Supporting Information for:

Kinetic Mechanisms Governing Stable Ribonucleotide Incorporation in Individual DNA Polymerase Complexes

Joseph M. Dahl, Hongyun Wang, José M. Lázaro,
Margarita Salas, and Kate R. Lieberman

Supporting Information Text

I. Theoretical expression of the normalized p/(1-p)

We consider the model shown in Figure 1a in the main text. Let

p_{U1} = equilibrium probability of the pre-translocation state (upper amplitude)
p_{U2} = equilibrium probability of the exonuclease state (upper amplitude)
p_{L1} = equilibrium probability of the unbound post-translocation state (lower amplitude)
p_{L2} = equilibrium probability of the d/rGTP bound post-translocation state (lower amplitude)

The observed probability of lower amplitude, p, is the sum over the 2 lower amplitude state.

$$p = p_{L1} + p_{L2} \tag{1}$$

At equilibrium, we have

$$p_{U2} = p_{U1} \cdot \frac{r_3}{r_4} \tag{2}$$

$$p_{U1} = p_{L1} \cdot \frac{r_2}{r_1} \tag{3}$$

$$p_{L2} = p_{L1} \cdot \frac{[\text{d/rGTP}]}{K_d} \tag{4}$$

We express everything in terms of p_{L1}.

$$p_{U1} = p_{L1} \cdot \frac{r_2 r_3}{r_1 r_4} \tag{5}$$

We solve for p_{L1} using that the total probability is 1.

$$p_{U1} + p_{U2} + p_{L1} + p_{L2} = 1 \tag{6}$$

$$\Longrightarrow \quad p_{L1}\left(\frac{r_2 r_3}{r_1 r_4} + \frac{r_2}{r_1} + 1 + \frac{[\text{d/rGTP}]}{K_d}\right) = 1 \tag{7}$$

$$\Longrightarrow \quad p_{L1} = \frac{1}{\left(\frac{r_2 r_3}{r_1 r_4} + \frac{r_2}{r_1} + 1 + \frac{[\text{d/rGTP}]}{K_d}\right)} \tag{8}$$

The quantity p/(1-p) has the expression

$$\frac{p}{1-p} = \frac{p_{L1} + p_{L2}}{1 - (p_{L1} + p_{L2})} = \frac{1 + \frac{[\text{d/rGTP}]}{K_d}}{\frac{r_2 r_3}{r_1 r_4} + \frac{r_2}{r_1}} \tag{9}$$

At [d/rGTP] = 0, the quantity p/(1-p) is

$$\left.\left(\frac{p}{1-p}\right)\right|_{[\mathrm{d/rGTP}]=0} = \frac{1}{\frac{r_2 r_3}{r_1 r_4} + \frac{r_2}{r_1}} \tag{10}$$

The theoretical expression for the normalized p/(1-p) is

$$\frac{\left(\frac{p}{1-p}\right)}{\left.\left(\frac{p}{1-p}\right)\right|_{[\mathrm{d/rGTP}]=0}} = 1 + \frac{[\mathrm{d/rGTP}]}{K_d} \tag{11}$$

The study of normalized p/(1-p) depends on the finding that the translocation rates r_1 and r_2 are not affected by the dNTP, which was established in a previous study [1]. The normalized p/(1-p) is an equilibrium property employed as an initial scan of the effects of mutants, dNTP binding and rNTP binding.

II. Survival probability and dwell time distribution of the lower amplitude

To study the survival probability and dwell time distribution, we consider the escape problem of exiting from the lower amplitude. Let

$S_{L1}(t)$ = probability of remaining in the lower amplitude beyond t and being in the unbound post-translocation state at time t

$S_{L2}(t)$ = probability of remaining in the lower amplitude beyond time t and being in the d/rGTP bound post-translocation state at time t

The survival probability of the lower amplitude is

$$S(t) = S_{L1}(t) + S_{L2}(t)$$

Based on the model shown in Figure 1a in the main text, $S_{L1}(t)$ and $S_{L2}(t)$ are governed by

$$\frac{d}{dt}\begin{pmatrix} S_{L1}(t) \\ S_{L2}(t) \end{pmatrix} = -\begin{pmatrix} \left(r_2 + k_{on}[\mathrm{d/rGTP}]\right) & -k_{off} \\ -k_{on}[\mathrm{d/rGTP}] & k_{off} \end{pmatrix}\begin{pmatrix} S_{L1}(t) \\ S_{L2}(t) \end{pmatrix} \tag{12}$$

