Supporting Material for Wang *et al.* submission to *ACS Nano*, entitled "Measurement and modeling of individual DNA-DNA polymerase complexes on a nanopore."

1 DNA primer/template substrate

Figure S1: The DNA primer/template substrate used in both capture and control experiments consists of a 79 mer template strand hybridized to a 23 mer 3′ H-terminated primer strand. Three abasic (1′,2′-H) residues shown as blue Xs in the template strand are a subset of those used in [1], and are used to amplify the difference in the current level amplitudes of two-level events. The abasic residues are positioned outside the region of the single-stranded template that interacts with KF to avoid affecting the biochemical properties of the polymerase-DNA complexes. The 20 residue binding site for the trans side tethering oligonucleotide used in the active control experiments (Figure 6) is shown in red letters.

2 Selection criteria for assigning type A/type B events

Logically, we define type B events as those that appear only in the presence of KF; the rest are type A events which appear both in the presence and in the absence of KF. We use the criteria below to classify observed events into type A and type B.

An event is assigned type B if it has two levels with an amplitude difference of at least 3 pA (the transition) and a pre-transition amplitude above 29 pA; otherwise, it is assigned type A. Note that all single-level events are assigned type A; a two-level event is assigned type A if the pre-transition level is below 29 pA and/or the amplitude difference between the two levels is less than 3 pA.

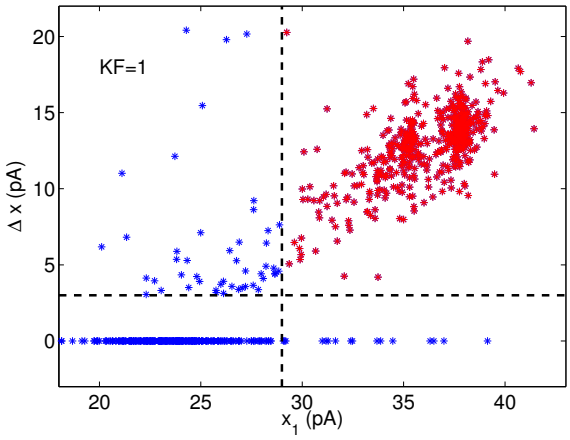
Mathematically, let

$$\begin{aligned} x_1 &= \text{the pre-transition amplitude} \\ \Delta x &= \text{the amplitude difference between the two levels} \end{aligned}$$

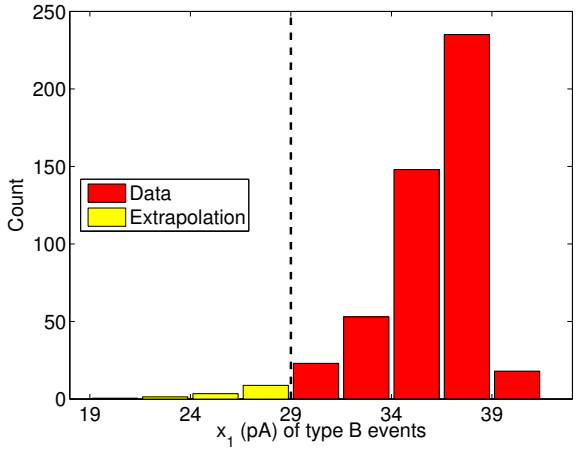
For convenience, we treat all single-level events as two level events by setting $\Delta x = 0$. The criteria for classifying type A/type B events are

$$\begin{aligned} \text{Type B}: &\quad \Delta x > 3 \text{ and } x_1 > 29 \\ \text{Type A}: &\quad \Delta x \leq 3 \text{ and/or } x_1 \leq 29 \end{aligned}$$

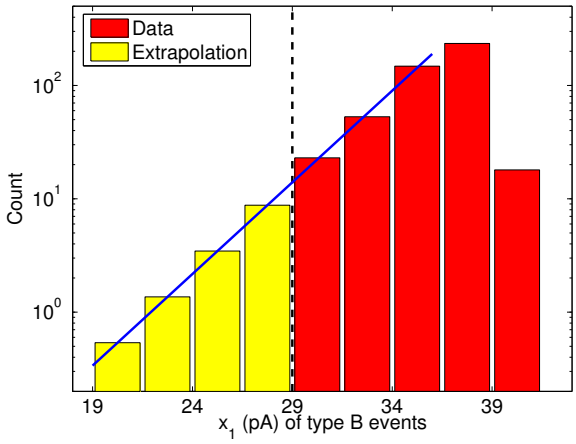
Figure S2(a) shows Δx vs x_1 of all events observed at 1μM DNA and KF. Events assigned as type B are colored red and events assigned as type A are colored blue.



