Supplemental Data for:

Direct Observation of Translocation in Individual DNA Polymerase Complexes

Joseph M. Dahl¹, Ai H. Mai¹, Gerald M. Cherf¹, Nahid N. Jetha⁴, Daniel R. Garalde³,
Andre Marziali², Mark Akeson¹, Hongyun Wang²*, and Kate R. Lieberman¹*

¹Department of Biomolecular Engineering, ²Department of Applied Mathematics and Statistics,
³Department of Computer Engineering, Baskin School of Engineering, University of California,
Santa Cruz, California, U.S.A. 95064
⁴Department of Physics and Astronomy, University of British Columbia, Vancouver, British
Columbia, Canada, V6T 1Z1

*corresponding authors: krlieberman@gmail.com, hongwang@soe.ucsc.edu
Section 1. Supplemental Data (Figures S1- S6)

Figure S1

A

phi29 DNAP-DNA, no d/ddNTPs

phi29 DNAP-DNA, 400 µM ddCTP

phi29 DNAP-DNA, 400 µM ddCTP, 100 µM dATP, 100 µM dTTP

B

Fraction in lower amplitude peak

0 µM dGTP, 10 µM dGTP, 40 µM dGTP, 200 µM dGTP

Figure S1
**Figure S1.** Effect of nucleotide substrates on the amplitude fluctuations of phi29 DNAP-DNA complexes captured atop the nanopore. (A) Representative segments of current traces are shown for complexes formed between wild type phi29 DNAP and the DNA 1 substrate shown in Figure 1B, captured at 180 mV applied potential, in the presence of (i) no added nucleotides, (ii) 400 µM ddCTP, or (iii) 400 µM ddCTP, and 100 µM each dATP and dTTP. (B) Plot of the average fraction in the lower amplitude peak for phi29 DNAP-DNA complexes captured at 180 mV in the presence of the indicated substrates. Fractions were determined from histograms of all amplitude data points fit to a two Gaussian function. Error bars indicate the standard error.

