SUPPLEMENTARY INFORMATION FOR

5-Formyl- and 5-carboxyl-cytosine reduce the rate and substrate specificity of RNA polymerase II transcription

Matthew W. Kellinger[1], Chunxiao Song[2], Jenny Chong[1], Xingyu Lu[2], Chuan He[2], and Dong Wang[1]

[1]Skaggs School of Pharmacy and Pharmaceutical Sciences, The University of California, San Diego, La Jolla, California, 92093, USA.

[2]Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637, USA.

Correspondence should be addressed to D.W. (dongwang@ucsd.edu).

This file includes:

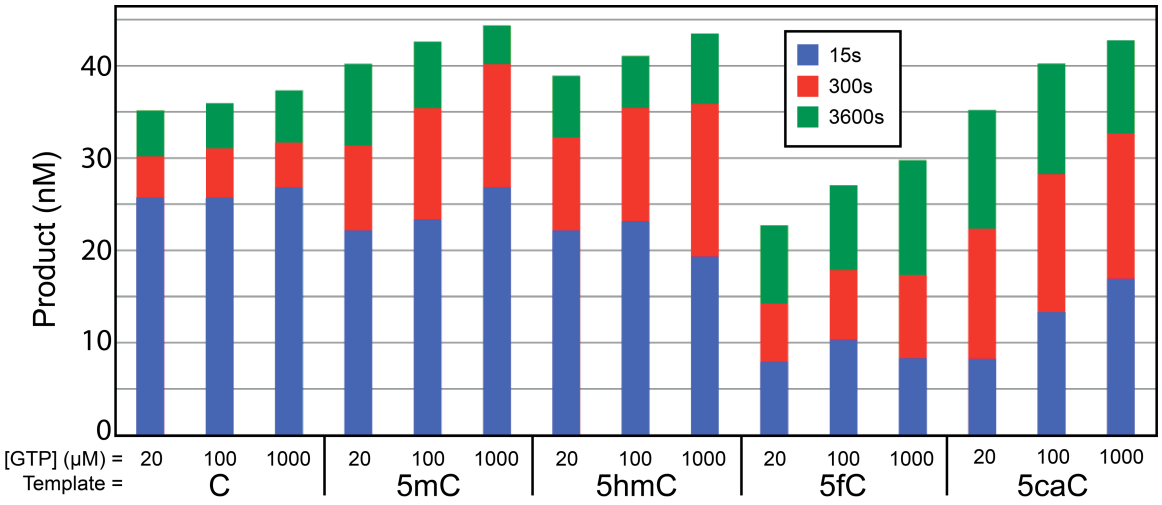
Supplementary Figure 1 to 5

Supplementary Table 1 to 3

Supplementary Note

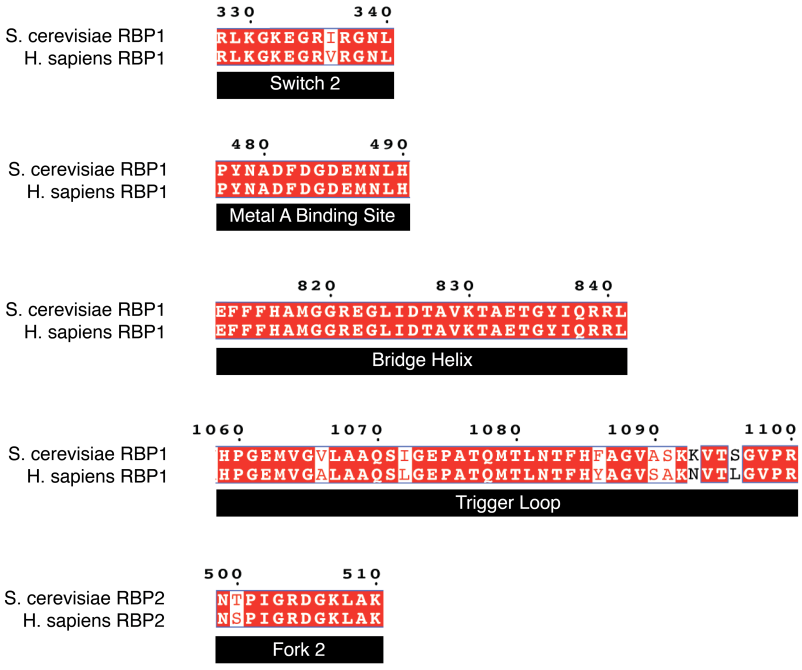
Supplementary References

Supplementary Figures
Supplementary Fig. 1



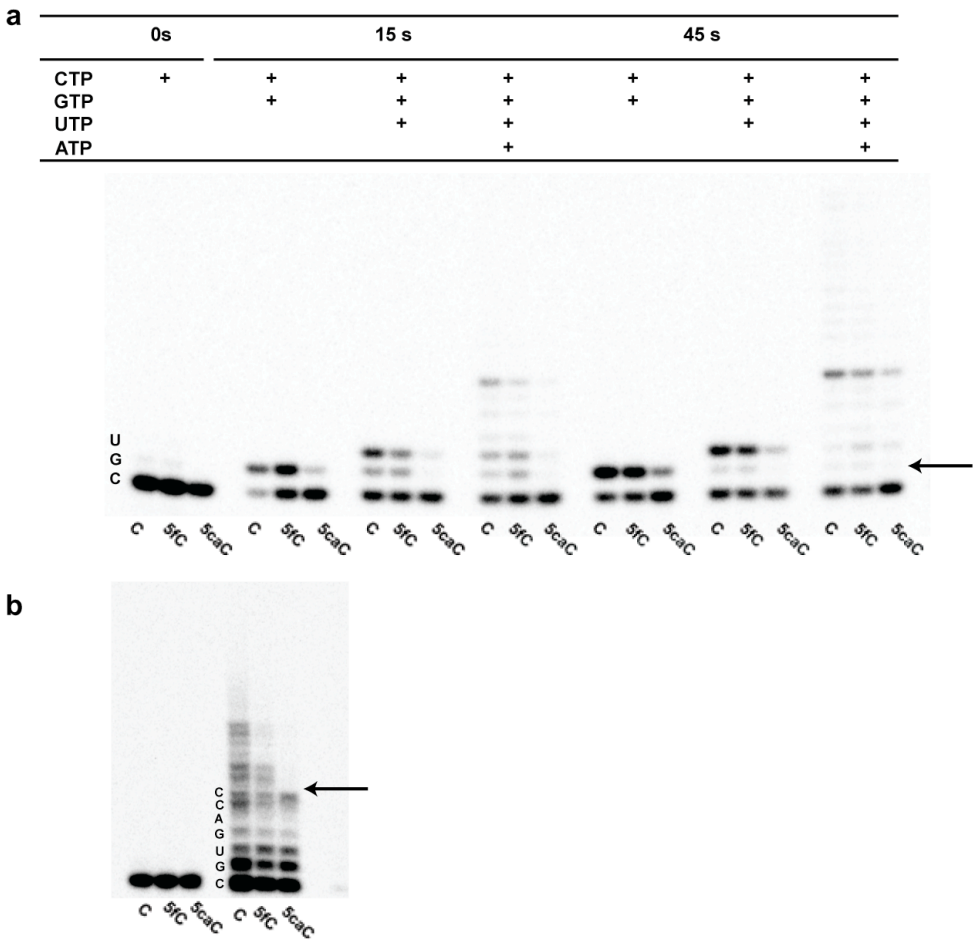
Supplementary Fig. 1. Time and concentration dependence of GTP incorporation on methylation templates. Scaffolds containing cytosine or each of modified cytosine templates were assayed for GTP product formation. Three different concentrations of GTP (20 μM, 100 μM, and 1000 μM) were allowed to react with 50 nM enzyme:scaffold complexes for 15 sec (blue), 5 min (red), or 60 min (green) before quenching. The rates of nucleotide incorporation for 5-formyl- (5fC) and 5-carboxyl-cytosine (5caC) templates are reduced compared with the C template.

Supplementary Fig. 2



Supplementary Fig. 2. Sequence alignments of human and yeast S. cerevisiae Pol II residues nearby active site. Human and yeast *S. cerevisiae* Pol II sequences of **(a)** metal A binding sites (Rpb1 477-490), **(b)** switch 2 region (Rpb1 328-340), **(c)** bridge helix (Rpb1 812-841), **(d)** trigger loop (Rpb1 1059-1100), and **(e)** fork loop 2 (Rpb2 499-510) are aligned with ClustalW.