The matrix has two distinct real eigenvalues.

$$\lambda_1 = \frac{\left(r_2 + k_{on}[\mathrm{d/rGTP}] + k_{off}\right) + \sqrt{\left(r_2 + k_{on}[\mathrm{d/rGTP}] + k_{off}\right)^2 - 4r_2 k_{off}}}{2} \tag{13}$$

$$\lambda_2 = \frac{\left(r_2 + k_{on}[\mathrm{d/rGTP}] + k_{off}\right) - \sqrt{\left(r_2 + k_{on}[\mathrm{d/rGTP}] + k_{off}\right)^2 - 4r_2 k_{off}}}{2} \tag{14}$$

Both $S_{L1}(t)$ and $S_{L2}(t)$, and the sum $S(t) = S_{L1}(t) + S_{L2}(t)$ have the general form

$$S(t) = a \cdot \exp(-\lambda_1 t) + b \cdot \exp(-\lambda_2 t) \tag{15}$$

The coefficients a and b are constrained by the 2 conditions below:

$$S(0) = 1 \tag{16}$$

$$S'(0) = -r_2 \tag{17}$$

Solving for a and b, we obtain

$$\begin{cases} a = \dfrac{r_2 - \lambda_2}{\lambda_1 - \lambda_2} \\ b = 1 - a \end{cases} \tag{18}$$

The survival probability is

$$S(t) = a \cdot \exp(-\lambda_1 t) + (1-a) \cdot \exp(-\lambda_2 t) \tag{19}$$

The probability density of the dwell time is

$$\rho(t) = -S'(t) = a \cdot \lambda_1 \exp(-\lambda_1 t) + (1-a) \cdot \lambda_2 \exp(-\lambda_2 t) \tag{20}$$

The survival probability and dwell time distribution of the upper amplitude state has the same mathematical structure.

III. Determining kinetic rates from observed dwell time samples

The observed dwell time samples represent the dwell time beyond the cut-off threshold t_C. Mathematically, the probability density for the observed dwell time samples is

$$\rho_{\text{Observed}}(t) = \frac{1}{S(t_C)} \rho(t + t_C) \tag{21}$$

$$= a_{\text{Observed}} \cdot \lambda_1 \exp(-\lambda_1 t) + (1 - a_{\text{Observed}}) \cdot \lambda_2 \exp(-\lambda_2 t) \tag{22}$$

where coefficient a_{Observed} is related to coefficient a as

$$a_{\text{Observed}} = \frac{a \cdot \exp(-\lambda_1 t_C)}{a \cdot \exp(-\lambda_1 t_C) + (1-a) \cdot \exp(-\lambda_2 t_C)} \tag{23}$$

Observed dwell time samples are fitted to the probability density (22) given above using the Maximum Likelihood Estimation (MLE) to calculate the intermediate parameters.

$$\left\{ \begin{matrix} \text{Observed dwell} \\ \text{time samples} \end{matrix} \right\} \xrightarrow{\text{MLE}} \{a_{\text{Observed}}, \lambda_1, \lambda_2\} \tag{24}$$

Coefficient a is calculated from a_{Observed} using the equation

$$a = \frac{a_{\text{Observed}} \cdot \exp(\lambda_1 t_C)}{a_{\text{Observed}} \cdot \exp(\lambda_1 t_C) + (1 - a_{\text{Observed}}) \cdot \exp(\lambda_2 t_C)} \tag{25}$$

Once the intermediate parameters $\{a, l_1, l_2\}$ are determined, we calculate the kinetic rates $\{r_2, k_{on}, k_{off}\}$ using the 3 equations below

$$r_2 = a \cdot \lambda_1 + (1-a)\lambda_2 \tag{26}$$

$$k_{on} = \frac{1}{[\text{d/rGTP}]} \cdot \frac{a(1-a)(\lambda_1 - \lambda_2)^2}{a \cdot \lambda_1 + (1-a)\lambda_2} \tag{27}$$

$$k_{off} = \frac{\lambda_1 \lambda_2}{a \cdot \lambda_1 + (1-a)\lambda_2} \tag{28}$$

These 3 equations are straightforward to verify using (13), (14) and (18).

