A Small Post-Translocation 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

ABSTRACT The RNA polymerase (RNAP) of bacteriophage T7 is a single subunit enzyme that can transcribe DNA to RNA in the absence of additional protein factors. In this work, we present a model of T7 RNAP translocation during elongation. Based on structural information and experimental data from single-molecule force measurements, we show that a small component of facilitated translocation or power stroke coexists with the Brownian-ratchet-driven motions, and plays a crucial role in nucleotide selection at pre-insertion. The facilitated translocation is carried out by the conserved Tyr639 that moves its side chain into the active site, pushing aside the 3’-end of the RNA, and forming a locally stabilized post-translocation intermediate. Pre-insertion of an incoming nucleotide into this stabilized intermediate state ensures that Tyr639 closely participates in selecting correct nucleotides. A similar translocation mechanism has been suggested for multi-subunit RNAPs involving the bridge-helix bending. Nevertheless, the bent bridge-helix sterically prohibits nucleotide binding in the post-translocation intermediate analog; moreover, the analog is not stabilized unless an inhibitory protein factor binds to the enzyme. Using our scheme, we also compared the efficiencies of different strategies for nucleotide selection, and examined effects of facilitated translocation on forward tracking.

INTRODUCTION

During gene transcription (1), RNA polymerases (RNAPs) act as molecular motors (2,3). They move processively along double-stranded (ds) DNA to synthesize a complementary RNA strand from the template DNA strand. The free energy fueling RNAP elongation comes from binding and incorporating nucleotides for RNA synthesis. RNAPs come in two molecular architectures: the single subunit RNAPs from certain viral and mitochondrial species (4), and the multi-subunit RNAPs from bacteria, eukaryotes, and archaea (5,6). The single subunit RNAPs share many biochemical properties with the more common multi-subunit RNAPs, the most essential of which is their two-magnesium ion catalysis mechanism that employs both RNA and DNA polymerases (7,8). However, there appear to be no structural or sequence similarities between the single and multi-subunit RNAPs. Instead, the single subunit RNAPs resemble, in both sequence and structure, the Family I DNA polymerases (DNAPs) (9,10).

The RNAP from bacteriophage T7 is a prototype single subunit RNAP (11–13). It can carry out all transcriptional functions, from initiation and elongation to termination, without additional protein partners. The molecular architecture of the T7 RNAP resembles the hand-like configuration of DNAPs (13,14), with the active site for nucleotide incorporation located on the palm of the hand. Among the single subunit RNAPs and some of DNAPs, the highly conserved O-helix from the fingers subdomain abuts the active site (see Fig. 1, a and b). The O-helix, along with the fingers subdomain, moves between open and closed conformations during each nucleotide addition or elongation cycle. Due to their small size and self-sufficiency, T7 RNAPs are widely used for synthesizing specific transcripts. This RNAP makes an ideal model system to study the transcriptional process in its simplest form and it has been investigated extensively in structural (15–18) and biochemical studies (19–22), and in single-molecule measurements as well (23–26).

During transcription elongation of RNAP, each nucleotide addition cycle can be viewed as taking place in two stages: polymerization and translocation. In the polymerization stage, a new RNA 3’-end is generated from an incoming NTP by phosphoryl transfer and pyrophosphate (PPI) dissociation. In the translocation stage, RNAP moves 1-nt forward on the dsDNA, breaking 1-bp downstream and reannealing 1-bp upstream. At the same time, the 3’-end of the RNA moves along with its template DNA, vacating the active site for the next incoming NTP (see Fig. 1 c). Concurrently, the 5’-end of the RNA is released 1-nt from the DNA-RNA hybrid, resulting in a dynamically growing RNA transcript. In addition to normal elongation and translocation, multi-subunit RNAPs can track either forward or backward (see Fig. 1 c). In forward-tracking (or hypertranslocation), the RNAP hops forward without synthesizing RNA. This can shorten the DNA-RNA hybrid and may lead to termination (22,27). In back-tracking, the 3’-end RNA is extruded (opposite to that in normal translocation).
into an exit channel as the RNAP moves backward along the dsDNA (28–31). Back-tracking, however, has not been detected in T7 RNAP.

Most experimental evidence suggests that RNAPs use a Brownian-ratchet mechanism of translocation (32–34) wherein the RNAP fluctuates back and forth equally fast (Brownian) until the incoming NTP binds, preventing backward movement (the ratchet), producing a net forward motion (20,24,35,36). Interestingly, single-molecule force measurements of T7 RNAP showed that there is a slight free energy bias (~1.3 k_BT) toward the post-translocated state compared with the pre-translocated state (24). However, experiments still favor a Brownian ratchet mechanism of translocation over a power-stroke mechanism—the latter requiring a significant free energy drop to drive translocation coupled with the chemical transition (37). Controversially, high-resolution structural studies of T7 RNAP intimate a power-stroke translocation mechanism in which PPI release triggers a rotational pivoting of the O-helix from the closed to open conformation, directly driving the RNAP translocation (13,16).

In this work, we examine the translocation mechanism more closely to resolve these conflicting views. The translocation dynamics take place on a millisecond timescale, too fast to be examined thoroughly in experiments but still much too slow for atomistic molecular dynamics simulations, which are limited to nano- to microseconds. To address this difficulty, we construct a semi-phenomenological model
of RNAP elongation combining single-molecule force measurements (24,25) with information from structural studies (16,17). To accomplish this, we regard published high-resolution structures as highly populated intermediate states in the enzymatic cycle. There are, however, more sparsely populated intermediate states that are unlikely to be captured by crystallography. This viewpoint allows us to extrapolate contemporary experimental knowledge to finer timescales, and to provide guidance for further detailed structure-dynamics studies.

To focus on elongation-translocation, we dissect translocation into two parallel paths as follows. The RNAP can either translocate without a free energy bias (as in pure Brownian motion) or the translocation can be locally biased (or facilitated) via pushing by the side chain of Tyr<sup>639</sup> that resides on the C-terminal of the O-helix. Tyr<sup>639</sup> is highly conserved and is essential for selecting rNTP over dNTP (38). In building the model, we utilize the same set of high-resolution structures from which the power-stroke mechanism was proposed (16). We nevertheless assume that PPI release precedes translocation, and that the O-helix opens partially upon PPI release. These assumptions are supported by recent molecular dynamics simulation studies on DNA polymerase I that is structurally homologous to T7 RNAP (39).

Importantly, the parallel-path scheme introduces an essential degree of freedom that links the two paths (Brownian and facilitated), i.e., fluctuation of the Tyr<sup>639</sup> side chain IN and OUT of the active site. This degree of freedom is important for both translocation and transcription fidelity. Under this scheme, the facilitated path terminates in a locally stabilized post-translocated intermediate wherein Tyr<sup>639</sup> stays close to the pre-insertion site (i.e., IN). In this position, Tyr<sup>639</sup> can screen incoming nucleotides, thus ensuring transcription fidelity. Without this energy stabilization or bias, Tyr<sup>639</sup> would frequently fluctuate (i.e., OUT) far from the pre-insertion site, impairing nucleotide selection and hence transcription fidelity.

Another important property of the parallel-path scheme is that it provides a common representation of the translocation-elongation kinetics for both single and multi-subunit RNAPs. Our model of T7 RNAP translocation is similar in some respects to the two-pawl-ratchet model previously proposed for a bacterial multi-subunit RNAP (36). In Fig. 1b, we show molecular views of the essential structural elements involved in translocation/elongation for both the single subunit T7 RNAP (left) and the multi-subunit RNAP (right). In the multi-subunit RNAPs (40,41), the bridge-helix alternates between bent and straight configurations (36,42), similar to Tyr<sup>639</sup> IN and OUT, respectively. The bending of the bridge-helix also appears to assist RNAP translocation, as suggested for Tyr<sup>639</sup>. Moreover, in the multi-subunit RNAPs the trigger loop folds and unfolds during each cycle, similar to the O-helix closing and opening in T7 RNAP. Hence, the two types of polymerases can carry out analogous functions by different structural elements. We show that, under the common parallel-path scheme, the two systems show comparable—but different—reaction topologies.