While dGTP (complementary to n=0 for the DNA1 substrate) causes a concentration dependent increase in the fraction in the lower amplitude peak, the noncomplementary ddNTP and dNTPs had no effect on the oscillations. The addition of ddCTP to the nanopore chamber when DNA substrates bearing ddCMP-terminated primers are used permits longer experiments to be conducted, by allowing the polymerase to restore the 3´-H terminal residue if it is exonucleolytically excised in the bulk phase (Lieberman et al, 2010). While occasionally excision and re-addition of ddCMP also occurs while complexes reside atop the pore, these instances are readily discerned based upon ionic current amplitude changes and the segments of current trace that include them are excluded from analysis of the oscillations.
Figure S2
**Figure S2.** The physical nature and direction of the fluctuations between the two amplitude states. (A) DNA substrates used to determine the physical cause of phi29 DNAP-DNA complex amplitude oscillations. The template strands of these substrates featured (i) poly dAMP from +4 to +29, or (ii) poly abasic residues from +5 to +29. The substrates were otherwise identical to DNA 1, shown in Figure 1B. (B) Amplitude histograms (blue curves) shown with the fit to a single Gaussian function (red curves) for representative phi29 DNAP-DNA complexes, formed with the substrates shown in panel S2A and captured at 180 mV applied potential. Current trace segments from within the measured events are shown as insets in the upper right of each histogram panel, and the concentration of dGTP present in the cis chamber is indicated in the lower right of each panel. Average peak amplitudes for single Gaussian fits (given as mean ± standard error, number of complexes) for complexes formed with the substrate in Figure S2A, i were: 0 µM dGTP, 25 ± 0.03 pA, 10; 10 µM dGTP, 24.8 ± 0.04 pA, 12; 20 µM dGTP, 24.9 ± 0.02 pA, 11; 200 µM dGTP, 25.1 ± 0.05 pA, 11. For complexes formed with the substrate in Figure S2A, ii, the average peak amplitudes for single Gaussian fits were: 0 µM dGTP, 46.5 ± 0.09 pA, 13; 5 µM dGTP, 46.3 ± 0.08 pA, 13; 40 µM dGTP, 47.1 ± 0.07 pA, 9; 200 µM dGTP, 47 ± 0.06 pA, 11. (C) Hairpin DNA substrate featuring a 14 base-pair duplex region and a single-stranded template region of 35 nucleotides. The primer strand terminates with a 3´-ddCMP residue, and the template strand contains five consecutive abasic (1´,2´-H) residues spanning positions +13 to +17 (indicated as red Xs). When complexes formed with this substrate are captured the abasic reporter in the template strand resides below the limiting aperture of the nanopore lumen (Lieberman et al, 2010). This is in contrast to the DNA 1 (Figure 1B), in which the abasic residues span template positions +8 to +12, and reside above the limiting aperture in captured complexes. (D) Amplitude histograms (blue curves) fit to a two Gaussian function (red curves) for representative phi29 DNAP-DNA complexes, formed with the substrate shown in panel A and captured at 180 mV applied potential. Current trace segments from within the measured events are shown as insets in the upper right of each histogram panel, and the concentration of dGTP present in the nanopore chamber is indicated in the lower right of each panel. Binary complexes formed with this DNA substrate and captured at 180 mV fluctuate between an lower amplitude level centered at ~ 32 pA and an upper amplitude level centered at ~ 35 pA. The average fraction in the upper amplitude state (the state stabilized by dGTP for this DNA substrate) for complexes captured under each condition in the experiment shown (given as
mean ± standard error, number of complexes) was: 0 µM dGTP, 0.02 ± 0.002, 9; 5 µM dGTP, 0.11 ± 0.038, 7; 10 µM dGTP, 0.21 ± 0.04, 7; 20 µM dGTP, 0.33 ± 0.036, 8; 40 µM dGTP, 0.69 ± 0.045, 8; 100 µM dGTP, 0.76 ± 0.023, 9; 200 µM dGTP, 0.78 ± 0.047, 6; 400 µM dGTP, 0.83 ± 0.013, 11; 600 µM dGTP, 0.83 ± 0.008, 12; 800 µM dGTP, 0.83 ± 0.007, 10.
Figure S3
**Figure S3.** Amplitude fluctuations of phi29 DNAP-DNA complexes formed with the DNA substrates used to determine the distance of the template strand movement. Representative current traces segments for phi29 DNAP complexes formed DNA substrates that feature five consecutive abasic residues embedded in a template strand that consists otherwise of poly dCMP from position +5 to +34. In (A) the abasic residues span template positions +7 to +11; in (B) they span positions +8 to +12; and in (C) they span positions +9 to +13. Traces shown are for binary complexes (left panels; no added dGTP) and for complexes captured in the presence of 40 µM dGTP (right panels); for each of the three substrates the lower amplitude state is stabilized by the addition of dGTP.
**Figure S4.** Active-site proximal DNA sequences affect the equilibrium between the two amplitude states. (A) DNA 3 substrate, which differs from the DNA 1 substrate in Figure 1B in the T-A base pair (primer-template) at -2 of the duplex (highlighted in purple) and in the dAMP residue at template position n=0 (highlighted in blue). (B) Amplitude histograms (blue curves) fit to a two Gaussian function (red curves) for representative phi29 DNAP-DNA complexes, formed with the DNA 3 substrate and captured at 180 mV applied potential. Current trace segments from within the measured events are shown as insets in the upper right of each histogram panel, and the concentration of dTTP present in the nanopore chamber is indicated in the lower right of each panel. The average fraction in the lower amplitude state for complexes captured under each condition in the experiment shown (given as mean ± standard error, number of complexes) was: 0 µM dTTP, 0.77 ± 0.005, 35; 20 µM dTTP, 0.86 ± 0.009, 21; 40 µM dTTP, 0.95 ± 0.003, 15; 100 µM dGTP, 0.97 ± 0.003, 17; 400 µM dGTP, 0.99 ± 0.001, 15. (C) The effect of active-site proximal DNA sequences on wild type phi29 DNAP catalyzed 3'-5' exonuclease in buffer containing 150 mM KCl. A denaturing 22% polyacrylamide gel shows the products of phi29 DNAP catalyzed exonucleolytic cleavage in buffer containing 15 mM K-Hepes, pH 8.0, 0.15 M KCl, 1 mM EDTA, and 1 mM DTT, for the two substrates shown in Figure 4D, i (lanes 1-7) or Figure 4D, iii (lanes 8-14). Mixtures containing 1 µM DNA substrate and 1.25 µM wild type phi29 DNAP were pre-incubated at 21°C for 15 minutes prior to initiation of the exonuclease reaction by the addition of 10 mM MgCl₂. Reaction times at 21°C are indicated.
**Figure S5**

