1. Histogram of Instantaneous Rates

Figure 1e represents data collected at 11 nM of HIV RT concentration and 3.0 pN of stretching force. The Gaussian fit has a peak at 1.2 ± 0.16 nt/s, and lognormal fit has a mean at 26.6 ± 3.92 nt/s (see below for error estimation). Instantaneous rates of "inactive" raw trajectories (DNA-tethered beads that do not show any enzymatic activity) are calculated in a similar fashion. The histogram is plotted in Figure s1 with a simple Gaussian fit with a center at 0.047 ± 0.058 nt/s.

2. Enzyme dwell time for sequence dependent correlation analysis

See Figure s2 (a) and (b) as an example. We converted total 64 trajectories x(t) into dwell time t(x) by using 250-bp moving window. The average of 64 dwell time traces was used for correlation analysis. In order for this analysis to be valid, we ensured that each enzymatic trace contains a well-defined initiation site for enzymatic activity in the template.

3. Hairpin distribution on λ DNA

We examined the hairpin distribution on λ DNA and observed that hairpins in the initial part of the DNA (< 2500 nt) are distinct from hairpins in the other region. First, the average hairpin size in the initial region is large and the number of hairpins in the same region is small (note: in Fig. 2a, the number of bases in hairpin stems is not very different from the initial part of DNA and the rest of the DNA). Second, hairpins that are AT rich in the initial portion of the stem (and therefore easy to disrupt or may fluctuate with smaller stems than initially predicted) are concentrated in the initial part of the template (See Fig. s3). We speculate that (1) presence of a fewer hairpins and (2) concentration of hairpins with AT rich in the stem result in a less correlated pattern shown in <2500 nt (Fig. 2 and 5a).

4. HIV-1 RT does not wait for complete opening of each hairpin by thermal fluctuations

We were interested in determining whether the duration of slow synthesis phase involves the time for which an enzyme waits until a hairpin opens completely by thermal fluctuations. We investigated this possibility by performing a correlation analysis between experimental dwell time data and the lifetime of hairpins in closed form. The hairpin lifetime is defined as the average timescale for hairpin opening due to thermodynamic fluctuations and is calculated from $e^{-\Delta G_{formation}/k_bT} / k_{formation}$, where $k_{formation}$ is the hairpin formation rate. Kinetic studies of hairpin folding have shown that the hairpin formation rate ($k_{formation}$) scales with the size of hairpin loop (N_{loop}) and stem (N_{stem}) because hairpin formation is related to the probability of end-to-end collision for the DNA strand [1,2]. We used the scaling relationship $k_{formation} \propto (N_{loop} + N_{stem})^{-\alpha}$ ($\alpha = 2.3$), which fit previous experimental folding data well [1,2], and we calculated hairpin lifetimes as a function of template sequence by summing local lifetimes within a 250-bp moving window. In this manner, hairpin lifetimes represent the average time required for an enzyme to pause upstream of a stable hairpin before the hairpin is opened by thermal fluctuations. Using this analysis, we found that there is essentially no correlation between RT dwell times and hairpin lifetimes. While $k_{formation}$ displays a small variation among hairpins, there was a large variation in $e^{-\Delta G_{formation}/k_bT}$. For example, the strongest hairpin, which has 19 bases in stem and 17 bases in loop, shows 13 orders of magnitude larger $e^{-\Delta G_{formation}/k_bT}$ than the weakest hairpin, yielding poor correlation between RT dwell time and hairpin lifetimes. Also, its estimated lifetime is 5000 times longer than the second most stable hairpin (22-base stem and 21-base loop). We do not observe such a large variation among the durations of slow synthesis phase. Hence, it is unlikely that kinetics of thermal hairpin opening limits the duration of slow synthesis phase.