Below we first calibrate the model with single-molecule force measurements and then address steady-state properties of T7 RNAP elongation. Details on how we constructed the model shown in Fig. 2 can be found in the Supporting Material. The focus of our work is to explain the functional role of facilitated translocation as the RNAP ratchets along DNA during elongation. Our model also shows how efficient nucleotide selection can be achieved during the elongation cycle, and examines whether facilitated translocation affects forward tracking. The model also suggests that the functional role of translocation is somewhat different for multi-subunit RNAPs.

### MODEL AND RESULTS

#### Fitting elongation rates with single-molecule measurements

To calibrate our model, we fit numerical simulations of our model with single-molecule force measurement data for T7 RNAP elongation (24,25). The measurements were taken at different NTP concentrations, against a mechanical force 5 ≤ F ≤ 15 pN opposing forward translocation of RNAP. The experimental data had been fitted to a simplified three-state scheme, composed of pre-translocated, post-translocated, and NTP-loaded states. The fitting gave a maximum elongation rate (v<sub>max</sub>) ~130 nt/s. It also indicated a ~1.3 k<sub>B</sub>T free energy bias during translocation to the post-translocated state (24).

Using the kinetic scheme shown in Fig. 2, we can also fit the force measurement data, as shown in Fig. 3a (see the Supporting Material for technical details). The most essential feature of the model, compared with the simpler scheme used for fitting experiments (24), is that the translocation is now split into two parallel paths. The path (I → II) with the Tyr<sup>639</sup> side chain positioned out of the active site (OUT) is assumed to be diffusive so that the RNAP can move forward and backward equally fast. The frequency of the movements is ~10<sup>3</sup> s<sup>-1</sup> so that at high NTP concentrations, translocation takes place much faster than the O-helix closing that happens at ~10<sup>2</sup> s<sup>-1</sup>. At high NTP concentrations, the O-helix closing is assumed to be the rate-limiting step. Thus, even though the load force slows down translocation, it may not much affect the overall elongation rate. Along the alternate pathway (I' → II'), however, the Tyr<sup>639</sup> side chain inserts into the active site (IN) at the beginning (I'), pushing the 3' end of the RNA out of the active site by the end (II'). Hence, Tyr<sup>639</sup> directly assists the movement of the RNA-DNA hybrid. According to our fitting, the free energy bias along I' → II' is ~3 k<sub>B</sub>T (see Table S1 in the Supporting Material: E<sub>I</sub> - E<sub>II</sub> = α + β, with α assumed ~1 k<sub>B</sub>T and
Steady-state probability distributions of elongation intermediates

We obtained the steady-state probability distributions for all intermediate states in the elongation cycle by simulating RNAP elongation for a long time (100 s for each trajectory), or equivalently by solving the master equation (see the Supporting Material). The probabilities at various NTP concentrations for these intermediate states are shown in Fig. 3b.

At high NTP concentrations (588 μM by default to compare with experimental data), the RNAP elongates fast (~100 nt/s). Under this condition the most populated state is the pre-insertion state III (P_{III} ~ 57%). Those states that are moderately populated include: the stabilized post-translocated state II' with Tyr^{639} IN (P_{II'} ~ 16%), the NTP insertion or substrate state IV (P_{IV} ~ 11%), and the product state V (P_{V} ~ 9%). Other intermediates are only marginally populated: the pre-translocated states I or I' (P_{I} ~ 3% and P_{I'} ~ 1%) and the nonstabilized post-translocated state II with the Tyr^{639} OUT (P_{II} ~ 2%). Indeed, those more populated intermediates (II', III, IV, and V in Fig. 2) had all been crystallized in the previous structural studies of T7 RNAP (16,17). This consistently shows that the captured structures are almost always those abundantly populated intermediates.

In particular, the pre-insertion state III populates most at high NTP concentrations. By contrast, the stabilized NTP insertion state IV is populated much less. This is due to the slow transition III → IV and the fast transition IV → V, such that the populations leave III slowly while passing through IV quickly. This property allows sufficient time for nucleotide selection upon pre-insertion (at III). Note that the stabilized NTP insertion state IV can become the most populated state if a noncatalytic analog of NTP is provided (16,17) (as IV → V inhibited).

Importantly, our results always show that P_{II'} > P_{II} at post-translocation (see Fig. 3b), which is essential for maintaining transcription fidelity. Because the Tyr^{639} side chain fluctuates IN and OUT rapidly (>10^4 s^{-1}), II' and II intermediates stay close to local equilibrium. By fitting the single-molecule experimental data we obtained the free energy difference between II and II' (or I and I'), because $E_I - E_{II'} - E_{II} \equiv \beta - 2k_B T > 0$, giving a population bias of P_{II'}/P_{II} ~ 7–8. This explains why the post-translocation structure II' with Tyr^{639} IN was detected and crystallized (16), but not the structure II with Tyr^{639} OUT (see Fig. 2).

In the pre-insertion state III, the Tyr^{639} side chain was also captured IN rather than OUT (16,17). From the structural comparisons shown in Fig. 1b (left), we note that the Tyr^{639} side chain is close to the pre-insertion site when it is IN the active site (II'), whereas the side chain stays...
far from the pre-insertion site when it is OUT (II). Hence, the performance of nucleotide selection by Tyr^{639} at the pre-insertion site relies strongly on whether it populates more in the II' or II state, or on the population bias $P_{II'} > P_{II}$. When $P_{II'} \leq P_{II}$ (i.e., without the post-translocation energy bias), the pre-inserted nucleotide would not be screened well because Tyr^{639} frequently stays (OUT) far from the pre-insertion site.

### Comparing the ratcheting scheme with multi-subunit RNAPs

As mentioned earlier, using the parallel-path scheme allows us to describe the translocation-elongation kinetics for single and multi-subunit RNAPs in a unified way. The IN and OUT configurations of Tyr^{639} in the single subunit T7 RNAP correspond to the bent and straight configurations, respectively, of the bridge-helix in the multi-subunit RNAP (36,42,43). Moreover, the closing and opening of the O-helix in T7 RNAP correspond, respectively, to folding and unfolding of the trigger loop in the multi-subunit RNAPs.

There are distinctive features, however, that vary between the two ratcheting schemes (see Fig. S1 in the Supporting Material):

First, our structural examination of T7 RNAP reveals that binding or pre-insertion of an incoming nucleotide can take place without steric hindrance whether the Tyr^{639} side chain is IN (II') or OUT (II) of the active site (see Fig. 1 b and Fig. 2, and see Fig. S1 a). Whereas for the multi-subunit RNAP, the nucleotide pre-insertion seems to take place only when the bridge-helix is straight (II) (see Fig. S1 b). The bending region of the bridge-helix (pink surface) has steric clashes with the pre-inserted nucleotide (van der Waals spheres inside the yellow oval), as demonstrated in Fig. 1 b, right.

Second, the configuration II' is more stabilized than II as we have found for T7 RNAP, likely due to stacking of the side chain of Tyr^{639} with the end base pair of the RNA-DNA hybrid. Whereas in the multi-subunit RNAP, there is no evidence that the bent configuration of the bridge-helix (II') is more stabilized than the straight one (II). In fact, a post-translocated state with the bridge-helix straight (II) was captured structurally (43,44), implying that II is likely more stabilized than II' in the multi-subunit RNAP. A high-energy translocation intermediate captured in complex with an inhibitor is consistent with this idea (45).

