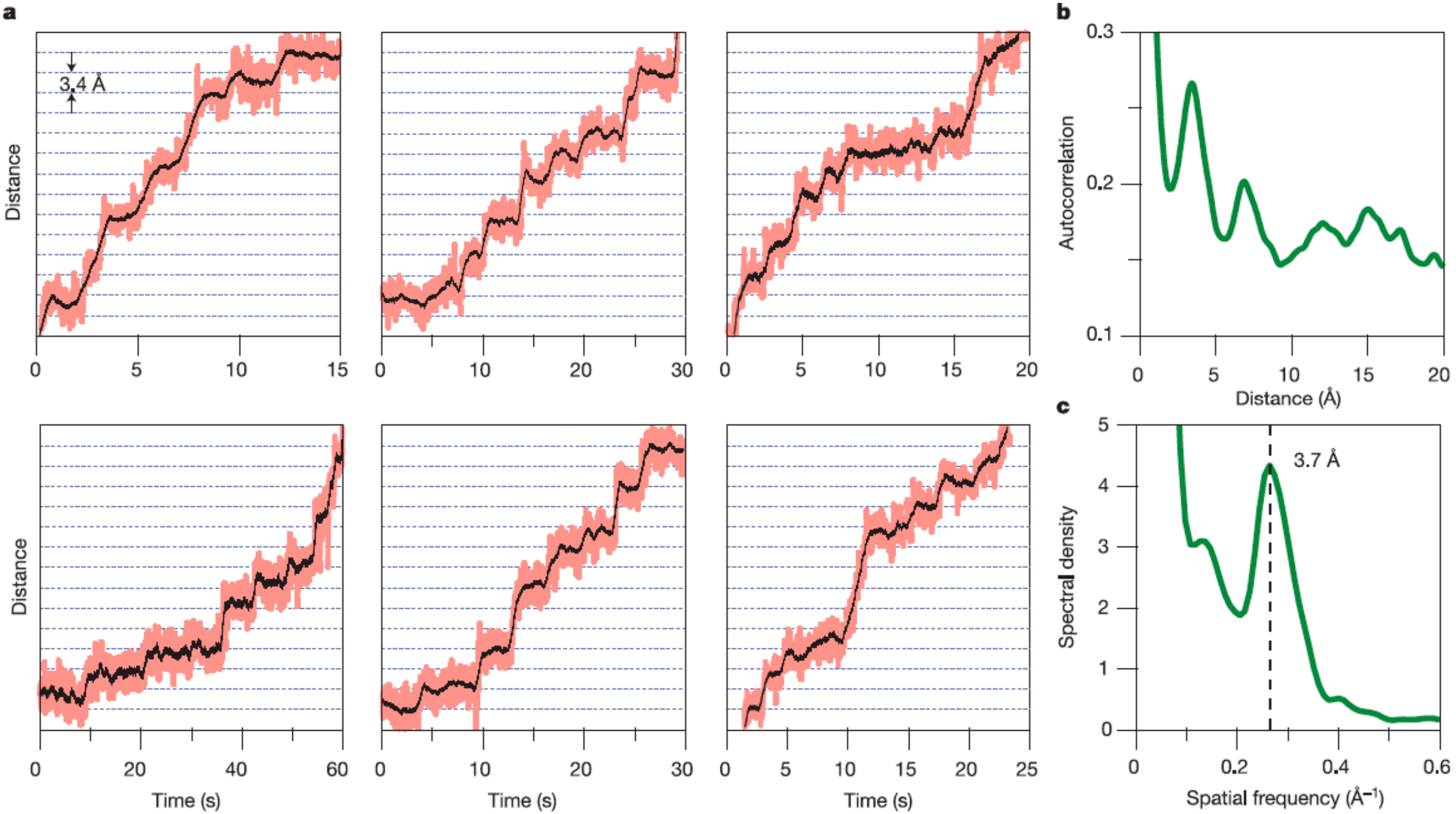




60x speed
1 sec = 1 min

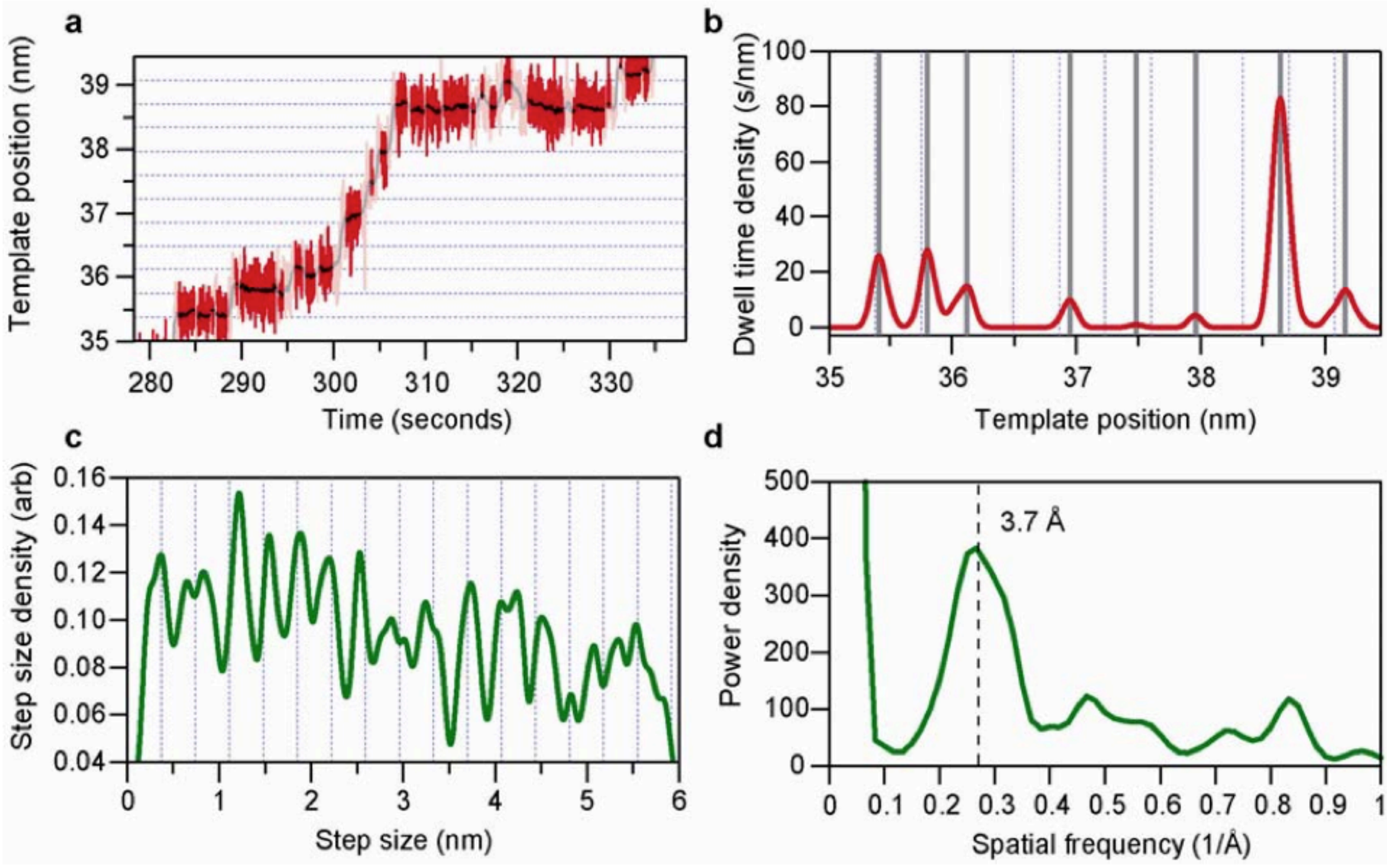


RNAP advances in single bp steps (3.4 Å)

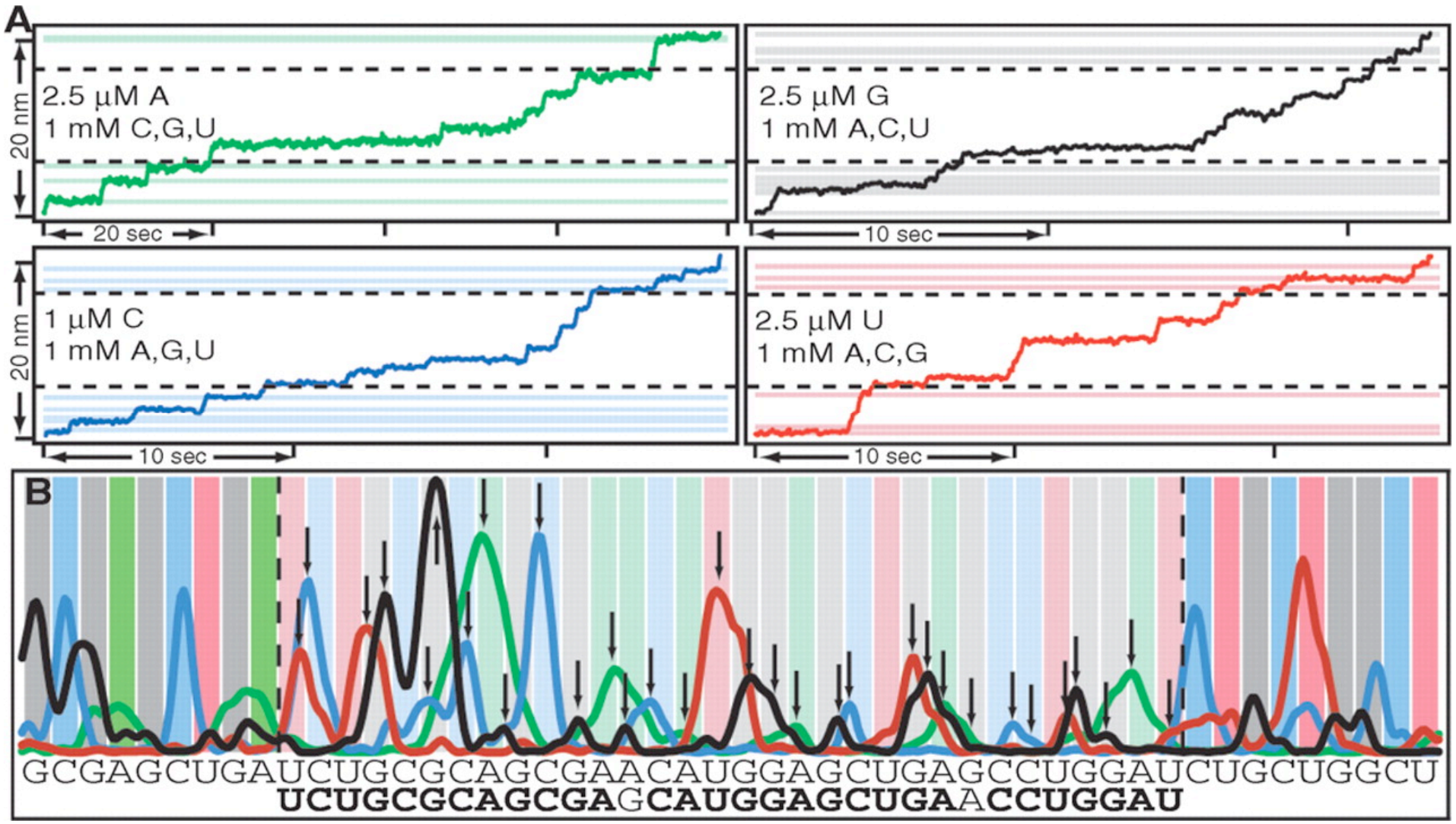


Abbondanzieri et al. *Nature* (2005)

Automated analysis finds same step size



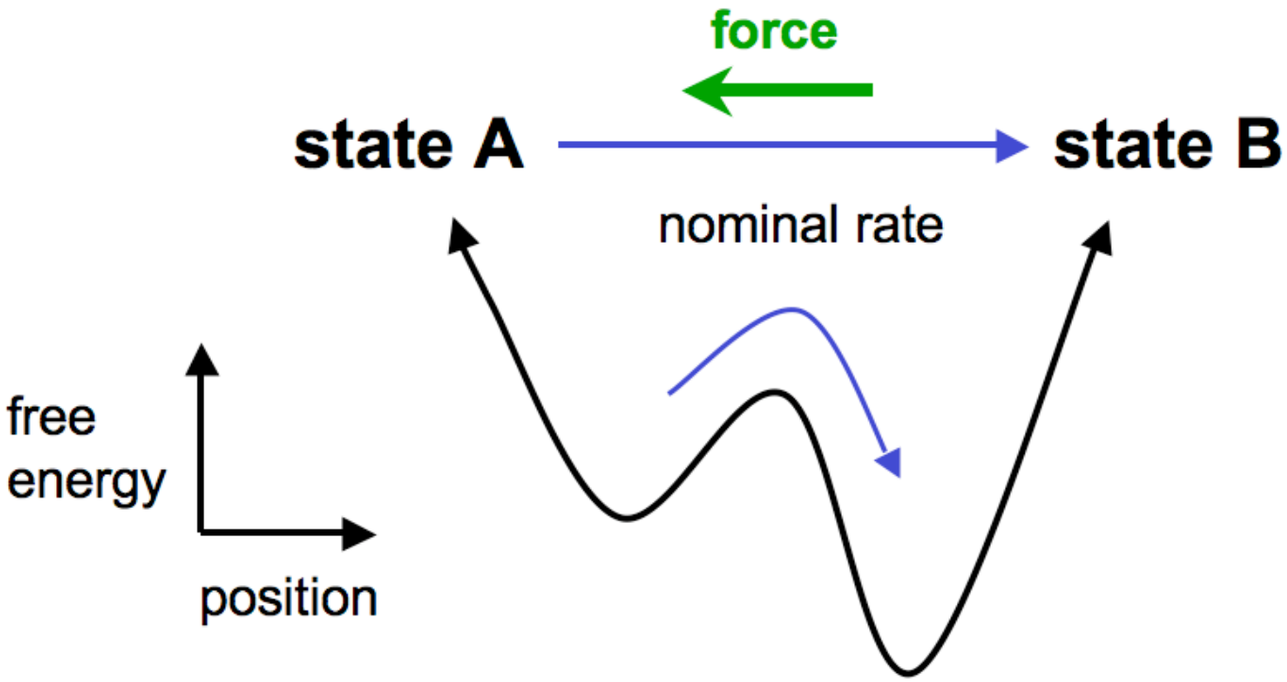
Using single proteins to sequence DNA



We make two mistakes out of 32 bases = 94% correct!

Greenleaf et al. *Science* (2006)

Force as a control variable

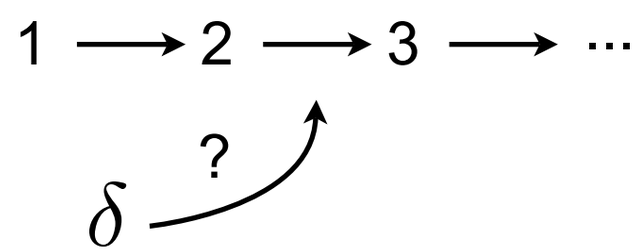


$$k(F) = k(F=0)e^{F\delta/k_BT}$$

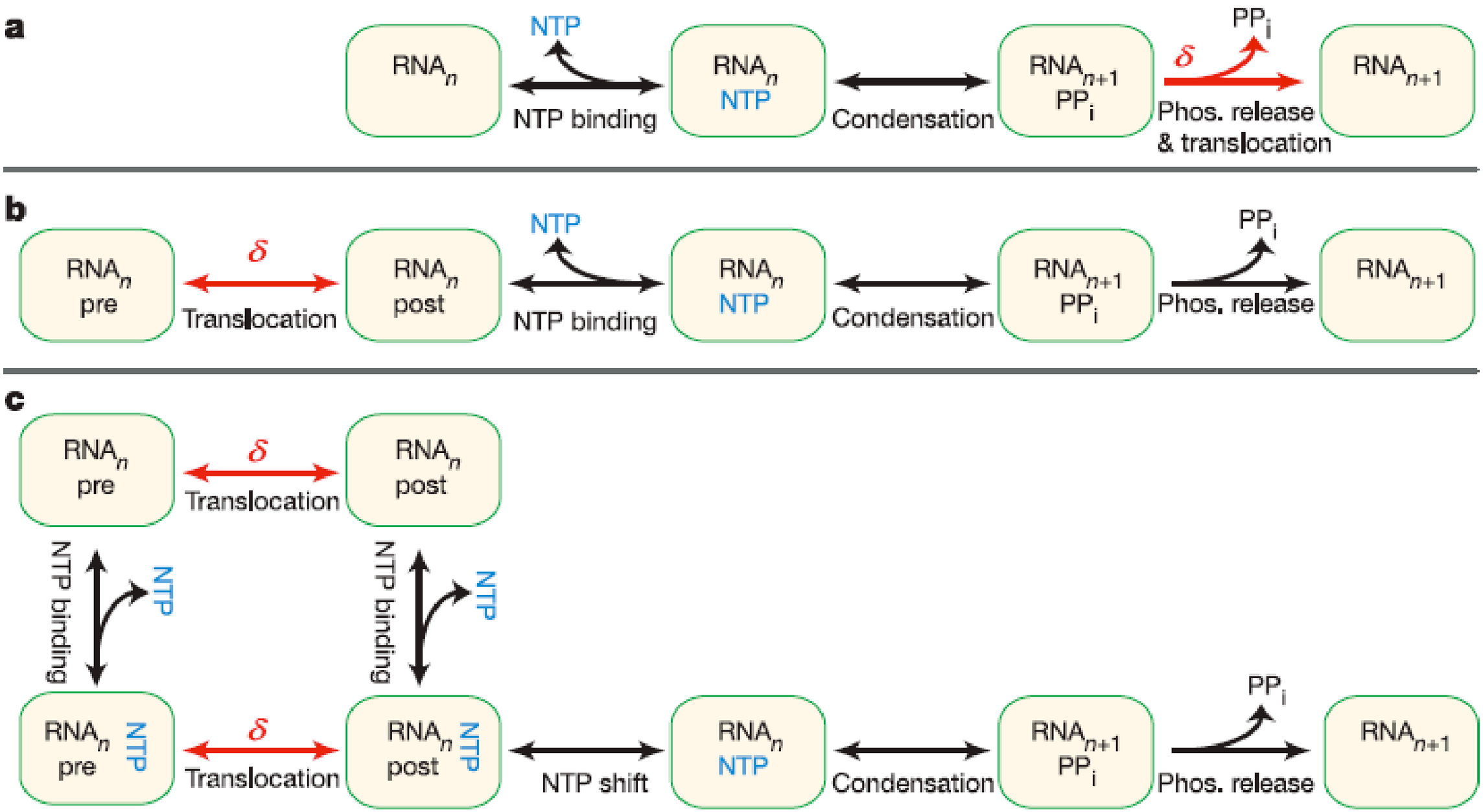
Use force as a control variable like temperature or substrate concentration etc...

Force dependent reactions involve physical motion and are slowed due to the energy required to move against the force.

$$v(F) = \frac{v_{max}}{1 + exp\left[-\frac{(F-F_{1/2})\delta}{k_BT}\right]}$$

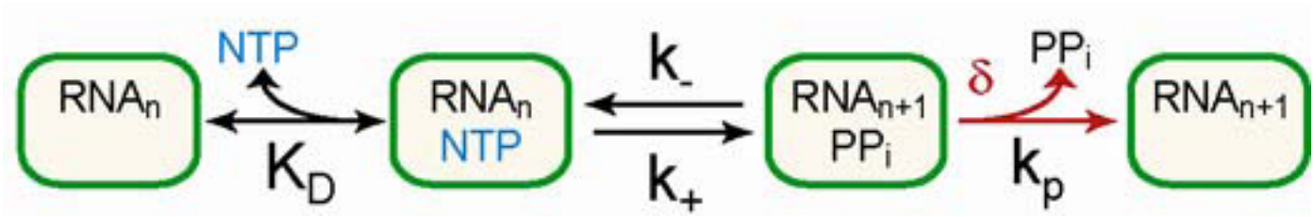


Three competing models of elongation



a	b	c
Yin and Steitz *Cell* (2004)	Guajardo and Sousa *J Mol Biol* (1997)	Zhang and Burton *J Mol Biol* (2004) Gong et al. *Mol Cell* (2005)

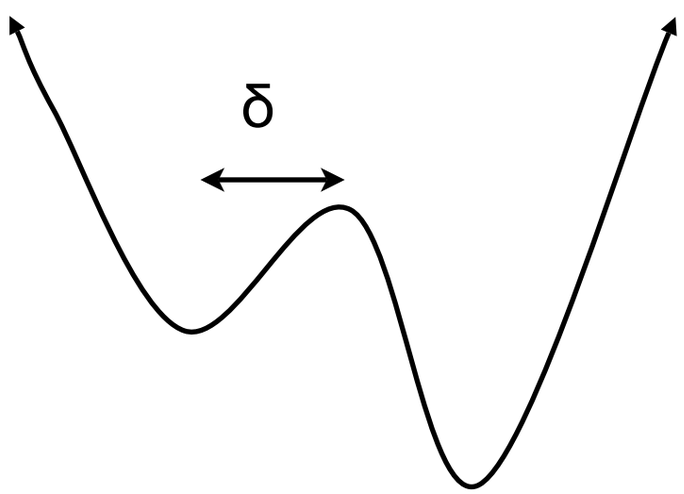
Phosphate-release power stroke



$$k_p(F) = k_p e^{F\delta/k_B T}$$

$$v(F,[NTP]) = \frac{k_p k_+}{k_+ e^{-F\delta/k_B T} + (k_- e^{-F\delta/k_B T} + k_p)(1 + K_D/[NTP])}$$