5. Chi square methods

After discounting the possibilities of enzyme dissociation-rebinding and thermal hairpin opening as major kinetic steps, we were assured that the dominant kinetics of the slow synthesis phase represents enzymatic

strand displacement activity. Consequently, we fit our data with a model whereby HIV-1 RT performs primer extension rate (*a*) at regular bases and strand displacement rate (*b*) at hairpin stem bases.
We performed chi square fitting or weighted least square fitting [3]. Our model predicts enzyme dwell time,

$$T(x_i) = \frac{250 - stem(x_i)}{a} + \frac{stem(x_i)}{b}$$

where stem(x_i) is the number of bases in hairpin stems within moving window x_i along the template, *a* is primer extension rate (nt/s), and *b* is strand displacement rate (nt/s). We found parameters (*a*, *b*), which minimizes chi-square

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{t(x_i) - T(x_i)}{\sigma_i}\right)^2$$

where t(x_i) is enzyme dwell time in 250-bp moving window, and σ_i is standard error in t(x_i). We used standard deviation in t(x_i) in 64 traces for the standard error, σ_i. Uncertainties in the (*a*, *b*) estimates are calculated as:

$$\sigma_a^2 = \sum_{i=1}^{n} \sigma_i^2 \left(\frac{\partial a}{\partial t_i}\right)^2$$

$$\sigma_b^2 = \sum_{i=1}^{n} \sigma_i^2 \left(\frac{\partial b}{\partial t_i}\right)^2$$

Goodness of fit is calculated from:

$$Q\left(\frac{n-2}{2}, \frac{\chi^2}{2}\right)$$

where Q is incomplete gamma function [3]. Q value for the fitting was 0.005, which asserts that the fit is acceptable. See Figure s4 for fitting. The correlation of the two curves are 0.75. The largest discrepancy occurs where the most pronounced hairpin is located in λ DNA. We suspect that it is likely that hairpin structure may switch during strand displacement synthesis as reported by Suo, et al [4,5] and appearance of unpredicted hairpins will increase the dwell time of HIV-1 RT in certain regions of the template.

6. Calculation of the force dependent primer extension rate (k_{pe})

Instantaneous rates in raw trajectories are calculated by a least-square fitting approach as described in the analysis of plateau durations. Histogram of all instantaneous rate values includes a dominant peak around 0 nt/s due to plateaus and residual tail around 20-25 nt/s. We calculated the average primer extension rate by excluding the peak due to effective plateaus from the original histogram. First, we fit Gaussian function to the histogram. The center of the Gaussian is around 0 nt/s and optimized such that when original histogram is subtracted with the Gaussian fit, population of around 0 nt/s is minimized. Then, the residual population is fit with a lognormal function and the average of the residual population (due to DNA synthesis on single-stranded region of template) is calculated from the mean of lognormal function. This is how we calculated the average rate of polymerization, 22.1 nt/s at 3.7 pN, and other data points shown in Figure 6b. All the fitting was done with curve fitting function in IgorPro.
Error in estimating the average rate of DNA polymerization – We calculated the error in the estimate of the average rate (at a given force) in a way Maier, B. *et al* used in their DNA polymerase experiments [6]. There could be two contributions in the error: (i) statistical error from the broadness of rate distribution, and (ii) a systematic error from the accuracy of the assay while converting the measured length changes to the number of nucleotides synthesized. Hence, the error follows the equation:

$$\sigma_{final}^2 = \frac{\sigma_v^2}{J} + \left\{\frac{\langle v \rangle \varepsilon}{\Delta l(F)}\right\}^2$$

where σ_v^2 is variation in the rate distribution, J is the number of independent measurements, ε is the precision of the assay, $\Delta l(F)$ is the difference in length between ssDNA and dsDNA under a given force, F. We find that the second term is negligible compared to the first term in our experiments.
The variation in the rate distribution is the broadness of the lognormal function.

$$\text{Lognormal function: } y = y_0 + A\exp\left[-\left(\frac{\ln x - \ln x_0}{w_0}\right)^2\right]$$

$$\text{Mean: } \exp\left(\frac{3}{4}w_0^2\right)\cdot x_0$$

$$\text{Variation: } x_0^2\cdot\left(e^{2w_0^2} - e^{1.5w_0^2}\right)$$

However, the lognormal fit to the DNA polymerization rate histogram depends on (i) window size in the least-square fitting approach for calculating instantaneous rate in trajectories, (ii) initial Gaussian fit to the rates around 0 nt/s, and (iii) subsequent lognormal fit to the residual population of polymerization rate in the histogram. We find that variations in (i)-(iii) by reasonable amounts result in less than 2 nt/s variation in the estimate of the average rate, and 2 nt/s variation is uniform over all F points.