**A**

**DNA 1: -2 bp = G-C; n = 0 = C**

- enzyme
  - wt
  - N62D
  - D12A/D66A

<table>
<thead>
<tr>
<th></th>
<th>phi29 DNAP-DNA, no dNTP</th>
<th>phi29 DNAP-DNA, 40 µM dGTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>N62D</td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>D12A/D66A</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
</tbody>
</table>

**B**

**DNA 3: -2 bp = T-A; n = 0 = A**

<table>
<thead>
<tr>
<th></th>
<th>phi29 DNAP-DNA, no dNTP</th>
<th>phi29 DNAP-DNA, 40 µM dTTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
</tr>
<tr>
<td>N62D</td>
<td><img src="image9" alt="Graph" /></td>
<td><img src="image10" alt="Graph" /></td>
</tr>
<tr>
<td>D12A/D66A</td>
<td><img src="image11" alt="Graph" /></td>
<td><img src="image12" alt="Graph" /></td>
</tr>
</tbody>
</table>
**Figure S5.** Comparison of current traces for complexes formed with wild type phi29 DNAP (wt) and the N62D and D12A/D66A mutants of phi29 DNAP. Segments from representative events for complexes formed between (A) the DNA 1 substrate (Figure 1B), or (B) the DNA 3 substrate (Figure 4A, ii), with the enzyme variants indicated in blue text on the left of the figure. Complexes were captured at 180 mV applied potential. The left column of panels in both (A) and (B) shows traces for binary complexes (no added dNTPs). In the right column, complexes were captured in the presence of 40 μM dGTP (A) or 40 μM dTTP (B), the dNTP substrates complementary to n=0 for the DNA 1 and DNA 3, respectively.
**Figure S6.** Control reactions for pyrophosphorolysis experiments shown in Figure 5E and 5F. Denaturing gel showing reactions catalyzed by the D12A/D66A mutant of phi29 DNAP with the DNA substrates shown in Figure 4D, i (lanes 1 and 2) or Figure 4D, iii (lanes 3 and 4). Mixtures containing 1 µM DNA substrate, 1.25 µM D12A/D66A phi29 DNAP, and 400 µM dCTP were pre-incubated at 21ºC for 15 minutes, after which 10 mM MgCl$_2$ was added to the reactions shown in lanes 2 and 4. Reactions were incubated at 30 ºC for 180 minutes, followed by a chase for 12 minutes at 21ºC with 250 µM each dATP, dGTP, dTTP.
Supplemental Experimental/Analysis Procedures

The 4-state model used in the analysis of data

From the histogram of electric current amplitudes measured in the absence of dNTP, we can clearly identify two clusters of current amplitudes, centered respectively at $I_1$ and $I_2$. A representative histogram at voltage = 150 mV is shown below.

![Histogram of measured current amplitude](image)

The two current amplitude clusters imply two states of the DNAP-DNA complex: the upper amplitude state and the lower amplitude state. In the main text, we identify the upper amplitude state (center at $I_1$ in the histogram above) as the pre-translocation state and identify the lower amplitude state (center at $I_2$) as the post-translocation state. In our mathematical model, we consider the two conformational states:

- **State 1:** pre-translocation conformation
- **State 2:** post-translocation conformation

The transition between the two conformational states corresponds to translocation of DNA with respect to the enzyme.

In the presence of dGTP (complementary to the dCMP residue at n=0 of the template strand), the centers ($I_1$ and $I_2$) of the two current amplitude clusters remain unchanged but their relative populations are affected by the dGTP concentration. This finding indicates that the binding of dGTP does not change the conformational state significantly in the translocation direction to produce a third conformational state, which would lead to a third current amplitude level. Instead, the binding of dGTP changes the equilibrium between the two conformational states. We introduce two chemical states to model the effect of dGTP binding/dissociation:

- **State B:** DNAP-DNA binary complex (without dGTP bound)
- **State T:** DNAP-DNA-dGTP ternary complex

The transition between the two chemical states corresponds to the binding and dissociation of dGTP from the DNAP-DNA complex. Thus, overall, we have 4 states:

- **State B1:** DNAP-DNA binary complex in pre-translocation conformation
State T1: DNAP-DNA-dGTP ternary complex in pre-translocation conformation
State B2: DNAP-DNA binary complex in post-translocation conformation
State T2: DNAP-DNA-dGTP ternary complex in post-translocation conformation

We use the 4-state model shown in the diagram below to study the equilibrium properties of the DNAP-DNA complex.