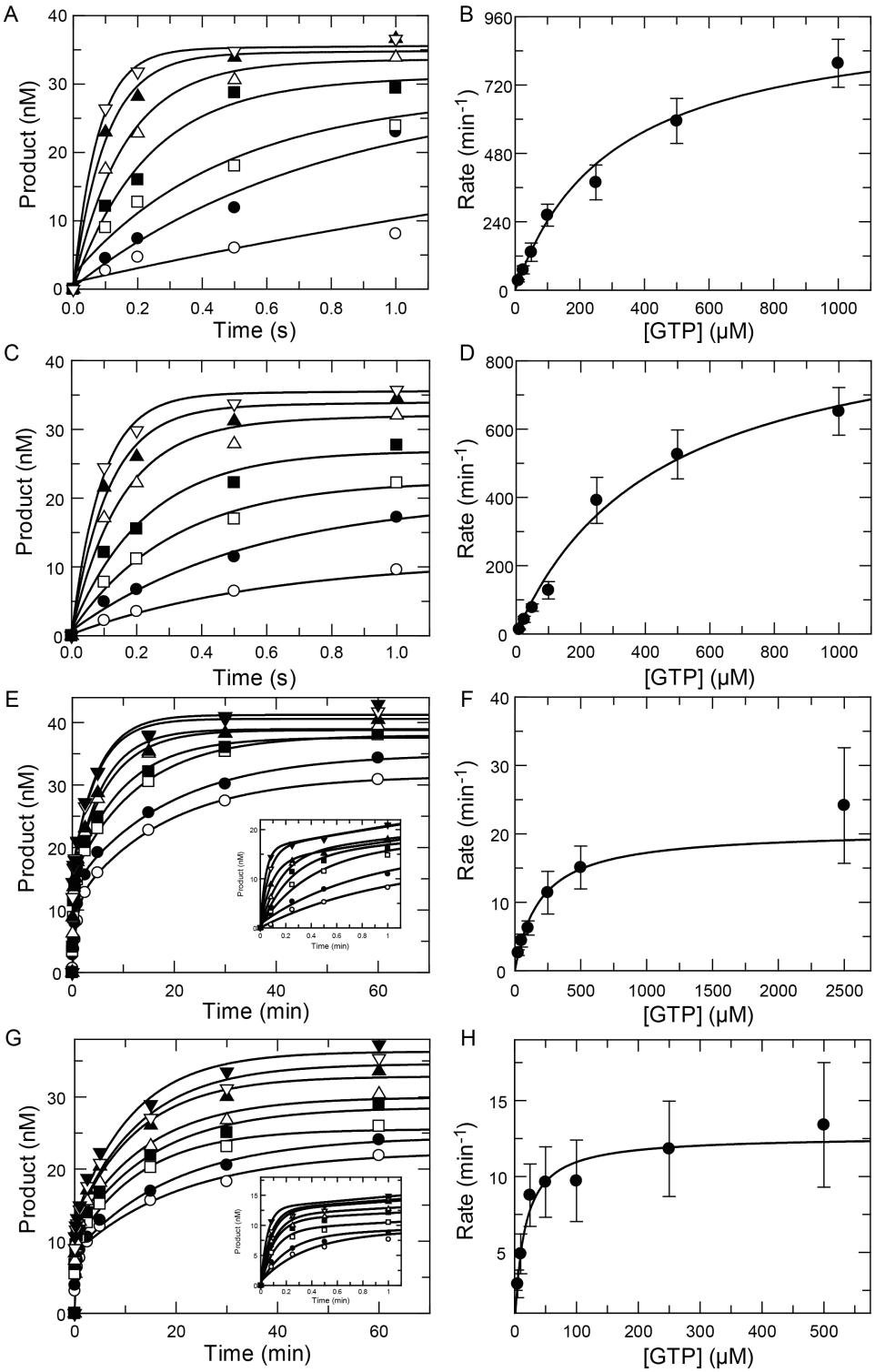
Supplementary Fig. 3



Supplementary Fig. 3. NTP incorporation on methylation templates with a full transcription bubble. **(a)** The Pol II elongation complex with scaffold B containing cytosine or each of the modified cytosine templates was assayed for product formation with incubation of 5 μM CTP, CTP/GTP mixture, CTP/GTP/UTP mixture, or NTP, respectively. The reactions were quenched after 15-sec or 45-sec incubation at room temperature. The band position