In Eqs. S6 and S7 in the Supporting Material, we derive approximate formulas for the elongation rates of both types of RNAPs. Correspondingly, Fig. 3 c shows the curves of elongation rate, $v$ versus the post-translocation free energy bias, $\beta$ (or $\beta + \gamma$ in general when $\gamma \equiv E_{II'} - E_{II} \neq 0$) for each type of the RNAPs. For T7 RNAP, we see that increasing the post-translocation free energy bias $\beta$ beyond $2k_BT$ does not improve the elongation rate much (e.g., <5% increase) as...
the rate is already close to its saturation value. On the other hand, in comparison to a pure Brownian ratchet, the 3 $k_B T$ ($\alpha + \beta$) free energy bias along $\Gamma \rightarrow \Gamma'$, or an overall $\sim 1.3 \ k_B T$ for both paths, does improve the elongation rate somewhat (e.g., $\sim 14\%$ increase).

For multi-subunit RNAPs, however, it is likely that $\beta < 0$ (E$_{\Gamma}$ > E$_{\Gamma'}$; see Brueckner et al. (43) and Fig. 1b in that work as well), with the bridge-helix more stabilized in the straight ($\Gamma$) than the bent ($\Gamma'$) configuration at post-translocation. In this case, the elongation rate still stays close to its saturation. Remarkably, the elongation rate can be significantly reduced when the bent configuration ($\Gamma'$) of the bridge-helix becomes more stabilized than the straight conformation ($\Gamma$) by binding some inhibitory factor such that $\beta > 0$ (E$_{\Gamma}$ > E$_{\Gamma'}$; see Fig. 3c).

Efficient nucleotide selection before stable insertion

Because T7 RNAP lacks a proof-reading mechanism, accurate nucleotide selection is crucial for transcription fidelity. For a wild-type T7 RNAP, experiments estimated that transcriptional errors occur at an average frequency $\sim 10^{-4}$ (46). If there were only one checkpoint for nucleotide selection in T7 RNAP, a free energy difference between wrong and right nucleotides at this checkpoint would be $\Delta E \sim -k_B T \ln(10^{-4}) \sim 10 \ k_B T$ to achieve this accuracy. Such a large energy disparity can hardly be maintained in a fluctuating protein cavity such as the active site of T7 RNAP. The site is not buried deeply inside the protein, and cocrystallized water molecules are observed close to the site (16) and the water molecules would smear the energy difference between the right and wrong nucleotides (47). Therefore, T7 RNAP would likely have two or more steps, or checkpoints, dedicated to its nucleotide selection.

Using our elongation scheme, we compared the energy efficiencies of several individual selection mechanisms, i.e., the error rates of elongation under different selection methods, using a constant energy penalty for differentiating between right and wrong nucleotides (see the Supporting Material for details). We find that selection methods (No. 1 and No. 2) that take place before the rate-limiting step (nucleotide insertion $III \rightarrow IV$) are more efficient than (No. 3 and No. 4) that take place after. For example, using an energy of $\sim 5 \ k_B T$ to reject the wrong nucleotide upon pre-insertion (No. 1) or to inhibit its insertion rate (No. 2) can achieve an error rate $\sim 10^{-2}$, whereas using the same amount of energy to reject the wrong nucleotide right after insertion (No. 3) or to inhibit the chemical reaction rate (No. 4) gives an error rate $\sim 10^{-1}$ (see Table S2). The property is closely related to where the rate-limiting transition is located in the elongation cycle. If one makes the chemical transition $IV \rightarrow V$ rate-limiting, then selections (Nos. 1–3) that take place before $IV \rightarrow V$ become similarly efficient, whereas the one after (No. 4) is still the least efficient.

This happens because a rate-limiting forward transition presumably crosses a large activation barrier that can further hinder the backward transition. The hindered backward transition makes releasing the wrong nucleotide difficult.

Our results also show that, when there is no selection against wrong nucleotides at pre-insertion, or if the selection at pre-insertion costs an energy only at the thermal fluctuation level ($1 \sim 2 \ k_B T$), the elongation rate becomes quite low. For example, the rate drops below 30 nt/s when single selection happens but not at pre-insertion (selections Nos. 2–4, see Table S2), which is much lower than the error-free rate $\sim 100$ nt/s. These results suggest that the pre-insertion step is the most crucial checkpoint in T7 RNAP for achieving efficient nucleotide selection as well as maintaining a sufficiently high elongation rate.

Forward tracking with facilitated translocation

In addition to studying the effect of facilitated translocation on the nucleotide selection at pre-insertion, we also examined whether the facilitated translocation affects forward tracking. We assume that in the stabilized post-translocated state $\Gamma''$, T7 RNAP can randomly switch between regular elongation and forward tracking, i.e., moving forward without synthesizing RNA (illustrated in Fig. 1c). Consequently, the RNA-DNA hybrid shortens and association between RNAP and the DNA weakens. Forward tracking likely takes place frequently at terminator regions causing intrinsic termination (22,27).

We examined the probability of T7 RNAP forward tracking around two terminator regions: one is $\phi$, a late terminator found in the T7 genome (48), and the other is a threonine (thr) attenuator in Escherichia coli (49). To take into account sequence dependence during translocation (see the Supporting Material), we evaluated the free energy difference between the post ($\Gamma$) and pre-translocated states ($\Gamma'$, $\gamma$, by a sum of energies calculated from unzipping/ripping of downstream/upstream of the dsDNA and RNA-DNA hybrid (50). Similarly we evaluated $\beta$ by the translocation energy difference plus an additional energy contribution from Tyr$^{639}$ stabilization at $\Gamma''$. We found that the forward tracking can take place at high efficiency at the terminator and at a very low efficiency at nonterminator regions, if one allows the forward rate to decrease exponentially with the stabilization energy of the RNA-DNA hybrid upstream. The forward tracking efficiency was determined by counting the number of extensive forward tracking events (beyond 3 nt) for every 100 trial events.

We then compared the forward-tracking efficiency (with the facilitated path I$'$ ↔ $\Pi''$) with that of a pure Brownian ratchet scheme (without the facilitated path I$'$ ↔ $\Pi''$). We found that the forward tracking efficiencies are almost identical in both cases. For example, when the Brownian ratchet case is tuned with $48 \pm 3\%$ forward tracking efficiency at the thr attenuator, the efficiency is $51 \pm 5\%$, and further
increase of the post-translocation energy bias ($\beta + \gamma$) hardly improves the efficiency ($\sim 53 \pm 5\%$ maximum). In all these cases, the forward tracking efficiency at the nonterminator region remains at the same low value ($\sim 2\%$ for equally long regions). Similar trends were found also for the T-terminator. Hence, compared with the pure Brownian translocation case, the facilitated translocation in T7 RNAP does not enhance the likelihood of forward tracking. Note that the forward tracking may not be the mechanism for intrinsic termination (51,52), so the conclusion does not extend in general to termination.