7. Calculation of ΔG_F

Contribution of template stretching force in effective base pairing energy (ΔG_{bp}), ΔG_F, is sum of (1) work performed by force, F to open one base and (2) loss of entropy upon stretching two newly opened bases [7]. We denote z(F) the length of one ssDNA base at a stretching force of F, and z(F) corresponds to the force extension curve of ssDNA normalized with the total number of bases (ref. Fig. 1b). The loss of entropy comes from the area under the ssDNA force extension curve.

$$\Delta G_F = F\cdot 2z(F) - \int_0^{z(F)} F'2dz(F') = 2\int_0^F z(F')dF'$$

While force extension curve for dsDNA can be readily expressed by worm-like chain polymer model [8], force extension curve for ssDNA is rather complex and is dependent on the base sequence and buffer conditions [9,10]. There is no analytic expression for force extension curve of ssDNA experimentally measured in our assay (Fig. 1b). Hence, we calculated ΔG_F by summing the area under the ssDNA force extension curve.

8. Notes on Fig. 6

In Figure 6a, we assumed average dwell time within the window of 250 bp is composed of duration of primer extension (fixed rate of 18.7 ± 6.0 nt/s) and strand displacement. Having the number of stem bases within the window, we could estimate duration and rate of strand displacement within the window. GC ratio is calculated from the sequence of stem bases within the window.

We estimated uncertainty in strand displacement rate using errors in average dwell time (σ_i) and primer extension rate. We repeated calculation of strand displacement rate with different average dwell time and primer extension rate within their error range. The calculated uncertainty interval of strand displacement rate is taken as error bar in Figure 6a.

The comparison of our data with the model developed to explain the unzipping activity of DNA helicases is based on several assumptions. We assumed that translocation and addition of one base by HIV-1 RT is preceded by duplex junction opening by *n* bases. Junction opening kinetics may be explained by recent theoretical model [11], which assumes the same base pairing energy throughout *n* bases (neglecting sequence effects within the step size). We also neglected possible interaction between secondary structures or DNA secondary structure switching during strand displacement synthesis [4,5].

9. Force dependent strand displacement rate

We estimated force dependent strand displacement rate at around 3700 nt template position in order to avoid sequence dependent variation. The rate is estimated based on the same assumptions used for estimation of strand displacement rate vs GC ratio. Force dependent primer extension rate, calculated from histogram of instantaneous rate, was used to estimate the strand displacement rate around 3700 nt. Error bars in Figure 6c are the sum of the statistical error and propagation error from force dependent primer extension rate.