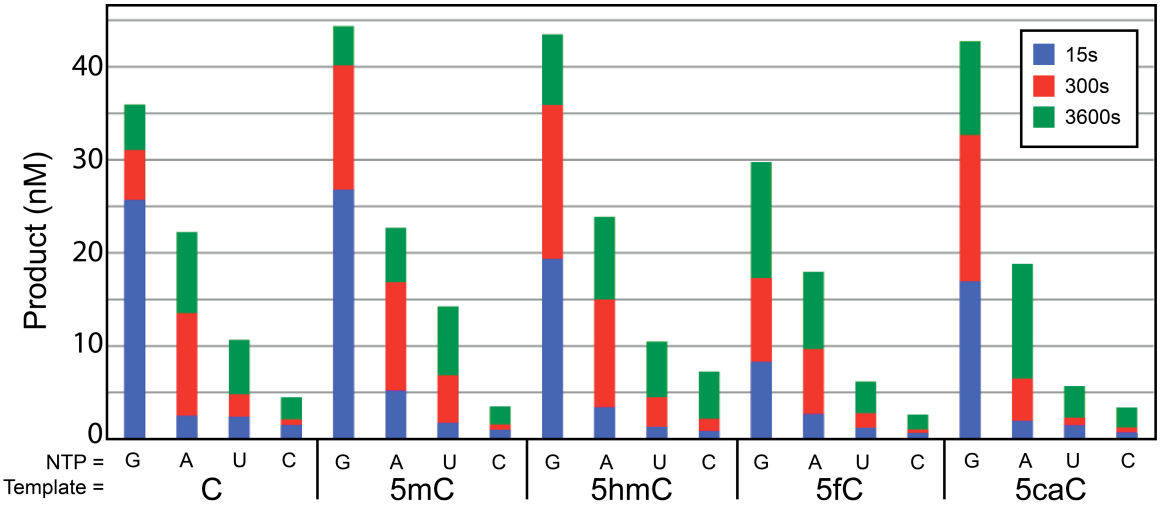
corresponding to GTP incorporation opposite to modification site was depicted with an arrow in the gel. **(b)** The Pol II elongation complex with scaffold C containing cytosine or each of modified cytosine template was assayed for product formation upon incubation with 5 μM NTP. The reactions were quenched after 5-min incubation at room temperature. The band position corresponding to GTP incorporation opposite to the modification site is depicted with an arrow in the gel.

Supplementary Fig. 4



Supplementary Fig. 4. Non-linear regression analysis of GTP incorporation concentration dependence. Enzyme:scaffold complexes (scaffold A, 50 nM) containing C, 5hmC, 5fC, or 5caC templating bases were reacted with various concentrations of GTP. Product formation at each GTP concentration was fit to (**Equation 1**) (A, C, E and G). The inlayed figures in E and G show the fast phase of GTP incorporation on a 1 sec timescale. The concentration dependence of the observed fast phase was fit to **Equation 2** to obtain k_{pol} and $K_{d,app}$ values for each of the templates (B, D, F, and H). GTP incorporation on the C and 5hmC templates (A and C) was fast and required the use of a rapid quench flow. 5-Formyl- and 5-carboxyl-cytosine templates resulted in much slower, biphasic GTP incorporation (E and G). Fitted values (shown as smooth lines) are reported in **Supplementary Table 1 and 2.** The concentrations of GTP (C and 5hmC template: 10, 25, 50, 100, 250, 500, and 1000 μM) are shown in symbols (from low to high): ○, ●, □, ■, △, ▲, ▽, respectively. The concentrations of GTP (5fC and 5caC template: 5, 10, 25, 50, 100, 250, 500, and 2500 μM) are also shown in symbols (from low to high): ○, ●, □, ■, △, ▲, ▽, ▼, respectively.

Supplementary Fig. 5



Supplementary Fig. 5. Time and concentration dependence of NTP incorporation on methylation templates. Scaffolds (scaffold A) containing cytosine or each modified cytosine template were assayed for product formation with incubation of 1 mM ATP, GTP, CTP, UTP, respectively. The reaction were allowed to react with 50 nM enzyme:scaffold complexes for 15 sec (blue), 5 min (red), or 60 min (green) before quenching.