DISCUSSION

A large translocation free energy bias is not an advantage for sequence detection

In our model, T7 RNAP translocation is driven largely—but not completely—by a Brownian ratchet mechanism. Each translocation step involves changes in DNA and RNA structures that depend on sequence stabilities. Therefore, translocation can be utilized by RNAP to sense and respond to sequence signals along dsDNA. When there is no structural element or energy output from RNAP to assist in translocation, its movements are diffusive. The RNAP moves forward and backward equally fast until an incoming NTP binds to prevent the backward movements. Because the RNAP itself does not induce a free energy bias along the translocation path, sequence variations on the DNA can readily be detected, and RNAP can respond to the variation by slowing down, tracking forward or backward. By contrast, a large translocation energy bias can reduce the activation barrier and accelerate the forward movements, while it can mask relatively small energy differences arising from different DNA sequences. In addition, the translocation is not rate-limiting during the elongation cycle, so a large translocation energy bias can hardly improve the overall elongation rate. In our translocation-elongation model of T7 RNAP, for example, the sequence signal is encoded in the translocation energy $\gamma$ that affects the elongation rate (see Eq. S6 in the Supporting Material). When the post-translocation free energy bias ($\beta + \gamma$) becomes large, the elongation rate increases to saturation independent of the sequence signal ($\gamma$). In multi-subunit RNAPs, a large post-translocation free energy bias simply stalls the elongation (see Eq. S7 in the Supporting Material and Fig. 3 c). Hence, a large translocation energy bias of RNAP does not improve elongation rate nor confer an advantage for sequence detection during transcription elongation.

A small post-translocation free energy bias in T7 RNAP aids in transcription fidelity

Our analysis of T7 RNAP data shows that a small free energy bias that stabilizes the post-translocated interme-
diate can aid in nucleotide selection at pre-insertion, without interfering with sequence detection.

Using experimental data from single-molecule force measurements (24,25), we identified a post-translocation free energy bias of $\sim 2 k_B T$ imposed by the RNAP. The structural elements that achieve this bias include, but are not necessarily limited to, Tyr$_{639}$ at the C-terminal end of the O-helix (see Fig. 1). The translocation can take place without a free energy bias (i.e., via Brownian path $I \rightarrow II$ in Fig. 2), during which the 3'-end of the RNA moves out of the active site without assistance (with the Tyr$_{639}$ side chain OUT). Alternatively, the side chain of Tyr$_{639}$ could squeeze IN the active site at the pre-translocated state when the active site is still occupied by the 3'-end of the RNA. Because the 3'-end of the RNA appears quite flexible (53), the squeezing ($I \rightarrow I'$) costs but a small amount of free energy (e.g., $\sim 1 k_B T$). The unstable intermediate ($I'$) can be easily driven by $-3 k_B T$ free energy bias toward a more stabilized post-translocated configuration ($II'$), in which the 3'-end of the RNA is pushed out of the active site, leaving only the side-chain Tyr$_{639}$ in occupancy. Indeed, this stabilized intermediate ($II'$) had been captured in high-resolution structural studies (16), and led to the suggestion that PPI release powers the translocation mechanism. Our studies indicate, however, that the intermediate ($II'$) results from the small power stroke exerted by Tyr$_{639}$, not from PPI release. In the post-translocated state, $II'$ is stabilized by the Tyr$_{639}$ side chain stacking with the base pair at the end of the RNA-DNA hybrid (see Fig. 2). Due to the side-chain fluctuation, there also exists a nonstabilized, and therefore lowly populated post-translocated state $II$, in which Tyr$_{639}$ is OUT. Because the energy changes among the translocation intermediates ($I$, $I'$, $II'$, and $II$) are small, the translocation mechanism appears to be largely a Brownian ratchet.

Although the post-translocation energy bias appears quite small, there is still a way to distinguish the biased translocation from the pure Brownian case. At high NTP concentrations, the post-translocated intermediate $II'$ is moderately populated during the elongation cycle. Its population grows as NTP concentration decreases, and becomes dominant at very low NTP concentrations (see Fig. 3 b). By comparison, in the pure Brownian ratchet scheme, both the pre-translocated ($I$) and the post-translocated states ($II$) become dominant and equally populated ($\sim 50\%$ each) at low NTP concentrations. Hence, to test whether the translocation is purely Brownian or slightly biased, one should examine whether there exist two equally populated conformations (Brownian) or one dominant conformation (biased) at very low NTP concentrations, close to equilibrium. Because sequence effects can locally bias state populations and interfere with accurate measurements, experimental tests should be conducted when RNAP transcribes homogeneous DNA sequences.

Taking into account both translocation paths, the overall free energy bias of translocation is $-1.3 k_B T$ (24). This small
post-translocation energy bias can actually slightly improve the elongation rate compared to a pure Brownian ratchet, whereas further increasing the energy bias improves the elongation rate very little (see Fig. 3 c). Thus, the major role of the facilitated translocation in T7 RNAP is not to improve the elongation rate, but to bring about a locally stabilized post-translocated intermediate (II') before NTP pre-insertion. In this intermediate state, the Tyr639 side chain inserts IN next to the pre-insertion site; its hydroxyl group can subsequently coordinate with a magnesium ion such as to discriminate between ribo- versus deoxy-ribo nucleotides (17) when the nucleotide pre-inserts. As the side chain of Tyr639 is OUT, its hydroxyl group moves farther (~6 Å; see Fig. 1 b) away from the pre-insertion site and cannot participate in nucleotide selection. In brief, by consuming only a small amount of energy to stabilize the Tyr639 side chain IN the active site, T7 RNAP can guarantee an ~80~90% chance of keeping the tyrosine abutting the pre-insertion site for nucleotide selection.

Our studies point out that there are two reasons why the pre-insertion is the most crucial checkpoint of nucleotide selection. First, it appears energetically more efficient to reject the wrong nucleotide before its slow insertion than afterward. This property depends on where the rate-limiting step is located in the nucleotide addition cycle: selection before the rate-limiting transition always seems more efficient than selection after. Second, to maintain both high fidelity (low error rate) and sufficiently high elongation rate very little (see Fig. 3 c). Thus, the major role of the facilitated translocation in T7 RNAP is not to improve the elongation rate, but to bring about a locally stabilized post-translocated intermediate (II') before NTP pre-insertion. In this intermediate state, the Tyr639 side chain inserts IN next to the pre-insertion site; its hydroxyl group can subsequently coordinate with a magnesium ion such as to discriminate between ribo- versus deoxy-ribo nucleotides (17) when the nucleotide pre-inserts. As the side chain of Tyr639 is OUT, its hydroxyl group moves farther (~6 Å; see Fig. 1 b) away from the pre-insertion site and cannot participate in nucleotide selection. In brief, by consuming only a small amount of energy to stabilize the Tyr639 side chain IN the active site, T7 RNAP can guarantee an ~80~90% chance of keeping the tyrosine abutting the pre-insertion site for nucleotide selection.

Our studies point out that there are two reasons why the pre-insertion is the most crucial checkpoint of nucleotide selection. First, it appears energetically more efficient to reject the wrong nucleotide before its slow insertion than afterward. This property depends on where the rate-limiting step is located in the nucleotide addition cycle: selection before the rate-limiting transition always seems more efficient than selection after. Second, to maintain both high fidelity (low error rate) and sufficiently high elongation rate, it is necessary to have an energy above the thermal fluctuation level to filter out the incorrect nucleotide at pre-insertion. Exactly how—and by how much—Tyr639 and other structural elements contribute to nucleotide selection awaits further studies.

In our further examination, forward tracking is allowed during elongation via the stabilized post-translocated intermediate (II'). The forward tracking seems to happen at high efficiencies at the terminator sequences. However, we have not identified any substantial effect that the facilitated translocation brings to the forward tracking compared to pure Brownian ratchet translocation (with forward tracking via II). Indeed, increasing the post-translocation energy bias along the facilitated path hardly affects the II' population in T7 RNAP. However, this property does not necessarily hold for multi-subunit RNAPs as increasing the post-translocation bias does substantially increase the population of II' as well as inhibit RNAP elongation.

