A Small Posttranslocation Energy Bias Aids Nucleotide Selection in T7 RNA Polymerase Transcription

Jin Yu and George Oster

Departments of Molecular and Cell Biology, and Environmental Science, Policy and Management, University of California, Berkeley, California

Supporting Information

Table of Contents

Constructing the model

We constructed a translocation and elongation scheme using information from high-resolution structures of T7 RNAP elongation complexes (1, 2). Our assumptions in constructing the model are listed explicitly below (denoted **A1** to **A9**). In main **Figure 2** we show the kinetic scheme. For easy illustration and comparison, we have inserted molecular images from Figure 3 of (1), showing close views around the active site for generally highly populated kinetic states (**II′**, **III**, **IV** and **V**).

Structural basis

In structural studies of RNAP, the product complex with PPi bound (state **V** in main **Figure 2**) was regarded as the pre-translocation state that follows directly after the chemical step of phosphoryl transfer (1). In this structure, the O-helix is in the same 'closed' conformation as in the substrate state (**IV**). It was suggested that, upon PPi dissociation, the O-helix undergoes a 'pivoting' rotation to its 'open' configuration, so that Tyr639 at the C-terminal end of the O-helix delivers a power stroke, pushing the 3′-end of the RNA out of the active site thus driving translocation (1).

However, the above interpretation carries some debatable assumptions as (i) the opening of the O-helix is a rigid-body rotation and (ii) the O-helix opening and PPi dissociation happen at the same time—and simultaneously with translocation. Recent molecular dynamics (MD) studies of DNAP Pol I (3), however, suggest that PPi release *precedes* translocation and triggers the opening transition of the O-helix. The O-helix is bent during the opening transition, with its N-terminal end (distal to the active site) opening first; the C-terminal end (proximal to the active site) of the O-helix remains stable during the initial opening, and later moves in concert with DNA translocation (3). Accordingly, in our model, we assume (**Assumption A1**) that PPi dissociation happens *before* translocation, and immediately after PPi release, there is a pre-translocated state (**I** or **I′** in main **Figure 2**) in which the O-helix is *partially* opened.

The high-resolution structures also show two configurations of Tyr639: one with the side chain outside the active site (OUT), and the other with the side chain inside the active site (IN). The OUT and IN configurations of Tyr 639 are presented respectively in the substrate/product complex (**IV** or **V**) and the post-translocation/pre-insertion complex (**II′** or **III**). In the IN configuration, the Tyr639 side chain has its aromatic ring partially stacked with the RNA-DNA hybrid base pair at the 3′-end RNA. Essentially, when Tyr639 is IN, it is *closer* to the pre-insertion site of the incoming nucleotide than it is in the OUT configuration (see **Figure 1b**). In our scheme, we assume (**Assumption A2**) that RNAP has *four* relevant configurations before and after translocation: **I** or **I′** pre-translocation, and **II** or **II′** post-translocation. In **I** and **II**, the Tyr639 is OUT, while in **I′** and **II′**, Tyr639 is IN.

Translocation energetics

We also assume (**Assumption A3**) that the RNAP can move diffusively along the translocation path **I** ↔ **II** (see main **Figure 2**). That is, the forward and backward rates of translocation are the same when Tyr 639 is OUT, or the free energy difference between **I** and **II** ($\gamma \equiv E_{II} - E_I$) is ~ 0. Note that when sequence effects are considered during elongation the free energy difference γ becomes sequence dependent and fluctuates about zero. In general, along path **I** ↔ **II**, the ratio between forward rate k_{2+} and back rate k_{2-} is: $k_{2+}/k_{2-} = e^{-\gamma/k_BT}$.

In the pre-translocated state (**I** and **I′**), the active site is occupied by the 3′-end of the RNA. In MD simulations, the 3′-end of the RNA appears highly flexible (4). Supposedly, if the Tyr 639 side chain is to 'squeeze into' the active site (OUT → IN or **I** → **I′**), it may only require a free energy at the level of thermal fluctuation (**Assumption A4**), i.e., $\alpha = E_{I'} - E_I \sim 1\ k_B T$. Moreover, NMR studies showed that the side chain of the tyrosine can flip/vibrate at a frequency ~10^4 per second (5). Hence we also assume (**Assumption A5**) that the fluctuations of Tyr 639 between IN and OUT are very fast and approach thermal equilibrium. The ratio between *in* (from **I** → **I′**) and *out* (from **I′** → **I**) rates, ω_{1+} and ω_{1-}, is modulated by α as: $\omega_{1+}/\omega_{1-} = e^{-\alpha/k_BT}$, with $\omega_{1+} \geq 10^4\ s^{-1}$.

In contrast, in the post-translocated states (**II** and **II′**), the 3′-end of the RNA has moved out of the active site. The stacking interaction between the aromatic ring of the Tyr639 and the end bp of the RNA-DNA hybrid makes the IN configuration (**II′**) energetically more stable than the OUT one (**II**). The stabilization appears essential for keeping Tyr 639 IN most of time when the incoming nucleotide binds into the pre-insertion site (2). The *out* rate from **II′** → **II**, ω_{2-}, is thus related to the *in* rate from **II** → **II′**, ω_{2+}, by $\beta+\gamma$ ($\beta \equiv E_{II} - E_I(II'') = E_{II} - E_I(II'') + E_{II} - E_{II}$): $\omega_{2+}/\omega_{2-} = e^{(\beta+\gamma)/k_BT}$. Consequently, the translocation path **I′** → **II′** is energetically favorable, as there is an energy drop of $E_{I'} - E_{II'} = E_{I'} - E_I + E_I - E_{II'} = \alpha + \beta > 0$. Hence, we call path **I′** → **II′** 'facilitated'. Along the path, the ratio between the forward rate $k_{2'+}$ and the back rate $k_{2'-}$ is: $k_{2'+}/k_{2'-} = e^{(\alpha+\beta)/k_BT}$

Nucleotide pre-insertion

As T7 RNAP finishes the translocation step, the O-helix opens fully at the post-translocated state (**II** or **II′**). Before the incoming NTP is stably inserted (in **IV**) into the active site, there

exists a pre-insertion intermediate state (1, 2) (**III**) in which the NTP binds to a pre-insertion site adjacent to the active (insertion) site. Structural examination shows that Tyr639 is captured with IN configuration in the pre-insertion structure **III (2)**; when Tyr639 is OUT (as in **IV** or **V**), the side chain moves *farther* from the pre-insertion site (see main **Figure 1**b). Thus, the NTP pre-insertion can be achieved without steric hindrance—either the Tyr639 is IN or OUT: **II′** → **III** or **II** → **III (Assumption A6**). However, pre-insertion at IN (**II′** → **III**) is more likely due to the local stability of **II′** (i.e. $E_{II'} < E_{II} \sim E_{I}$).