(a)



(b)



(c)

Figure S2: Observed events at 1μM DNA and KF in the *cis* chamber. (a) Plot of Δx (the amplitude difference between the two levels) vs. x_1 (the pre-transition amplitude) of all events. Events with $x_1 > 29$ and $\Delta x > 3$ are assigned as type B and are colored red (upper right quadrant). The rest of events are assigned as type A and are colored blue. (b) Histogram of x_1 for Type B events. Red bars show the histogram of events assigned as type B based on data. Yellow bars represent the part of the histogram that is truncated at $x_1 = 29$ (i.e, the histogram of type B events being misidentified as type A). Yellow-bar histogram is reconstructed by extrapolating the tail of the red-bar histogram. (c) Same histogram as in (b), but plotted in logarithmic scale to demonstrate the extrapolation of fitting an exponential function.

Ideally, each type of events would be characterized by an unique well distinguishable amplitude pattern. However, due to the thermal fluctuations affecting the DNA-enzyme

interaction and non-deterministic DNA-nanopore interactions, type A/type B events cannot be identified with 100% accuracy. If we label type B as "positive" and type A as "negative", the false positive rate is the percentage of type A events being misidentified as type B. At KF = 0 all events should be type A. Consequently, the false positive rate is calculated as the percentage of events being identified as type B at KF = 0. Our data yield a false positive rate of 0.6% (6 events identified as type B out of 1010 events in experiments with 1μM DNA and absent KF).

The false negative rate (the percentage of type B events being misidentified as type A) is less obvious from the data since at KF > 0, type A events are always present with unknown percentage. To estimate the false negative rate, we examine the histogram of x_1 for events identified as type B, shown as red bars in Figure S2(b). The histogram appears to be truncated at $x_1 = 29$, suggesting that some of true type B events yield measured values of $x_1 < 29$, and thus, those events are misidentified as type A. To reconstruct the part being truncated, we fit function $\exp(a(x_1 - 29) + b)$ to the tail of histogram at the left side and extend the histogram to $x_1 < 29$. The part of histogram reconstructed by extrapolation is shown as yellow bars in Figure S2(b). The fraction being truncated at $x_1 = 29$ is calculated to be 2.95%. So we estimate that the false negative rate due to the cut-off threshold for x_1 is approximately 3%.

Note that we were unable to find a known distribution that fits the whole histogram of x_1 for type B events. Instead, we found that the tail of the histogram at the left side decays approximately exponentially. In general, the estimation obtained by extrapolating just the tail part is less accurate and less reliable than the one based on fitting a known distribution to the whole histogram. Partly for that reason, we will not correct for the false negative identification caused by the cut-off threshold for x_1. A more important reason is that the false negative identification due to the cut-off threshold for the pre-transition *amplitude* likely overlaps with the false negative identification due to the cut-off threshold for the pre-transition *duration time*, which will be discussed in the next section. As can be seen in panel i of Figure 3(b) of the main text, the duration and the amplitude of the pre-transition level are positively correlated, especially in the region of short durations. Not surprisingly, in the next section we will see that the false negative identification rate due to the 0.2 ms cut-off threshold for the pre-transition duration time is also around 3%. Thus, to avoid correcting for one thing twice, in two different ways, we will correct only for the false negative identification due to the cut-off threshold for the pre-transition duration, which is determined accurately from the dwell time distribution in the next section.

3 Correction factor for the fraction of type B events

In the observed type B events, the time duration of the pre-transition signal is the dwell time of the enzyme-DNA complex (i.e., the time to dissociation of the enzyme from the DNA). We first use our data to demonstrate that the time duration of the pre-transition signal has an exponential distribution.

The time duration of the pre-transition signal is a random variable. Let

$$T = \text{the time duration of the pre-transition signal of a type B event.}$$

In experiments, a pre-transition signal is observed/recognized if its time duration is longer than a threshold t_0. In other words, only samples of T with values larger than t_0 are recorded; samples of T with values less than t_0 are undetected. Since we have no information about the distribution of T for $T < t_0$, we use the distribution of conditional random variable $(T-t_0)|\{T > t_0\}$ to infer the distribution of T. In particular, due to the memory-less property of an exponential distribution, if T has an exponential distribution, then the conditional random variable $(T - t_0)|\{T > t_0\}$ should have the same exponential distribution.

We calculate the conditional survival probability $Pr[(T - t_0) > t|\{T > t_0\}]$ as a function of t from measured samples of T. Figure S3 plots $\ln(Pr[(T - t_0) > t|\{T > t_0\}])$ vs t. The data points fit very well to a straight line through the origin point. That is, the conditional survival probability decays exponentially with t

$$Pr[(T - t_0) > t|\{T > t_0\}] = \exp\left(\frac{-t}{t_{mean}}\right)$$

which demonstrates that the conditional random variable $(T - t_0)|\{T > t_0\}$ is exponentially distributed with mean t_{mean}. A least square fitting yields $t_{mean} = 6.769$ ms. For capture experiments with a capture voltage of 180 mV, $t_0 = 0.2$ ms. Given that the observation cut-off threshold t_0 is much smaller than t_{mean}, it is reasonable to infer that random variable T (with both the observed samples and unobserved samples) is also exponentially distributed. Then it follows that random variable T and conditional random variable $(T - t_0)|\{T > t_0\}$ have the same exponential distribution

$$\Pr[T > t] = \exp\left(\frac{-t}{t_{mean}}\right)$$

In particular, we have $\langle T \rangle = t_{mean} = 6.769$ ms. The exponential distribution of the enzyme dwell time is consistent with the assertion that the dissociation of enzyme is dominated by a single kinetic transition.