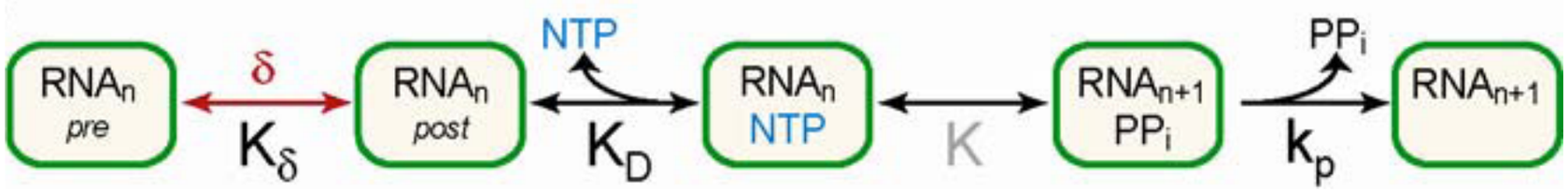
Here, δ is the distance to the transition state:



$$v_{max} = \frac{k_+[NTP]}{K_D + [NTP]}$$

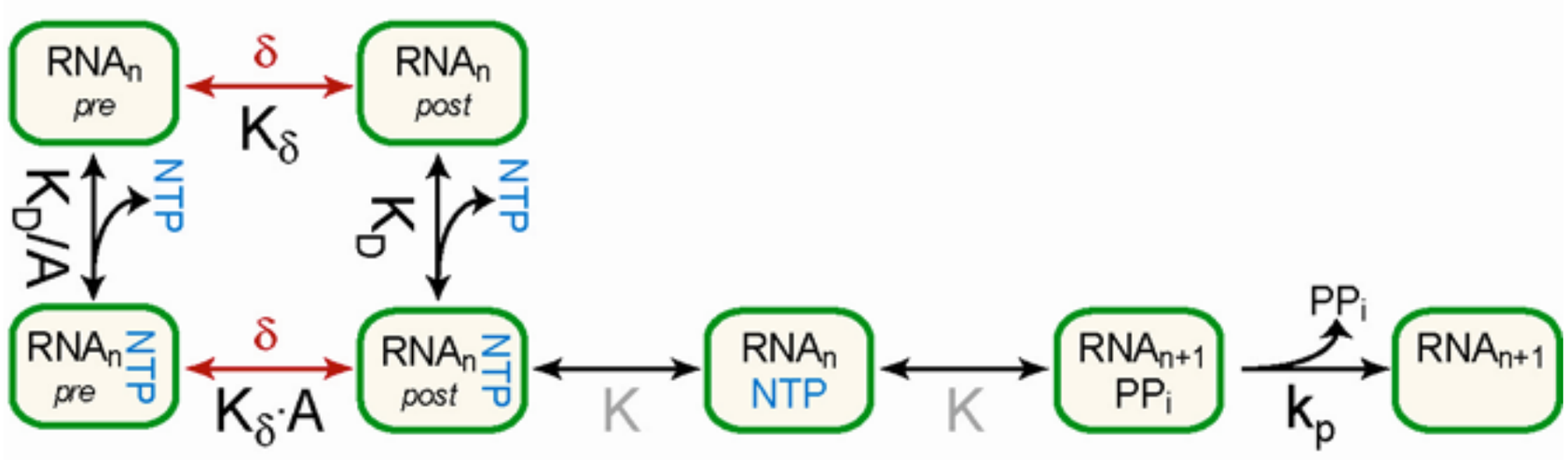
$$F_{1/2} = \frac{k_B T}{\delta} ln\left(\frac{v_{max} + k_-}{k_p}\right)$$

Two brownian ratchet models



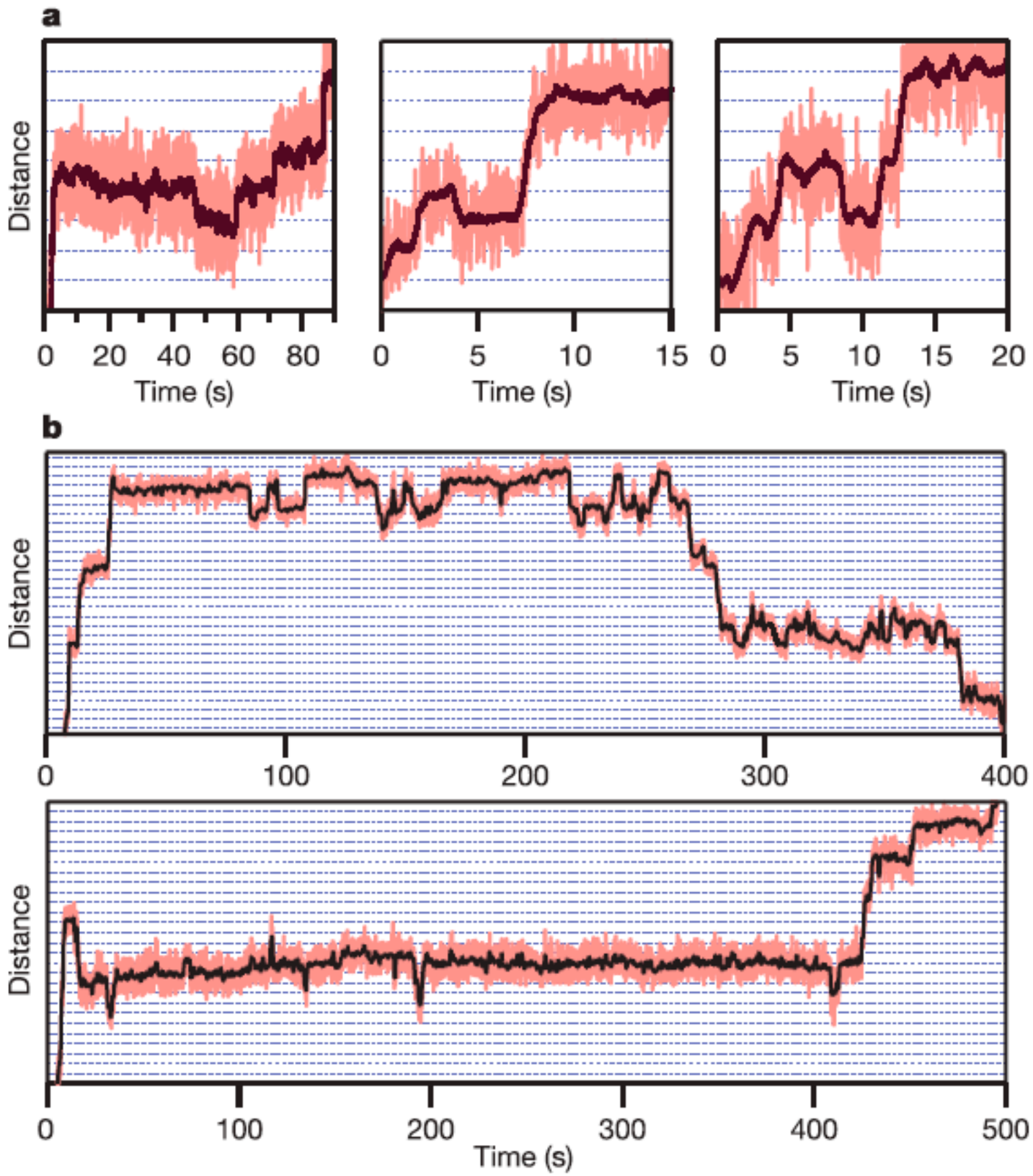
$$K_\delta(F) = K_\delta e^{-F\delta/k_BT}$$

$$v_{max} = \frac{k_p[NTP]}{K_D + [NTP]} \qquad F_{1/2} = \frac{k_BT}{\delta} ln\left(\frac{K_DK_\delta}{K_D + [NTP]}\right)$$



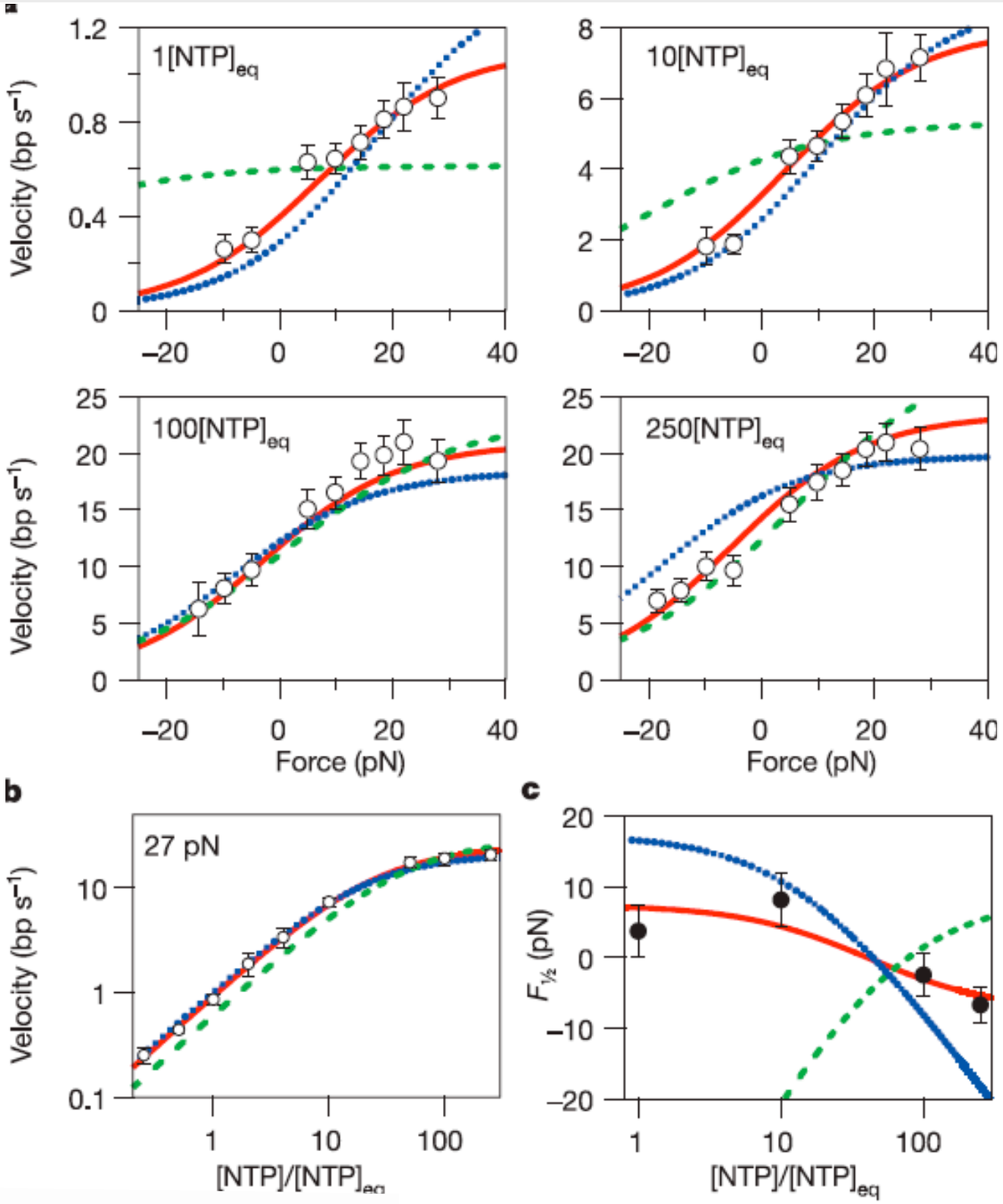
$$F_{1/2} = \frac{k_BT}{\delta} ln\left(\frac{K_DK_\delta + AK_\delta[NTP]}{K_D + [NTP]}\right)$$

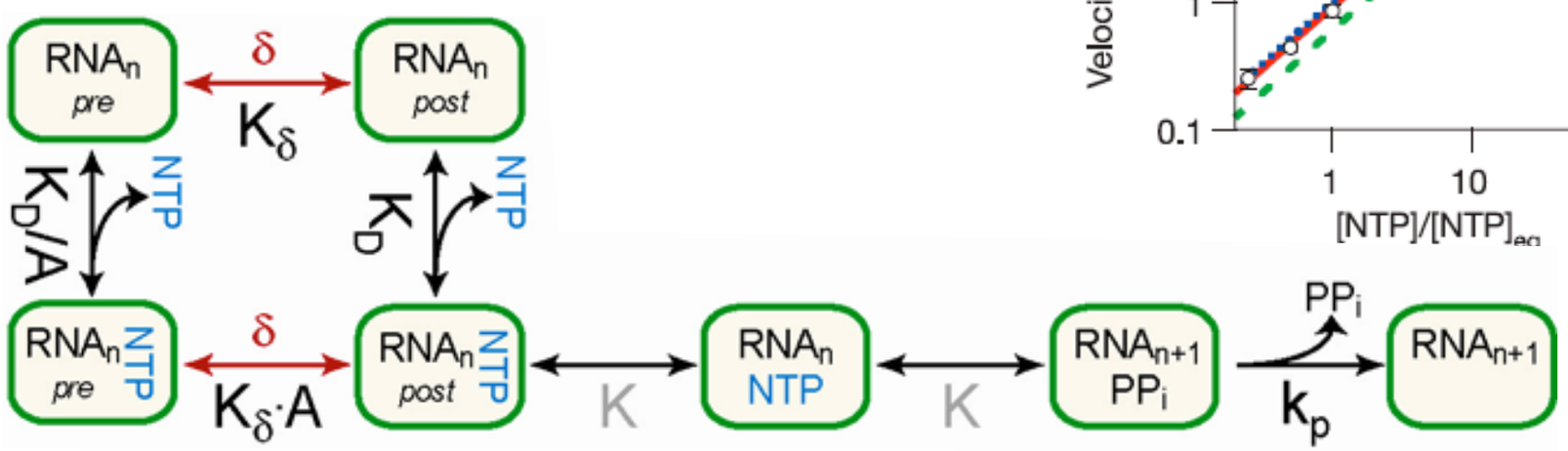
We remove backwards motion (a separate pathway)



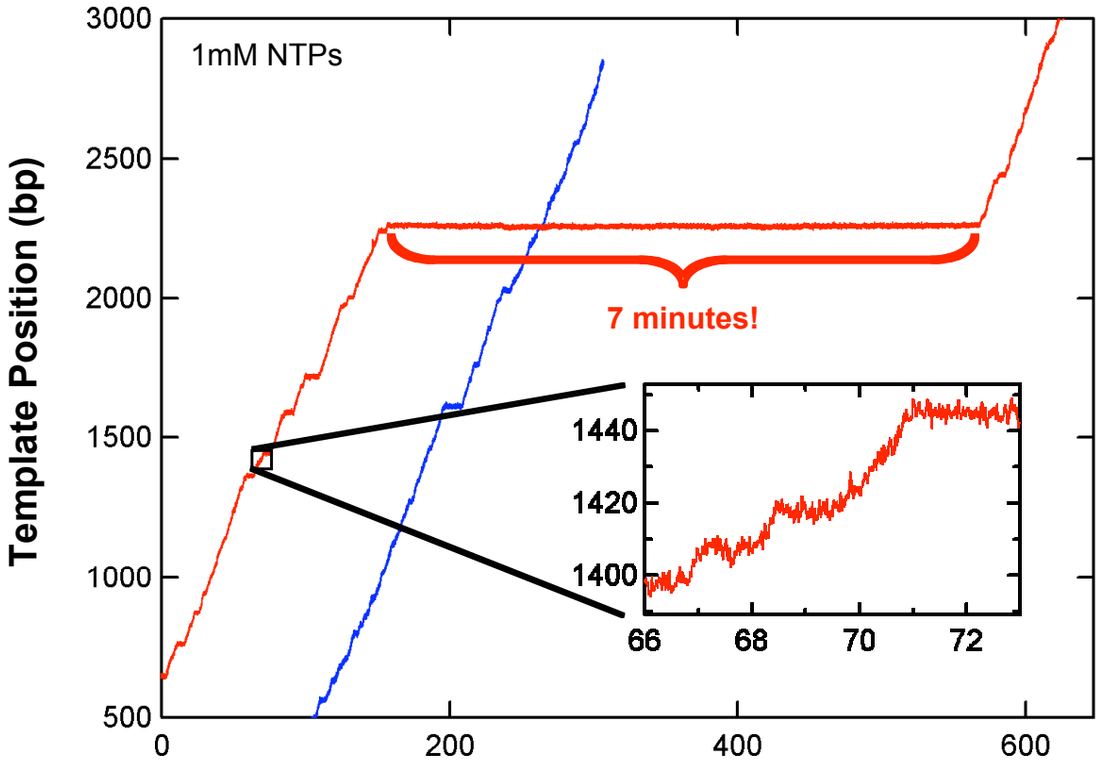
Force data rules out Power Stroke model

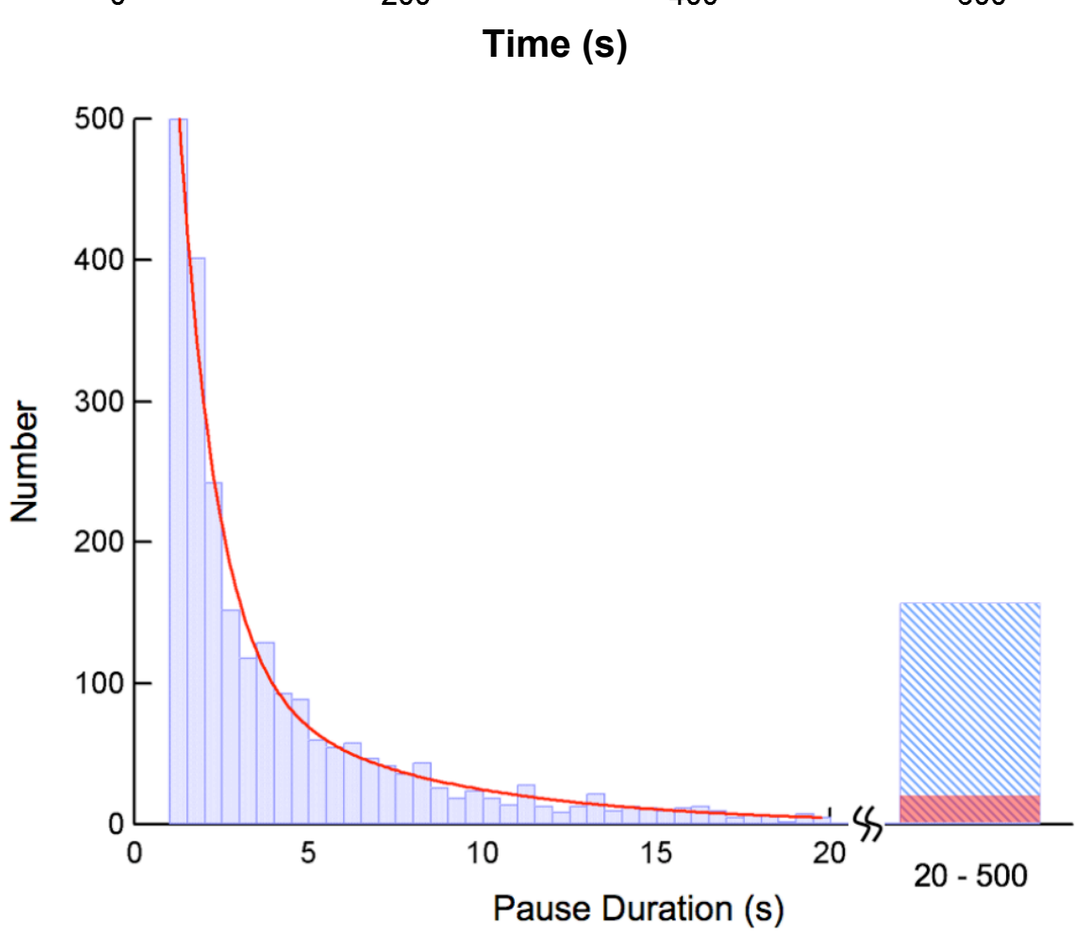
Power Stroke Model
Brownian Ratchet
Brownian Ratchet with secondary NTP binding site





Pauses occur on many time scales





Pausing occurs on many timescales and is the main method of regulation during elongation.

Pausing:

1.allows for the recruitment of factors (DNA repair etc.)