Note that in multi-subunit RNAPs, configurations **II′** and **II** correspond to the post-translocated state with the bridge helix bent and straight (6-8), respectively. Our structural examination shows that the bent configuration (**II′**) of the bridge helix has steric clashes with the pre-inserted nucleotide (see main **Figure 1**b). In the pre-insertion state, the bridge helix is captured straight (9). Hence, nucleotide pre-insertion in the multi-subunit RNAP can only happen through **II** → **III** but *not* **II′** → **III**.

Overall kinetics

In the pre-insertion state, the O-helix remains open. To reach the substrate insertion state **IV**, the O-helix closes, and the transition can be quite slow (1, 2, 10). We assume (**Assumption A7**) that the O-helix closing **III** →**IV** is the rate-limiting step in the nucleotide addition cycle (at a high enough NTP concentration). Following the nucleotide insertion, polymerization takes place quickly to produce the product complex, state **V**. Subsequently, PPi release from the product complex **V** leads to the pre-translocation state (**I** or **I′**) readying the system for the next translocation step. Assuming PPi concentration is low in the vicinity of the active site (**Assumption A8**) (≤ 0.1 μM e.g.) (11), then the pyrophosphorylase reaction, (i.e., the reverse of PPi release) is very slow.

This kinetic scheme can be described mathematically using master equation and solved for the steady state (see **SI**). Equivalently, one can numerically simulate the cycles using kinetic Monte-Carlo methods (12). The rate parameters used in the model are listed in **SI Table S1**: some of them are adopted from transient state kinetic measurements (10) and some of them can be tuned and fitted with the single molecule experimental data (11, 13). When forward tracking (or back tracking in multi-subunit RNAP) is considered (see main **Figure 1**c), we further assume (**Assumption A9**) that the forward tracking (or back tracking) proceeds via configuration **II′** (or **I′**) (35).

Using master equation approach

Following the kinetic scheme in **Figure 2** in main, we define the probability distribution of intermediate states **I**, **I′**, **II′**, **II**, **III**, **IV**, and **V** as $\vec{\Pi} = (P_I \; P_I' \; P_{II}' \; P_{II} \; P_{III} \; P_{IV} \; P_V)^T$. Hence, the master equation describing the kinetics can be written in a matrix form as:

$$\frac{d}{dt}\vec{\Pi} = M\,\vec{\Pi} \qquad \text{(S1)}$$

where M is a 7x7 transition matrix

$$M=\begin{pmatrix} -k_{1-}-k_{2+}-\omega_{1+} & \omega_{1-} & 0 & k_{2-} & 0 & 0 & k_{1+} \\ \omega_{1+} & -\omega_{1-}-k_{2'+} & k_{2'-} & 0 & 0 & 0 & 0 \\ 0 & k_{2'+} & -k_{2'-}-\omega_{2-}-k_{3'+} & \omega_{2+} & k_{3'-} & 0 & 0 \\ k_{2+} & 0 & \omega_{2-} & -k_{2-}-\omega_{2+}-k_{3+} & k_{3-} & 0 & 0 \\ 0 & 0 & k_{3'+} & k_{3+} & -k_{3'-}-k_{3-}-k_{4+} & k_{4-} & 0 \\ 0 & 0 & 0 & 0 & k_{4+} & -k_{4-}-k_{5+} & k_{5-} \\ k_{1-} & 0 & 0 & 0 & 0 & k_{5+} & -k_{5-}-k_{1+} \end{pmatrix}$$

In particular, the forward and backward translocation rates along path **I**⇔ **II** and **I′**⇔ **II′**

are described as:

$k_{2+} = r_0$ $k_{2-} = r_0 e^{\gamma}$ EMBED Equation.3 $k_{2'+} = r_0 e^{\alpha}$ EMBED Equation.3 $k_{2'-} = r_0 e^{-\beta}$ EMBED Equation.3

. (S2)

The IN and OUT rates of the Tyr 639 side chain at pre- and post-translocated states are:

$\omega_{1+} = \omega_0$ $\omega_{1-} = \omega_0 e^{\alpha}$ EMBED Equation.3 $\omega_{2+} = \omega_0 e^{\gamma}$ EMBED Equation.3 $\omega_{2-} = \omega_0 e^{-\beta}$ EMBED Equation.3

. (S3)

As defined in the main text, $\alpha = E_{I'} - E_I$, $\beta = E_I - E_{II'}$, $\gamma = E_{II} - E_I$ are free energy differences (unit: k_BT) among pre- and post-translocated states **I**, **I′**, **II′**, and **II**. The above convention ensures that the transition rates follow detailed balance at thermal equilibrium.

At the steady state (non-equilibrium in general), the probability distribution $\Pi^s = (P_I^s \; P_{I'}^s \; P_{II'}^s \; P_{II}^s \; P_{III}^s \; P_{IV}^s \; P_V^s)^T$ satisfies:

$$M\,\Pi^s = \mathbf{0} \quad . \tag{S4}$$

Hence, one obtains the solution Π^s. The elongation rate v (nt/s) is proportional to the steady-state probability flux J

$$v = l_0 J \quad , \tag{S5}$$

with l_0 the periodic length 1-nt, and

$J \equiv P_V^s k_{1+} - P_I^s k_{1-} = P_{IV}^s k_{5+} - P_V^s k_{5-} = P_{III}^s k_{4+} - P_{IV}^s k_{4-} = P_{II}^s k_{3+} - P_{III}^s k_{3-} + P_{II}^{s'} k_{3'+} - P_{III}^s k_{3'-} = P_I^s k_{2+} - P_{II}^s k_{2-} +$

.

Corresponding to the master equation description, one can generate trajectories of elongation by running kinetic Monte-Carlo simulations (12). The steady state can be quickly reached, and the simulation make it easy to monitor a variety of properties, such as error rates, forward tracking events, etc..