Let \((p_{B1}, p_{B2}, p_{T1}, p_{T2})\) denote the equilibrium probabilities of states (B1, B2, T1, T2).

Let \(p_1 = p_{B1} + p_{T1}\) and \(p_2 = p_{B2} + p_{T2}\).

In experiments, states B1 and T1 yield the same current amplitude. So do states B2 and T2. From the measured current amplitudes, we can only distinguish the two conformational states, not the two chemical states. In particular, the relative populations of the two current amplitude clusters in the measured time trace tell us only \(p_1\) and \(p_2\). In other words, only \(p_1\) and \(p_2\) are measurable in experiments while probabilities \((p_{B1}, p_{B2}, p_{T1}, p_{T2})\) are not directly measurable in experiments.

**The statistical method for determining \(I_1, I_2, p_1\) and \(p_2\) from data**

Recall that \(I_1, I_2, p_1\) and \(p_2\) are

- \(I_1\): center of amplitude cluster of pre-translocation state (states B1 + T1)
- \(I_2\): center of amplitude cluster of post-translocation state (states B2 + T2)
- \(p_1\): probability of pre-translocation state (states B1 + T1)
- \(p_2\): probability of post-translocation state (states B2 + T2)

Since there are only two conformational states in the model, we have

\[ p_1 = 1 - p_2 \]

Suppose the current amplitude cluster of conformational state \(j\) is a Gaussian distribution with mean \(= I_j\) and variance \(= \sigma^2\). Here we use the same variance for the two clusters. Both the assumption of Gaussian distribution and the assumption of two clusters having the same variance are confirmed by experimental data.
The overall distribution of the current amplitude is the superposition of two Gaussians, each representing a conformational state. The measured current amplitude has the probability density
\[
\rho(x|I_1, I_2, p_1, \sigma) = (1 - p_2) \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{(x - I_1)^2}{2\sigma^2} \right) + p_2 \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{(x - I_2)^2}{2\sigma^2} \right)
\]

The overall distribution contains 4 unknown parameters: \((I_1, I_2, p_2, \sigma)\). We are going to use the maximum likelihood estimation (MLE) to determine these 4 parameters from data.

Let \( \{X_i, i = 1, 2, \ldots, N \} \) be the measured samples of current amplitude.

The log-likelihood function is defined as
\[
l(I_1, I_2, p_2, \sigma|\{X_i, i = 1, 2, \ldots, N\}) = \sum_{i=1}^{N} \log \rho(X_i|I_1, I_2, p_2, \sigma)
\]

The estimated values of \((I_1, I_2, p_2, \sigma)\) is calculated by maximizing the log-likelihood function:
\[
(I_1, I_2, p_2, \sigma) = \arg \min_{I_1, I_2, p_2, \sigma} l(I_1, I_2, p_2, \sigma|\{X_i, i = 1, 2, \ldots, N\})
\]

In experiments, data are naturally divided into about 20 sets of roughly the same size. This situation has two advantages:

1) We can avoid the solution of MLE on a huge data set; instead we solve MLE individually on each of the 20 large (but computationally manageable) data sets; and

2) It allows us to calculate the statistical uncertainty in the estimated parameter values in a convenient way, which we describe below.

Suppose we have \( m \) data sets of the same size
\[
D^{(k)} = \{X_i^{(k)}, i = 1, 2, \ldots, N\}, \quad k = 1, 2, \ldots, m
\]

From each data set, we use the MLE to calculate one estimated value for parameter \( q \). Here we use \( p \) to denote an abstract parameter, which can represent \( p_2 \) or \( I_1 \) or \( I_2 \).

\[
\hat{q}^{(k)} = \text{MLE}\left(D^{(k)}\right)
\]

Based on \( \{\hat{q}^{(k)}, k = 1, 2, \ldots, m\} \), we calculate a new estimated value for parameter \( q \) and we calculate the statistical uncertainty of the new estimated value. We use the sample mean as the new estimated value for \( q \).