IV. Transition in the slope of log(survival probability)

We examine the slope of log(S(t)).

$$-\frac{d}{dt} \log(S(t)) = \frac{-1}{S(t)} S'(t) = r_2 \frac{S_{L1}(t)}{S(t)} \tag{29}$$

Quantity $Q(t) = \frac{S_{L1}(t)}{S(t)}$ is the fraction of the survival probability in the unbound post-translocation state, from which the complex can fluctuate to the pre-translocation state (and thus, exit from the lower amplitude level).

At time $t = 0$, Q(0) = 1. In the presence of dGTP (or rGTP), as t increases quantity Q converges to Q(∞), a constant smaller than 1. Correspondingly, the slope of log(S(t)) decreases from r_2 at t = 0 to $r_2Q(\infty)$ at large time.

We study the time evolution of quantity Q(t). Based on the time evolution equation (12) for $S_{L1}(t)$ and $S_{L2}(t)$, we derive

$$\frac{d}{dt}Q(t) = \frac{S_{L1}{}'(t)}{S(t)} - S_{L1}(t)\frac{S'(t)}{S^2(t)} \tag{30}$$

$$= \frac{S_{L1}{}'(t)}{S(t)} - Q\cdot\frac{S'(t)}{S(t)} \tag{31}$$

$$= k_{off} - \left(r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off}\right)Q + r_2Q^2 \tag{32}$$

The governing equation for Q(t) is

$$\frac{d}{dt}Q(t) = F(Q) \tag{33}$$

where F(Q) has the expression

$$F(Q) = k_{off} - \left(r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off}\right)Q + r_2Q^2 \tag{34}$$

F(Q) has 2 stationary points:

$$F(Q) = r_2(Q - z_1)(Q - z_2) \tag{35}$$

$$z_1 = \frac{\left(r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off}\right) + \sqrt{\left(r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off}\right)^2 - 4r_2k_{off}}}{2r_2} \tag{36}$$

$$z_2 = \frac{\left(r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off}\right) - \sqrt{\left(r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off}\right)^2 - 4r_2k_{off}}}{2r_2} \tag{37}$$

Mathematically, z_1 and z_2 have the properties:

$$z_1 > \frac{\left(r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off}\right) + \sqrt{\left(r_2 - k_{on}\left[\text{d/rGTP}\right] - k_{off}\right)^2}}{2r_2} \geq 1 \tag{38}$$

$$z_2 < \frac{\left(r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off}\right) - \sqrt{\left(r_2 - k_{on}\left[\text{d/rGTP}\right] - k_{off}\right)^2}}{2r_2} \leq 1 \tag{39}$$

$$F'(z_1) = r_2(z_1 - z_2) = \sqrt{\left(r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off}\right)^2 - 4r_2k_{off}} > 0 \tag{40}$$

$$F'(z_2) = r_2(z_2 - z_1) = -\sqrt{\left(r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off}\right)^2 - 4r_2k_{off}} < 0 \tag{41}$$

It follows that z_1 is unstable and z_2 is stable. Consequently, Q(∞) is the same as z_2.

Function F(Q) is not linear. As a result, the decrease of Q from 1 to z_2 is not a simple exponential decay with a constant rate. F(Q) is quadratic with $F''(Q) = 2r_2 > 0$. In $[z_2, 1]$, F(Q) satisfies

$$F(Q) \geq F'(z_2)(Q - z_2) \tag{42}$$

Note that both sides of the inequality above are negative, corresponding to that Q decreases from 1 to z_2 in the dynamical system.
The inequality implies that the decrease of Q is slower than a simple exponential decay with rate $-F'(z_2)$, which is a constant and which is bounded by

$$-F'(z_2) = \sqrt{\left(r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off}\right)^2 - 4r_2k_{off}} \;\le\; r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off} \qquad (43)$$

Therefore, we conclude that the rate of Q relaxing to z_2 is bounded by $r_2 + k_{on}\left[\text{d/rGTP}\right] + k_{off}$.
The observed survival probability is

$$S_{\text{Observed}}(t) = \frac{S(t + t_C)}{S(t_C)} \qquad (44)$$

$$\Longrightarrow \quad \log\left(S_{\text{Observed}}(t)\right) = \log\left(S(t + t_C)\right) - \log\left(S(t_C)\right) \qquad (45)$$

Thus, the plot of $\log\left(S_{\text{Observed}}(t)\right)$ is the just the plot of $\log\left(S(t)\right)$ for $t > t_C$, horizontally and vertically shifted. The most important point is that in $\log\left(S_{\text{Observed}}(t)\right)$ we don't have the information of $\log\left(S(t)\right)$ for $t < t_C$. In particular, if any transition occurs in $\log\left(S(t)\right)$ in the time period $[0, t_C]$, we won't see it in the plot $\log\left(S_{\text{Observed}}(t)\right)$.