Deriving elongation rates for two ratchet schemes

Below we derive approximate solutions of elongation rates for T7 and multi-subunit RNAPs, using two slightly different ratchet schemes shown in **Figure S1**. To focus on rate dependence on translocation energetics, we simplify the elongation scheme (see main **Figure 2**), leaving only pre-translocated states (**I** and **I′**), post-translocated states (**II** and **II′**), and a NTP loaded state (**III***). Transition **III*** to **I** represents a convolution of NTP insertion (slow), phosphoryl transfer, and PPi release (nearly irreversible at low [PPi]), hence, we assume **III*** $\rightarrow$ **I** slow (rate-limiting, see **A7**) and irreversible ($k^*_{1+} = k_{cat}$ and $k^*_{1-} \approx 0$; see **A8**). We also assume Ty639 side chain fluctuates much faster (see **A5**) than the RNAP translocation rates, so that **I** and **I′** or **II** and **II′** are close to equilibrium ($P_{I'} \approx e^{-\alpha/k_BT} P_I$ and $P_{II'} \approx e^{(\beta+\gamma)/k_BT} P_{II}$).

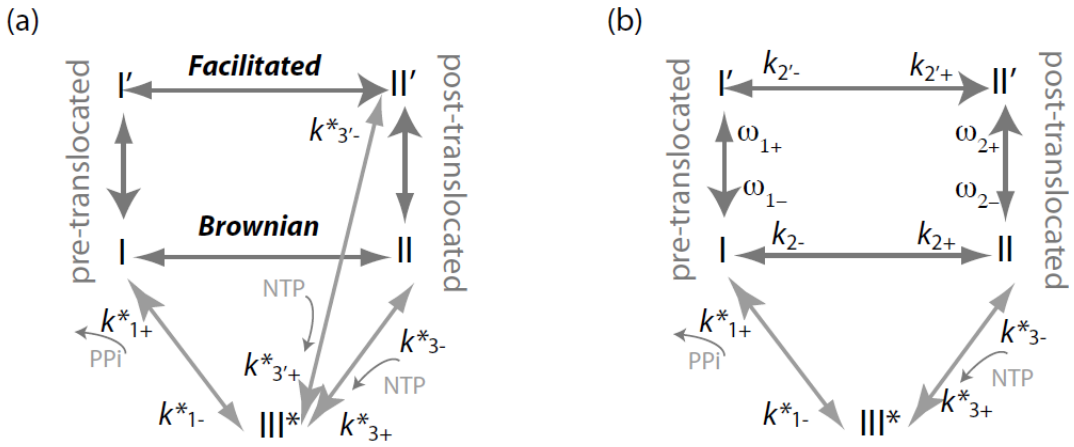

Figure S1 Ratcheting schemes with two parallel translocation paths in RNAP transcription elongation. After NTP loading, a generic 'catalytic' transition **III***$\rightarrow$**I** happens, which is slow and nearly irreversible. (a) In T7 RNAP, NTP can bind either **II′** (Tyr639 IN) or **II** (Tyr639 OUT) to **III***. (b) For a multi-subunit RNAP, NTP binds only at **II** (bridge helix straight) but not **II′** (bridge helix bent).

For T7 RNAP (**Figure S1a**), NTP binds at either **II′** (Tyr639 IN) or **II** (OUT), i.e., through either **II′**$\rightarrow$ **III*** or **II**$\rightarrow$ **III*** (see **A6**). Since the binding rate is likely diffusion-limited, $k^*_{3+} = k^*_{3'+} = k^0_{NTP} \cdot [NTP]$. For unbinding, $k^*_{3-} = k^*_{3'-} e^{-(\beta+\gamma)/k_BT}$, as $E_{II} - E_{II'} = \beta + \gamma$. Solving the master equation of the simplified scheme, we obtained $P_I \approx e^{\gamma/k_BT} P_{II}$ and $P_{II} \approx \dfrac{k^*_{3-} + k^*_{3'-} + k^*_{1+}}{k^*_{3+} + k^*_{3'+} e^{(\beta+\gamma)/k_BT}} P_{III^*}$, so that

$v = l_0 k_{cat}[NTP] / ((1 + e^{\gamma/k_BT} + e^{(\gamma-\alpha)/k_BT} + e^{(\gamma+\beta)/k_BT})(k^*_{3-} + k^*_{3'-} + k_{cat}) / (k^0_{NTP} + k^0_{NTP} e^{(\beta+\gamma)/k_BT}) + [NTP])$.
Using $K'_D \equiv k^*_{3'-} / k^0_{NTP}$, and $K_D = K'_D [1 + e^{-2(\beta+\gamma)/k_BT}] / [1 + e^{-(\beta+\gamma)/k_BT}]$ (average of dissociation **III*** $\rightarrow$ **II′** and **III*** $\rightarrow$ **II**), one obtains:

$$v = \frac{l_0 k_{cat}[NTP]}{(1 + \frac{e^{\gamma/k_BT} + e^{(\gamma-\alpha)/k_BT}}{1 + e^{(\gamma+\beta)/k_BT}})[\frac{k_{cat}}{k^0_{NTP}} + K_D \frac{(1 + e^{-(\gamma+\beta)/k_BT})^2}{1 + (e^{-(\gamma+\beta)/k_BT})^2}] + [NTP]} \quad \text{(S6)}$$

where l_0 = 1 nucleotide (nt) is the periodic length of translocation, k_{cat} is the effective maximal rate of 'catalysis' (~ 130 s^{-1} for T7 RNAP (11)), k^0_{NTP} is the NTP binding /pre-insertion rate (fitted to ~ 2 μM^{-1} s^{-1}, see section below) , and K_D is the dissociation constant at the NTP binding/pre-insertion site, measured at ~ 80 μM (10). The translocation energy associated with Brownian motions is defined $\gamma \equiv E_{II} - E_I \sim 0$ with fluctuations around zero caused by DNA sequence effects. Since the 3′-end of the RNA is quite flexible (4), we assume that $\alpha \equiv E_{I'} - E_I \sim 1\ k_BT$ (thermal fluctuation level), so that it is easy to 'squeeze' Tyr639 IN when the active site is still occupied by the 3′-end of the RNAP at pre-translocation. The key parameter is the *post-translocation free energy bias* $\beta \equiv E_I - E_{II'} \sim E_{II} - E_{II'} > 0$, which is fitted to 2 ± 1 k_BT (see section below). Note that when DNA sequence effects are considered, γ and -β are affected identically while β + γ (=$E_{II} - E_{II'}$), the post-translocation free energy bias (γ≠0), still keeps sequence independent. Accordingly, when the load force is applied to the RNAP as in the single molecule experiments (11, 13), the force increases γ and −β identically, while β + γ is force-independent.