Reference

1. Bonnet, G., Krichevsky, O. & Libchaber, A. (1998). Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc. Natl Acad. Sci. USA* **95**, 8602-8606.
2. Jung, J. & van Orden, A. (2005). Folding and unfolding kinetics of DNA hairpins in flowing solution by multiparameter fluorescence correltaion spectroscopy. *J. Phys. Chem. B* **109**, 3648-3657.
3. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical recipes in C*, Cambridge University Press, New York.
4. Suo, Z. & Johnson, K. A. (1997). Effect of RNA secondary structure on the kinetics of DNA synthesis catalyzed by HIV-1 reverse transcriptase. *Biochemistry* **36**, 12459-12467.
5. Suo, Z. & Johnson, K. A. (1998). DNA secondary structure effects on DNA synthesis catalyzed by HIV-1 reverse transcriptase. *J. Biol. Chem.* **273**, 27259-27267.
6. Maier, B., Bensimon, D. & Croquette, V. (2000). Replication by a single DNA polymerase of a stretched single-stranded DNA. *Proc. Natl Acad. Sci. USA* **97**, 12002-12007.
7. Lionnet, T., Spiering, M. M., Benkovic, S. J., Bensimon, D. & Croquette, V. (2007). Real-time observation of bacteriophage T4 gp41 helicase reveals an unwinding mechanism. *Proc. Natl Acad. Sci. USA* **104**, 19790-19795.
8. Bustamante, C., Marko, J. F., Siggia, E. D., and Smith, S. (1994). Entropic elasticity of lambda-phage DNA. *Science* **265**, 1599.
9. Dessinges, M. N., Maier, B., Zhang, Y., Peliti, M., Bensimon, D. & Croquette, V. (2002). Stretching single stranded DNA, a model polyelectrolyte. *Physical Review Letters* **89**, 248102.
10. Zhang, Y., Zhou, H. & Ou-Yang, Z.-C. (2001). Stretching single-stranded DNA: Interplay of electrostatic, base-pairing, and base-pair stacking interactions. *Biophys. J* . **81**, 1133-1143.
11. Betterton, M. D. & Julicher, F. (2005). Opening of nucleic-acid double strands by helicases: active versus passive opening. *Phys. Rev. E.* **71**, 011904.

Fig. s1

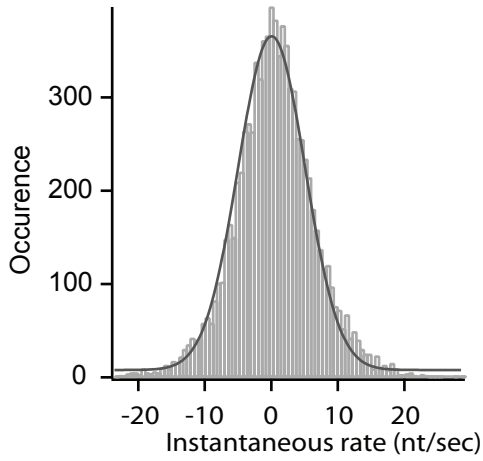

Figure s1 | Histogram of instantaneous slopes from "inactive" DNA traces. A simple Gaussian fit has a center at 0.047 ± 0.058 nt/sec

Fig. s2

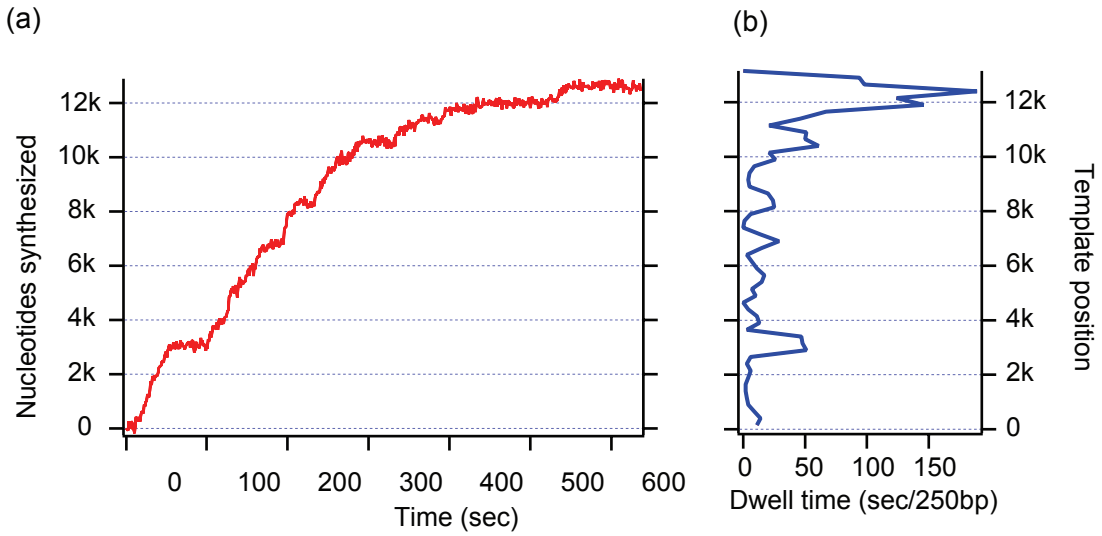

Figure s2 | DNA sequence dependent analysis of enzyme dwell time (a) A raw time trajectory of HIV-1 RT DNA polymerization. (b) Dwell time distribution of trace shown in (a).