2.serves as a precursor for termination and arrest

3.used during proofreading

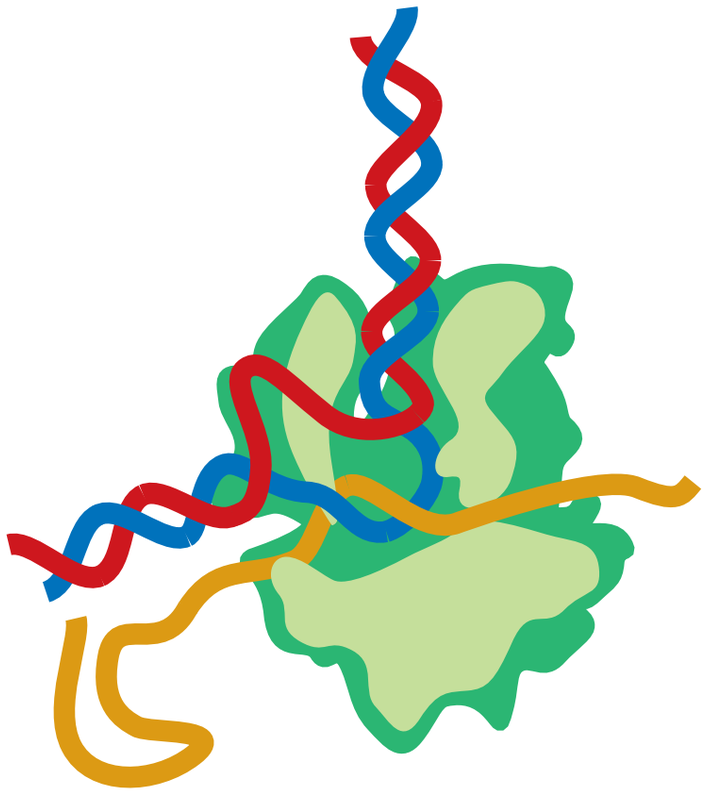
4.couples transcription to translation in prokaryotes

5.couples transcription to splicing and polyadenylation in eukaryotes

6.transcription factors can modulate pausing to control the overall rate of RNA synthesis

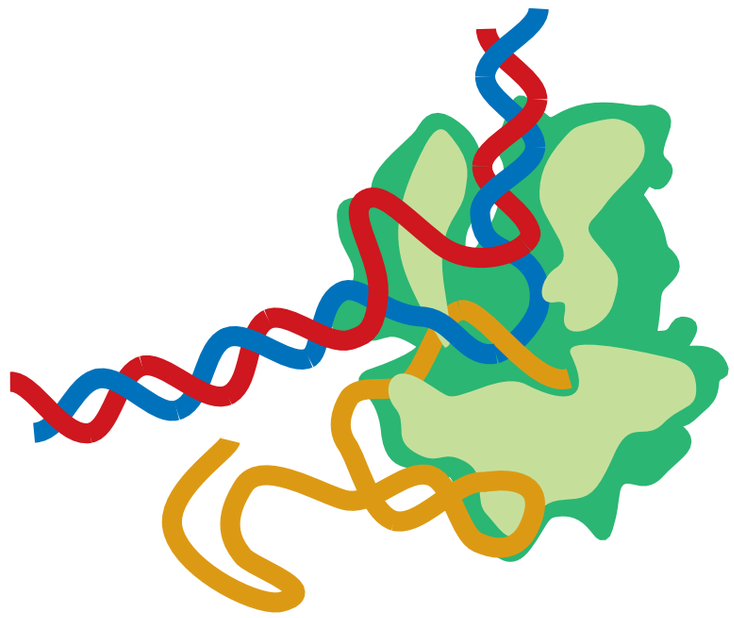
Two main mechanisms of pausing

Backtracking Pause



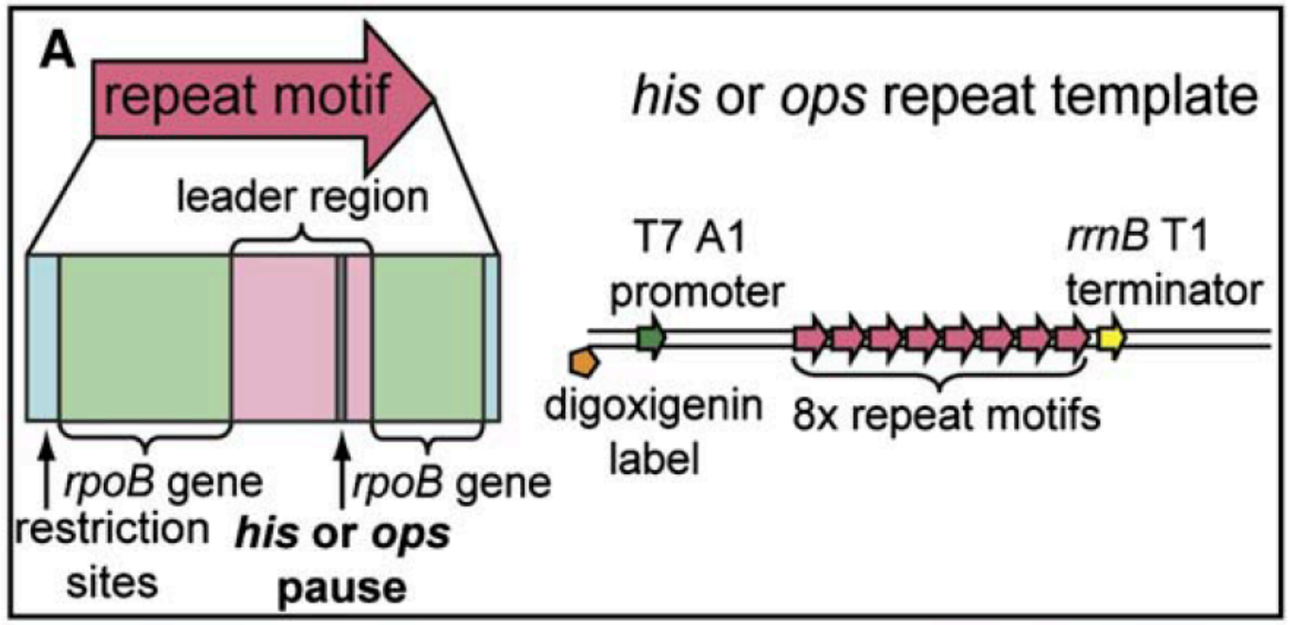
Formed by weak DNA:RNA hybrid (?)
Used to allow factor recruitment
Example: *ops* pause in *E coli* -- Backtracking leads to binding of RfaH factor that suppresses early termination

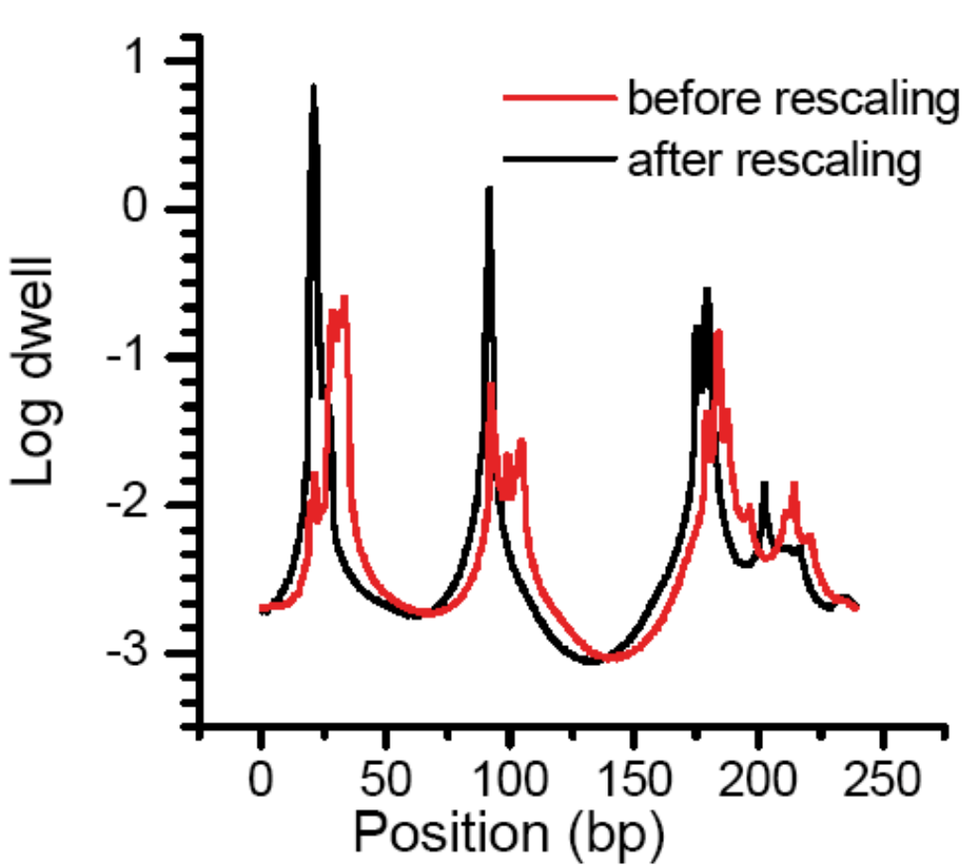
Hairpin Pause

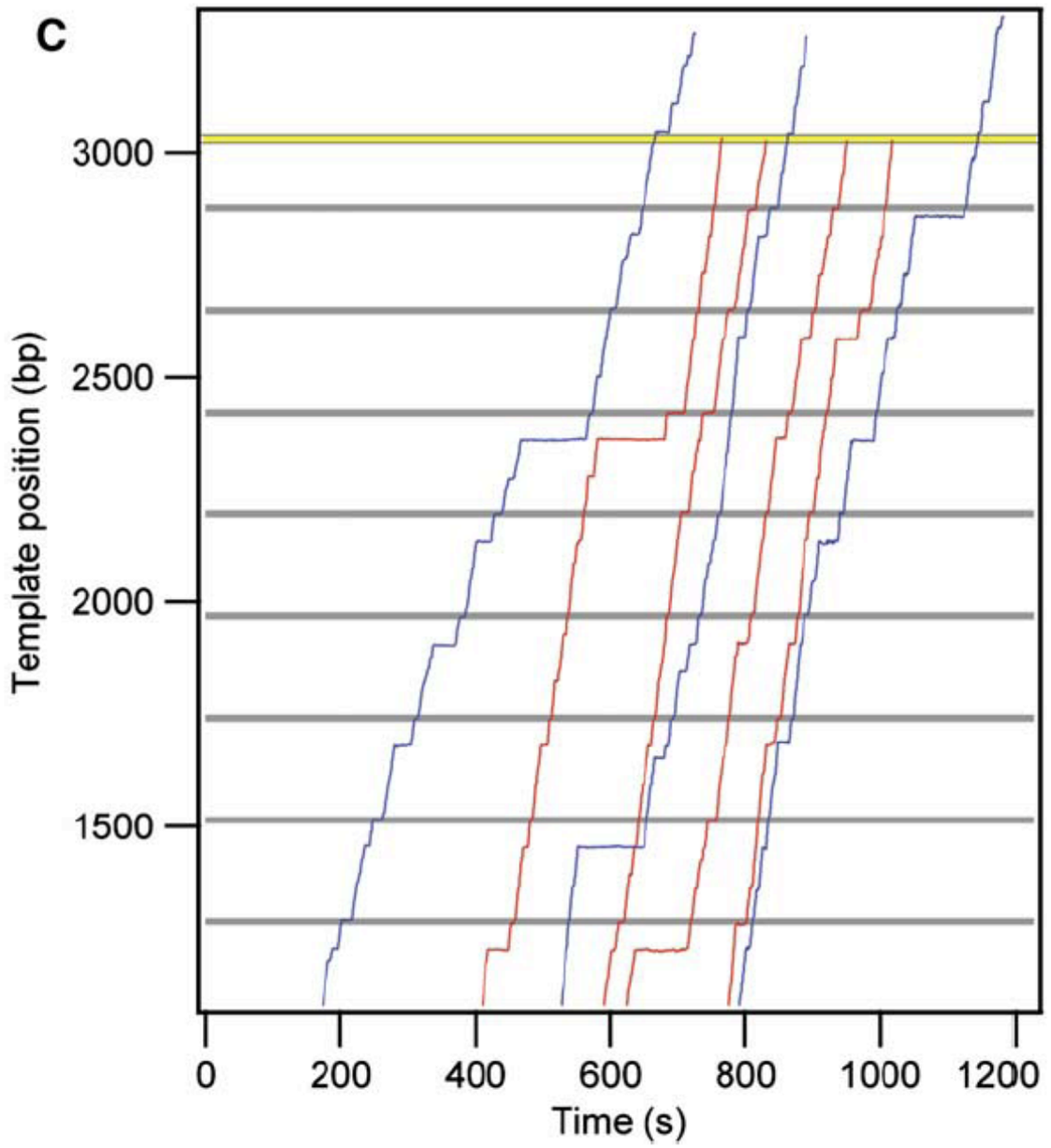


Secondary structure in RNA strains RNAP causing a pause (not clear?)
Found in leader region of operons in bacteria to synchronize RNAP with ribosomes during attenuation. Modification of secondary structure by factors can regulate this
Example: *his* pause near beginning of histidine operon in *E coli*

A repeating pause sequence

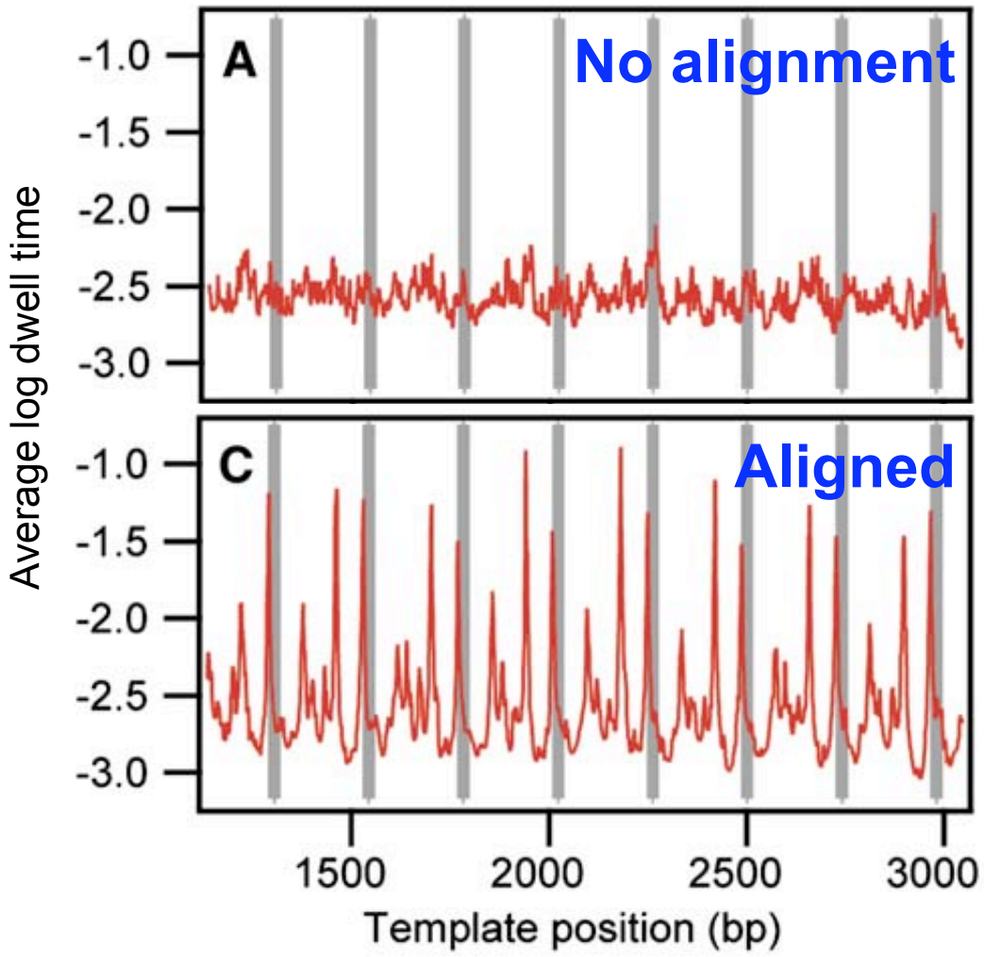


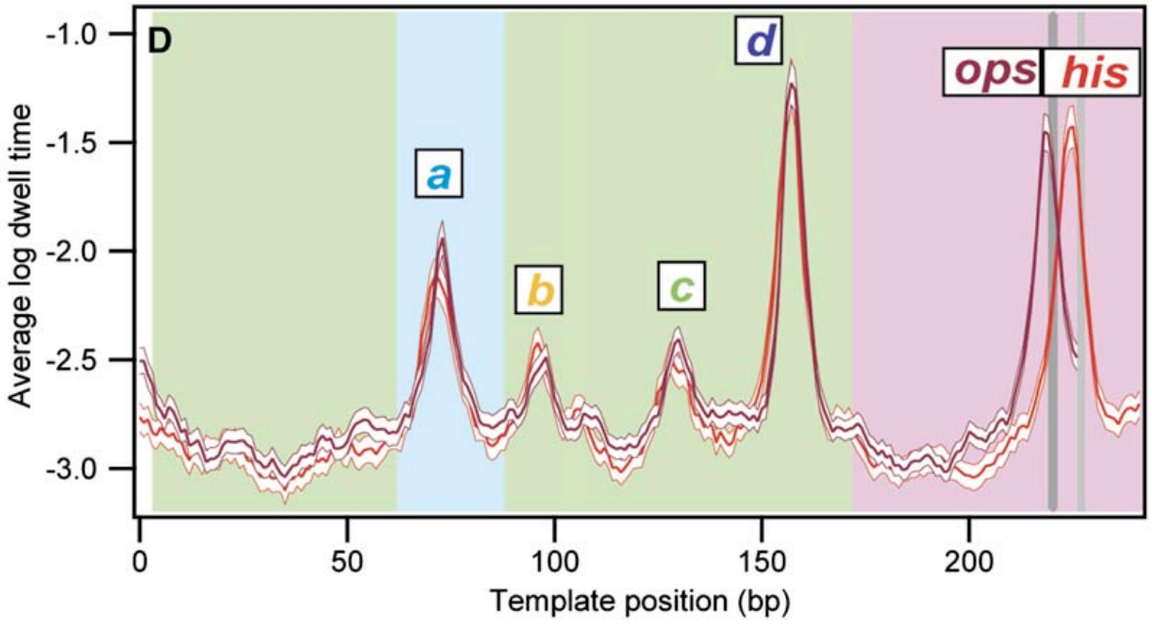


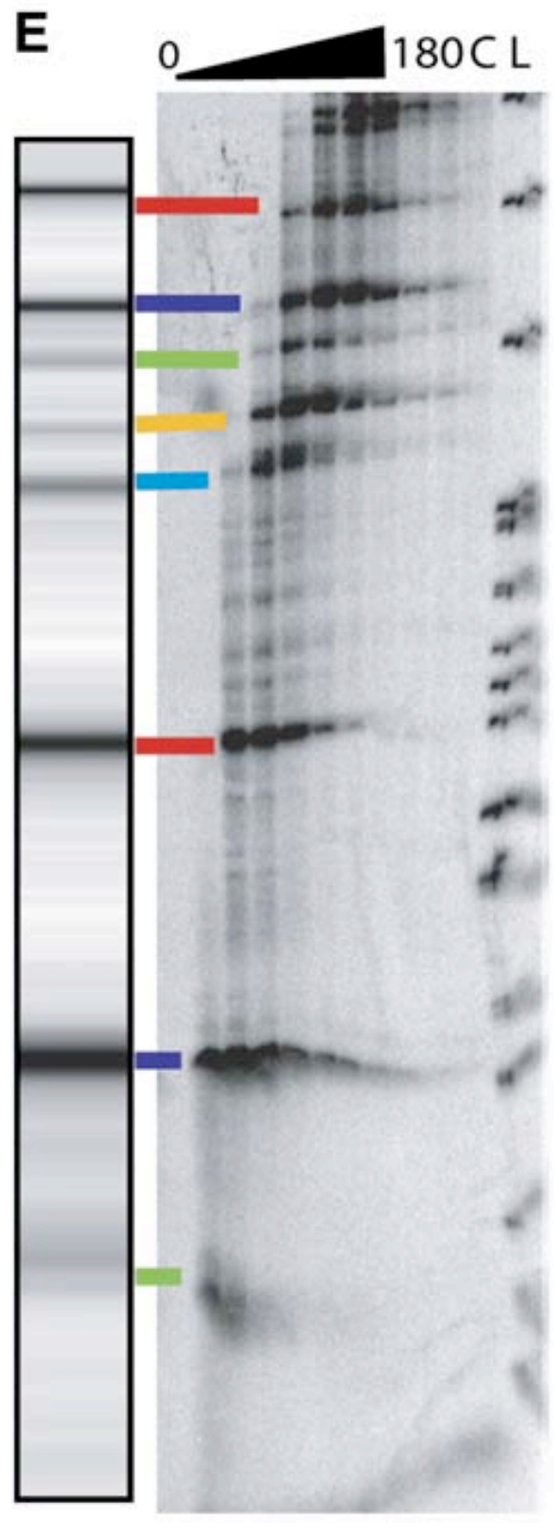


Repeat pause can be used as fiducial references to study sequence dependance.

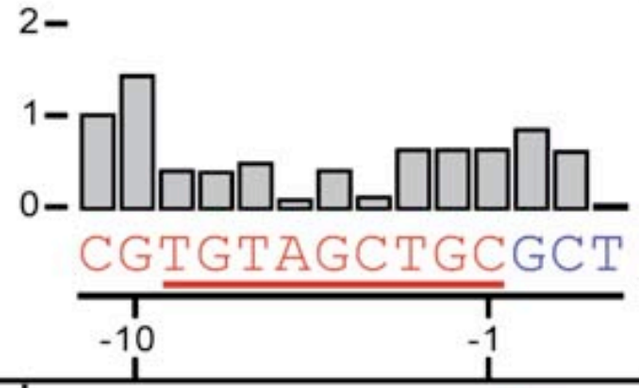
Aligned single molecule data shows many sequence dependent pauses and agrees with bulk data