For the multi-subunit RNAP (**Figure S1b**), NTP binds without steric hindrance only at **II** (bridge helix straight), i.e., through **II**$\rightarrow$ **III*** but not **II′**$\rightarrow$ **III***. We obtained $P_{II} \approx \frac{k^*_{3-} + k^*_{1+}}{k^*_{3+}} P_{III^*}$, so that
$v = l_0 k_{cat}[NTP] / ((1 + e^{\gamma/k_BT} + e^{(\gamma-\alpha)/k_BT} + e^{(\gamma+\beta)/k_BT})(k^*_{3-} + k_{cat}) / k^0_{NTP} + [NTP])$. With $K_D \equiv k^*_{3-} / k^0_{NTP}$ one obtains:

$$v = \frac{l_0 k_{cat}[NTP]}{(1 + e^{\gamma/k_BT} + e^{(\gamma-\alpha)/k_BT} + e^{(\gamma+\beta)/k_BT})(\frac{k_{cat}}{k^0_{NTP}} + K_D) + [NTP]} \quad \text{(S7)}$$

Note that the above calculation has not considered pauses or back-tracking pathways for the multi-subunit RNAP during elongation.

If we cast the formula in the form of a Michaelis-Menten expression:
$v = v_{\max}[NTP] / (K_M + [NTP])$ we see that, even though the translocation is not rate limiting in the elongation cycle, its local energetics can affect the elongation rate through the 'apparent Michaelis constant' K_M. For T7 RNAP, we see in main **Figure 3**c that increasing the post-

translocation free energy bias β beyond 2 k_BT does not improve the elongation rate much (e.g. < 5% increase) as the rate is already close to its saturation value (as $\beta + \gamma \to +\infty$): $v^s \to l_0 k_{cat}[NTP] \Big/ (\frac{k_{cat}}{k_{ET}} + K_D + [NTP])$. On the other hand, in comparison to a pure Brownian ratchet ($\gamma = 0$, $\alpha \to +\infty$, and $\beta \to \infty$, i.e., path **I** ↔**II** only), the 3 k_BT free energy bias ($\alpha + \beta$) along **I′** → **II′**, or an overall ~ 1.3 k_BT for both paths, does improve the elongation rate somewhat (e.g. ~ 14% increase).

For multi-subunit RNAPs, however, it is likely that β or $\beta + \gamma < 0$; the elongation rate still stays close to its saturation value (as $\beta + \gamma \to -\infty$): $v^s = l_0 k_{cat}[NTP] / ((1 + e^{\gamma/k_BT} + e^{(\gamma-\alpha)/k_BT})(\frac{k_{cat}}{k_{ET}} + K_D) + [NTP])$. When the bent configuration (**II′**) of the bridge helix is stabilized by some inhibitory factor relative to the straight conformation (**II**) such that $\beta + \gamma > 0$ ($E_{II} > E_{II'}$), the elongation rate can be significantly reduced, due to the exponential term $e^{(\beta+\gamma)/k_BT}$ in the apparent K_M (see Eq. S7).

Setting and tuning parameters

Parameter	Definition	Default value	Notes	Extras
r_0	RNAP translocation rate from **I** to **II** (see Eq. S2)	5000 s^{-1}	> 10^3 s^{-1}; translocation non-rate-limiting	Elongation rate v insensitive to r_0
ω_0	Oscillation rate for Tyr 639 side chain from **I** to **I′** (see Eq. S3)	50000 s^{-1} or even larger	Tyrosine side-chain flipping rate on the order ~ 10^4 s^{-1}(5)	v insensitive to ω_0; very low ω_0 can induce pauses though
α	Free energy for Y639 side chain move IN when the active site is still occupied with 3′ RNA, $\alpha \equiv E_{I'} - E_I > 0$	1 k_BT	Assumption **A4**: a small amount at thermal fluctuation level	v insensitive to α
β	Free energy difference $\beta \equiv E_I - E_{II}'$; called 'post-translocation bias' as $\gamma \sim 0$	2 k_BT	The most essential tuning parameter, tuned for apparent K_M in current scheme (see text)	v *is* sensitive to β; ranged between 1.4 to 2.5 k_BT as tuning average v within ±5 nt/s (as k^0_{NTP} =2 $\mu M^{-1}s^{-1}$)
γ	Translocation free energy (under diffusion)	0	Assumption **A3**: Brownian motion along path I ⇔II	non-zero and fluctuating under DNA sequence

	$\gamma \equiv E_{II} - E_I$			effects
k^0_{NTP}	NTP binding rate constant: $k_{3+} = k_{3'+} = k^0_{NTP} \cdot [N$	2 $\mu M^{-1}s^{-1}$	1~10 $\mu M^{-1}s^{-1}$ for some known NTPases; tuned for apparent K_M in the single –path scheme (see text)	v *is* sensitive to k^0_{NTP}; ranged at 1.4 ~ 2.5 $\mu M^{-1}s^{-1}$ as tuning average v within ±5 nt/s
$k_{3'-}$	NTP unbinding rate via **III** → **II′**	178 s^{-1}	$K'_D \equiv \frac{k_{3'-}}{k^0_{NTP}}$ tuned to $\sim 89\ \mu M$ by β	$K'_D \sim K_D \frac{1+e^{-\beta-\gamma}}{1+(e^{-\beta-\gamma})^2}$ where $K_D = 80\ \mu M$ (10)
k_{3-}	NTP unbinding rate via **III** → **II** $k_{3-} = e^{-(\beta+\gamma)} k_{3'-}$	24 s^{-1}	$k_{3-} = k_{3'} e^{-\beta-\gamma}$ since $E_{II} - E_{II'} = \beta + \gamma$	Both $k_{3'-}$ and k_{3-} depend on β
k_{4+}	NTP insertion / O-helix closing rate	220 s^{-1}	Measured in (10) and assumed rate-limiting (see Assumption **A6**)	v_{max} *is* sensitive to k_{4+}
k_{4-}	Reverse rate for NTP insertion	210 s^{-1}	Choose below 220 s^{-1}; tuned for v_{max}	Constrained by v_{max} as well as total free energy (see Eq. S9)
k_{5+}	Catalytic rate of polymerization reaction	1000 s^{-1}	Choose ~ 10^3 s^{-1}, fast; tuned for v_{max}	Constrained by v_{max}
k_{5-}	Reverse rate for polymerization reaction	135 s^{-1}	$\ln\left(\frac{k_{3+}}{k_{3-}}\right)$ ~ 2 k_BT likely small (e.g. ~ 1 k_BT for F_1 ATPase (14))	Constrained by total free energy constraint
k_{1+}	PPi release rate	1200 s^{-1}	$K_d \equiv \frac{k_{1-}}{k^0_{PPi}}$ ~ 1200 μM measured in (10)	Constrained by v_{max}
k_{1-}	Reverse (PPi binding rate $k_{1-} = k^0_{PPi} \cdot$ [PPi]	0.1 s^{-1}	Use [PPi]~ 0.1 μM ([PPi] low (11))	Set k^0_{PPi} ~ 1 $\mu M^{-1}s^{-1}$, the same order as k^0_{NTP}

Table S1 Parameter setting for current elongation scheme of T7 RNAP.