What about translocation and ratcheting in the multi-subunit RNAPs?

In comparison with the single subunit T7 RNAP, we suggest a slightly different ratcheting scheme under our parallel-path scheme for the multi-subunit RNAPs (see Fig. S1). In the multi-subunit RNAPs, bending and straightening of the bridge helix are analogous to IN and OUT fluctuations of the Tyr639 side chain in T7 RNAP. In this alternative ratcheting scheme, however, there are two essential features that distinguish the multi-subunit RNAPs from T7 RNAP:

First, pre-insertion of nucleotide can only happen when the bridge helix is straight (II → III allowed), but not when it is bent (II' → III forbidden). This is because the bent bridge helix sterically occludes the pre-insertion site (see Fig. 1 b, right). In T7 RNAP, by contrast, the nucleotide can pre-insert at either IN (II') or OUT (II) configuration of Tyr639 without steric hindrance.

Second, the bridge helix bent configuration in the post-translocated state (II') does not appear to be stabilized as is the corresponding IN configuration of Tyr639 in T7 RNAP. Indeed, the bent configuration seems to be less stable than the straight configuration of the bridge helix in the post-translocated state because crystal structures of both the post-translocated state (II) and the pre-insertion state (III) of the multi-subunit RNAP were caught with the straight form of the bridge helix (43,44). Whereas in T7 RNAP, it is Tyr639 IN that was structurally captured (II' and III) but not OUT (16).

Nevertheless, the bent configuration of the bridge helix has been structurally identified when an inhibitory factor was bound to the multi-subunit RNAP. Previously, a toxin-bound translocation intermediate (α-amanitin) was structurally resolved (45). The bridge helix is bent in this structure with the trigger loop in a wedged conformation. The captured structure is likely a stabilized form of a high-energy intermediate in translocation. This intermediate corresponds nicely to II' in our model (Fig. 2). According to our interpretation, the post-translocation bias β + γ (≡ EII'− EII) < 0 is originally set for the multi-subunit RNAP, while the toxin stabilizes the II' intermediate such that β + γ > 0 (as in T7 RNAP). However, because the II' configuration cannot bind incoming nucleotide in the multi-subunit RNAP, it competes with the nucleotide for the binding (pre-insertion) site. In multi-subunit RNAPs, only a few kBT stabilization of II' can lower the effective concentration of NTP significantly and cause a large drop of the elongation rate (see Fig. 3 c). Therefore, II' appears to be a target intermediate state for inhibitory factors of elongation in the multi-subunit RNAPs: once the factor stabilizes the state (i.e., by increasing the post-translocation free energy bias), the elongation activity is greatly reduced. Notably, in recent high-resolution studies of a multi-subunit RNAP from a bacterial species (54), a ratcheted state has been identified that is in complex with a transcription inhibitor (Gfh1). This inhibitor protein occludes the channel for NTP entry into the active site. The bridge helix is kinked in the ratcheted state. Accordingly, this structure appears to be a stabilized form of II' as well. Once II' becomes more stabilized than the other post-translocated intermediate II (bridge helix straight), II' dominantly populates at low NTP concentration.
In summary, we have dissected the ratchet mechanism of T7 RNAP during its transcription elongation. A small post-translocation free energy bias (~2 k_BT) is identified and its functional role is proposed: The energy bias creates a locally stabilized post-translocated intermediate in which the highly conserved Tyr^{639} stays close to the pre-insertion site for nucleotide selection. Without this post-translocation energy bias (equivalently, local stabilization) the tyrosine side chain frequently moves away from the pre-insertion site so that nucleotide selection is less effective, lowering transcription fidelity. Further studies will be necessary to determine if other structural elements participate in nucleotide selection at pre-insertion, how subsequent nucleotide selection proceeds, and whether a similar property exists for other single subunit polymerases.

Remarkably, one can identify analogous functional elements and build an analogous but distinct ratcheting scheme for the multi-subunit RNAPs. The variations of structural features of the RNAPs lead to different reaction topologies. The post-translocation intermediate with the bent bridge-helix does not allow NTP binding. Nor is it destabilized unless some inhibitory factor binds and enhances the post-translocation energy bias. Hence, bending of the bridge-helix, in addition to coordinating backtracking, appears to support inhibitory control (36).

SUPPORTING MATERIAL

Additional methodology, with equations, two figures, two tables, and references (57–62), is available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(11)05457-9.

J.Y. is supported by National Science Foundation grant No. DMS 1062396.

G.O. was supported by National Science Foundation grant No. DMS 0414039.

REFERENCES


Supporting Information

Table of Contents

Constructing the model ................................................................................................................. 1
Using master equation approach ........................................................................................................ 3
Deriving elongation rates for two ratchet schemes ........................................................................ 4
Setting and tuning parameters ......................................................................................................... 7
Comparing nucleotide selection strategies ................................................................................... 10
Simulating sequence-dependent translocation .............................................................................. 11

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 (γ ≡ 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:

$$\frac{k_{2+}}{k_{2-}} = e^{−γ/k_BT}.$$ 

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 ≡ I or I’), it may only require a free energy at the level of thermal fluctuation (Assumption A4), i.e., $α ≡ E_{I′} − E_{I} ≈ 1 k_BT$. 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, $ω_{2+}$ and $ω_{2-}$, is modulated by $α$ as:

$$\frac{ω_{2+}}{ω_{2-}} = e^{−α/k_BT} , \text{ with } ω_{2+} ≥ 10^4 \text{ s}^{-1}.$$ 

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 the 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 $β + γ (≡ E_{II} − E_{II′} = E_{II} − E_{I′} + E_{II′} − E_{I})$: $ω_{2+}/ω_{2-} = e^{(β + γ)/k_BT}$. Consequently, the translocation path I’ → II’ is energetically favorable, as there is an energy drop of $E_{I′} - E_{II′} = E_I - E_{I′} + E_{II} - E_{II′} = α + β > 0$. 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:

$$\frac{k_{2+}}{k_{2-}} = e^{(α + β)/k_BT}.$$ 

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 1b). 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. EII′ < EII ~ EII′).

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 1b). 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 1c), 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 $\Pi = (P_I, P}_{I′}, P_{II}, P_{III}, P_{IV}, P_{V})^T$. Hence, the master equation describing the kinetics can be written in a matrix form as:

$$\frac{d\Pi}{dt} = M \Pi \quad \text{(S1)}$$

where $M$ is a 7x7 transition matrix.
As defined in the main text, tracking events, etc., by running kinetic Monte-Carlo simulations (12). The steady state can be quickly reached, and corresponding to the master equation description, one can generate trajectories of elongation.