V. Review and summary of relevant results from our previous studies

A. The assignment of the biochemical identity of states for specific Φ29 DNAP kinetic transitions to the measured ionic current signal has been established:

1) Transitions across the translocation step were assigned to the fluctuations of ionic current between two amplitude levels [2]. The assignment was based upon i) two amplitude levels are observed only when an abasic reporter group is inserted in the template strand to couple the current amplitude to the DNA displacement inside the nanopore lumen; in the absence of an abasic reporter group in the template strand, only one amplitude level is observed. Crucially, these experiments ruled out the possibility that the observed two amplitude levels are attributed to conformational changes in the binary complex that alter the extent to which the complex impedes ion flow into the nanopore; ii) the distance of the DNA movement; and iii) the direct comparison of the amplitudes in the fluctuations to amplitudes traversed during processive synthesis catalyzed by Φ29 DNAP complexes atop the nanopore.

2) The dNTP binding was assigned to the emergence of a second dwell time cluster in the lower amplitude [1]. The assignment was based upon i) the strict dependence of the second dwell time cluster on the complementarity of dNTP to the templating base; ii) the dependence of the second dwell time cluster on the complementary dNTP concentration.

3) The primer strand transfer between the polymerase and exonuclease sites was assigned to the presence of a second dwell time cluster in the upper amplitude [3]based upon the behavior of Φ29 DNAP mutants known to perturb the primer strand transfer. Examination of ionic current traces recorded during exonucleolysis catalyzed by Φ29 DNAP complexes atop the nanopore verified that there is no unique amplitude associated with the state in which the primer strand occupies the exonuclease site.

B. We developed experimental methods to observe the DNA displacement in the transitions between the pre-translocation state to the post-translocation state [4]. Combining this experimental capability and a mathematical method, we studied the detailed kinetic relation of translocation and dNTP binding. We found that the translocation and dNTP binding are sequential. Specifically, dNTP binding can occur only in the post-translocation state; dNTP binding locked the complex in the post-translocation state; fluctuation back to the pre-translocation state can occur only after the dissociation of dNTP. We reconstructed rates of translocation and dNTP binding simultaneously from data [1].

The forward translocation rate, the reverse translocation rate, and the nucleotide triphosphate asscociation and dissociation rates are measured simultaneously in our studies. If an apparent dNTP binding rate is measured without resolving the translocation states, the apparent rate contains the effects of translocation rates and the dNTP binding rate. We further studied the effect of voltage (force) and dNTP concentrations on these rates. We found that the dNTP has no effect on the translocation rates and the voltage (force) has no effect on dNTP association or dissociation rates [1]. In contrast, the apparent dNTP binding rate measured without resolving the translocation states will be affected by the voltage (force) since the voltage changes the translocation rates. The actual binding rates in the post-translocation state can be affected by alterations in the structure of the nucleotide triphosphate, or by mutations in the enzyme.

C. The kinetic rates r_1, r_2, k_{on} and k_{off} are reconstructed from dwell time distributions. In the presence of dNTP, the dwell time distribution of the lower amplitude is a combination of 2 exponential modes. Mathematically, both modes vary with dNTP concentration. The mode with long dwell time is affected much more prominently by dNTP concentration. The long dwell time mode is caused by binding and repeated binding of dNTP. Given a set of dwell time samples, there is no clear-cut threshold for distinguishing which sample belongs to which mode.

In our method, we fit all dwell time samples to a combination of 2 exponential modes to obtain 3 intermediate parameters: decay rate of the first mode, decay rate of the second mode and the relative fraction of the first mode. From these 3 intermediate parameters and the fitting results for the upper amplitude dwell time samples, we calculate kinetic rates r_1, r_2, k_{on} and k_{off}. This set of 4 kinetic rates is calculated from data for each individual voltage and each individual dNTP concentration. In the calculation, there is no priori assumption about the independence of rates on voltage or dNTP. Rather, the conclusion that r_1 and r_2 are independent of dNTP and the conclusion that k_{on} and k_{off} are independent of voltage are established by comparing rates calculated from data corresponding to different voltages or different dNTP concentrations [1]. This mathematical method thus avoids the ambiguous task of classifying an individual dwell time sample as corresponding to “no dNTP binding”, “dNTP binding once”, or “repeated dNTP binding”.