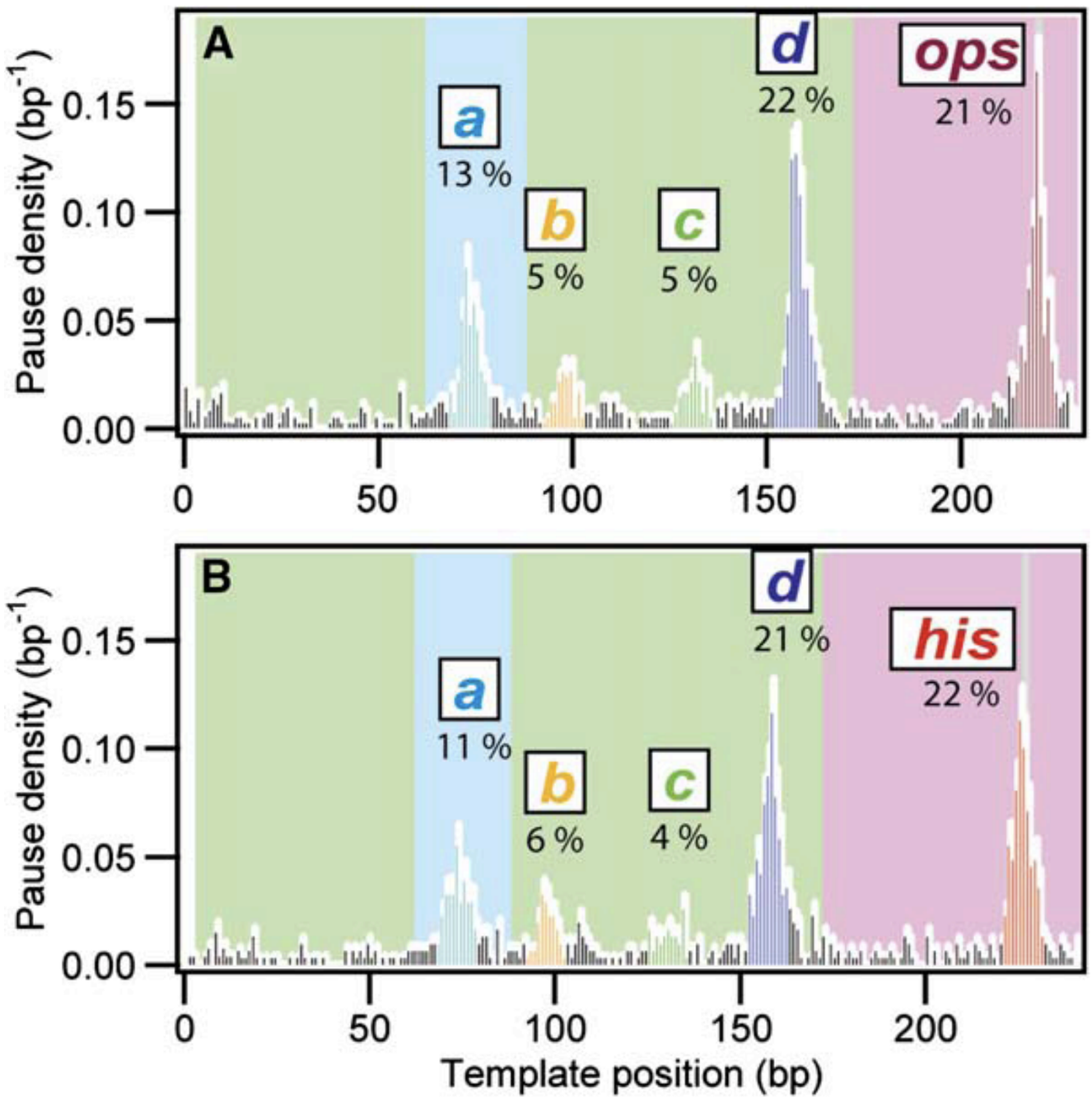


Sequence similarities between the pauses

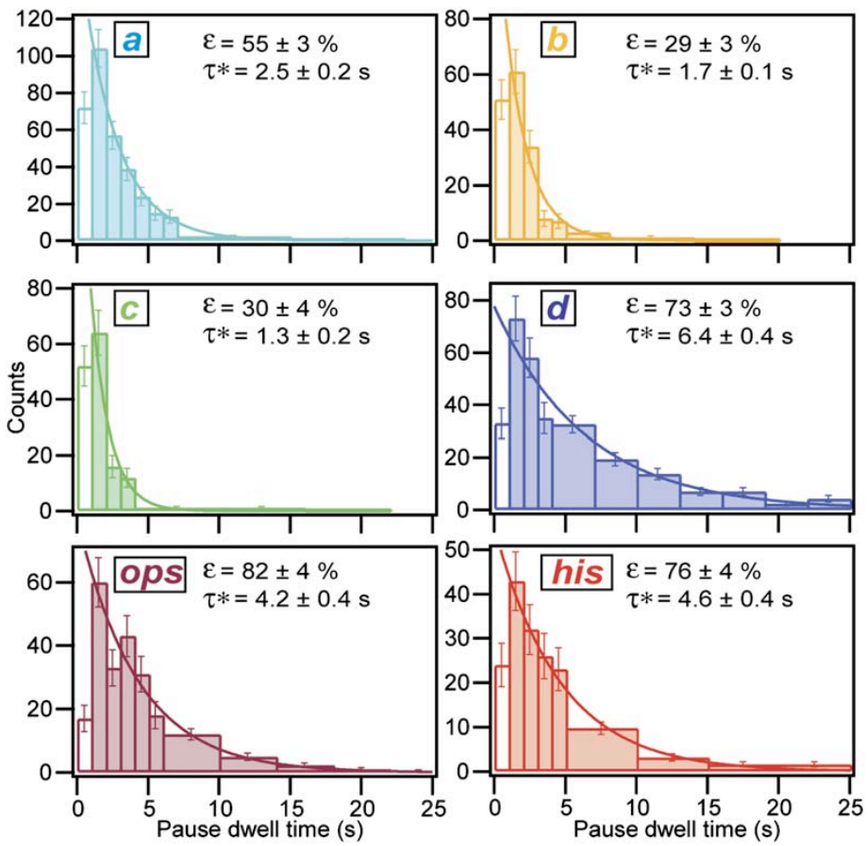
		Translocation state (bp)
Information (bits) / Consensus sequence		
a	CGGGTAGATCCGCC	0.70 ± 0.50
b	GGTGAAACCGCAWC	-0.25 ± 0.50
c	GGTAAAGTGTACGT	-0.20 ± 0.50
d	CGTATCACTGCGCG	0.40 ± 0.50
ops1	CGGTAGTCTGTGCG	0.75 ± 0.25
ops2	GTAGTCTGTGCGCT	-1.25 ± 0.25
his	CGATGTGTGCTGGA	0.00 ± 0.25

No large displacement forward or backward
No pre or post translocation

Pause density varies greatly over the template



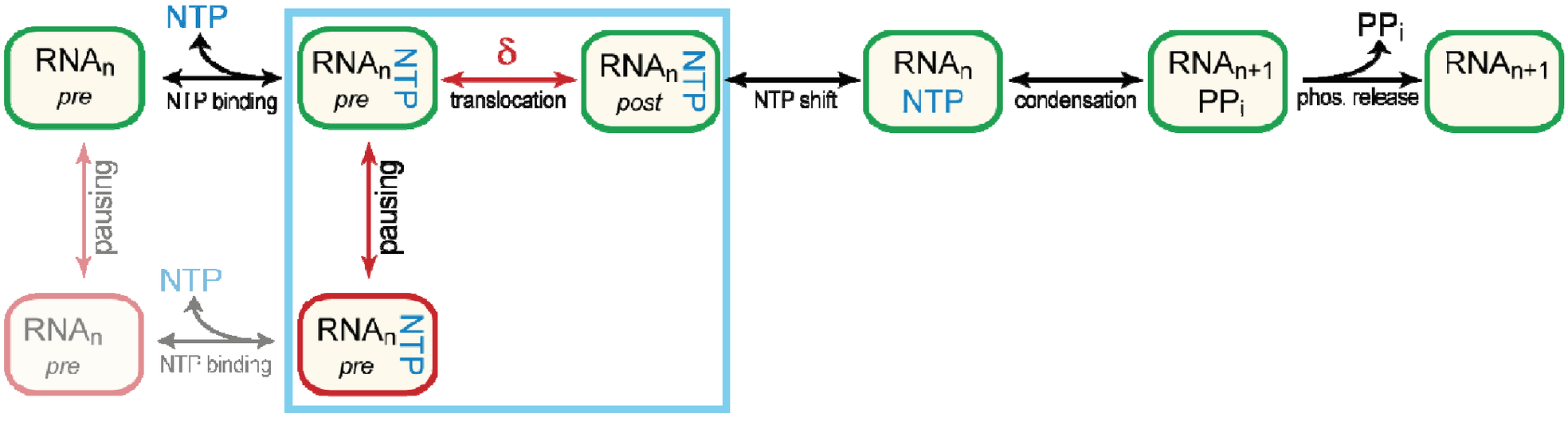
All 6 pauses exhibit the same corrected lifetime



D	Pausing kinetics		
	ε	τ* (s)	τ (s)
a	55 ± 3 %	2.5 ± 0.2	1.1 ± 0.1
b	29 ± 3 %	1.7 ± 0.1	1.2 ± 0.1
c	30 ± 4 %	1.3 ± 0.2	0.9 ± 0.1
d	74 ± 3 %	6.4 ± 0.4	1.8 ± 0.2
ops	82 ± 4 %	4.2 ± 0.4	0.8 ± 0.2
his	76 ± 4 %	4.6 ± 0.4	1.1 ± 0.2

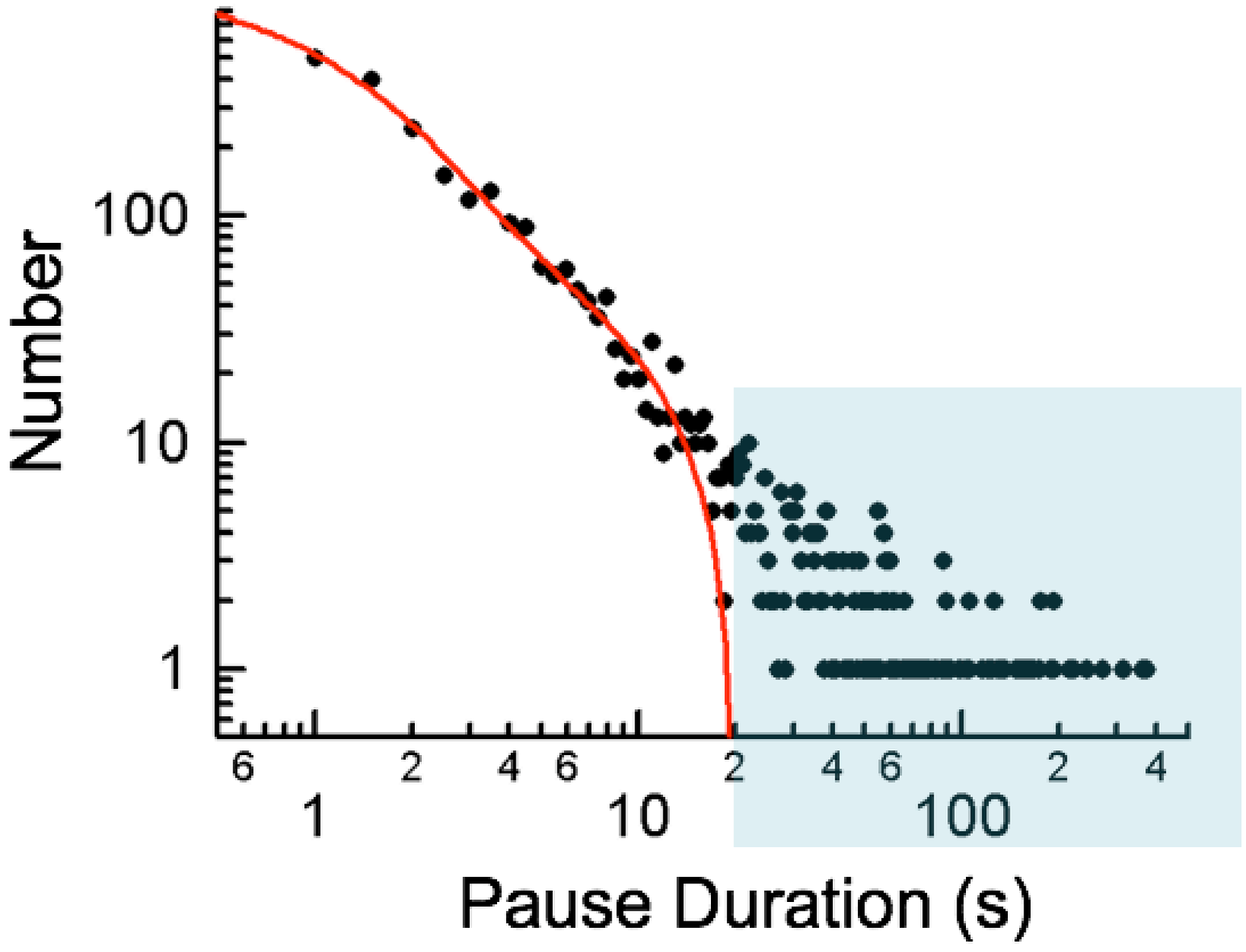
$$\tau^* = \frac{\tau}{1-\epsilon}$$

Efficiency not 100% -> Pausing must be off pathway!
When corrected for the efficiency, all six pauses have the same lifetime
They may all be the same off-pathway intermediate that leads to the regulatory pauses?



Backtracking pauses

The *ops* pause data I just showed had no backtracking. Why?



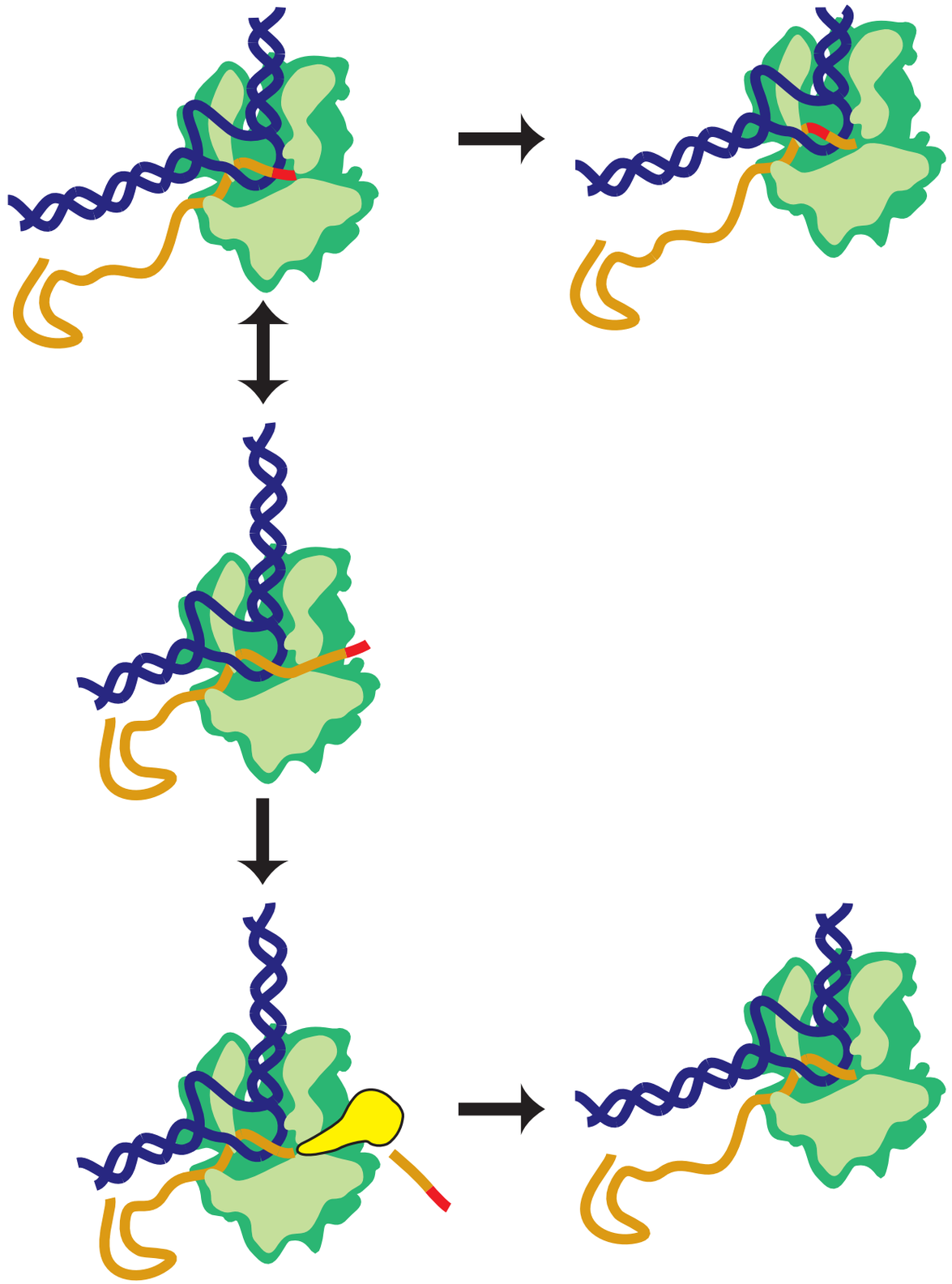
Does RNAP have a proofreading mechanism?

- Error rates *in vitro* : 10^{-3}- 10^{-4}

- Error rates *in vivo* : 10^{-5}-10^{-6}

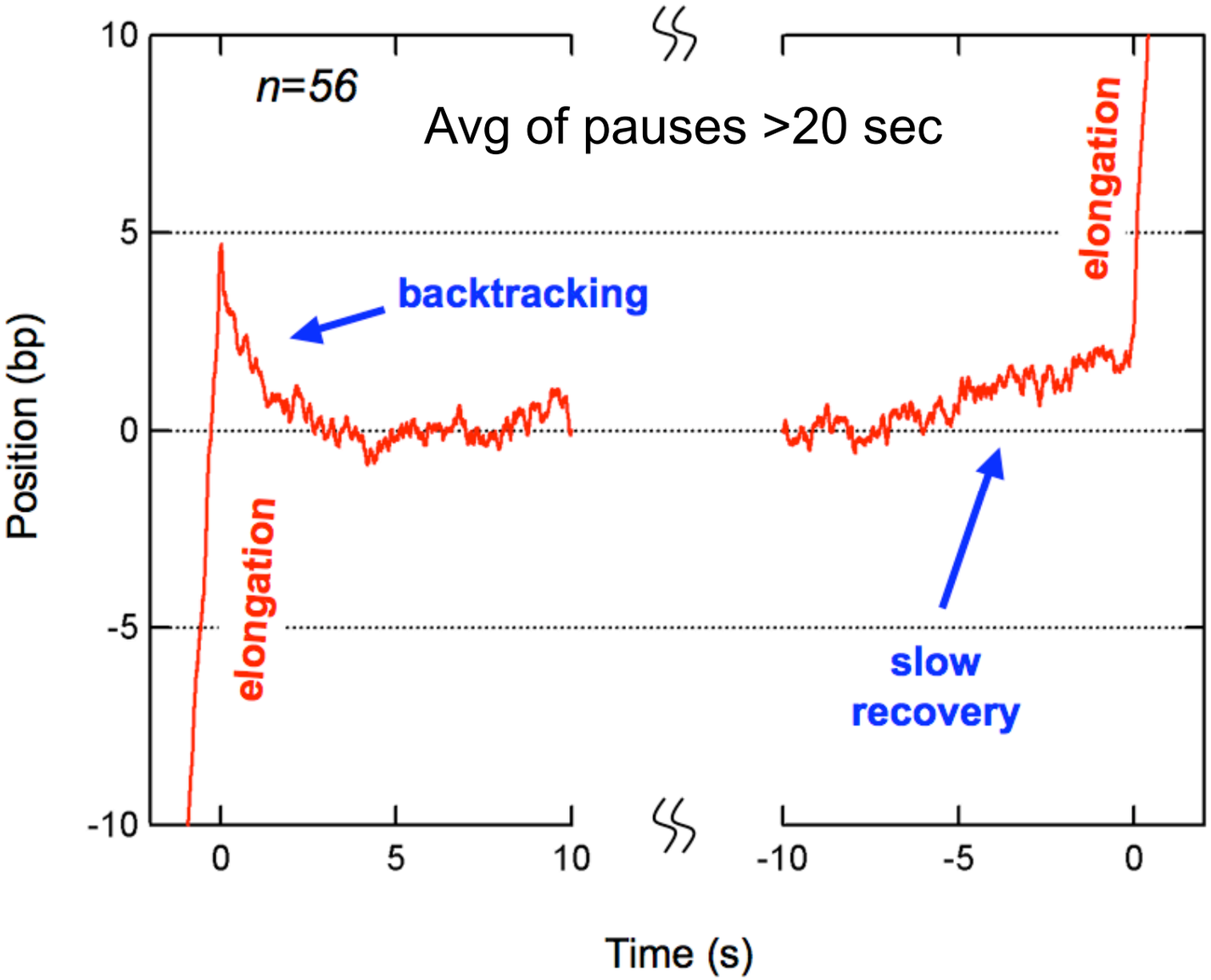
Erie et al., *Science* (1993)

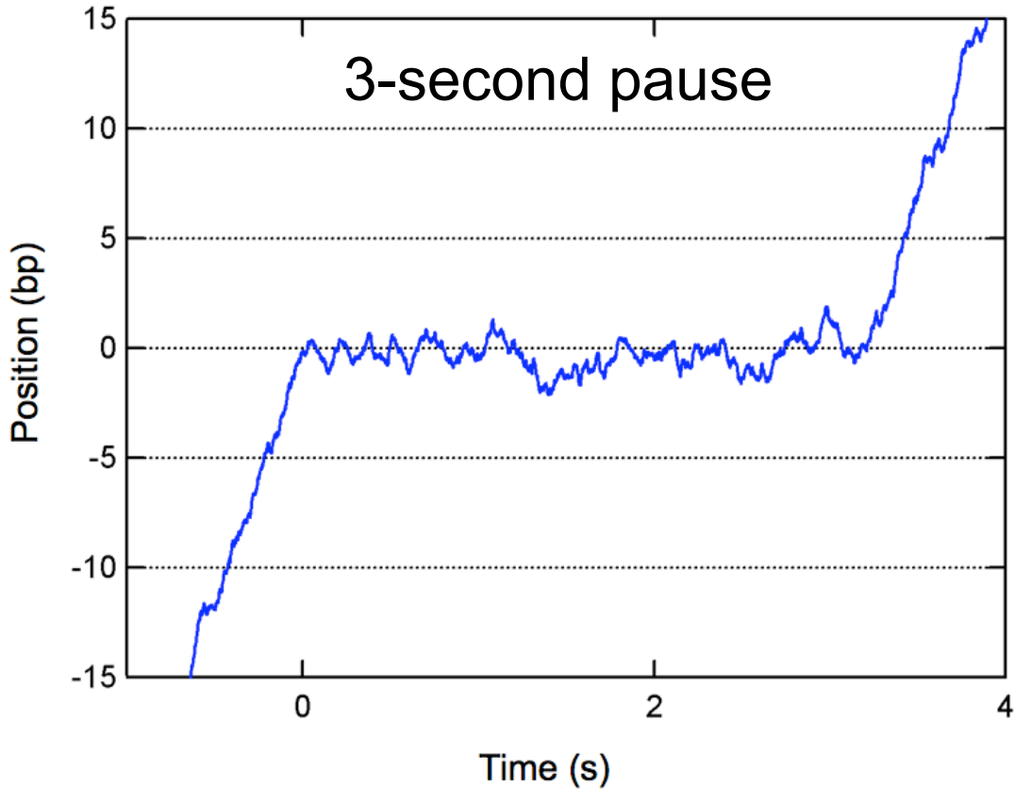
Consensus model of RNAP proofreading



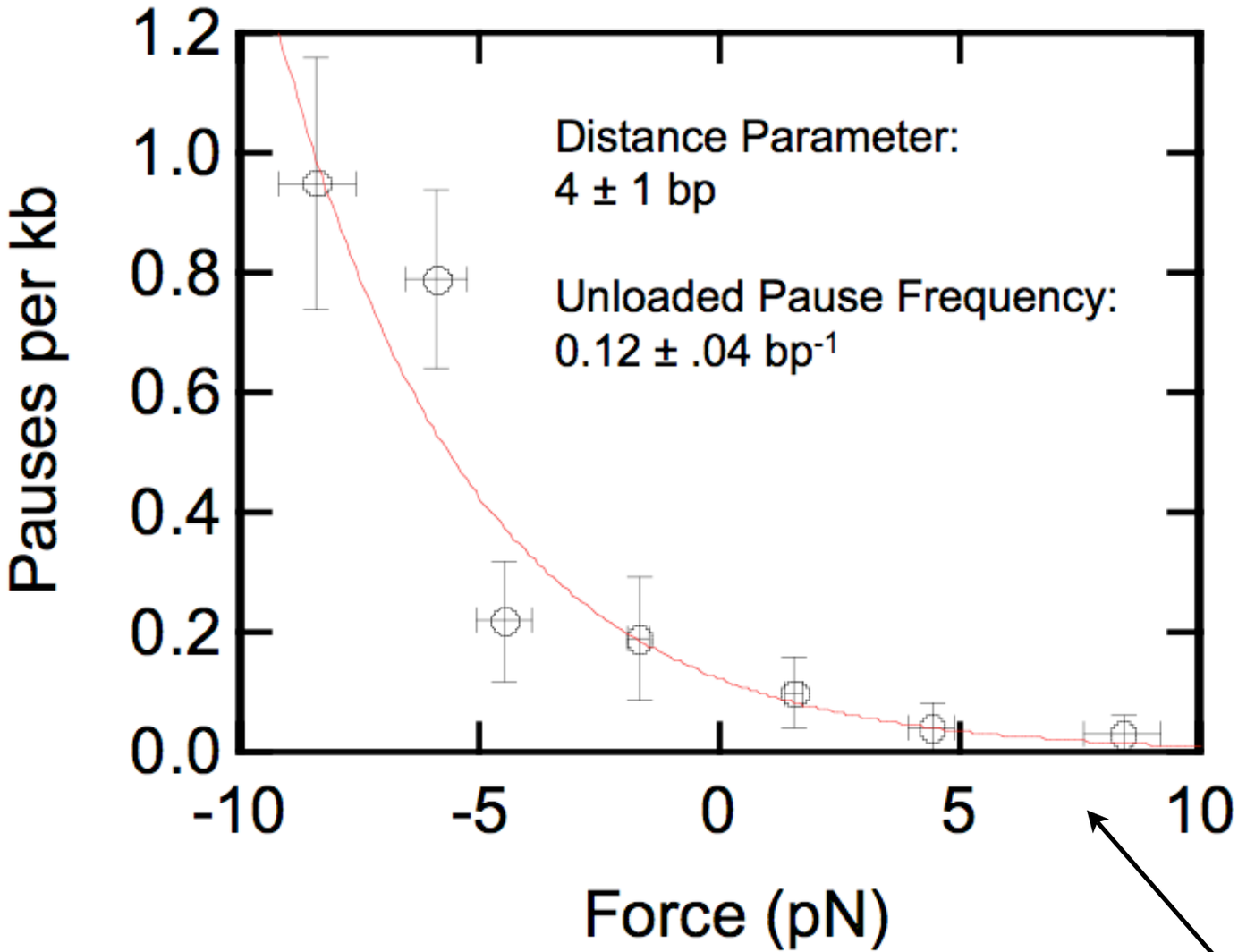
Erie et al., *Science* (1993)
Jeon and Agarwal *PNAS* (1996)
Thomas et al. *Cell* (1998)