The above table lists all parameters we used for T7 RNAP elongation-translocation. The maximum elongation rate v_{max} had been fitted to ~ 130 nt/s from experiments(11). Parameters $k_{4\pm}$ $k_{5\pm}$ were tuned accordingly together under the constraints of v_{max} and free

energy consumption (see below Eq. S9), and were kept consistent with the rate-limiting assumption (**A6**). These parameters are not quite relevant to the translocation part. Parameter k^0_{NTP} was first estimated according to the three-state (single translocation path) scheme from (11). Note that the experimentally fitted elongation rate can be written as:

$$v \approx \frac{l_0 k_{cat}[NTP]}{(1+e^{r/k_BT})(\frac{k_{cat}}{k^0_{NTP}}+K_D)+[NTP]} \qquad \text{(S8)}$$

where $l_0 k_{cat} = v_{max}$. The dissociation constant of NTP at the binding/ pre-insertion site K_D was later on measured as ~ 80 μM (10). Next, β along with K'_D were tuned in current scheme to fit the experimental results (11). Indeed, we estimate K'_D, the dissociation constant through path **III** → **II′** by: $K'_D \sim K_D \frac{1+e^{-(\beta+\gamma)/k_BT}}{1+(e^{-(\beta+\gamma)/k_BT})^2}$. We regard the measured K_D as the average of that through path **III** → **II′** (K'_D with a 'weight' 1) and that through path **III** → **II** ($K'_D e^{-(\beta+\gamma)}$ EMBED Equation.3 , with a 'weight' $e^{-(\beta+\gamma)}$ EMBED Equation.3) in T7 RNAP. Since K_D had been measured experimentally (10), K'_D can be determined by β. It turns out that β is the most essential tuning parameter in current model. β was tuned 1.4 ~2.5 k_BT when k^0_{NTP} is set at 2 $\mu M^{-1}s^{-1}$ (if we allow the average rate vary within 5 nt/s), hence, we use $\beta \sim 2$ k_BT. Varying k^0_{NTP} (1.4 ~ 2.5 $\mu M^{-1}s^{-1}$, determined using the three-state single-path scheme) allows β adopt values from about 1 k_BT to a little over 3 k_BT. Hence we present current estimation of β as 2±1 k_BT.

Note that in the single molecule experiments, the standard errors of the mean values of the transcription rates are largely within 5 nt/s (~85% data points (13)). With a 5 nt/s rate variation (as mentioned above), we can fit β to 1.4 ~ 2.5 k_BT, so that the overall or average translocation energy bias (simply estimated as (α+β)/2) is about 1 ~ 2 k_BT. Even if the standard errors of the rates rise close to 10 nt/s (15% data points (13)), the overall translocation energy bias can still be tuned in between 0 ~ 3 k_BT, showing that the energy bias exists but is small according to the experimental data.

The standard free energy consumption for a NTP addition cycle, as estimated from (10, 15), sets an important constraint for above parameters:

$$\Delta G_0 = k_BT(\ln\frac{k^0_{NTP}}{k_{3-}} + \ln\frac{k_{4+}}{k_{4-}} + \ln\frac{k_{5+}}{k_{5-}} + \ln\frac{k_{1+}}{k^0_{PPI}}) \approx 4 \sim 7\ k_BT \qquad \text{(S9)}$$

With above default values of the parameters, $\Delta G_0 \sim 6.6\ k_BT$. At a relatively high concentration of NTP (588 μM) and a low concentration of PPi (0.1 μM), the free energy drop for each step

is, for example, NTP pre-insertion (**II**→**III**): $\Delta G^{NTP} = k_B T \ln \frac{k^0_{NTP}[NTP]}{k_{3-}} \sim 3.9\ k_B T$ (note that for **II′**→**III**, ΔG^{NTP} is 2 k_BT less as that was spent in translocation); NTP insertion (**III**→**IV**): $\Delta G^{insert} = k_B T \ln \frac{k_{4+}}{k_{4-}} \sim 0.05\ k_B T$ (almost no free energy change); the polymerization /catalysis (**IV**→**V**): $\Delta G^{catalysis} = k_B T \ln \frac{k_{5+}}{k_{5-}} \sim 2\ k_B T$; PPi release (**V**→**I**): $\Delta G^{PPI} = k_B T \ln \frac{k_{1+}}{k^0_{PPI}[PPi]} \sim 9.4\ k_B T$. Hence, $\Delta G_{total} = \Delta G_0 + \mathrm{k}_B T \ln \frac{[NTP]}{[PPi]} \sim 15\ \mathrm{k}_B T$ at the default NTP concentration.

Comparing nucleotide selection strategies

In the nucleotide addition cycle, the nucleotide selection starts at the pre-insertion state (**III**) (2). Upon pre-insertion, a wrong nucleotide is more likely to be rejected than a correct one. We call this selection method #1: the wrong nucleotide has a larger *un*binding rate, k_{3-} or $k_{3'-}$, than the correct nucleotide (see **Table S2**). Following pre-insertion, insertion of the nucleotide (**III** → **IV**) takes place slowly as the O-helix closes. If a wrong nucleotide is not rejected upon pre-insertion, it may further slow down the O-helix closing (selection method #2: decreased k_{4+}) or lead to a less stabilized insertion state (selection method #3: increased k_{4-}). Moreover, inserting a wrong nucleotide into the active site may prevent an appropriate configuration for phosphoryl transfer at the RNA 3′-end. Thus the chemical transition (**IV** →**V**) can be slowed (selection method #4: decreased k_{5+}). Some or all of these mechanisms may contribute to the nucleotide selection.