In particular, the forward and backward translocation rates along path \( I \rightleftharpoons II \) and \( I' \rightleftharpoons II' \) are described as:

\[
  k_{2+} = r_0 e^{\gamma k_B T} \quad k_{2'} = r_0 e^{\alpha k_B T} \quad k_{2-}' = r_0 e^{-\beta k_B T} .
\]  \hfill (S2)

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

\[
  \omega_{1+} = \omega_0 e^{\alpha k_B T} \quad \omega_{2+} = \omega_0 e^{\gamma k_B T} \quad \omega_{2-} = \omega_0 e^{-\beta k_B T} .
\]  \hfill (S3)

As defined in the main text, \( \alpha \equiv E_{I'} - E_I \), \( \beta \equiv E_I - E_{II'} \), \( \gamma \equiv E_{II} - E_I \) are free energy differences (unit: \( k_B T \)) among pre- and post-translocated states \( I, I', II, II' \). 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 \( \Pi_s = (P_{I}^s, P_{II}^s, P_{III}^s, P_{IV}^s, P_{V}^s)^T \) satisfies:

\[
  M \Pi_s = 0 .
\]  \hfill (S4)

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

\[
  v = l_0 J ,
\]  \hfill (S5)

with \( l_0 \) the periodic length 1-nt, and

\[
  J \equiv P_{V}^s k_{1+} - P_{I}^s k_{1-} = P_{II}^s k_{5+} - P_{III}^s k_{5-} = P_{IV}^s k_{4+} - P_{I}^s k_{4-} = P_{II}^s k_{3+} - P_{III}^s k_{3-}. \]

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* → I slow (rate-limiting, see A7) and irreversible (k⁺^I = kcat and k⁻^I = 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 (PI^I_e = e^(-E_I)/kBT PI and PII^I_e = e^(-E_{II}^I)/kBT PII). Figure S1 Ratcheting schemes with two parallel translocation paths in RNAP transcription elongation. After NTP loading, a generic ‘catalytic’ transition III*→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'→III* or II→III* (see A6). Since the binding rate is likely diffusion-limited, k⁺^III* = k⁺^NTP*[NTP]. For unbinding, k⁻^III* = k⁻^III*e^(-(βγ)/k_BT), as E_{II} - E_{II'} = β + γ. Solving the master equation of the simplified scheme, we obtained P_I = e^{(-γ)/k_BT}P_I and P_H = e^{((γ+β)/k_BT)}P_H, so that

v = l_0 k_{cat}[NTP]/(1 + e^{-(γ)/k_BT} + e^{(-γ+α)/k_BT} + e^{((γ+β)/k_BT)}(k^+_III + k^-_III + k_{cat}))

Using K_D = k^-_NTP/k^-_NTP, and K_D = K_D'[1 + e^{-(α+β)/k_BT}]/[1 + e^{-(α+β)/k_BT}] (average of dissociation III* → II' and III* → II), one obtains:

v = l_0 k_{cat}[NTP]

(1 + e^{-(γ)/k_BT} + e^{(-γ+α)/k_BT})/(1 + e^{(γ+β)/k_BT})(k_{cat}^{-1} + k_NTP^{-1} + K_D[1 + e^{(-α-β)/k_BT}]^2)[NTP]
where \( l_0 = 1 \) nucleotide (nt) is the periodic length of translocation, \( k_{\text{cat}} \) is the effective maximal rate of ‘catalysis’ (\( \sim 130 \text{ s}^{-1} \) for T7 RNAP (11)), \( k^0_{\text{NTP}} \) is the NTP binding /pre-insertion rate (fitted to \( \sim 2 \mu\text{M}^{-1} \text{s}^{-1} \), see section below), and \( K_D \) is the dissociation constant at the NTP binding/pre-insertion site, measured at \( \sim 80 \mu\text{M} \) (10). The translocation energy associated with Brownian motions is defined \( \gamma \equiv E_H - E_I \sim 0 \), with fluctuations around zero caused by DNA sequence effects. Since the 3’-end of the RNA is quite flexible (4), we assume that \( \alpha \equiv E_I - E_{I'} \sim 1 \, k_B T \) (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 \( \pm 2 \, k_B T \) (see section below). Note that when DNA sequence effects are considered, \( \gamma \) and \( -\beta \) are affected identically while \( \beta + \gamma (=E_{II} - E_{II'}) \), the post-translocation free energy bias \( (\gamma \neq 0) \), 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 \( \gamma \) and \( -\beta \) identically, while \( \beta + \gamma \) 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

\[
P_H = \frac{k^*_3 + k^*_4}{k^*_3} \, P_{II^*},
\]

so that

\[
v = \frac{l_0 k_{\text{cat}} [\text{NTP}]}{(1 + e^{\gamma k_B T} + e^{(\gamma - \alpha) k_B T} + e^{(\gamma + \beta) k_B T}) (k^*_3 + k^*_{\text{cat}}) / k^0_{\text{NTP}} + [\text{NTP}]).
\]

With \( K_D \equiv k^*_3 / k^0_{\text{NTP}} \), one obtains:

\[
v = \frac{l_0 k_{\text{cat}} [\text{NTP}]}{(1 + e^{\gamma k_B T} + e^{(\gamma - \alpha) k_B T} + e^{(\gamma + \beta) k_B T}) (\frac{k_{\text{cat}}}{k^0_{\text{NTP}}} + K_D) + [\text{NTP}].
\]

(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:

\[
v = v_{\text{max}} [\text{NTP}] / (K_M + [\text{NTP}])
\]

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 3c that increasing the post-translocation free energy bias \( \beta \) beyond \( 2 \, k_B T \) does not improve the elongation rate much (e.g. \(< 5\% \) increase) as the rate is already close to its saturation value (as \( \beta + \gamma \rightarrow +\infty \)):

\[
v' \rightarrow l_0 k_{\text{cat}} [\text{NTP}] / (\frac{k_{\text{cat}}}{k_{\text{ET}}} + K_D + [\text{NTP}]).
\]

On the other hand, in comparison to a pure Brownian ratchet \( (\gamma = 0, \alpha \rightarrow +\infty \), and \( \beta \rightarrow -\infty \), i.e., path I ↔II only), the \( 3 \, k_B T \) free energy bias \( (\alpha + \beta) \) along I’ →II’, or an overall \( \sim 1.3 \, k_B T \) for both paths, does improve the elongation rate somewhat (e.g. \( \sim 14\% \) increase).
For multi-subunit RNAPs, however, it is likely that $\beta$ or $\beta + \gamma < 0$; the elongation rate still stays close to its saturation value (as $\beta + \gamma \to -\infty$):

$$v' = l_0 k_{cat} [\text{NTP}] / ((1 + e^{\gamma/k_BT} + e^{(\gamma-\alpha)/k_BT})(k_{cat} + K_D) + [\text{NTP}]).$$