\[
\hat{q} = \frac{1}{m} \sum_{k=1}^{m} \hat{q}^{(k)}
\]

The standard error of \( \hat{q} \) is approximately
\[
SE(\hat{q}) = \frac{1}{\sqrt{m}} \text{stdev}\{\hat{q}^{(k)}, k = 1, 2, \ldots, m\}\]
where \( \text{stdev}\left\{ \hat{q}^{(k)}, k = 1, 2, \ldots, m \right\} \) is the sample standard deviation of \( \{\hat{q}^{(k)}, k = 1, 2, \ldots, m\} \)

\[
\text{stdev}\left\{ \hat{q}^{(k)}, k = 1, 2, \ldots, m \right\} = \sqrt{\frac{1}{m-1} \left( \hat{q}^{(k)} - \hat{q} \right)^2}
\]

Throughout the manuscript, whenever its standard error is available, an estimated parameter value is reported in the form

\[ q = \hat{q} \pm SE(\hat{q}) \]

Given sufficient amount of data, the values of \( \{I_1, I_2, p_2, \sigma\} \) can be determined with an adequate accuracy. Hence, we shall view parameters \( \{I_1, I_2, p_2\} \) as measurable in experiments. In particular, the ratio \( p_2/p_1 = p_2/(1-p_2) \) is measurable in experiments. We shall study the ratio \( p_2/p_1 \) as a function of voltage and dGTP concentration.

**A note on the notation:**

In the main text we use \( p \) to denote the probability of lower amplitude state, which corresponds to probability \( p_2 \) in the full description of the model in this document. So the quantity \( p/(1-p) \) in the main text corresponds to the ratio \( p_2/p_1 \) in this document.

**Theoretical expression of \( p_2/p_1 \) predicted by the 4-state model**

We introduce the free energy differences between the two conformational states as model parameters. Let

\[ \Delta G_B = \text{(free energy of state B2)} - \text{(free energy of state B1)} \text{ at voltage 0} \]

\[ \Delta G_T = \text{(free energy of state T2)} - \text{(free energy of state T1)} \text{ at voltage 0} \]

At equilibrium, the probabilities of states and the free energy differences are connected by the Boltzmann relation

\[ \frac{p_{B2}}{p_{B1}} = \exp\left(\frac{-\Delta G_B}{k_B T}\right) \text{ at voltage 0} \]

\[ \frac{p_{T2}}{p_{T1}} = \exp\left(\frac{-\Delta G_T}{k_B T}\right) \text{ at voltage 0} \]

Since the two conformational states differ by a translocation along the direction of voltage, the presence of an applied voltage changes the free energy difference between the two conformational states. The change in the free energy difference is proportional to the pulling force induced by the voltage, which, in turn, is proportional to the voltage. Let \( \alpha \cdot V \) be the effect of voltage \( V \) on the free energy difference. The coefficient \( \alpha \) is proportional to the displacement of DNA in the transition between the two conformational states. Later we will investigate whether or not the coefficient \( \alpha \) changes with the dGTP concentration.

At voltage \( V \), we have
The ratio
\[ \frac{p_{b2}}{p_{b1}} = \exp\left( \frac{-\Delta G_b - \alpha \cdot V}{k_B T} \right) \] at voltage V

\[ \frac{p_{t2}}{p_{t1}} = \exp\left( \frac{-\Delta G_T - \alpha \cdot V}{k_B T} \right) \] at voltage V

We introduce a separate binding affinity of dGTP for each conformational state. Let

\[ K_d^{(1)} = \text{binding affinity of dGTP when the DNAP-DNA complex is fixed} \]

in conformational state 1 (pre-translocation state)

\[ K_d^{(2)} = \text{binding affinity of dGTP when the DNAP-DNA complex is fixed} \]

in conformational state 2 (post-translocation state)

At equilibrium, the probabilities of states and the binding affinities are related by

\[ \frac{p_{t1}}{p_{b1}} = \frac{[dGTP]}{K_d^{(1)}} \]

\[ \frac{p_{t2}}{p_{b2}} = \frac{[dGTP]}{K_d^{(2)}} \]

Here we assume that \( K_d^{(1)} \) and \( K_d^{(2)} \) are independent of the voltage. Note that \( K_d^{(j)} \) is the binding affinity of dGTP when the DNAP-DNA complex is fixed in conformational state \( j \). The assumption that \( K_d^{(1)} \) and \( K_d^{(2)} \) are independent of the voltage does not imply that the overall apparent binding affinity is independent of the voltage. As a matter of fact, the voltage can alter the overall apparent binding affinity by changing the equilibrium probabilities of the two conformational states even if the binding affinity for each conformational state is independent of the voltage. So even under the assumption that \( K_d^{(1)} \) and \( K_d^{(2)} \) are independent of the voltage, the model is still very flexible in allowing the voltage to affect the dGTP binding and dissociation indirectly.