Average behavior shows backtracking and recovery





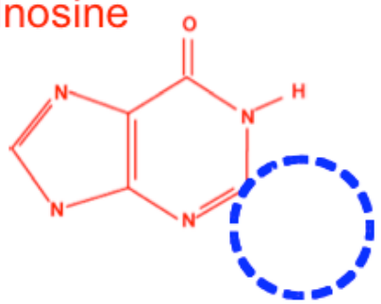
Backtracking is force dependent



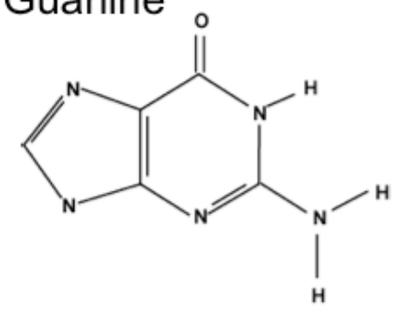
At >7pN, backtracking would have been hard to see in Herbert et al.'s experiments.

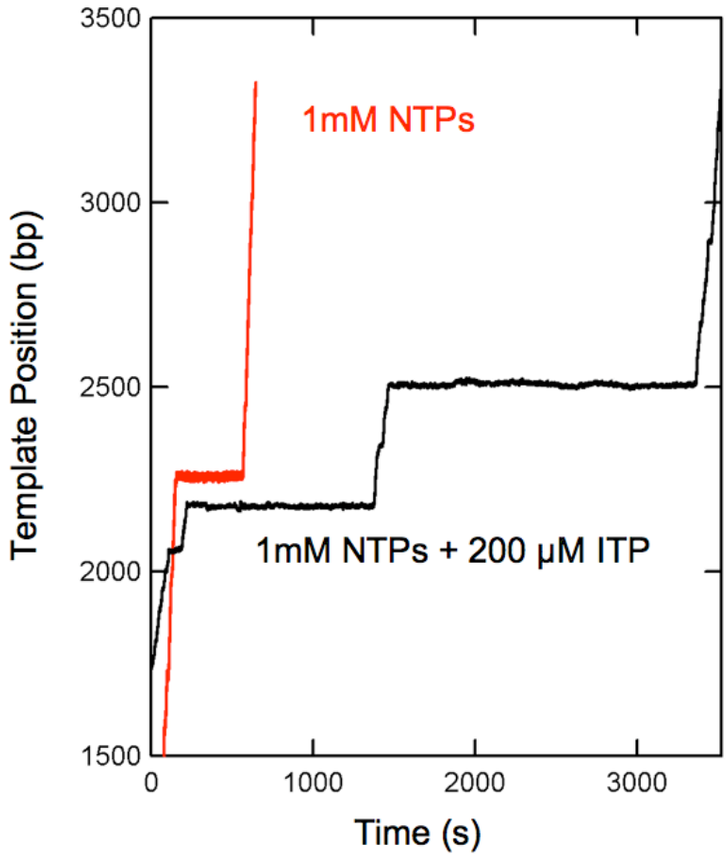
ITP increases pause number and duration

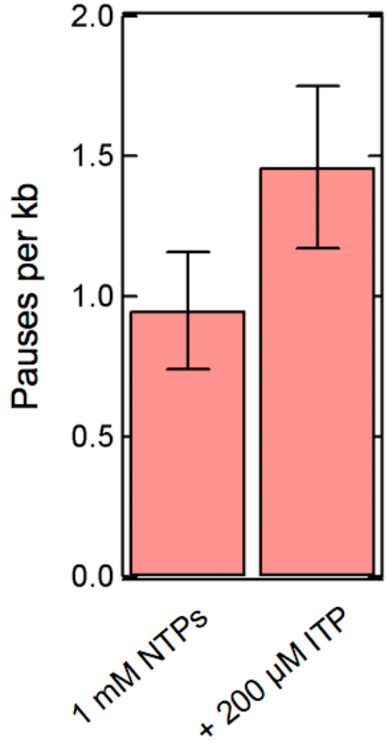
Inosine

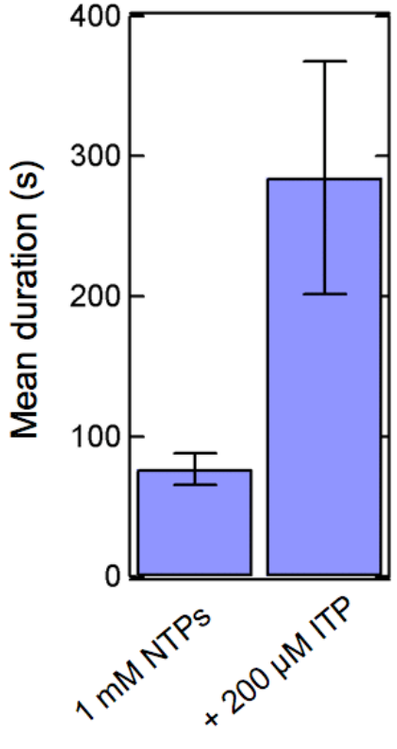


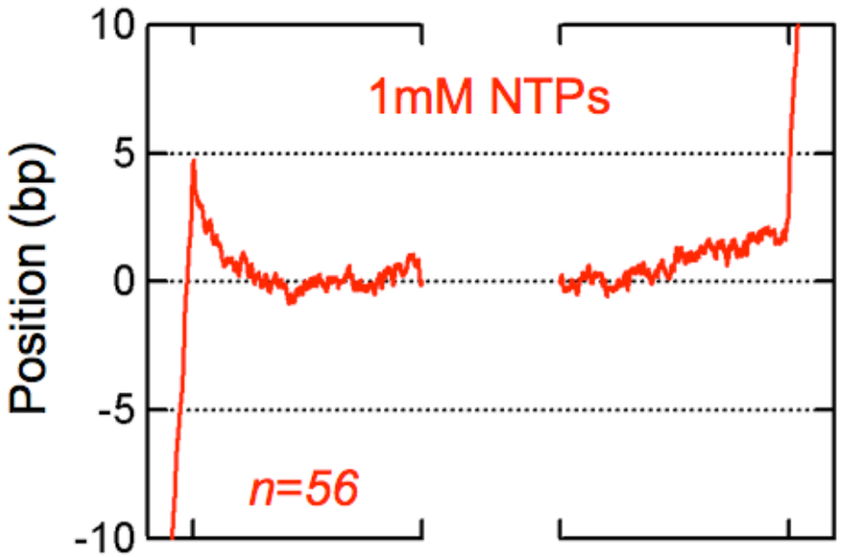
Guanine

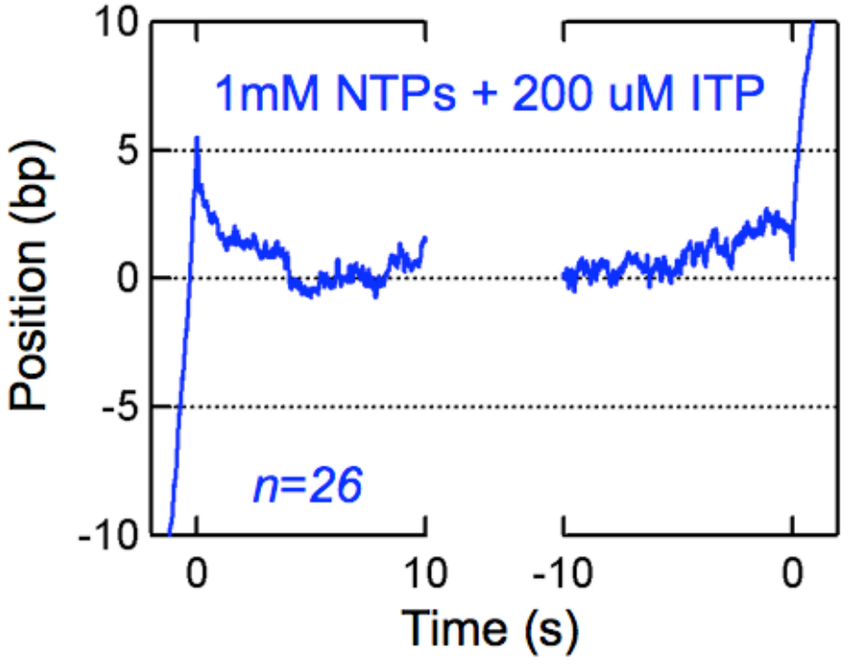




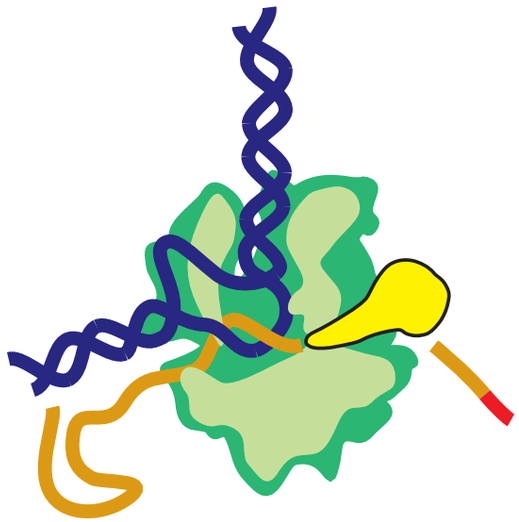








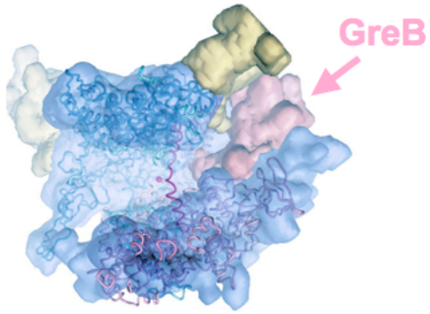
Cleavage reduces the duration of long pauses

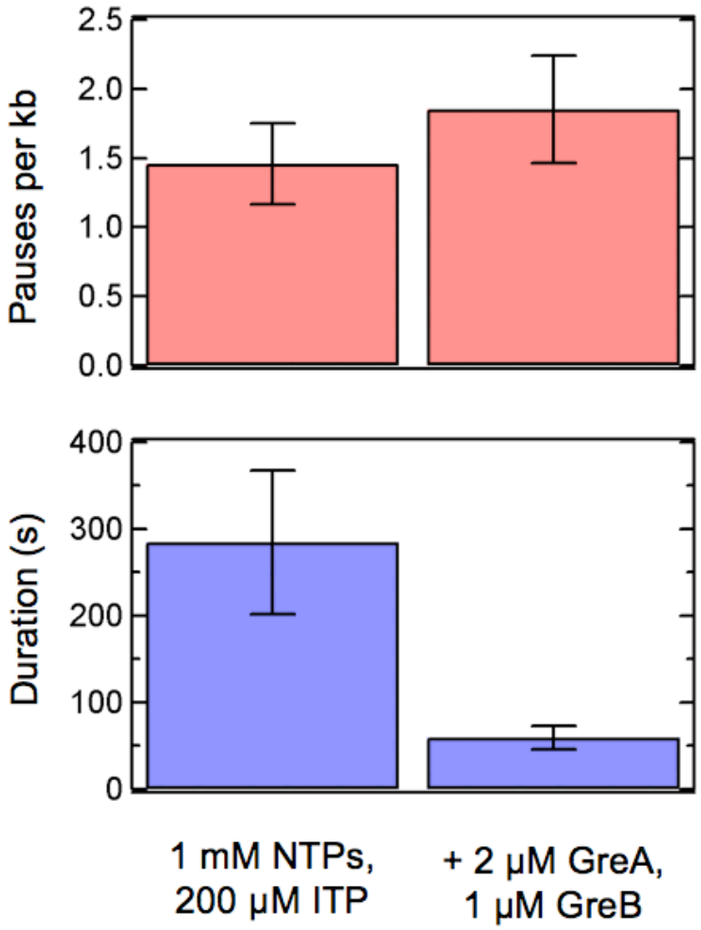


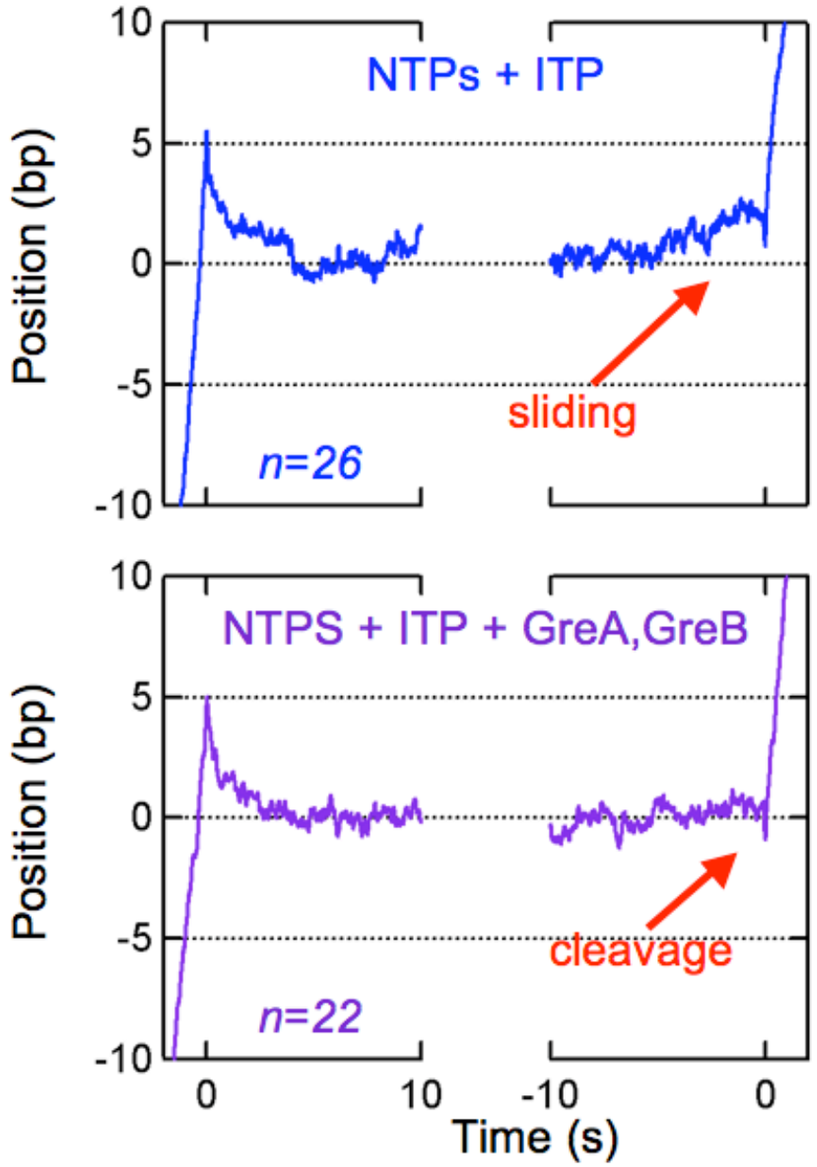
RNAP has an intrinsic cleavage ability at the polymerization active site that is stimulated by Gre proteins or high pH

Functional analogs of GreA and GreB have been found in over 60 organisms, including TFIIS in eukaryotes

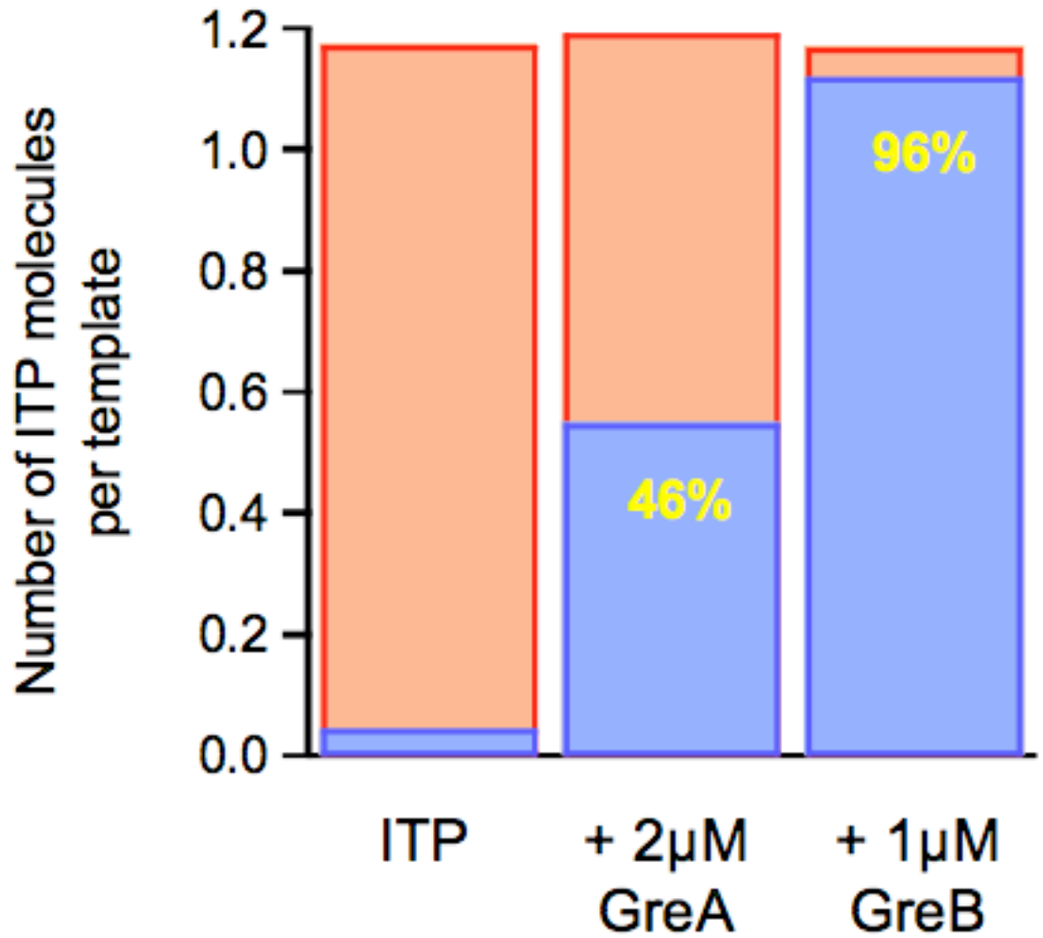
Opalka et al., *Cell* (2003)