Supplementary Tables

Supplementary Table 1. Kinetic parameters for GTP incorporation against modified cytosine templates

Template	k_{pol} (min^{-1})	$K_{d,app}$ (μM)	$k_{pol}/K_{d,app}$ ($\mu M^{-1}min^{-1}$)	Relative Specificity[a]	Fold Change[b]
C	980 ± 80	300 ± 37	3.3 ± 0.5	100 ± 15	1
5hmC	960 ± 100	440 ± 100	2.2 ± 0.5	67 ± 18	1.5 ± 0.5
5fC	20 ± 3	190 ± 40	0.11 ± 0.03	3.3 ± 1.0	30 ± 10
5caC	12.7 ± 0.7	16 ± 2	0.79 ± 0.11	24 ± 5	4.2 ± 1.1

[a]Relative Specificity = 100 x $(k_{pol}/K_{d,app})_{template}$ / $(k_{pol}/K_{d,app})_C$

[b]Fold Change = $(k_{pol}/K_{d,app})_C/(k_{pol}/K_{d,app})_{template}$

Supplementary Table 2. Biphasic GTP incorporation

Template	θ_1 Fast Phase (%)	θ_2 Slow Phase (%)
C	77 ± 3	23 ± 1
5hmC	78 ± 4	22 ± 2
5fC	39 ± 4	61 ± 2
5caC	35 ± 5	65 ± 2

Supplementary Table 3. Pol II substrate discrimination of GTP over ATP on 5fC and C templates

Template	NTP	$k_{pol}/K_{d,app}$ ($\mu M^{-1}min^{-1}$)	Discrimination[a]	Fold Change[b]
C	GTP	3.3 ± 0.5	(2.8 ± 0.5) x 10^4	31 ± 8
	ATP	(1.2 ± 0.1) x 10^{-4}		
5fC	GTP	0.11 ± 0.03	(9 ± 3) x 10^2	
	ATP	(1.2 ± 0.3) x 10^{-4}		

[a]Discrimination = $(k_{pol}/K_{d,app})_{GTP}$ / $(k_{pol}/K_{d,app})_{ATP}$
[b]Fold Change = $Discrimination_C$ / $Discrimination_{5fC}$

Supplementary Note

The NTP concentration dependence of product formation for all five forms of cytosine residues was first examined using 20 μM, 100 μM, and 1000 μM GTP, respectively. The time dependence at each NTP concentration was resolved by quenching reactions after 15 sec, 5 min, and 60 min (**Supplementary Fig. 1** and **Fig. 5**). GTP incorporation against a cytosine template nears completion at 15 sec shows no difference across various concentrations (from 20 μM to 1 mM). Together, these findings indicate that correct GTP:dC incorporation is too fast to be resolved on the min timescale. In sharp contrast, a significantly different pattern of GTP incorporation was observed for 5fC and 5caC templates. GTP incorporation against the 5fC template showed little product formation at the 15 sec time point. Higher amounts of product formation were observed with much longer incubation. Notably, slow GTP incorporation was not observed for C, 5mC, or 5hmC. This finding indicates that GTP incorporations for 5fC and 5caC have a significantly slower k_{pol} compared with GTP for C template.

Meaning of apparent K_d depends on rate-limiting step. Single turnover reactions at various GTP concentrations were then performed to determine the maximum rate of nucleotide polymerization (k_{pol}) and an *apparent* dissociation constant ($K_{d,app}$) for C, 5hmC, 5fC and 5caC templates, respectively. Polymerization rate constants (k_{pol}) for 5fC and 5caC (**Supplementary Table 1**)

were reduced by 49-fold and 77-fold, respectively, in comparison with the cytosine template. *Apparent* K_d values were also reduced by 1.6-fold and 19-fold, respectively (**Supplementary Table 1**). The observation of "tighter" *apparent* GTP binding compared with a cytosine template initially seems counter-intuitive. However, recent advances in polymerase kinetics explain this apparent paradox.

The specificity constant (k_{cat}/K_m) was derived using well-established mechanistic assumptions applied throughout the DNA and RNA polymerase fields to allow $k_{pol}/K_{d,app}$ to equal to k_{cat}/K_m[1-3]. These assumptions (required due to experimental limitations) include: nucleotide binding is a single-step rapid equilibrium, chemistry (k_{pol}) is the rate-limiting step in product formation, and the steps following chemistry (pyrophosphate release and translocation) are fast. Recent studies investigating nucleotide binding to DNA polymerases have shown that the assumption of rapid equilibrium of NTP binding is valid for altered or mismatched nucleotide incorporation events, but becomes misleading for correct nucleotide incorporations[4-7]. This difference is the reason the term *apparent* is used when describing the binding constant derived from single turnover polymerase analysis.