When the bent configuration $\text{(II')}$ of the bridge helix is stabilized by some inhibitory factor relative to the straight conformation $\text{(II)}$ such that $\beta + \gamma > 0$ ($E_{\text{II}} > E_{\text{II'}}$), 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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
<th>Default value</th>
<th>Notes</th>
<th>Extras</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_0$</td>
<td>RNAP translocation rate from $\text{I to II}$ (see Eq. S2)</td>
<td>$5000 \text{ s}^{-1}$</td>
<td>$&gt; 10^3 \text{ s}^{-1}$; translocation non-rate-limiting</td>
<td>Elongation rate $v$ insensitive to $r_0$</td>
</tr>
<tr>
<td>$\omega_0$</td>
<td>Oscillation rate for Tyr 639 side chain from $\text{I to I'}$ (see Eq. S3)</td>
<td>$50000 \text{ s}^{-1}$ or even larger</td>
<td>Tyrosine side-chain flipping rate on the order $\sim 10^4 \text{ s}^{-1}$ (5)</td>
<td>$v$ insensitive to $\omega_0$; very low $\omega_0$ can induce pauses though</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Free energy for Y639 side chain move IN when the active site is still occupied with 3' RNA, $\alpha \equiv E_{\text{I'}} - E_{\text{I}}&gt;0$</td>
<td>$1 \text{ k_BT}$</td>
<td>Assumption A4: a small amount at thermal fluctuation level</td>
<td>$v$ insensitive to $\alpha$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Free energy difference $\beta \equiv E_{\text{I}} - E_{\text{II}}$; called ‘post-translocation bias’ as $\gamma \to 0$</td>
<td>$2 \text{ k_BT}$</td>
<td>The most essential tuning parameter, tuned for apparent $K_M$ in current scheme (see text)</td>
<td>$v$ is sensitive to $\beta$; ranged between 1.4 to 2.5 $\text{k_BT}$ as tuning average $v$ within $\pm 5 \text{ nt/s}$ ($\text{as} k^0_{\text{NTP}} = 2 \mu\text{M}^{-1}\text{s}^{-1}$)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Translocation free energy (under diffusion) $\gamma \equiv E_{\text{II}} - E_{\text{I}}$</td>
<td>$0$</td>
<td>Assumption A3: Brownian motion along path $\text{I} \Leftrightarrow \text{II}$</td>
<td>non-zero and fluctuating under DNA sequence effects</td>
</tr>
<tr>
<td>$k^0_{\text{NTP}}$</td>
<td>NTP binding rate constant; $k_{3^+} = k_{3'+} = k^0_{\text{NTP}} \cdot [\text{NTP}]$</td>
<td>$2 \mu\text{M}^{-1}\text{s}^{-1}$</td>
<td>1~10 $\mu\text{M}^{-1}\text{s}^{-1}$ for some known NTPases; tuned for apparent $K_M$ in the single-path scheme (see text)</td>
<td>$v$ is sensitive to $k^0_{\text{NTP}}$; ranged at 1.4 ~ 2.5 $\mu\text{M}^{-1}\text{s}^{-1}$ as tuning average $v$ within $\pm 5 \text{ nt/s}$</td>
</tr>
<tr>
<td>(k_{3^-})</td>
<td>NTP unbinding rate via III → II'</td>
<td>178 s⁻¹</td>
<td>(K_D' = \frac{k_{3^-}}{k_{NTP}^0}) tuned to (\sim 89\ \mu M) by (\beta)</td>
<td>(K_D' \approx K_p \frac{1 + e^{-\rho - \gamma}}{1 + (e^{-\rho - \gamma})^2}) where (K_D = 80\ \mu M) ((10))</td>
</tr>
<tr>
<td>(k_{3^-})</td>
<td>NTP unbinding rate via III → II</td>
<td>24 s⁻¹</td>
<td>(k_{3^-} = k_{3^-}' e^{-\beta - \gamma}) since (E_{II} = E_{II'} = \beta + \gamma)</td>
<td>Both (k_{3^-}) and (k_{3^-}') depend on (\beta)</td>
</tr>
<tr>
<td>(k_{4^+})</td>
<td>NTP insertion / O-helix closing rate</td>
<td>220 s⁻¹</td>
<td>Measured in ((10)) and assumed rate-limiting (see Assumption A6)</td>
<td>(v_{max}) is sensitive to (k_{4^+})</td>
</tr>
<tr>
<td>(k_{4^-})</td>
<td>Reverse rate for NTP insertion</td>
<td>210 s⁻¹</td>
<td>Choose below 220 s⁻¹; tuned for (v_{max})</td>
<td>Constrained by (v_{max}) as well as total free energy (see Eq. S9)</td>
</tr>
<tr>
<td>(k_{5^+})</td>
<td>Catalytic rate of polymerization reaction</td>
<td>1000 s⁻¹</td>
<td>Choose (\sim 10^4) s⁻¹, fast; tuned for (v_{max})</td>
<td>Constrained by (v_{max})</td>
</tr>
<tr>
<td>(k_{5^-})</td>
<td>Reverse rate for polymerization reaction</td>
<td>135 s⁻¹</td>
<td>(\ln \left(\frac{k_{5^+}}{k_{5^-}}\right) \approx 2k_B T) likely small (e.g. (\sim 1) k_B T for F₁ ATPase ((14)))</td>
<td>Constrained by total free energy constraint</td>
</tr>
<tr>
<td>(k_{1^+})</td>
<td>PPI release rate</td>
<td>1200 s⁻¹</td>
<td>(K_d = \frac{k_{1^-}}{k_{PPI}^0} = 1200\ \mu M) measured in ((10))</td>
<td>Constrained by (v_{max})</td>
</tr>
<tr>
<td>(k_{1^-})</td>
<td>Reverse (PPI binding rate) (k_{1^-} = k_{PPI}^0 \cdot [PPI])</td>
<td>0.1 s⁻¹</td>
<td>Use [PPI] (\sim 0.1) (\mu M) (PPI low ((11)))</td>
<td>Set (k_{PPI}^0 \approx 1\ \mu M^{-1} s^{-1}), the same order as (k_{NTP}^0)</td>
</tr>
</tbody>
</table>

**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 \(\sim 130\) nt/s from experiments \((11)\).

Parameters \(k_{4^+}, k_{5^+}\) 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 \(k_{NTP}^0\) 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:

\[
v = \frac{l_0 k_{cat}[NTP]}{(1 + e^{r/k_B T})(\frac{k_{cat}}{k_{NTP}^0} + K_D) + [NTP]}\]

\((S8)\)
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 \( \sim 80 \mu M \) (10). Next, \( \beta \) 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 \( \text{III} \rightarrow \text{II}' \) by: 
\[
K'D' \sim K_D - \frac{1}{1 + e^{-(\beta + \gamma)/k_BT}}. 
\]
We regard the measured \( K_D \) as the average of that through path \( \text{III} \rightarrow \text{II}' \) (with a ‘weight’ \( \beta \)) and that through path \( \text{III} \rightarrow \text{II} \) (with a ‘weight’ \( \frac{\beta + \gamma}{k_BT} \)) in T7 RNAP. Since \( K_D \) had been measured experimentally (10), \( \beta \) can be determined by \( \beta \). It turns out that \( \beta \) is the most essential tuning parameter in current model.

\( \beta \) was tuned \( 1.4 \sim 2.5 \text{ k}_B \text{T} \) when \( k_{NTP}^0 \) is set at \( 2 \mu M^{-1}s^{-1} \) (if we allow the average rate vary within 5 nt/s), hence, we use \( \beta \sim 2 \text{ k}_B \text{T} \). Varying \( k^0_{NTP} \) \( (1.4 \sim 2.5 \mu M^{-1}s^{-1}) \), determined using the three-state single-path scheme) allows \( \beta \) adopt values from about \( 1 \text{ k}_B \text{T} \) to a little over \( 3 \text{ k}_B \text{T} \). Hence we present current estimation of \( \beta \) as \( 2 \pm 1 \text{ k}_B \text{T} \).

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 \( \beta \) to \( 1.4 \sim 2.5 \text{ k}_B \text{T} \), so that the overall or average translocation energy bias (simply estimated as \( (\alpha + \beta)/2 \)) is about \( 1 \sim 2 \text{ k}_B \text{T} \). 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 \( 0 \sim 3 \text{ k}_B \text{T} \), 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:

\[
\Delta G_0 = k_BT (\ln \frac{k^0_{NTP}}{k_{3-}} + \ln k_{4+} + \ln k_{5-} + \ln k_{6+}) = 4 \sim 7 \text{ k}_B \text{T} \quad \text{(S9)}
\]

With above default values of the parameters, \( \Delta G_0 \sim 6.6 \text{ k}_B \text{T} \). At a relatively high concentration of NTP (588 \( \mu M \)) and a low concentration of PPI (0.1 \( \mu M \)), the free energy drop for each step is, for example, NTP pre-insertion (\( \text{II} \rightarrow \text{III} \)): \( \Delta G^{\text{NTP}} = k_BT \ln \frac{k^0_{NTP}[\text{NTP}]}{k_{3-}} \sim 3.9 \text{ k}_B \text{T} \) (note that for \( \text{II}' \rightarrow \text{III} \), \( \Delta G^{\text{NTP}} \) is \( 2 \text{ k}_B \text{T} \) less as that was spent in translocation); NTP insertion (\( \text{III} \rightarrow \text{IV} \)): \( \Delta G^{\text{insert}} = k_BT \ln \frac{k_{i+}}{k_{i-}} \sim 0.05 \text{ k}_B \text{T} \) (almost no free energy change); the polymerization /catalysis (\( \text{IV} \rightarrow \text{V} \)): \( \Delta G^{\text{catalysis}} = k_BT \ln \frac{k_{6+}}{k_{5-}} \sim 2 \text{ k}_B \text{T} \); PPI release (\( \text{V} \rightarrow \text{I} \)):

\[
\Delta G^{\text{ppi}} = k_BT \ln \frac{k_{i+}}{k^0_{ppi}[\text{PPI}]} \sim 9.4 \text{ k}_B \text{T} \].
\]