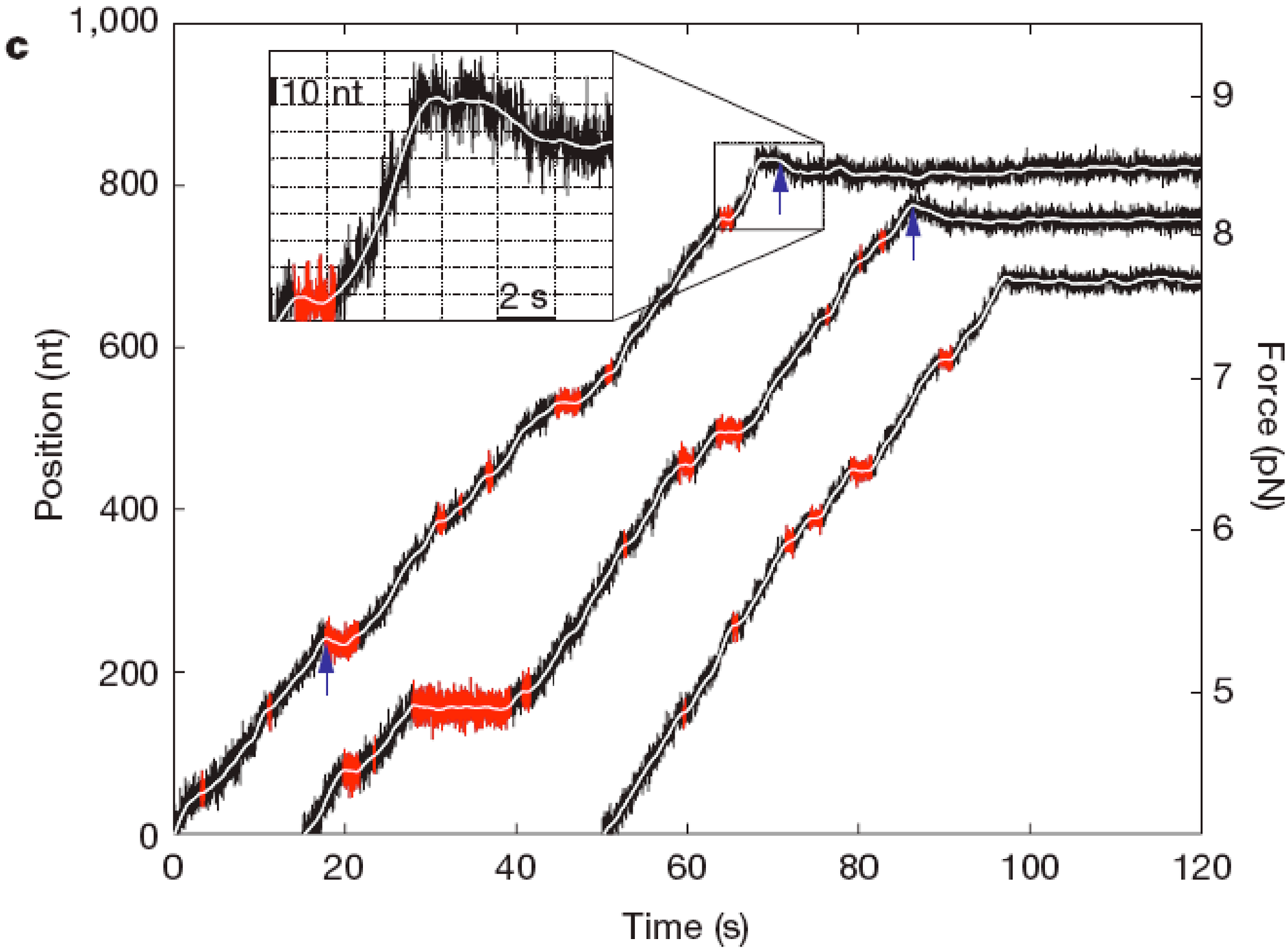




Cleavage removes inosine from the transcript

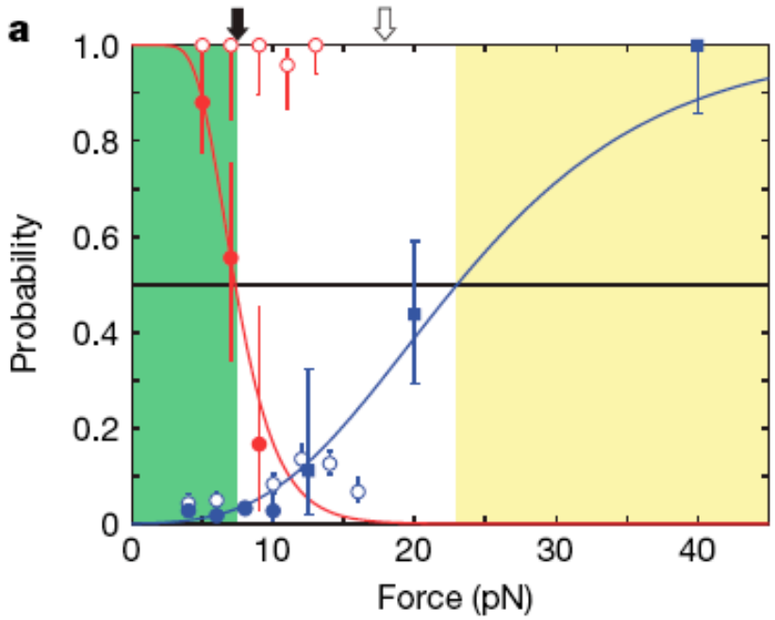


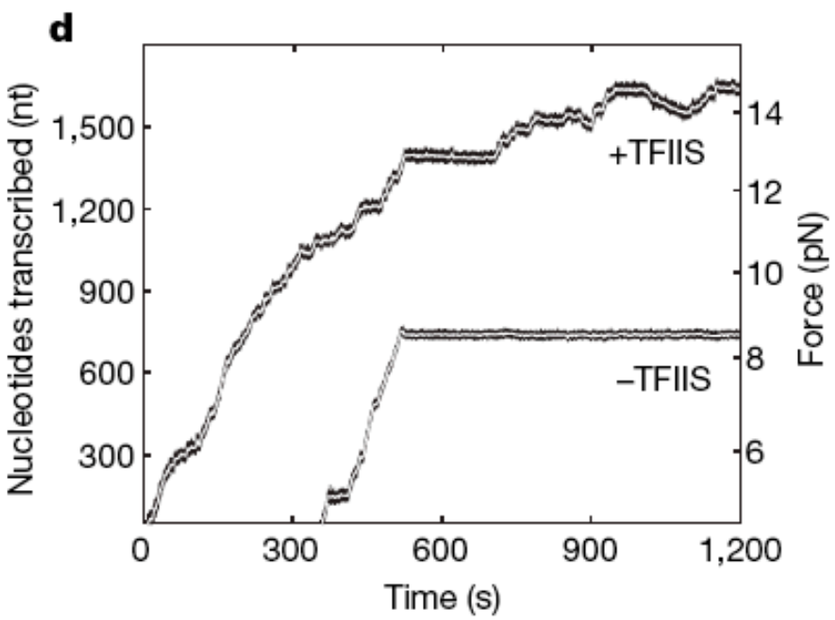
Eukaryotic RNAP II acts the same way

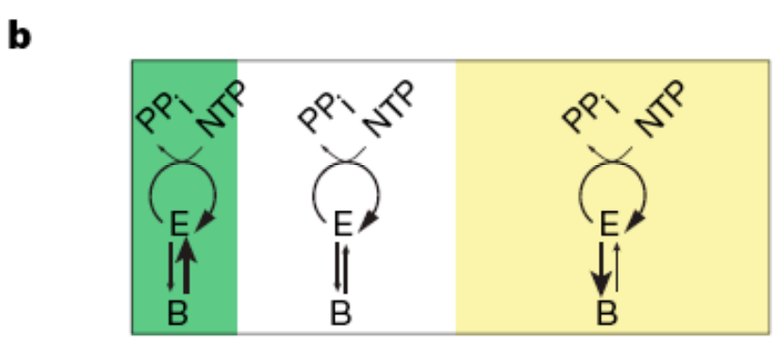


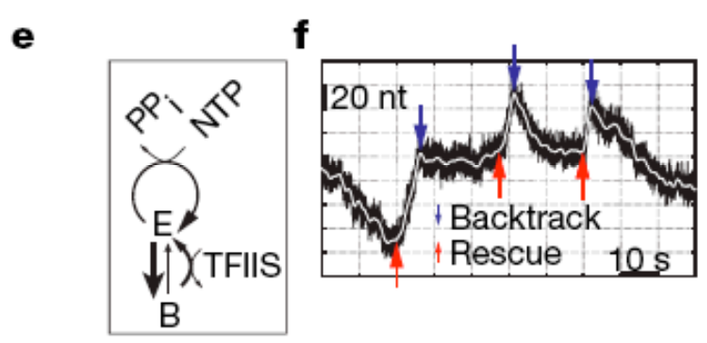
Galburt et al. *Nature* (2007)

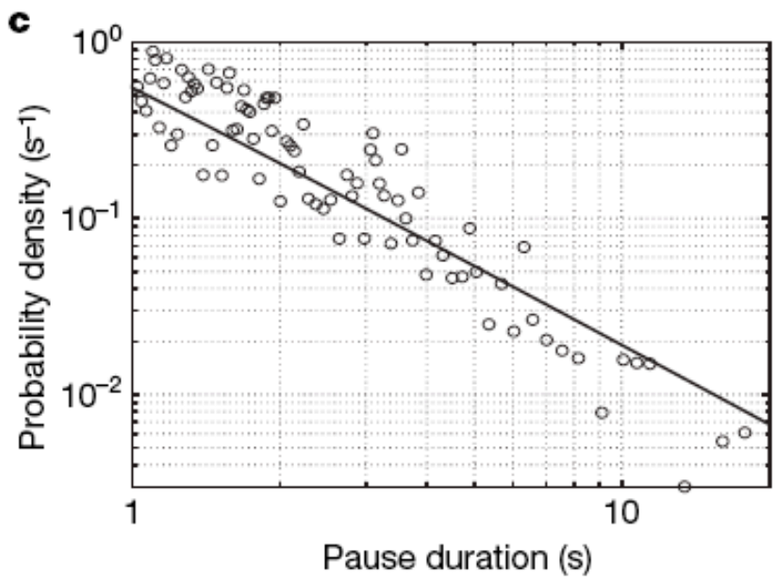
TFIIS is necessary to sustain high force

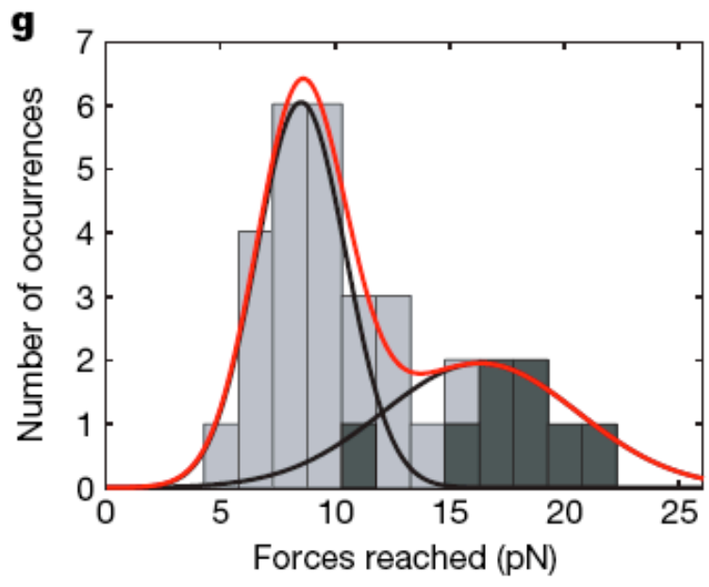








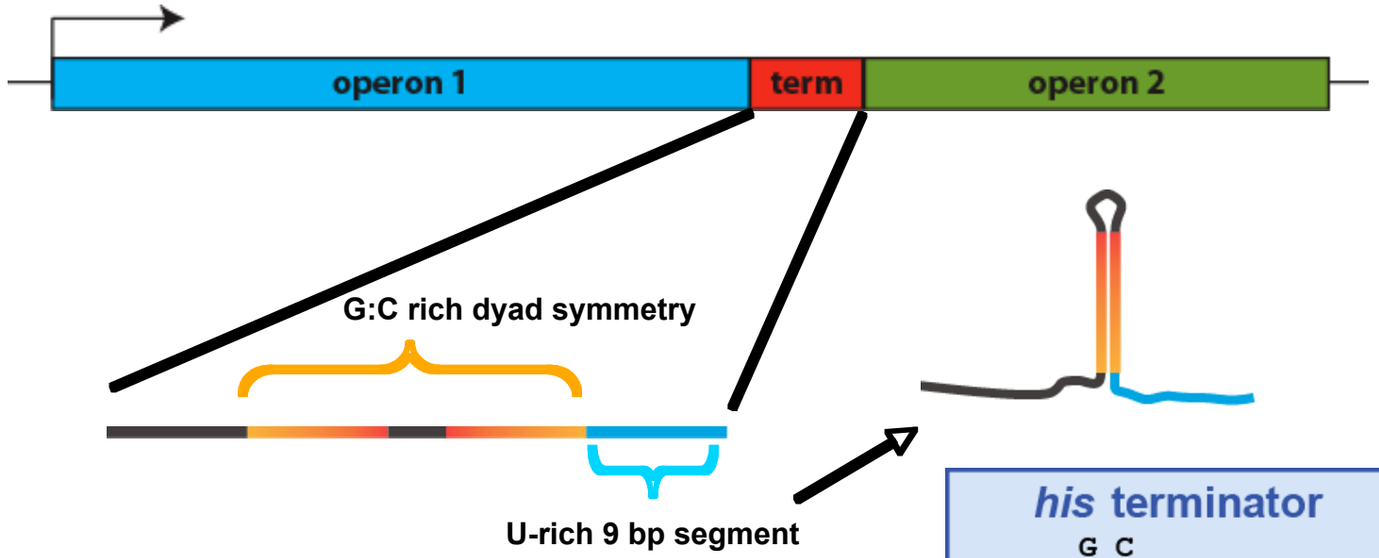




Max force doubles with TFIIS

Backtracking limits force in Pol II

The RNA's role in pausing and termination



t500 terminator

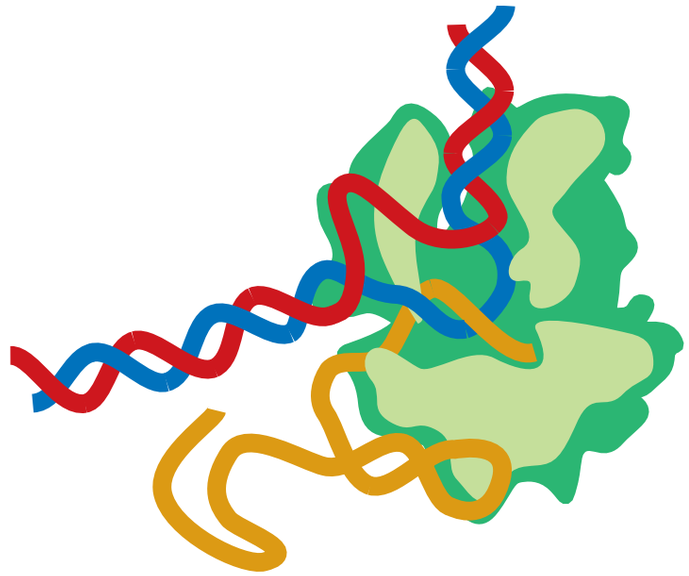
```
  A A
G   A
 C-G
 C-G
 G-C
 C-G   TE = 98%
 C-G
 C-G
 G-C
…CAAA UUUUCUGU-OH 3′
```

λtR2 terminator

```
 G U A A
G     U
U     C
  C-G
  G-C
  G-C   TE = 35%
  U-A
  C-G
  C-G
  G-C
  G-C
…AACA UUUUUAUU-OH 3′
```

***his* terminator**

```
   G C
 U    A
  A-U
  G-C
  A-U
  A-U
  G-C
  G-C
  C-G
  C-G
  C-G   TE = 74%
  C-G
  C-G
  G-C
…GAGAAA UUUUUUUUU-OH 3′
```



Two main features:

1. RNA hairpin
2. U-rich section (U-tract)

Will Greenleaf

Matthew Larson

Models of intrinsic hairpin termination

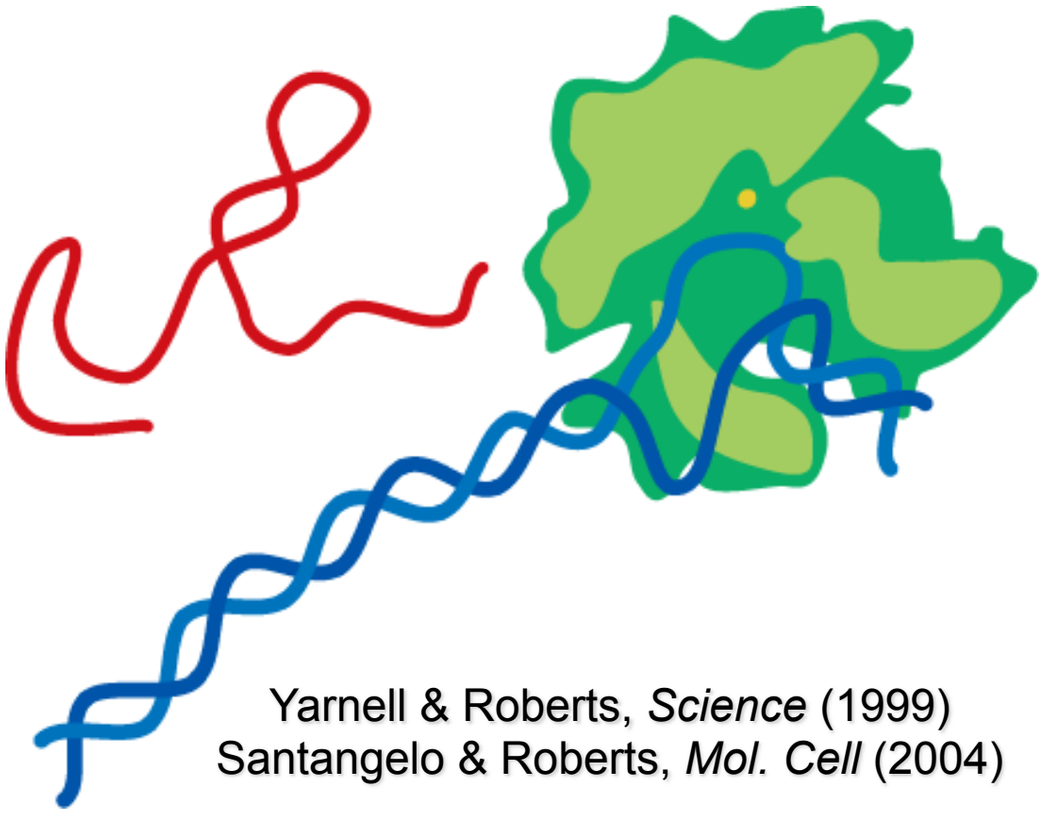
U-rich ‘slippery’ sequence:

leads to RNAP pausing *and/or* forms an unstable RNA:DNA hybrid

Forward Translocation Model

hairpin drives RNAP downstream without transcript elongation

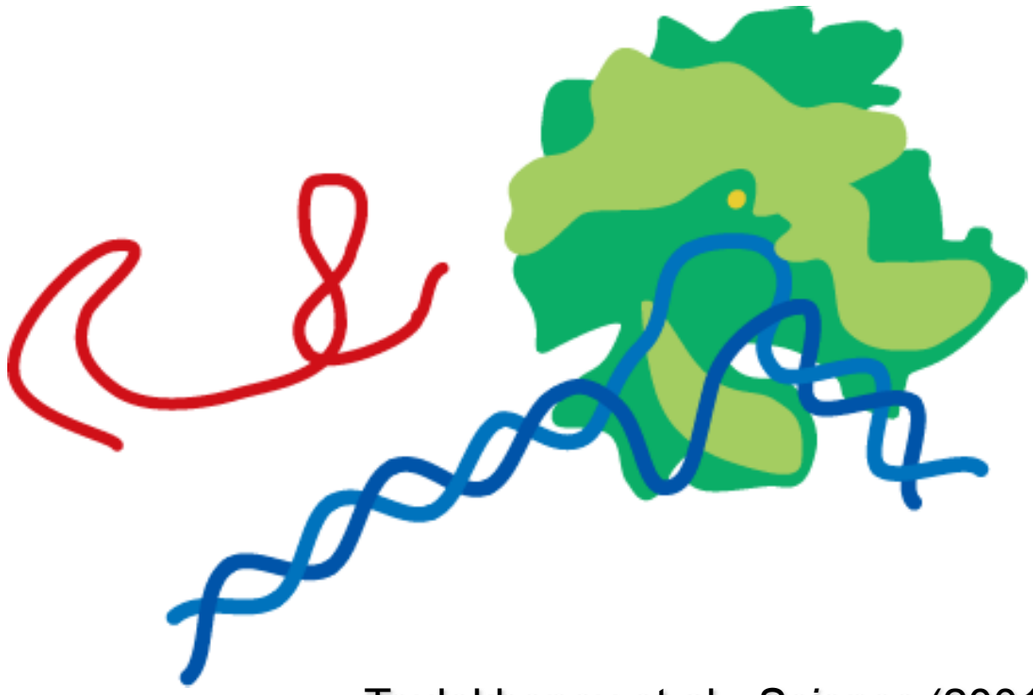
RNA stays “in register” with DNA



Yarnell & Roberts, *Science* (1999)
Santangelo & Roberts, *Mol. Cell* (2004)

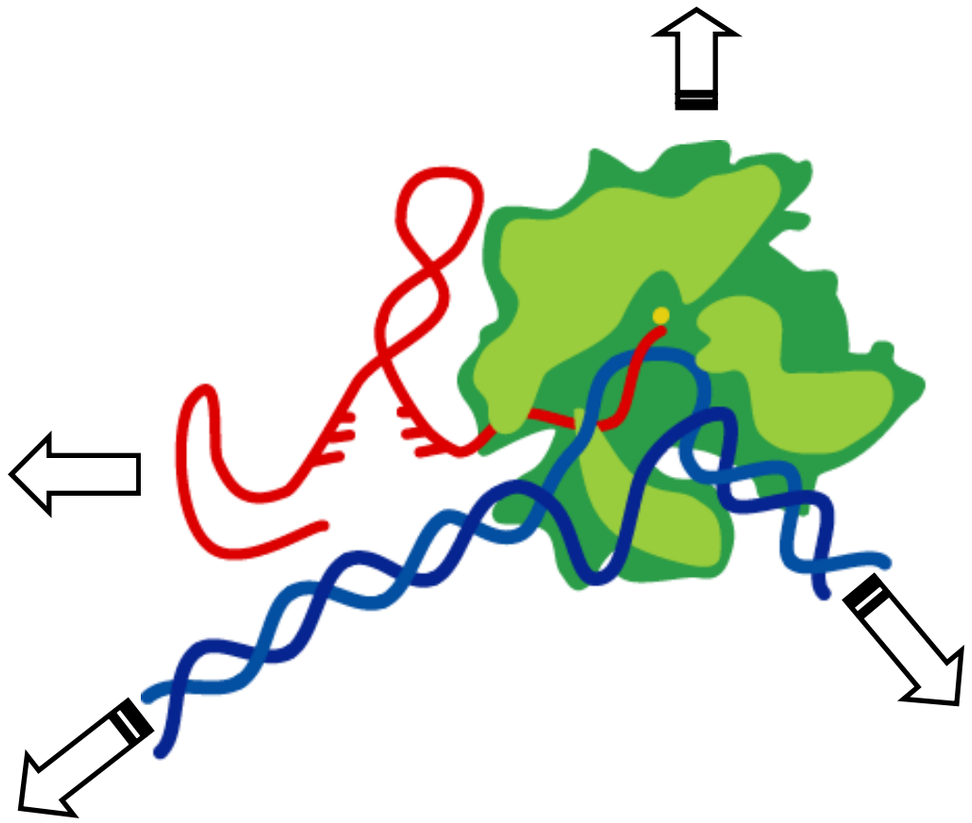
Allosteric Model

hairpin induces a conformational rearrangement of RNAP



Toulokhonov et al., *Science* (2001)