The condition of detailed balance dictates that the 4 model parameters \((\Delta G_b, \Delta G_T, K_d^{(1)}, K_d^{(2)})\) are constrained by the relation

\[ \frac{K_d^{(1)}}{K_d^{(2)}} = \exp\left( \frac{\Delta G_b - \Delta G_T}{k_B T} \right) \]

The ratio \( p_2/p_1 \) has the expression

\[ \frac{p_2}{p_1} = \frac{p_{b2} + p_{t2}}{p_{b1} + p_{t1}} = \frac{p_{b2}}{p_{b1}} \cdot \frac{1 + \frac{p_{t2}}{p_{t1}}}{1 + \frac{p_{t1}}{p_{b1}}} = \exp\left( \frac{-\Delta G_b - \alpha \cdot V}{k_B T} \right) \cdot \frac{1 + \frac{[dGTP]}{K_d^{(2)}}}{1 + \frac{[dGTP]}{K_d^{(1)}}} \]
Since the ratio $p_2/p_1$ can be calculated from measured current amplitudes, below we shall treat $p_2/p_1$ as the data and fit the mode to the data to determine parameters \( \Delta G_B, K^{(1)}_d, K^{(2)}_d \).

**Fitting the 4-state model to the data**

We fit the theoretical expression of $p_2/p_1$ to the measured values of $p_2/p_1$ at various voltages and dGTP concentrations.

**Voltage titration in the absence of dGTP:**

Taking the log of the theoretical expression of $p_2/p_1$ at \([\text{dGTP}] = 0\), we have

\[
\log \left( \frac{p_2}{p_1} \right) = -\frac{\Delta G_B}{k_B T} - \frac{\alpha}{k_B T} \cdot V
\]

That is, the quantity $\log \left( \frac{p_2}{p_1} \right)$ is a linear function of voltage $V$.

![Graph showing data points and a straight line fit](image)

The plot above shows the data points of \( V, \log \left( \frac{p_2}{p_1} \right) \) at \([\text{dGTP}] = 0\) and the straight line fitting to the data. In the plot, the error bars represent the standard errors of \( \log \left( \frac{p_2}{p_1} \right) \), which is approximately calculated using the Taylor expansions as described below.

\[
\log (q + \Delta q) = \log q + \log \left( 1 + \frac{\Delta q}{q} \right) = \log q + \frac{\Delta q}{q}
\]
\[
\frac{p + \Delta p}{1 - (p + \Delta p)} = \frac{p}{1 - p} + \frac{1}{(1 - p)^2} \Delta p
\]

\[
SE \left( \log \left( \frac{p_2}{p_1} \right) \right) = SE \left( \log \left( \frac{p_2}{1 - p_2} \right) \right) = \frac{1 - p_2}{p_2} \cdot \frac{1}{(1 - p_2)^2} \cdot SE (p_2) = \frac{1}{p_2 (1 - p_2)} \cdot SE (p_2)
\]

From the least square fitting, we obtain the parameter values

\[
\Delta G_B = 4.182 \ k_B T
\]

\[
\alpha = 3.45 \times 10^{-2} \ k_B T / mV
\]

**Voltage titration at various dGTP concentrations:**

\[
\log \left( \frac{p_2}{p_1} \right) = \log \left( \frac{1 + \left[ \frac{dGTP}{K_d^{(2)}} \right]}{1 + \left[ \frac{dGTP}{K_d^{(1)}} \right]} \right) = \frac{\Delta G_B}{k_B T} - \frac{\alpha}{k_B T} \cdot V
\]

The plot below shows the result of fitting straight lines to data at [dGTP] = 0, 10 \( \mu \)M, 40 \( \mu \)M, 200 \( \mu \)M and 1200 \( \mu \)M.