Here we examine which kinetic steps in the elongation scheme (see **Figure 2**) are more efficient than others in selecting correct nucleotides. We define quantities $\eta_i^- \equiv \frac{k^w_{i-}}{k^c_{i-}}$ (i = 3, 4) and $\eta_i^+ \equiv \frac{k^w_{i+}}{k^c_{i+}}$ (i = 4, 5) for individual selection mechanisms, with 'w/c' labeling the rate for the wrong/correct nucleotide. When the wrong nucleotide destabilizes the pre-insertion (**III**) or the substrate insertion state (**IV**), it gets an enhanced rate of rejection from the state comparing to that of the correct nucleotide, so that η_3^- or η_4^- becomes larger than 1. On the other hand, when the wrong nucleotide has a reduced rate of the insertion or polymerization, η_4^+ or η_5^+ becomes smaller than 1. In **Table S2**, we list error rates calculated from simulations for different selection mechanisms, i.e., using different sets and values of $\eta_i^- > 1$ and/or $\eta_i^+ < 1$. Equal concentrations of four nucleotides are considered in the simulation. We show individual error rates at the pre-insertion, insertion, and the product state (p^{III}_{error}, p^{IV}_{error}, p^{V}_{error}), with the final error rate $p_{error} = p^{V}_{error}$. Originally, when there exists no nucleotide selection for the RNAP (case #**0**), all $\eta_i^{\pm}$ are equal to 1. As for each cycle only one of four nucleotides matches the template, $p_{error} \sim 75\%$.

For each case (#**1** to **8)**, we choose η_i^- = 10 to 100 and/or η_i^+= 0.1 to 0.01. η_i^-=10 means that for a backward transition starting from state *i*, the wrong nucleotide is less stabilized at state *i* than the correct one, or faces a lower backward activation barrier than the correct one (by $\Delta\Delta G^0 \cdot \ln 10$ ~ 2.3 k_BT). While η_i^+=0.1 means that for a forward transition toward state *i*, the wrong nucleotide is more stabilized at *i*-1 (the state prior to *i*) or faces a higher forward activation barrier than the correct one (also by $\Delta\Delta G^0$).

First we compare cases #**1** to **4.** The individual selection mechanisms use the same amount of energy in selection ($2\Delta\Delta G^0$). One sees that the selection against the wrong nucleotide at the pre-insertion (#**1**) or during its insertion (#**2**) gives a lower (one tenth) error rate, i.e., more efficient, than the other two (#**3** and **4**) implemented after the nucleotide being inserted (#**3** and 4). Also, case #**1** allows a higher elongation rate than case #**2** to **4**.

Index	$\eta_i^- \equiv \frac{k_{i-}^w}{k_{i-}^c}$ $\eta_i^+ \equiv \frac{k_{i+}^w}{k_{i+}^c}$	Elongation rate (nt/s)	p_{error}^{III}	p_{error}^{IV}	p_{error}^{V}	Order of error rate
0	$\eta_i^{\pm}=1$	124	75%	74%	74%	10^{-1}~10^{0}
1	$\eta_3^-=100$	106	5.1%	5.2%	5.4 %	10^{-2}
2	$\eta_4^+=0.01$	27	84%	5.0%	5.2%	10^{-2}
3	$\eta_4^-=100$	34	84%	22%	23%	10^{-1}
4	$\eta_5^+=0.01$	18	84%	97%	22%	10^{-1}
5	$\eta_3^-=100$ $\eta_4^+=0.01$	100	5.2%	0.06%	0.07%	10^{-4}
6	$\eta_3^-=10$ $\eta_4^+=0.01$ $\eta_4^-=100$	78	35%	0.03%	0.01%	10^{-4}
7	$\eta_3^-=100$ $\eta_4^+=0.1$ $\eta_4^-=10$ $\eta_5^+=0.1$	99	5.1%	0.4%	0.04%	10^{-4}
8	$\eta_{3\to2}^-=10$ $\eta_{3\to2'}^-=100$ $\eta_4^+=0.01$	99	5.7%	0.05%	0.05%	10^{-4}

Table S2 Error rates of elongation simulated under different nucleotide selection mechanisms.

In case #**5** we show an efficient two-step selection mechanism that leads to an error rate ~10^{-4} as experimentally measured (16). In this case a high elongation rate is also maintained. Other cases with the same error rates either consume more energy (#**6** and **7**, i.e., less efficient) or have a low elongation rate (#**6**, when $\eta_3^- < 100$).

In case #**8** we consider that the 'strength' of the selection upon pre-insertion (**III**) varies depending on which states (**II** or **II′**) the wrong nucleotide being rejected to. If the nucleotide is rejected to **II** without Tyr 639 participation (or OUT), the selection can be weak, e.g., $\eta_{3\to2}^-=10$. If the rejection involves Tyr 639 'sensing' (**III** → **II′**), the rejection becomes

stronger, e.g., $\eta^{-}_{3\to 2'} = 100$. Note that in cases #**1** to **7**, however, we use $\eta^{-}_{3\to 2} = \eta^{-}_{3\to 2'} = \eta^{-}_{3}$, without considering this specific selection from Tyr 639.

Simulating sequence-dependent translocation

For each step of RNAP translocation, there are three sources of free energy changes that are sequence dependent: (a) Unwinding 1-bp DNA downstream of the transcription bubble, and rewinding 1-bp DNA upstream; (b) Unzipping 1-bp RNA-DNA hybrid upstream as 1-nt RNA transcript is released from the RNAP, and moving of 1-nt template DNA (unpaired, downstream) toward the active site adjacent to the 3′-end of the RNA; (c) Constant folding and unfolding (secondary and tertiary changes) of the RNA transcript. The RNAP-DNA/RNA interactions are assumed independent of sequences.
In current implementation, we only examine sequence effects of translocation at terminator right *after* formation of an RNA hairpin or loop (assuming that the elongation complex has not been perturbed during the hairpin/loop formation process). Hence, we ignore part (c) as the RNA hairpin or loop just forms and would not change shortly.

Below, we show calculations of translocation energetics including part (a) and (b) at two terminator sequences: T-Φ (17) and a threonine attenuator (pTZ19thr) (18). Both terminators are characterized by a stretch of consecutive T residues, and RNA transcripts ahead of the corresponding U-stretch can form a stem-and-loop or hairpin structure. The size of the RNAP is estimated at 20-nt length along DNA. The RNA-DNA hybrid is of 8-nt length. The ssDNA regions upstream and downstream of the RNA-DNA hybrid are estimated as 2-nt and 1-nt, respectively.