REFERENCES

(1) Lieberman, K. R., Dahl, J. M., Mai, A. H., Cox, A., Akeson, M., and Wang, H. (2013) Kinetic mechanism of translocation and dNTP binding in individual DNA polymerase complexes. *J Am Chem Soc 135*, 9149–9155.

(2) Dahl, J. M., Mai, A. H., Cherf, G. M., Jetha, N. N., Garalde, D. R., Marziali, A., Akeson, M.,

Wang, H., and Lieberman, K. R. (2012) Direct observation of translocation in individual DNA polymerase complexes. *J Biol Chem 287*, 13407–13421.

(3) Lieberman, K. R., Dahl, J. M., and Wang, H. (2014) Kinetic mechanism at the branchpoint between the DNA synthesis and editing pathways in individual DNA polymerase complexes. *J Am Chem Soc 136*, 7117–7131.

(4) Lieberman, K. R., Dahl, J. M., Mai, A. H., Akeson, M., and Wang, H. (2012) Dynamics of the translocation step measured in individual DNA polymerase complexes. *J Am Chem Soc 134*, 18816–18823.

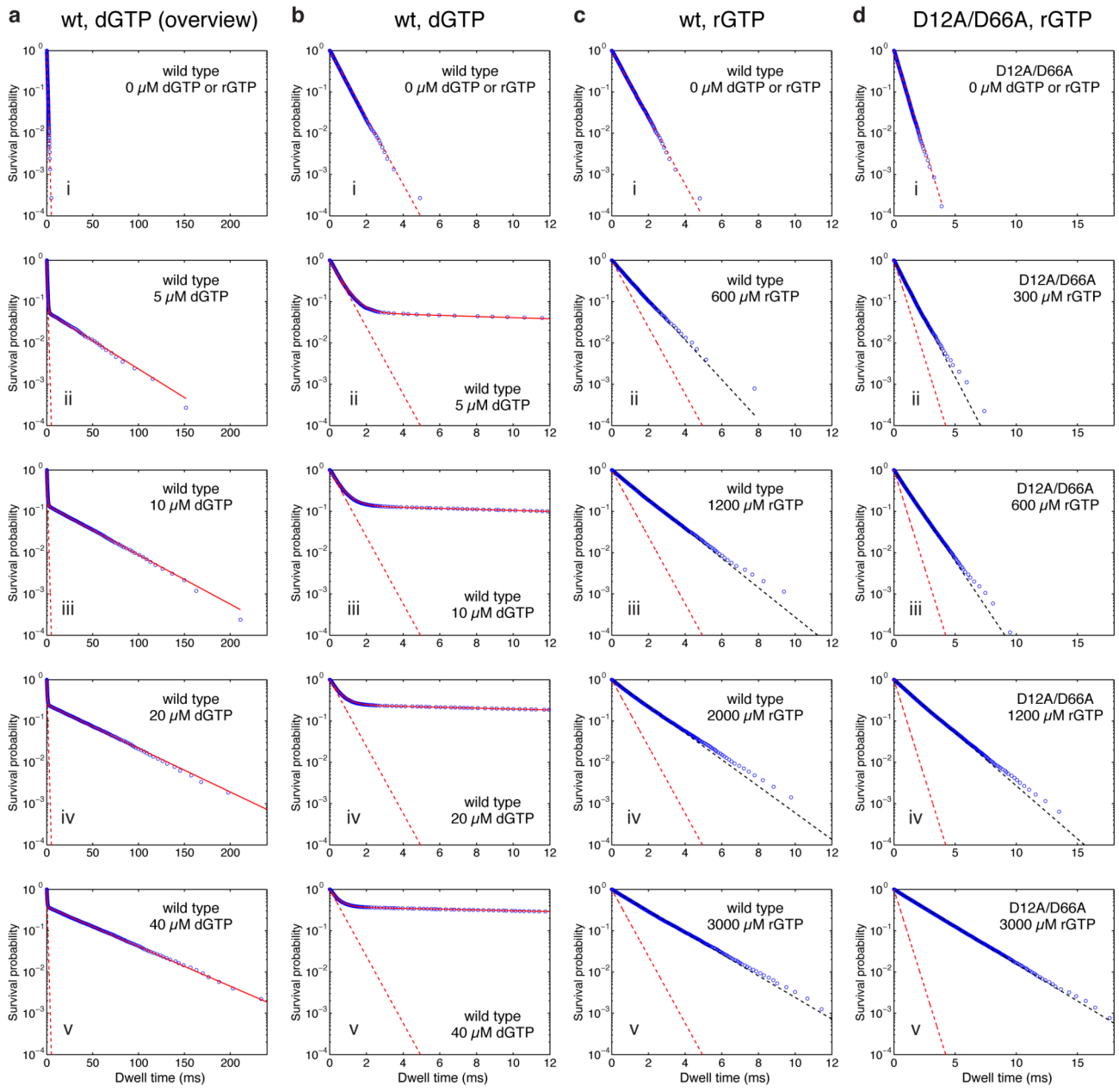

Figure S1. Effects of complementary dNTP or complementary rNTP on the lower amplitude dwell time distributions. Plots of log(survival probability) vs. dwell time for the lower amplitude level for complexes formed between DNA1-H_H and the wild type Φ29 DNAP (a-c) or the D12A/D66A mutant (d), showing the concentration-dependent effects of dGTP (a and b) or rGTP (c and d). In all plots, the dashed red line represents an exponential distribution with rate r_2, which is the constant slope of log(survival probability) at 0 μM dGTP or rGTP (see text), and which is obtained by fitting to the data at 0 μM dGTP or rGTP (panels a-d, i). The solid red fitting lines in panels a and b (ii-v) show the fit of the data to a model of two exponential modes. The dashed black fitting lines in

panels c and d (ii-v) show the fit of the data to an exponential distribution. The dwell time samples were extracted from data files collected when complexes were captured at 180 mV; each file yields ~8,000-80,000 dwell time samples for each amplitude level. In the plots, while 1 out of every 20 points is shown, the curves are fit to the full set of dwell time samples.