Hence, \( \Delta G_{\text{total}} = \Delta G_0 + k_BT \ln \frac{[\text{NTP}]}{[\text{PPi}]} \sim 15 \text{ k}_B \text{T} \) 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 unbinding rate, $k_{3-}$ than the correct nucleotide (see Table S2). Following pre-insertion, insertion of the nucleotide (III $\rightarrow$ 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 $\rightarrow$ 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 $\eta_i^- = \frac{k_{iw}}{k_{ic}}$ (i = 3, 4) and $\eta_i^+ = \frac{k_{i+}}{k_{i+}}$ (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 $\eta_3^-$ or $\eta_4^-$ becomes larger than 1. On the other hand, when the wrong nucleotide has a reduced rate of the insertion or polymerization, $\eta_4^+$ or $\eta_5^+$ 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 ($p_{error}$, $p_{error}$, $p_{error}$), with the final error rate $p_{error} = p_{error}$. 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 $\eta_i^- = 10$ to 100 and/or $\eta_i^+ = 0.1$ to 0.01. $\eta_i^-$=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$A_0^0 \equiv ln10\sim 2.3 k_B T$). While $\eta_i^+=0.1$ 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 $A_0^0$).

First we compare cases #1 to 4. The individual selection mechanisms use the same amount of energy in selection (2$A_0^0$). 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.
Table S2  Error rates of elongation simulated under different nucleotide selection mechanisms.

<table>
<thead>
<tr>
<th>Index</th>
<th>$\eta_i^{-} = \frac{k_i^w}{k_i^{c'}}$</th>
<th>$\eta_i^{+} = \frac{k_i^{c}}{k_i^{c'}}$</th>
<th>Elongation rate (nt/s)</th>
<th>$p_{error}^{III}$</th>
<th>$p_{error}^{IV}$</th>
<th>$p_{error}^{V}$</th>
<th>Order of error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\eta_i^{-} = 1$</td>
<td></td>
<td>124</td>
<td>75%</td>
<td>74%</td>
<td>74%</td>
<td>$10^{-1}$~$10^{0}$</td>
</tr>
<tr>
<td>1</td>
<td>$\eta_i^{-} = 100$</td>
<td>$\eta_i^{+} = 0.01$</td>
<td>106</td>
<td>5.1%</td>
<td>5.2%</td>
<td>5.4%</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>2</td>
<td>$\eta_i^{-} = 100$</td>
<td>$\eta_i^{+} = 0.01$</td>
<td>27</td>
<td>84%</td>
<td>5.0%</td>
<td>5.2%</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>3</td>
<td>$\eta_i^{-} = 100$</td>
<td>$\eta_i^{+} = 0.01$</td>
<td>34</td>
<td>84%</td>
<td>22%</td>
<td>23%</td>
<td>$10^{-1}$</td>
</tr>
<tr>
<td>4</td>
<td>$\eta_i^{-} = 100$</td>
<td>$\eta_i^{+} = 0.01$</td>
<td>18</td>
<td>84%</td>
<td>97%</td>
<td>22%</td>
<td>$10^{-1}$</td>
</tr>
<tr>
<td>5</td>
<td>$\eta_i^{-} = 100$</td>
<td>$\eta_i^{+} = 0.01$</td>
<td>100</td>
<td>5.2%</td>
<td>0.06%</td>
<td>0.07%</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>6</td>
<td>$\eta_i^{-} = 10$</td>
<td>$\eta_i^{+} = 0.01$</td>
<td>78</td>
<td>35%</td>
<td>0.03%</td>
<td>0.01%</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>7</td>
<td>$\eta_i^{-} = 100$</td>
<td>$\eta_i^{+} = 0.1$</td>
<td>99</td>
<td>5.1%</td>
<td>0.4%</td>
<td>0.04%</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>8</td>
<td>$\eta_{3 \rightarrow 2}^{-} = 10$</td>
<td>$\eta_{3 \rightarrow 2}^{+} = 0.01$</td>
<td>99</td>
<td>5.7%</td>
<td>0.05%</td>
<td>0.05%</td>
<td>$10^{-4}$</td>
</tr>
</tbody>
</table>

In case #5 we show an efficient two-step selection mechanism that leads to an error rate $\sim 10^{-4}$ 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., $\eta_{3 \rightarrow 2}^{-} = 10$. If the rejection involves Tyr 639 ‘sensing’ (III $\rightarrow$ II’), the rejection becomes stronger, e.g., $\eta_{3 \rightarrow 2}^{-} = 100$. Note that in cases #1 to 7, however, we use $\eta_{3 \rightarrow 2}^{-} = \eta_{4 \rightarrow 2}^{-} = \eta_5^+$, 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 $E_{DNA}^i$ 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:

$$
\Delta G_{DNA}^i = E_{DNA}^i - E_{DNA}^{i+12}
$$

(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 $-E_{hybrid}^{i+3}$; while the 1-nt DNA nucleotide adjacent to the 3′-end of the RNA is at position $i$+10. The free energy change is:

$$
\Delta G_{hybrid}^i = \frac{1}{2} E_{hybrid}^{i+10} - E_{hybrid}^{i+3}
$$

Therefore, taking into account of sequence effects for each step of translocation, the free energy change $\gamma \equiv E_{II} - E_i$ along path $I \rightarrow II$ includes the sum of the two free energy terms above:

$$
\gamma = \Delta G_{DNA}^i + \Delta G_{hybrid}^i.
$$

At the same time, we also consider the facilitated translocation path $I' \rightarrow II'$ (see Figure 2 in main text). Comparing to path $I \rightarrow II$, 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 $E_{III} - E_{II} = \frac{1}{2} E_{CG}^{i+10}$. Since $-\beta \equiv E_{III} - E_i = \gamma + \frac{1}{2} E_{CG}^{i+10}$ and $E_{CG}^{i+10} \sim -4 k_B T$, the estimation is consistent with our fitted results: $\beta \sim 2 k_B T, \gamma = 0$. 

12
In Figure S2, we show values of $\Delta G_{\text{DNA}}$, $\Delta G_{\text{hybrid}}$, $\gamma$ 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 $\Delta G_{\text{DNA}}$, $\Delta G_{\text{hybrid}}$, $\gamma$ 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 ($I \rightarrow II$) is chosen as a constant $r_0 \sim 5000 \text{ s}^{-1}$(see Table S1). Under the sequence effects, the rate is adjusted to
\[ r_f^i \sim 50000 \exp \left( -\frac{\Delta E}{k_B T} \right), \]

where \( \Delta E \) 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. \( \Delta E = E_{\text{hybrid}}^{i+3} \). 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 \( I' \rightarrow II' \), the forward tracking is allowed at \( II' \). In the pure Brownian ratchet case (with \( I \rightarrow II \) 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 \( r_0^i \). We used an adjustable factor \( c_f \) (< 1) to modulate the forward tracking rate as \( c_f r_0^i \). For example, in simulation at T-\( \Phi \), \( c_f = 0.1 \) gives \( \sim80\% \) 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**