(a) The free energy for unwinding or rewinding of DNA is calculated from mfold (19) (at T=27 °C, and $[Na^+]$ = 1M), taking into account both base-pairing and nearest-neighbor stacking effects. Assuming that the free energy for stabilizing two neighboring bps is E^i_{DNA} at position *i*, then rewinding 1-bp (at position *i*) and unwinding 1-bp downstream (at position *i* +12, due to the ssDNA regions and the 8-nt RNA-DNA hybrid) result in a free energy change:

$$\Delta G^i_{DNA} = E^i_{DNA} - E^{i+12}_{DNA} \qquad \text{(S10)}$$

(b) The free energy for unzipping RNA-DNA hybrid is calculated using parameters from (20) (Table 3 in (20) at T=27 °C, and $[Na^+]$ = 1M). Introducing 1-nt unpaired nucleotide on the template DNA adjacent to the 3′-end of the RNA brings about some stacking stabilization, which we estimate as half of the stabilizing free energy for an RNA-DNA bp as in (21). When the upstream DNA rewinding is at position *i*, the RNA-DNA hybrid unzipping happens at position *i*+3 (2-nt ssDNA region in between), with free energy cost $-E^{i+3}_{hybrid}$; while the 1-nt DNA nucleotide adjacent to the 3′-end of the RNA is at position *i*+10. The free energy change is:

$$\Delta G^i_{hybrid} = \frac{1}{2} E^{i+10}_{hybrid} - E^{i+3}_{hybrid} \qquad \text{(S11)}$$

Therefore, taking into account of sequence effects for each step of translocation, the free energy change $\gamma \equiv E_{II} - E_I$ along path **I** → **II** includes the sum of the two free energy terms above: $\gamma = \Delta G^i_{DNA} + \Delta G^i_{hybrid}$.

At the same time, we also consider the facilitated translocation path **I′** → **II′** (see **Figure 2** in main text). Comparing to path **I** → **II**, there is an additional stabilizing effect at post-translocation, due to Tyr 639 side chain partially stacking with the RNA-DNA hybrid bp adjacent to the active site. In current implementation, we estimate the stabilization energy as if the Tyr 639 side ring mimics C (cytosine)- base of the RNA. Accordingly, we count the stacking interaction the half of free energy for an RND-DNA hybrid (CG) bp at this position $E'_{II} - E_{II} = \frac{1}{2} E^{i+10}_{CG}$. Since $-\beta \equiv E_{II'} - E_I = \gamma + \frac{1}{2} E^{i+10}_{CG}$ and $E^{i+10}_{CG} \sim -4\ k_B T$, the estimation is consistent with our fitted results: $\beta \sim 2\ k_B T, \gamma = 0$.

In **Figure S2**, we show values of ΔG^i_{DNA}, ΔG^i_{hybrid}, γ and $-\beta$ calculated around terminator sequences of T-Φ and pTZ19thr. From the free energy diagrams, one cannot see significant energy rises at the terminator site, or any identifiable energy barrier of translocation that can destabilize the RNAP. The calculation does not seem to support a 'thermodynamic' mechanism of intrinsic termination (22), though the mechanism cannot be ruled out as current calculation does not consider the effect of RNA hairpin/loop formation.

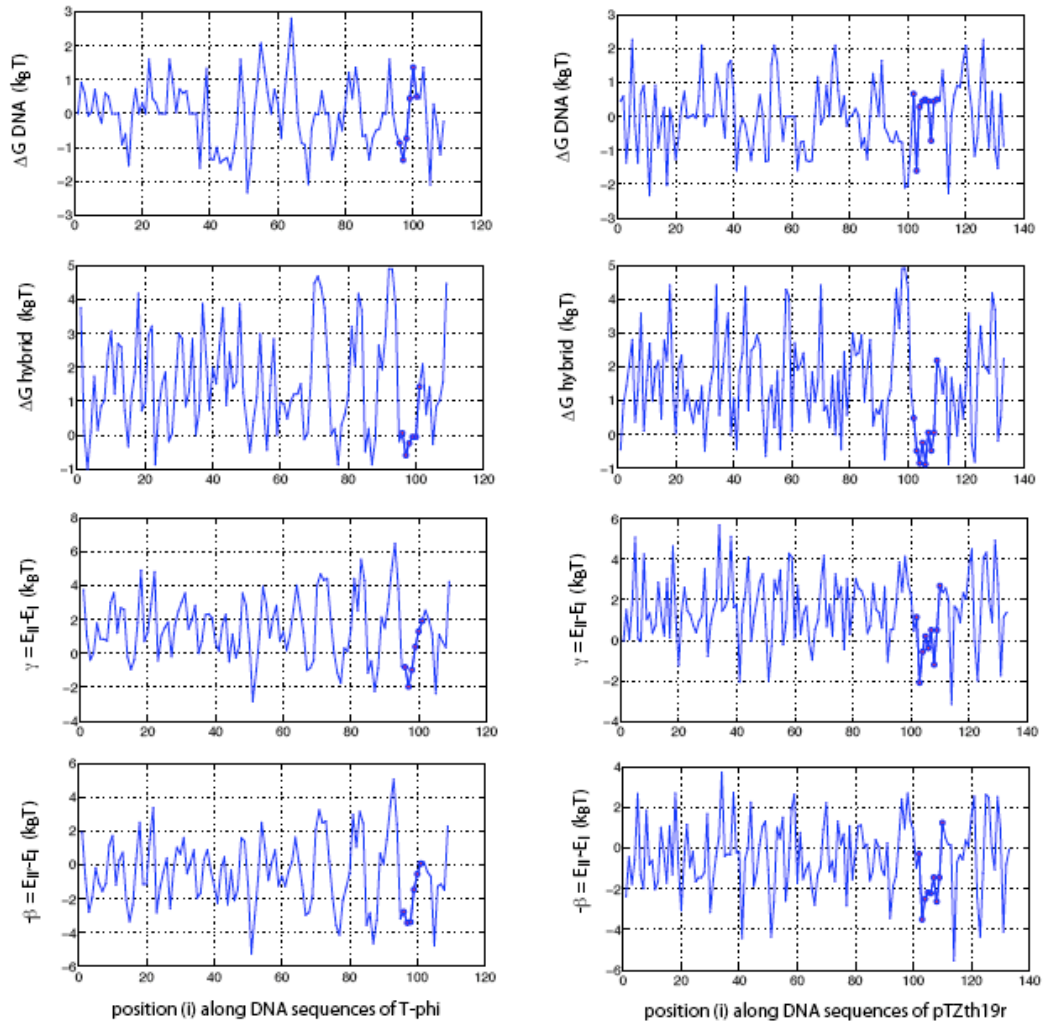

Figure S2 Free energy changes for sequence-dependent translocation of T7 RNAP along DNA. Shown are values of ΔG^{i}_{DNA}, ΔG^{i}_{hybrid}, γ and $-\beta$, calculated respectively for sequences of T-Φ and pTZ19thr (after RNA loop/hairpin formation). The positions where the RNAP reaches at T-stretch (as U-stretch RNA starts to be released from hybrid) are highlighted with circles.