The fitting lines for different dGTP concentrations are nearly parallel to each other and differ only by a shift in the vertical direction. To further illustrate this finding, we plot the slope coefficient \( \alpha \) as a function of [dGTP] in the panel below.
Recall that the slope coefficient $\alpha$ is proportional to the DNA displacement associated with the transition between the two conformational states. The finding that slope coefficient $\alpha$ is independent of [dGTP] is consistent with the assertion that the binding of dGTP does not change the two individual conformational states significantly in the translocation direction. In particular, the distance of translocation associated with the transitions between the two conformational state is not affected by the dGTP binding and the dGTP binding does not result in a third conformational state along the direction of translocation. Rather, the binding of dGTP simply changes the equilibrium between the two conformational states.

dGTP titration at various voltages:

Recall the theoretical expression of $p_2/p_1$

$$\frac{p_2([dGTP])}{p_1([dGTP])} = \exp\left(\frac{-\Delta G_B - \alpha \cdot V}{k_BT}\right) \cdot \frac{1 + \frac{[dGTP]}{K_d^{(2)}}}{1 + \frac{[dGTP]}{K_d^{(1)}}}$$

Combining the quantity $p_2/p_1$ in the absence of dGTP and that in the presence of dGTP, we obtain

$$\frac{\frac{p_2([dGTP])}{p_1([dGTP])}}{\frac{p_2([dGTP] = 0)}{p_1([dGTP] = 0)}} = \frac{1 + \frac{[dGTP]}{K_d^{(2)}}}{1 + \frac{[dGTP]}{K_d^{(1)}}}$$

We shall call the quantity on the left hand side the normalized $p_2/p_1$, which is measurable (it is calculated from measurable quantities). The quantity on the right hand side depends only on [dGTP]. For the convenience of discussion, we shall label is as $Q([dGTP])$. Quantity $Q([dGTP])$
is independent of the voltage. Thus, the model predicts that the normalized $p_2/p_1$ should be independent of the voltage.

Before we plot the data of normalized $p_2/p_1$, we examine the theoretical behavior of $Q([\text{dGTP}])$. Since the binding of dGTP stabilizes conformational state 2 (the post-translocation state), we expect $K_d^{(2)} < K_d^{(1)}$.

In the regime of $[\text{dGTP}] < K_d^{(1)}$, quantity $Q([\text{dGTP}])$ is linear in [dGTP]:

$$Q([\text{dGTP}]) = \frac{1 + \frac{[\text{dGTP}]}{K_d^{(2)}}}{1 + \frac{[\text{dGTP}]}{K_d^{(1)}}} = 1 + \frac{[\text{dGTP}]}{K_d^{(2)}}$$

In the regime of $[\text{dGTP}] \gg K_d^{(1)}$, quantity $Q([\text{dGTP}])$ is a constant independent of [dGTP]:

$$Q([\text{dGTP}]) = \frac{1 + \frac{[\text{dGTP}]}{K_d^{(2)}}}{1 + \frac{[\text{dGTP}]}{K_d^{(1)}}} \approx \frac{K_d^{(1)}}{K_d^{(2)}}$$

In the regime of $[\text{dGTP}] \sim K_d^{(1)}$, quantity $Q([\text{dGTP}])$ transitions from a linear function of [dGTP] to leveling off at constant $\frac{K_d^{(1)}}{K_d^{(2)}}$.

The data points of the normalized $p_2/p_1$ vs [dGTP] are shown in the plot below for various voltages.

![Plot showing normalized $p_2/p_1$ vs [dGTP]](image)

The plot demonstrates that the normalized $p_2/p_1$ is indeed independent of the voltage and is a linear function of [dGTP] up to [dGTP] = 1200 µM (the maximum dGTP concentration used in experiments). The fact that the normalized $p_2/p_1$ increases linearly with [dGTP] without saturation for dGTP concentrations up to 1200 µM implies that
$K_d^{(1)} > 1200 \mu M$

The slope of the linear function gives us the value of $K_d^{(2)}$, the binding affinity of dGTP for conformational state 2 (post-translocation state). A least square fitting yields

$K_d^{(2)} = 1.4 \mu M$

The very large value $K_d^{(1)} > 1200 \mu M$ of the dGTP binding affinity for conformational state 1 (pre-translocation state) is consistent with the assertion that dGTP cannot bind to the pre-translocation state. That is, dGTP can only bind to the post-translocation state.