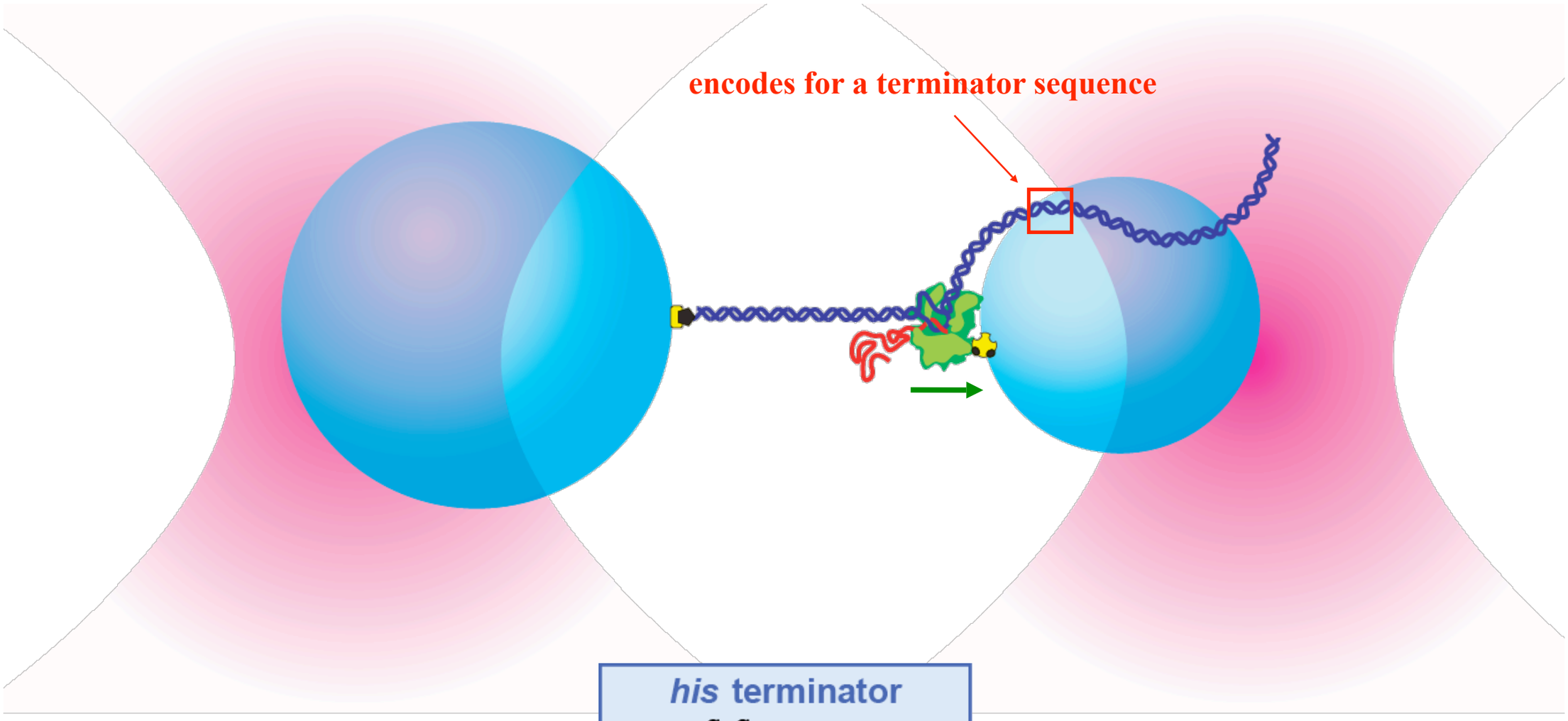
With force you can probe different parts of the system



You can:

1. determine the stability of the RNA:DNA hybrid
2. bias formation of secondary structure in the RNA
3. probe steps that involve enzyme motion along DNA or RNA

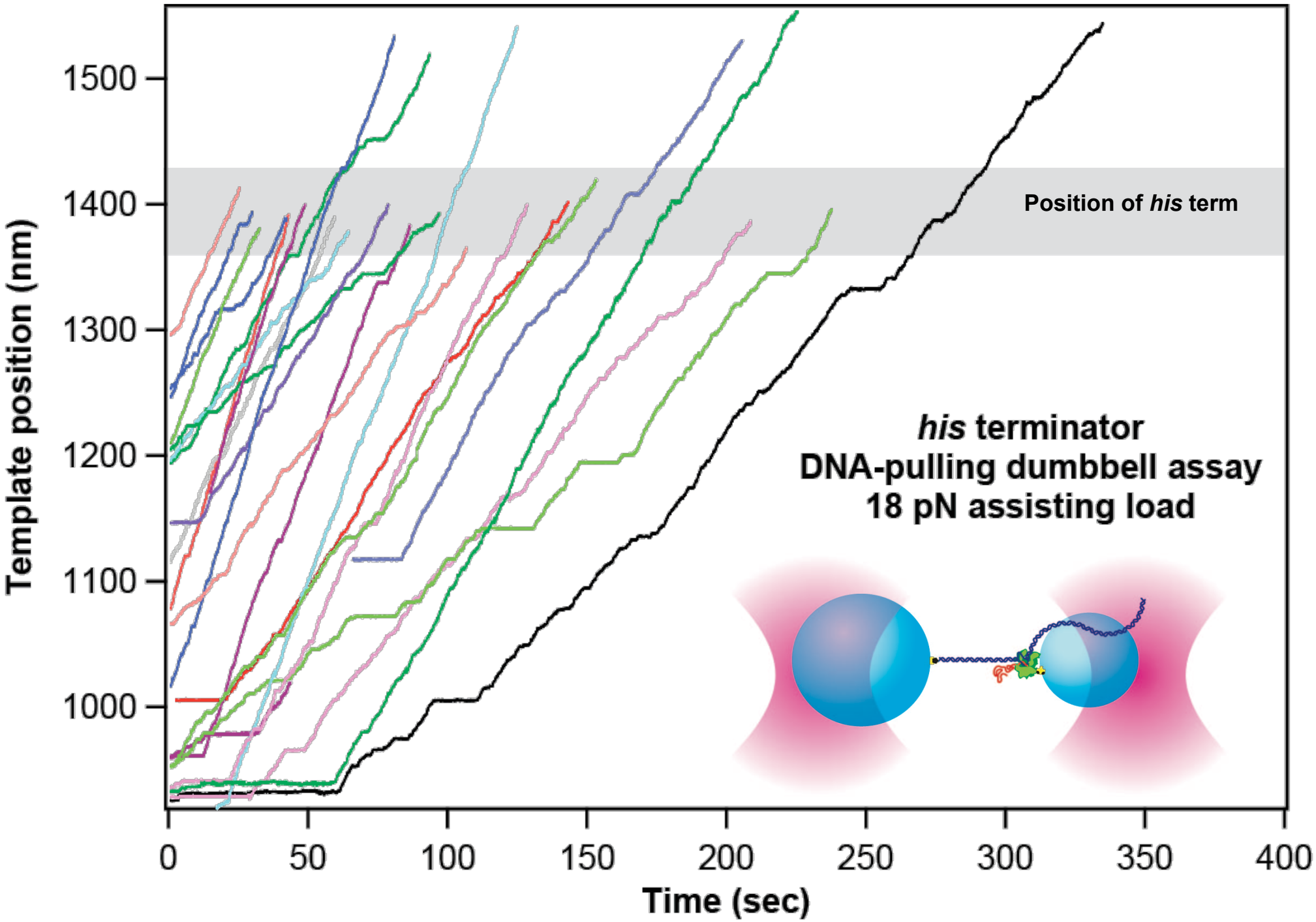
Pulling on the DNA tests translocation of the enzyme



***his* terminator**

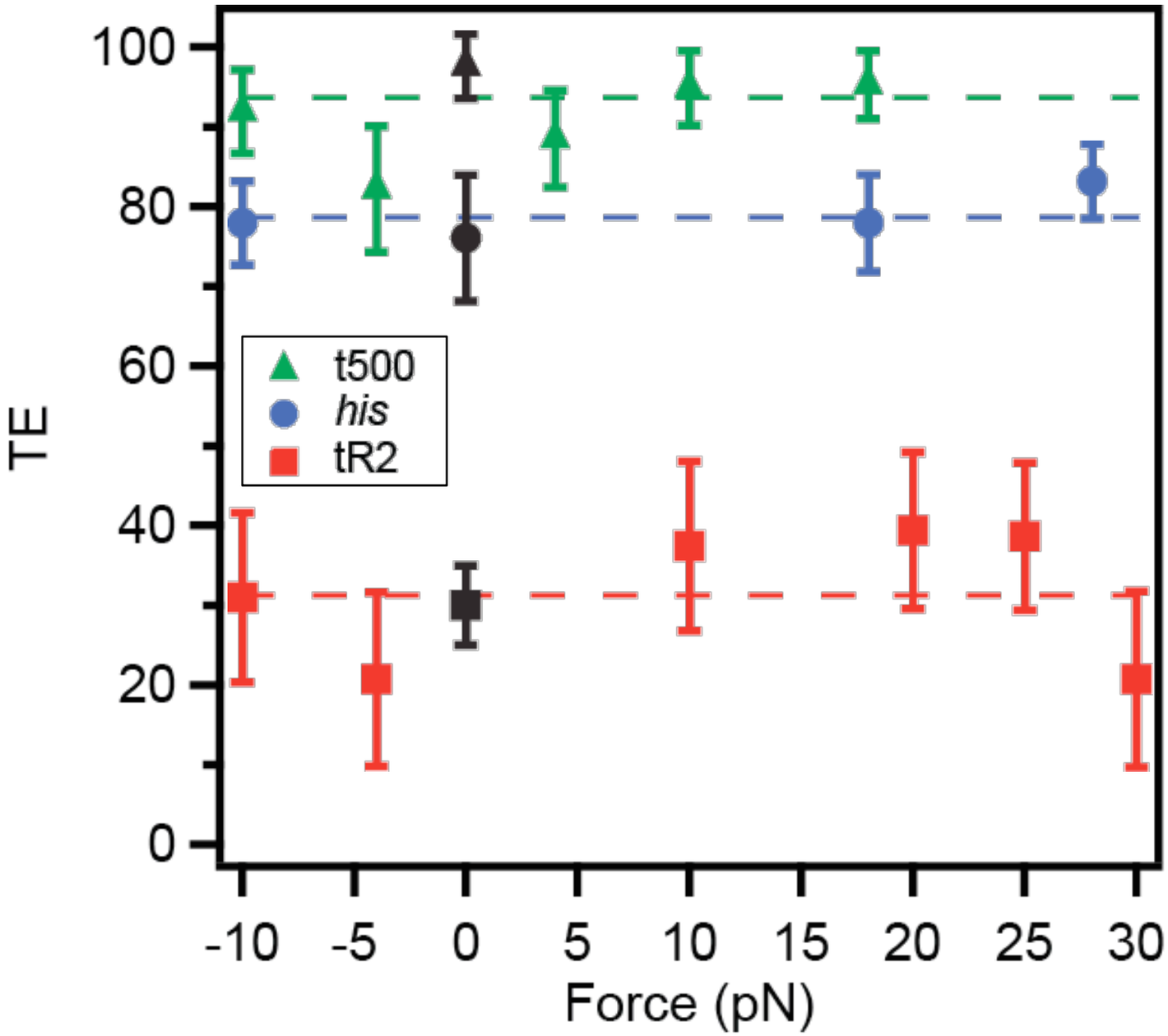
```
  G C
 U   A
  A-U
  G-C
  A-U
  A-U
  G-C
  G-C
  C-G
  C-G
  C-G
  C-G
  C-G
  G-C
…GAGAAA UUUUUUUUU-OH 3′
```

Elongation followed by termination

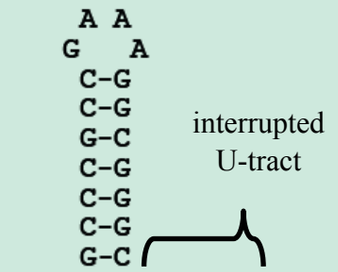


$$TE = \frac{term}{elongate + term} = 77\%$$

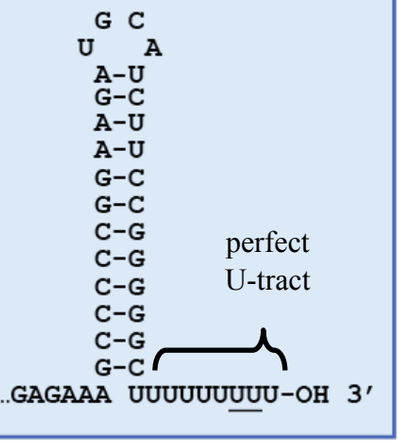
TE independent of force between DNA and RNAP



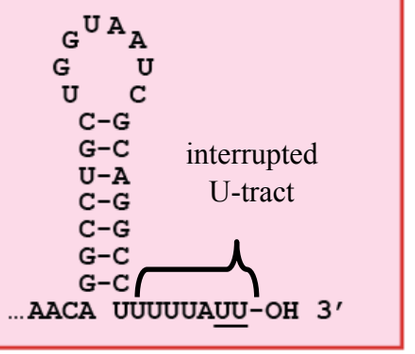
t500 terminator



***his* terminator**

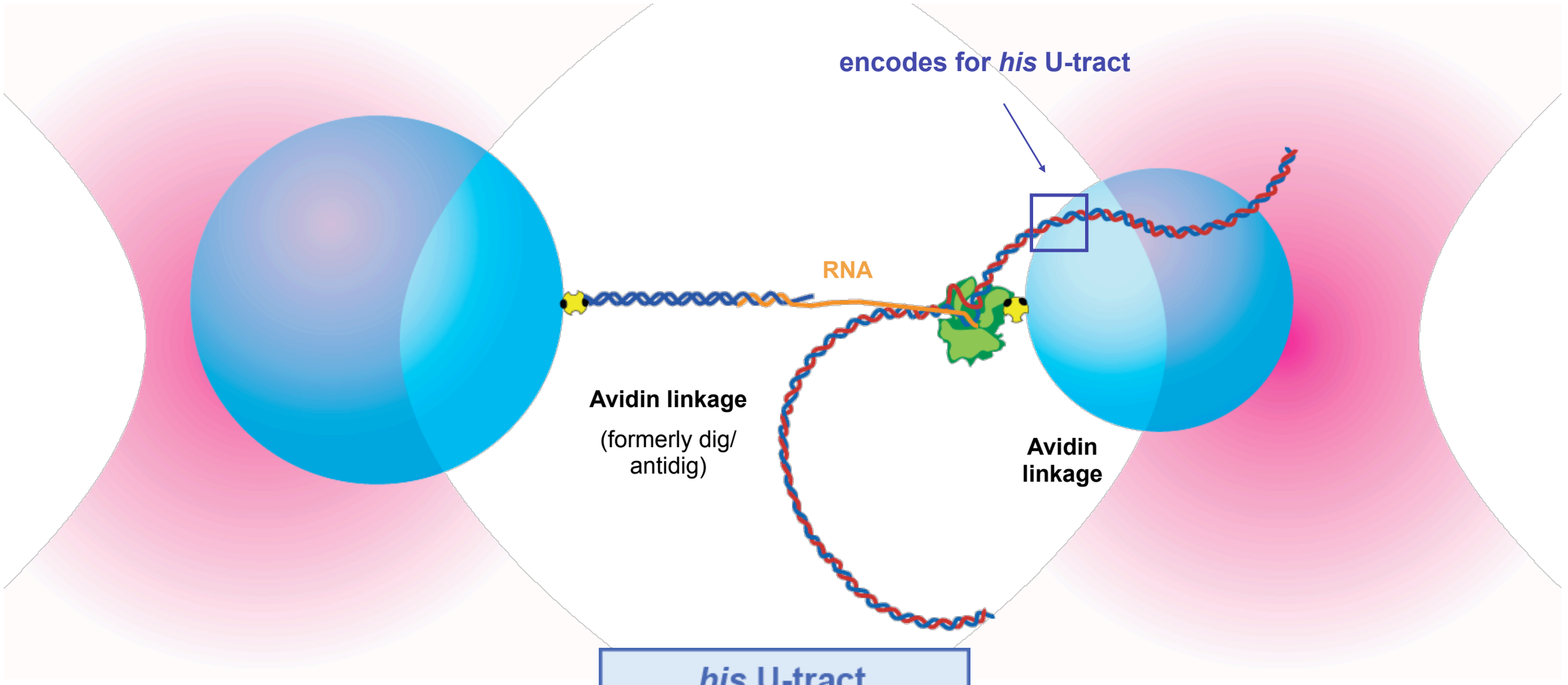


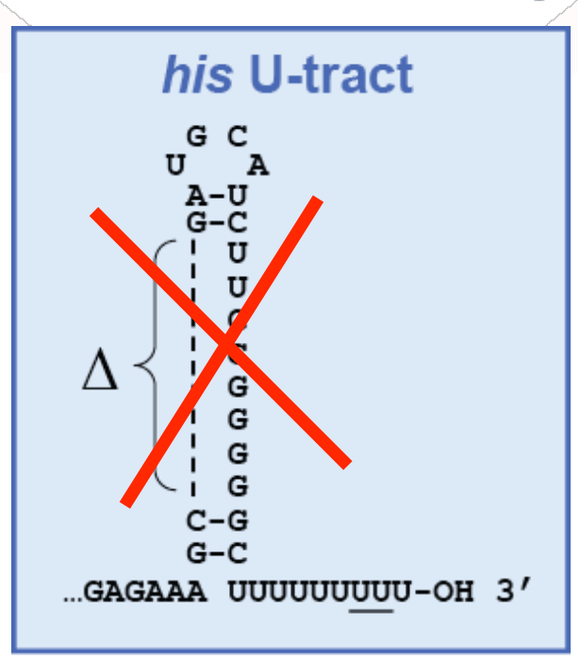
λtR2 terminator



No forward translocation!

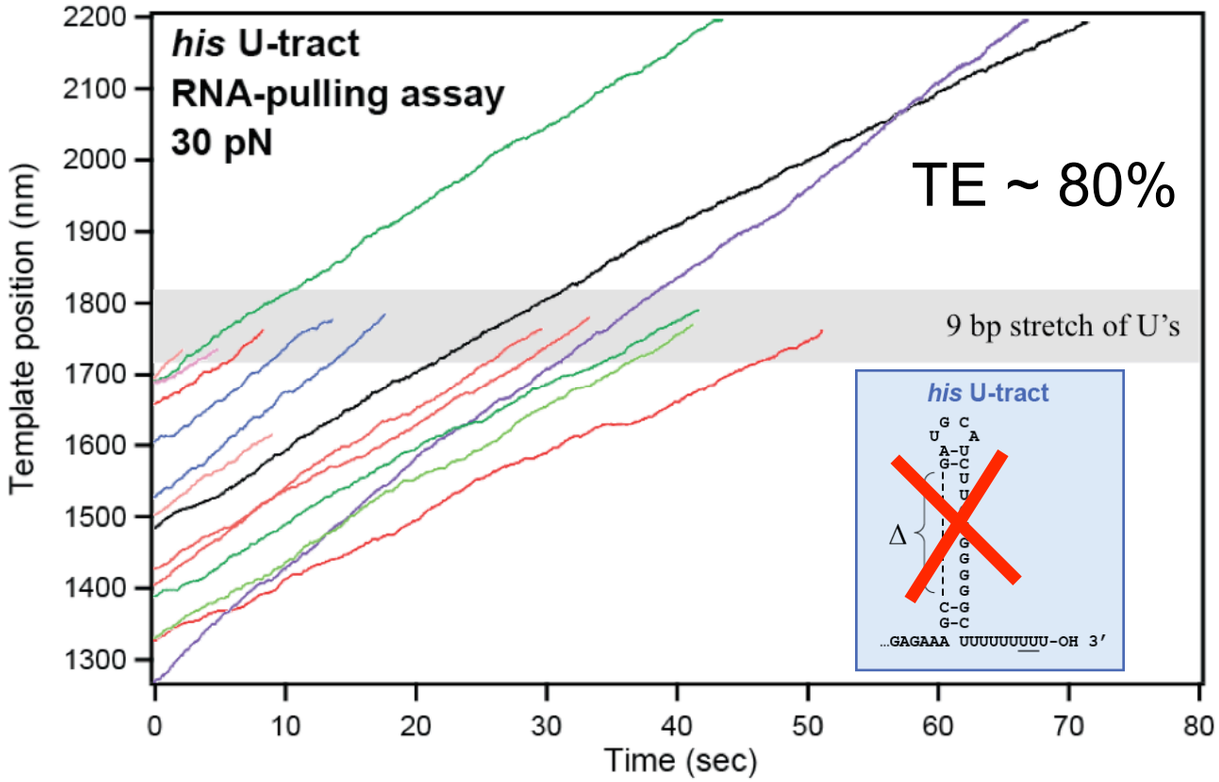
Pulling on the RNA

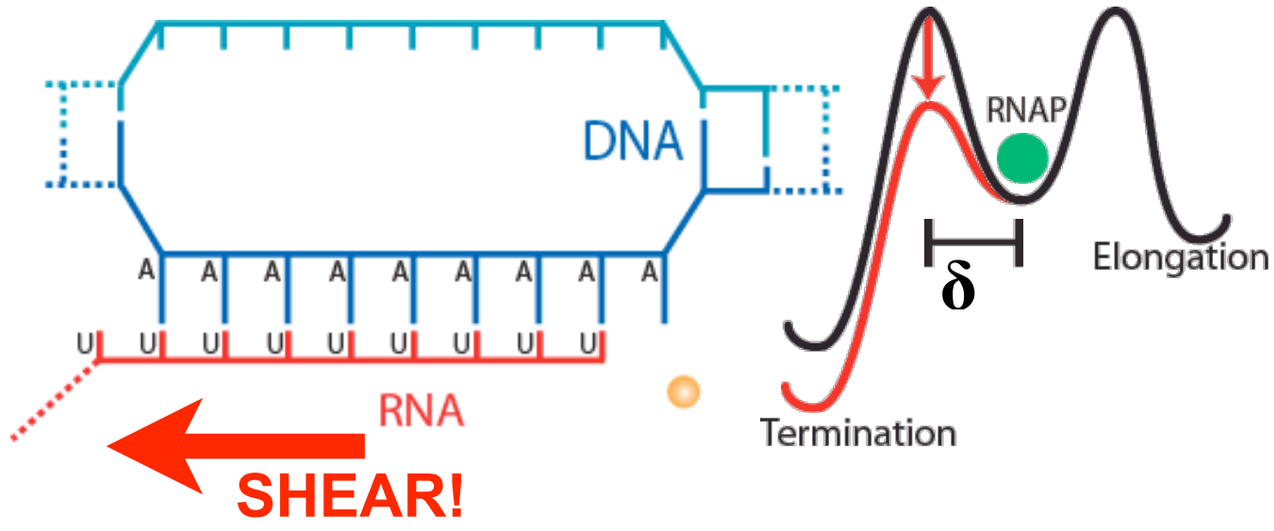




Modified from Dalal, Larson et al., *Mol. Cell* (2006)