The forward rate of translocation is accordingly adjusted for each step as DNA sequences vary. In the absence of the sequence effects, the forward rate (**I** → **II**) is chosen as a constant $r_0 \sim 5000\ s^{-1}$ (see **Table S1**). Under the sequence effects, the rate is adjusted to $r_0^i \sim 50000 \exp\left(-\frac{\Delta E}{k_B T}\right)$, where ΔE is an activation barrier with sequence dependences. Indeed, we found that the RNAP could forward track with high specificity at the terminator if one assumes the activation barrier comes from unzipping 1-bp of the RNA-DNA hybrid, i.e. $\Delta E = E^{i+1}_{hybrid}$. With this implementation, the RNAP can have a high efficiency of forward tracking at the terminator while maintain a low efficiency at non-terminator regions.

In the presence of the facilitated translocation path **I′ → II′**, the forward tracking is allowed at **II′**. In the pure Brownian ratchet case (with **I → II** only), the forward tracking is allowed at **II**. Note that the forward tracking rate is assumed to be proportional but smaller than the regular forward translocation rate r_0^f. We used an adjustable factor c_f (< 1) to modulate the forward tracking rate as $c_f r_0^f$. For example, in simulation at T-Φ, $c_f = 0.1$ gives ~80% forward tracking efficiency at the terminator. In simulation at threonine attenuator (pTZ19thr), we tuned $c_f = 0.035$ that gives 40~50% efficiency of forward tracking, matching the measured termination efficiency (18). The calculation is consistent with the idea that forward tracking can lead to intrinsic termination. Nevertheless, further studies are needed to examine the exact physical mechanism of the intrinsic termination.

Supporting References

1. Yin, Y. W., and T. A. Steitz. 2004. The Structural Mechanism of Translocation and Helicase Activity in T7 RNA Polymerase. Cell 116:393-404.
2. Temiakov, D., V. Patlan, M. Anikin, W. T. McAllister, S. Yokoyama, and D. G. Vassylyev. 2004. Structural Basis for Substrate Selection by T7 RNA Polymerase. Cell 116:381-391.
3. Golosov, A., J. Warren, L. Beese, and M. Karplus. 2010. The Mechanism of the Translocation Step in DNA Replication by DNA Polymerase I: A Computer Simulation Analysis. Structure 18:83-93.
4. Woo, H.-J., Y. Liu, and R. Sousa. 2008. Molecular dynamics studies of the energetics of translocation in model T7 RNA polymerase elongation complexes. Proteins: Structure, Function, and Bioinformatics 73:1021-1036.
5. Rice, D. M., R. J. Wittebort, R. G. Griffin, E. Meirovitch, E. R. Stimson, Y. C. Meinwald, J. H. Freed, and H. A. Scheraga. 1981. Rotational jumps of the tyrosine side chain in crystalline enkephalin. Hydrogen-2 NMR line shapes for aromatic ring motions in solids. Journal of the American Chemical Society 103:7707-7710.
6. Gnatt, A. L., P. Cramer, J. Fu, D. A. Bushnell, and R. D. Kornberg. 2001. Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 A Resolution. Science 292:1876-1882.
7. Bar-Nahum, G., V. Epshtein, A. E. Ruckenstein, R. Rafikov, A. Mustaev, and E. Nudler. 2005. A Ratchet Mechanism of Transcription Elongation and Its Control. Cell 120:183-193.
8. Brueckner, F., J. Ortiz, and P. Cramer. 2009. A movie of the RNA polymerase nucleotide addition cycle. Current Opinion in Structural Biology 19:294-299.
9. Kettenberger, H., K.-J. Armacher, and P. Cramer. 2004. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Molecular Cell 16:955-965.
10. Anand, V. S., and S. S. Patel. 2006. Transient State Kinetics of Transcription Elongation by T7 RNA Polymerase. The jouornal of biological chemistry 281:35677-35685.
11. Thomen, P., P. J. Lopez, and F. Heslot. 2005. Unravelling the Mechanism of RNA-Polymerase Forward Motion by Using Mechanical Force. Physical Review Letters 94:128102.

12. Gillespie, D. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics 22:403-434.
13. Thomen, P., P. J. Lopez, U. Bockelmann, J. Guillerez, M. Dreyfus, and F. Heslot. 2008. T7 RNA Polymerase Studied by Force Measurements Varying Cofactor Concentration. Biophysical Journal 95:2423-2433.
14. Senior, A. 1992. Catalytic sites of Escherichia coli F1-ATPase. Journal of Bioenergetics and Biomembranes 24:479-483.
15. Erie, D. A., T. D. Yager, and P. H. von Hippel. 1992. The Single-Nucleotide Addition Cycle in Transcription: a Biophysical and Biochemical Perspective. Annual Review of Biophysics and Biomolecular Structure 21:379-415.
16. Huang, J., L. G. Brieba, and R. Sousa. 2000. Misincorporation by Wild-Type and Mutant T7 RNA Polymerases: Identification of Interactions That Reduce Misincorporation Rates by Stabilizing the Catalytically Incompetent Open Conformation. Biochemistry 39:11571-11580.
17. Dunn, J. J., and F. W. Studier. 1983. Complete Nucleotide Sequence of Bacteriophage T7 DNA and the Locations of T7 Genetic Elements. Journal of Molecular Biology 166:477-535.
18. Jeng, S. T., J. F. Garder, and R. I. Gumport. 1992. Transcription Termination in Vitro by Bacteriophage T7 RNA Polymerase. The Journal of Biological Chemistry 267:19306-19312.
19. Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research 31:3406-3415.
20. Sugimoto, N., S. Nakano, M. Katoh, A. Matsumura, H. Nakamuta, T. Ohmichi, M. Yoneyama, and M. Sasaki. 1995. Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes. Biochemistry 34:11211-11216.
21. Bai, L., A. Shundrovsky, and M. D. Wang. 2004. Sequence-dependent Kinetic Model for Transcription Elongation by RNA Polymerase. Journal of Molecular Biology 344:335-349.
22. Yager, T. D., and P. H. Von Hippel. 1991. A thermodynamic analysis of RNA transcript elongation and termination in Escherichia coli. Biochemistry 30:1097-1118.