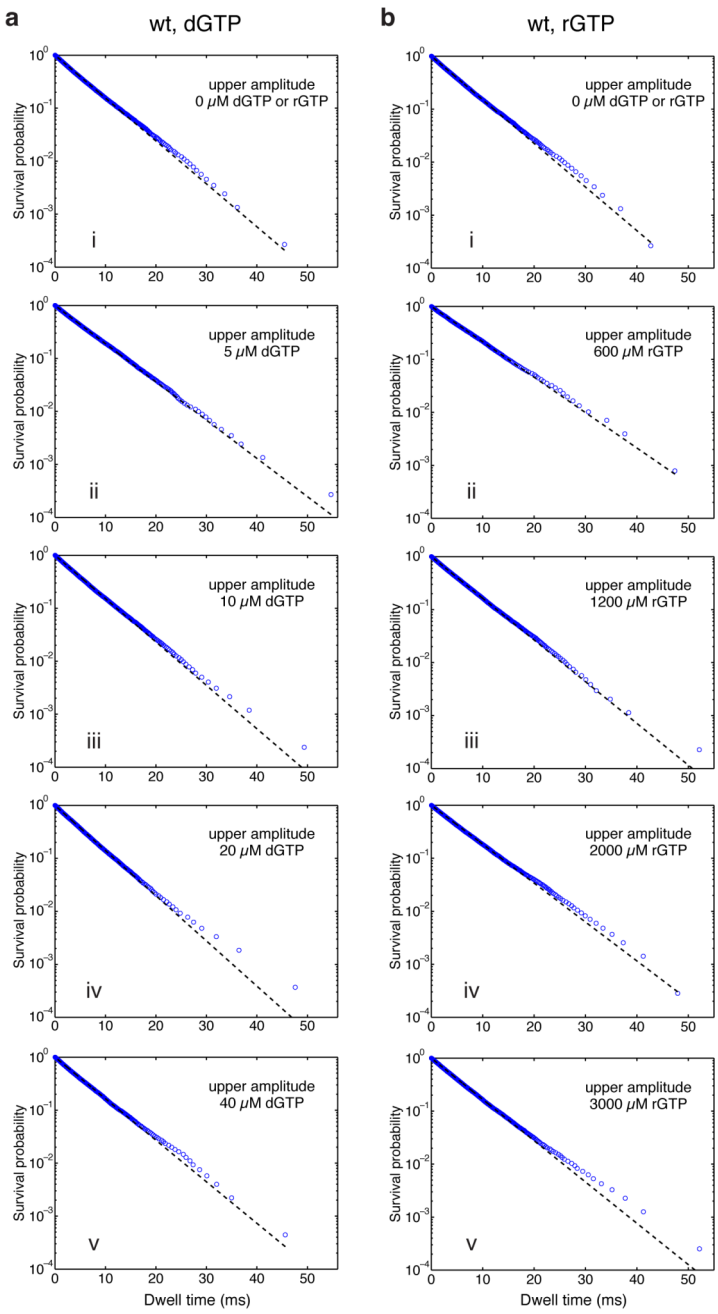

Figure S2. Effects of complementary dNTP or complementary rNTP on the upper amplitude dwell time distributions. Plots of log(survival probability) vs. dwell time for the upper amplitude level for complexes formed between DNA1-H_H and the wild type Φ29 DNAP, captured at 180 mV in the presence of increasing concentrations of (a) dGTP or (b) rGTP. The dashed black fitting lines show the fit to an exponential distribution. These data demonstrate that the presence of dNTP or rNTP does not affect the dwell time distribution of the upper amplitude level, confirming that dNTP or rNTP binds only after the translocation, when the complex is in the post-translocation state.

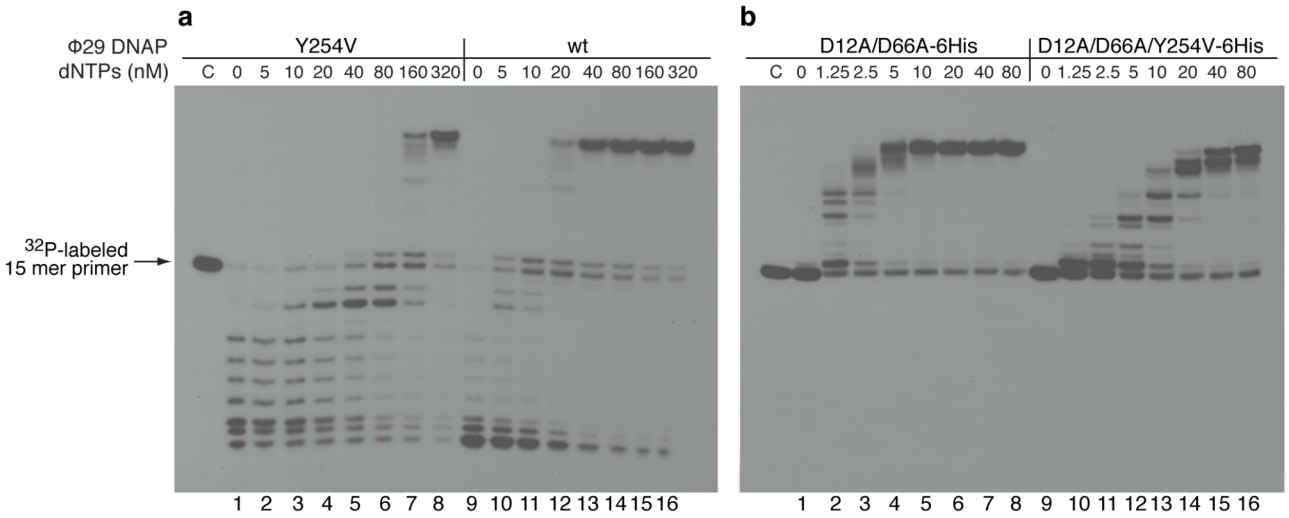

Figure S3. Effect of the Y254V mutation on the dNTP concentration-dependence of polymerization. (a) The functional coupling between DNA synthesis and exonucleolysis for the Y254V mutant Φ29 DNAP (lanes 1-8) or the wild type Φ29 DNAP (lanes 9-16) is shown as a function of the indicated concentration of dNTPs. (b) DNA synthesis catalyzed by the D12A/D66A-6His mutant (lanes 1-8) or the D12A/D66A/Y254V-6His mutant (lanes 9-16) as a function of the indicated concentration of dNTPs.

Reactions were conducted in 12.5 μl using 0.36 ng of a primer/template DNA substrate (^{32}P-labeled, 15 mer primer strand hybridized with a 33 mer template strand), in 50 mM Tris-HCl, pH 7.5, 1 mM DTT, 4% glycerol, 12.5 μg BSA, 20 ng of the indicated enzyme, and the indicated concentration of dNTPs. Reactions were initiated by the addition of $MgCl_2$ to a final concentration of 10 mM, incubated for 5 min at 30°C, and stopped with 3 μl of formamide buffer (95% formamide, 20 mM EDTA, 0.05% bromphenol blue and 0.05% xylene cyanol). Products were resolved by denaturing electrophoresis (20% PAGE/8M urea) in TBE buffer. Lanes labeled 'C' show the position of the primer strand in the unreacted DNA substrate.