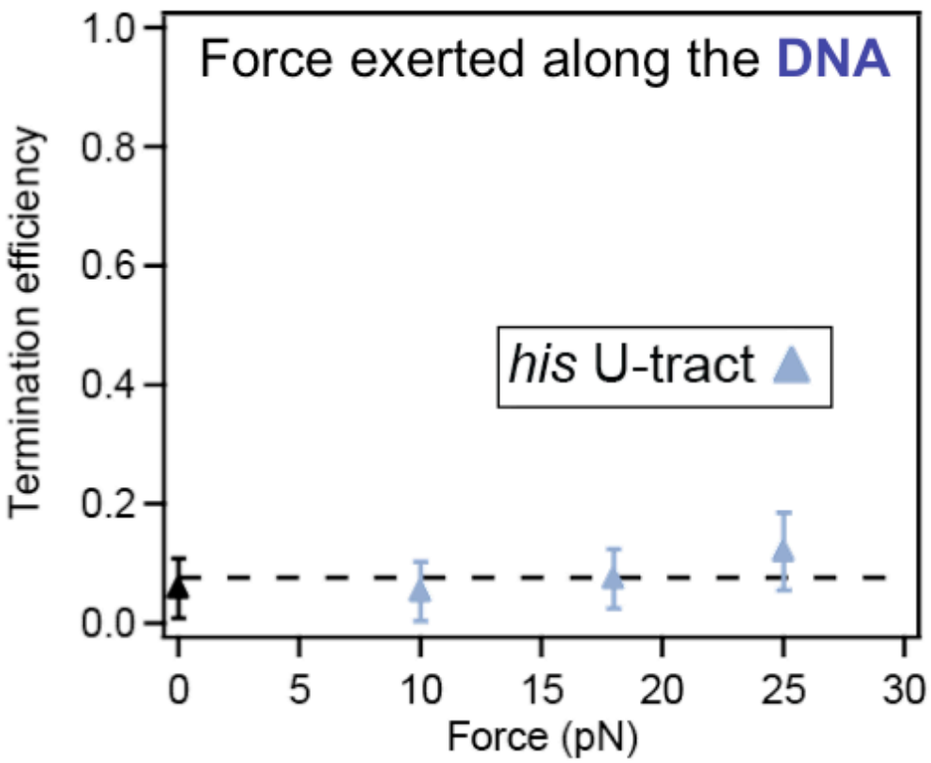
Termination occurs at the U-tract, is force dependent

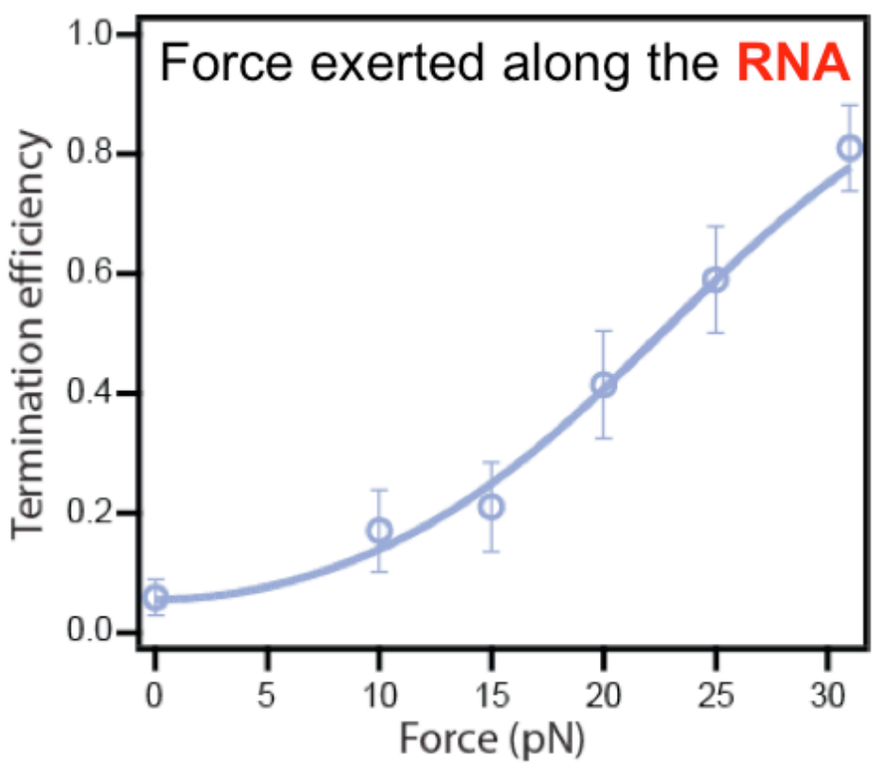




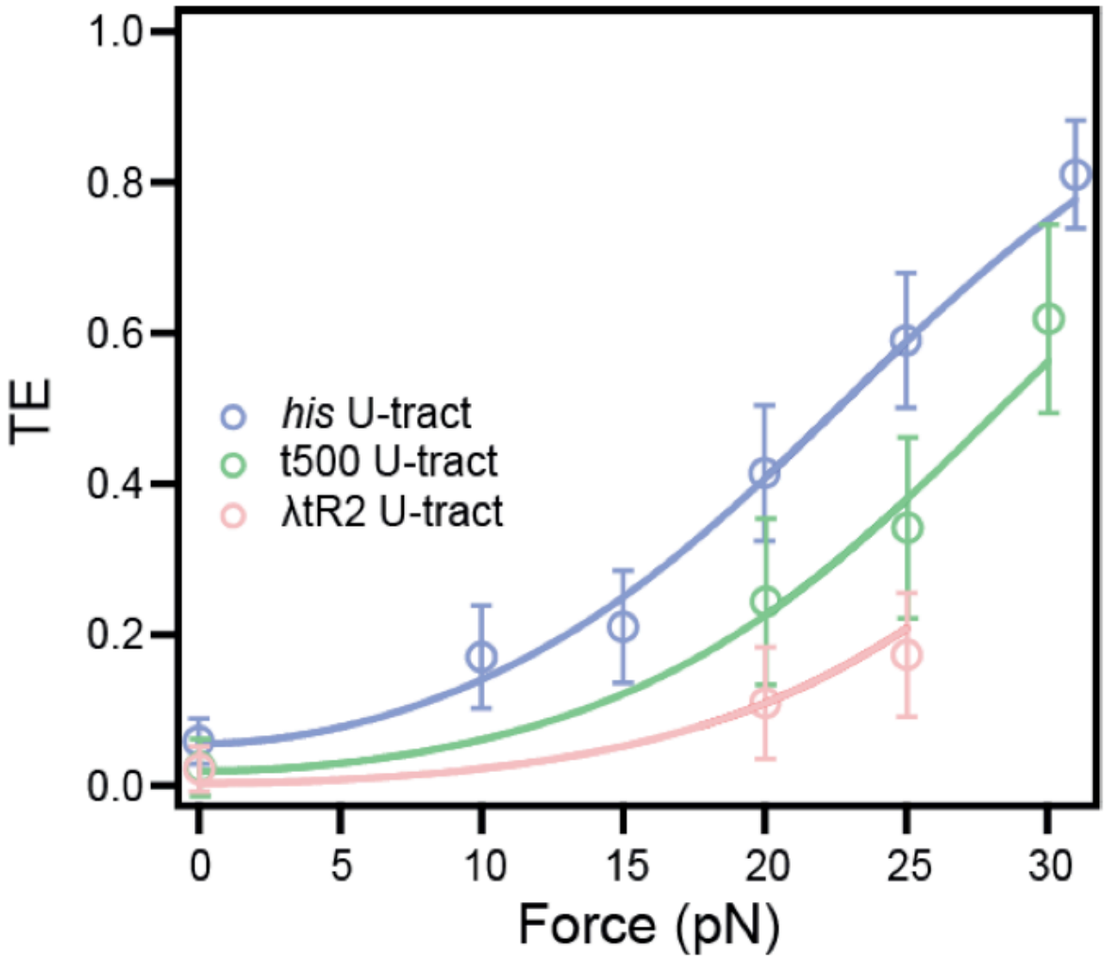
$$TE(F) = \frac{1}{1 + e^{(E_{shear} - F\delta)/k_B T}}$$

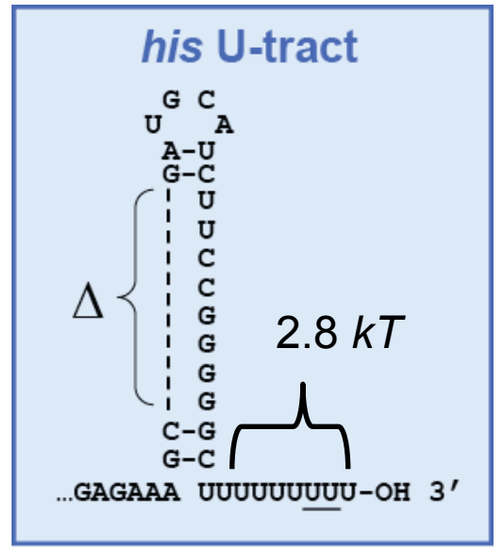
δ = 0.6nm = 1 bp

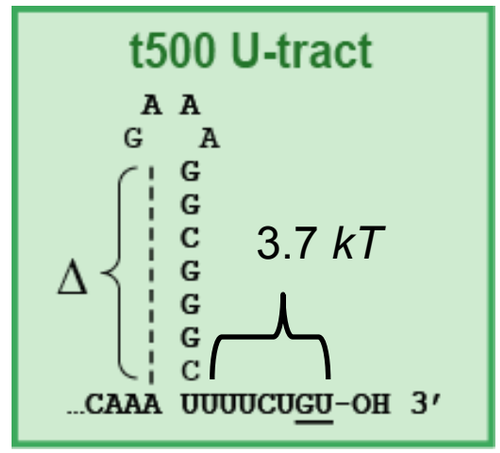


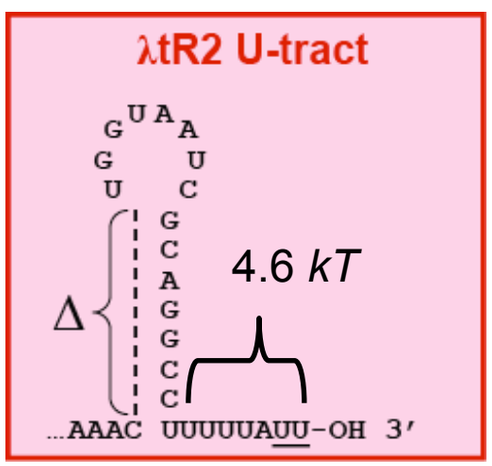


We find the same behavior in all 3 terminators







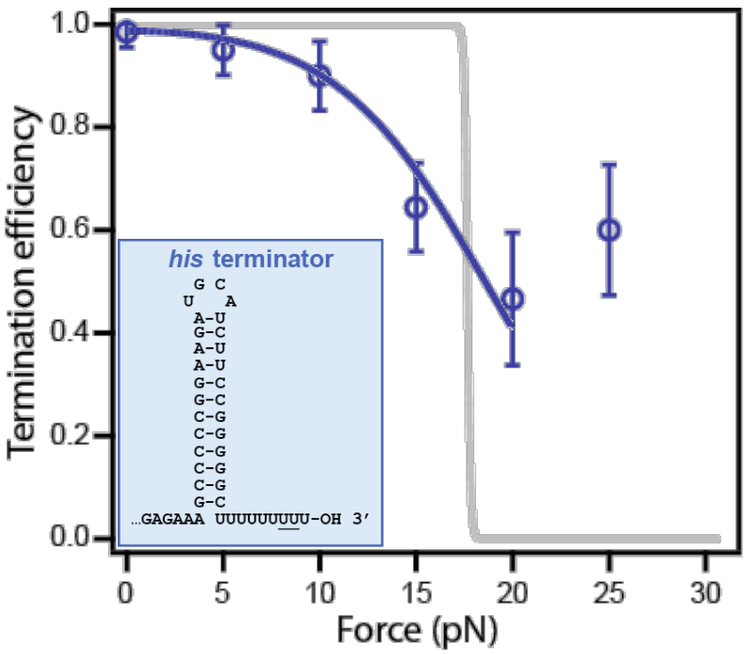


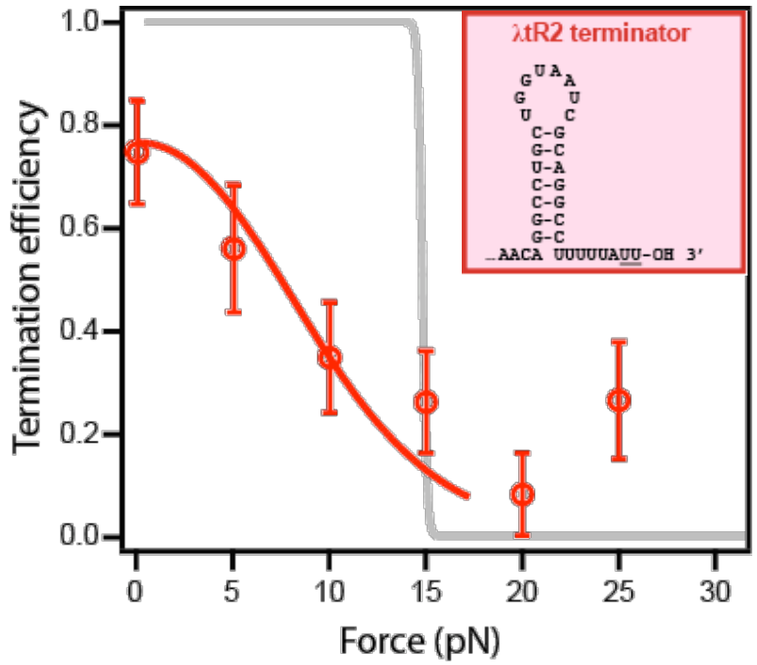
For all three the shear distance is the same, 1 bp

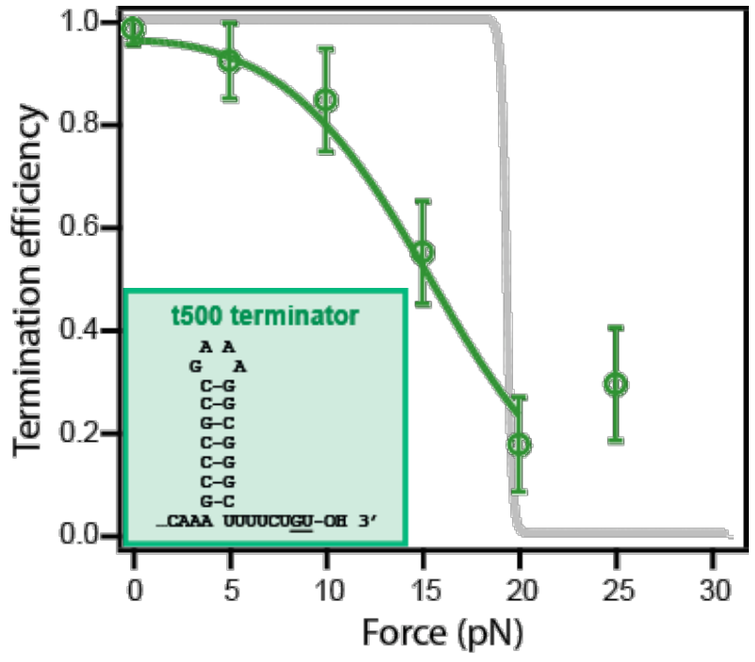
But, the shear energy barrier is different,
related to the sequence ...

If shearing causes termination, what is the role of the hairpin?

Pulling on the hairpin





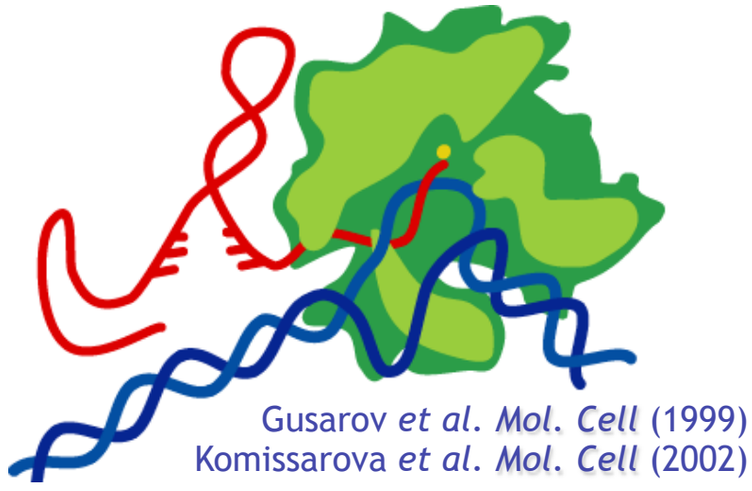


δ = 1.4nm ~ 2 bp

Unzipping the hairpin lowers the TE.

For these hairpins the last two bases are important for holding it together

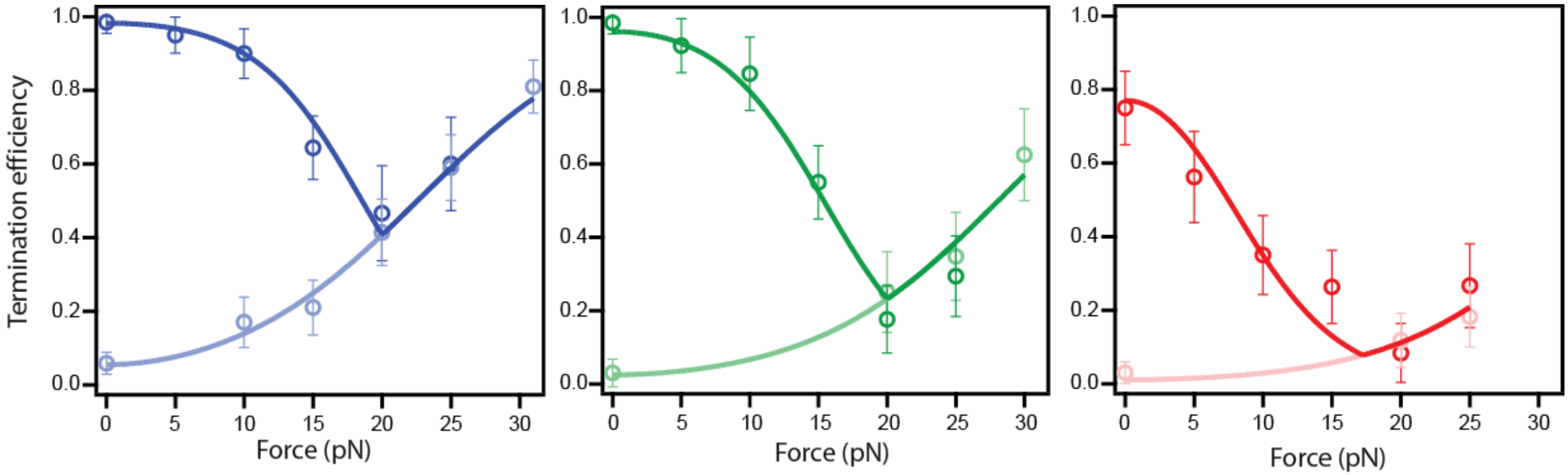
Hairpin zipping pulls the RNA out of the enzyme and causes termination...

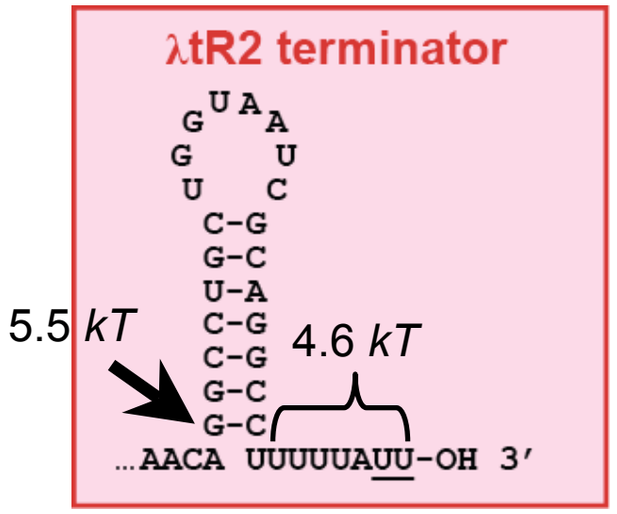


Gusarov *et al. Mol. Cell* (1999)
Komissarova *et al. Mol. Cell* (2002)

Simple model predicts TE and effect of load

$$\Delta E_{total} = [E_{shear} - F\delta_{shear}] - [E_{hairpin} - F\delta_{hairpin}]$$



δ_{shear}	0.6 nm ↔ 1 nt
$E_{shear}(his)$	2.8 *kT* (1.7 kcal/mol)
E_{shear}(t500)	3.7 *kT* (2.2 kcal/mol)
E_{shear}(λtR2)	4.6 *kT* (2.8 kcal/mol)
$\delta_{hairpin}$	1.4 nm ↔ 2 nt
$E_{hairpin}$	6.5 – 5.5 *kT*

These values match m-fold predictions