While the specificity constant assumption ($k_{pol}/K_{d,app}=k_{cat}/K_m$) is valid regardless of correct or incorrect incorporation, the meaning of *apparent* K_d changes. This change is the result of a more complete polymerase mechanism

that includes substrate induced fit (supported by structural differences in the Pol II Trigger Loop upon correct NTP binding). We note that due to current experimental limitations, measurement of multistep nucleotide binding kinetics for Pol II cannot be resolved. The greatly reduced k_{pol} values for incorrect nucleotide incorporation events allow the establishment of rapid equilibrium nucleotide binding. This equilibrium is the result of the faster rate of nucleotide dissociation prior to chemistry relative to incorporation. In these cases, the *apparent* K_d is an estimate of the true K_d. However, when chemistry is fast, the equilibrium binding is never established because the rate of NTP incorporation is greater than the rate of NTP release. Thus, the meaning of *apparent* K_d changes from an estimate of the true K_d to an estimate of the K_m (the ratio of product formation to substrate binding). Historically, this subtle difference in the meaning of *apparent* K_d has led to significant underestimation of the true dissociation constant for correct NTP binding. Due to the current inability to measure multistep nucleotide binding to Pol II, we have placed a greater emphasis on specificity constant measurements (which can be directly compared regardless of binding equilibrium) rather than the *apparent* K_d values, because these values are likely to report true K_d values for GTP:5fC and GTP:5caC (due to the reduction in k_{pol}) and a K_m for GTP:C and GTP:5hmC.

Increased biphasicity of product formation for 5fC and 5caC templates. We observed an increased biphasicity of product formation for 5fC and 5caC templates in contrast to C and 5hmC template (**Supplementary Table 2**). The

fraction of species producing fast (θ_1) and slow (θ_2) phases of GTP incorporation shifted due to the formyl and carboxyl substitutions. The cytosine and 5hmC template partitioned to heavily favor the fast phase of GTP incorporation.

The two incorporation phases of 5fC and 5caC may represent parallel incorporation pathways where GTP can bind and incorporate template bases either in the *anti*- or *syn*- conformations of 5fC or 5caC. The syn- conformation of 5fC and 5caC would position the formyl or carboxyl oxygens to interact with the incoming nucleotide. This model would predict different *apparent* K_d values for the fast and slow phase with predictably weaker binding for the slow phase. However, a second weak $K_{d,app}$ for the slow phase was not observed for 5fC or 5caC, requiring an alternative model.

The second (and favored) model to account for the biphasic incorporation kinetics suggests equilibrium between two states of Pol II complex (non-reactive conformation and poised conformation) that slowly interconvert. One population is poised for rapid GTP incorporation (observed as the fast phase of product formation, such as Pol II in post-translocation state); the second, slower phase represents a Pol II population that requires longer time for GTP incorporation (paused population, such as Pol II in the pre-translocation, frayed, or backtracked states). The slow phase of product formation then reflects a rate-limiting isomerization from a non-reactive conformation to the poised conformation that allows for fast GTP incorporation. The presence of 5fC and 5caC substantially

shifts the Pol II from an active population (poised for elongation) to a paused population. Further structural investigation will be required to test these models.

References

1. Johnson, K.A. *Annu Rev Biochem* **62**, 685-713 (1993).
2. Joyce, C.M. & Benkovic, S.J. *Biochemistry* **43**, 14317-24 (2004).
3. Kireeva, M.L. et al. *Mol Cell* **30**, 557-66 (2008).
4. Tsai, Y.C., Jin, Z. & Johnson, K.A. *Anal Biochem* **384**, 136-44 (2009).
5. Tsai, Y.C. & Johnson, K.A. *Biochemistry* **45**, 9675-87 (2006).
6. Kellinger, M.W. & Johnson, K.A. *Proc Natl Acad Sci U S A* **107**, 7734-9 (2010).
7. Kellinger, M.W. & Johnson, K.A. *Biochemistry* **50**, 5008-15 (2011).