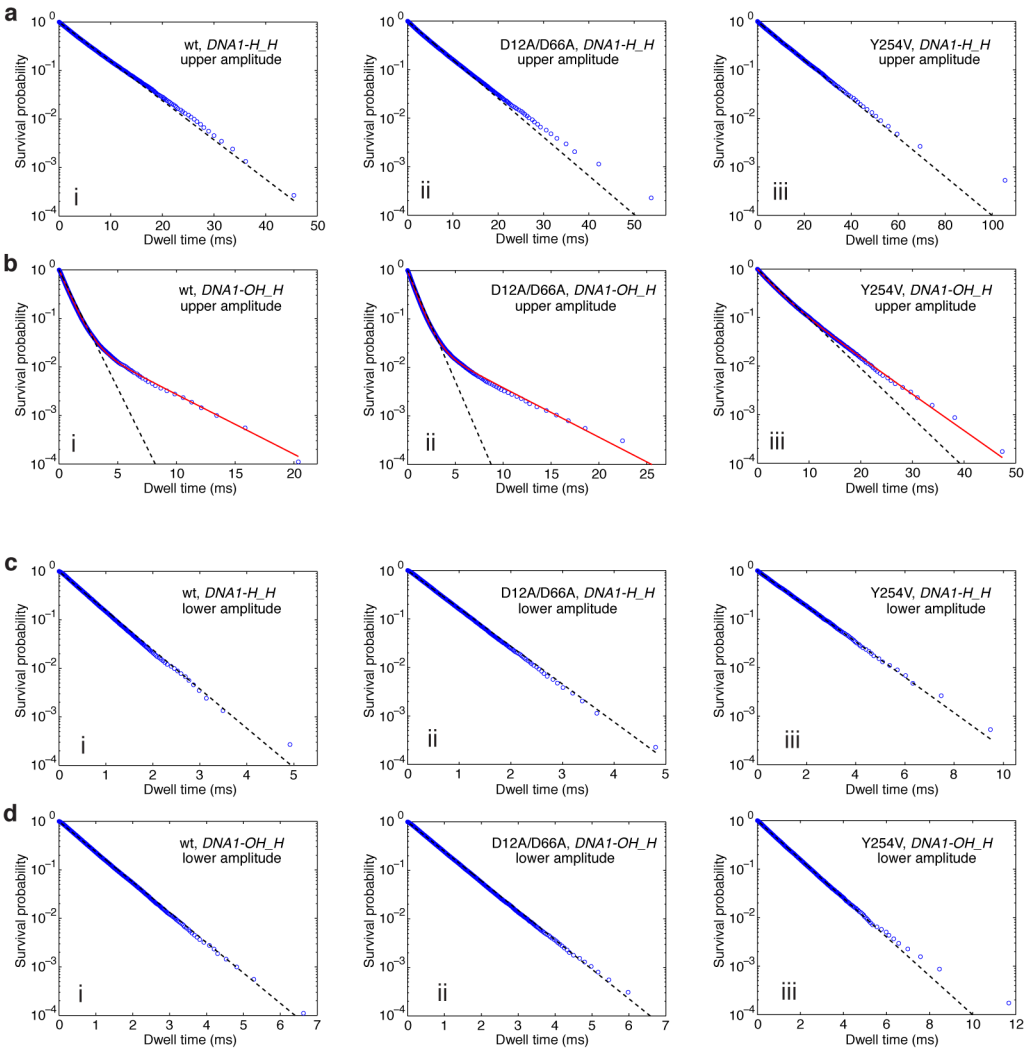

Figure S4. Effects of a primer terminal 2´-OH group on dwell time distributions at each of the two amplitude levels for wild type and mutant Φ29 DNAP binary complexes. Plots of log(survival probability) vs. dwell time for the upper amplitude (a and b), or the lower amplitude (c and d), for complexes formed between wild type (a-d, i), D12A/D66A (a-d, ii), or Y254V Φ29 DNAP (a-d, iii), and DNA1-H_H (a and c) or DNA1-OH_H (b and d). The dashed black fitting lines show the fit of the data to an exponential distribution. The solid red fitting lines in panels b, i-iii, show the fit of the data to a model of two exponential modes. Complexes were captured at 180 mV.