Fig. s3

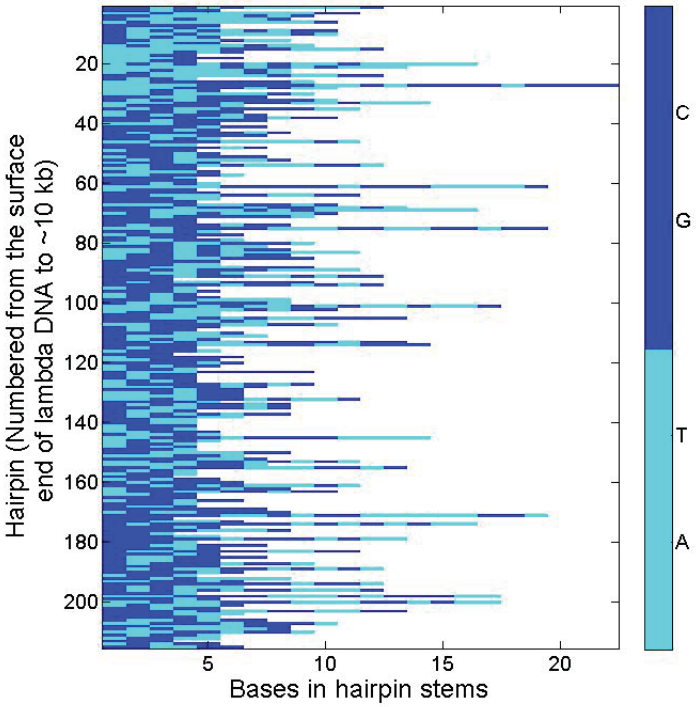

Figure s3 | Color map showing hairpin stem bases for each hairpin analyzed for correlation in Fig 2 and 5a. Each row represents a stem base sequence of a hairpin (AT are in grey and GC are in black). Hairpins are presented in sequence of their appearance in the lambda DNA. Hairpin 1-43 form in the initial part of DNA (<2500 nt). Clearly, several hairpins that start with AT rich bases are accumulated in the beginning of the DNA template.

Fig. s4

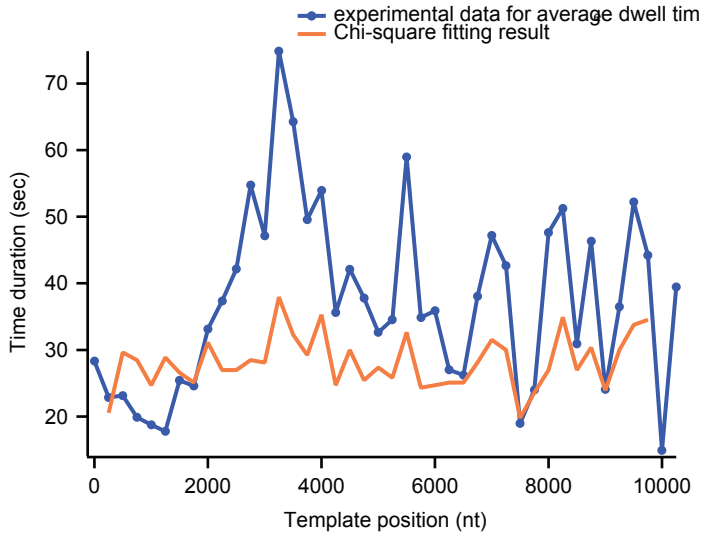

Figure s4 | Chi square fitting to experimental average dwell time Orange curve show the chi-square fitting result with primer extension rate at 18.7 nt/sec and strand displacement rate at 2.3 nt/sec