In experiments, if a true type B event has a pre-transition duration less than t_0, then that event is incorrectly detected as a type A event. As a result, a fraction of type B events is missed. To account the missed fraction, we need to calculate the correction factor $\eta > 1$, which is the ratio of the true total number of type B events to the number of detected type B events (with pre-transition signal duration longer than t_0). Mathematically, the probability of a type B event being correctly identified as type B is $\Pr[T > t_0]$. It follows that the correction factor η should be

$$\eta = \frac{1}{\Pr[T > t_0]} = \exp\left(\frac{t_0}{t_{mean}}\right)$$

For capture experiments with a capture voltage of 180 mV, $t_0 = 0.2$ ms, $t_{mean} = 6.769$ ms, and the correction factor is $\eta = 1.030$.

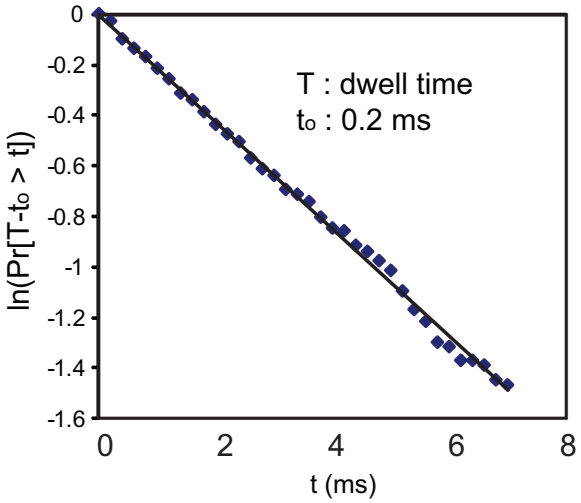


Figure S3: A single exponential model matches the distribution of time to dissociation of the enzyme when the DNA-enzyme complex is captured on the nanopore at 180 mV. The survival probability of conditional random variable $(T - t_0)|\{T > t_0\}$ is shown as a function of t where T is the time duration of the pre-transition signal of a type B event and t_0 is the observation threshold. Only samples of T with values larger than t_0 are observed/recorded in experiments. As a result, the data contain only information about the conditional random variable $(T - t_0)|\{T > t_0\}$. The data conclude that the conditional random variable $(T - t_0)|\{T > t_0\}$ is exponentially distributed with $t_{mean} = 6.769$ ms, and suggest that T has the same distribution. To count the missed type B events (samples of T with values less than t_0), a correction factor is calculated based on the distribution of T, as described in the text.

Let p[KF*DNA,observed] be the observed percentage of type B events, and p[KF*DNA,true] be the true percentage of type B events (including type B events that have a pre-transition duration shorter than 0.2 ms, and thus, are incorrectly identified as type A). The true percentage is calculated as the product of the observed percentage and the correction factor:

$$p[\text{KF*DNA,true}] = \eta \times p[\text{KF*DNA,observed}]. \quad (1)$$

In all modeling results, observed percentages of type B events are corrected using Eq. (1).

4 Capture Rate of DNA without and with saturating KF

In order to test if the measured fraction of type B events reflects the fraction of type B complexes in the bulk phase *cis* chamber, we examined the capture rate of DNA in experiments without and with KF.

Here we consider two species of DNA: DNA without an enzyme bound (unbound DNA) and DNA with an enzyme bound (DNA-enzyme complex). Unbound DNAs are the only species present in the bulk phase in the absence of KF whereas both unbound DNAs and DNA-enzyme complexes are in the bulk phase in the presence of KF. A DNA-enzyme complex

is geometrically larger than an unbound DNA. One possible situation is that a DNA-enzyme complex has a smaller diffusion coefficient than an unbound DNA does. If that is the case, then the capture rate of DNA enzyme complexes per concentration will be smaller than that of unbound DNAs. Consequently, the overall capture rate (the rate of capturing either species) should be smaller in the presence of KF than the overall capture rate in the absence of KF. Another possible situation is that the size difference between unbound DNAs and DNA-enzyme complexes does not appreciably affect the diffusion coefficient appreciably. This possibility is also reasonable because a DNA is a slender object and the diffusion coefficient of a slender object is mainly affected by its length instead of by the dimensions perpendicular to its length. If the two DNA species have the same diffusion coefficient, then the overall capture rate should be the same in the presence and in the absence of KF. In addition, the percentage of DNA-enzyme complexes in all DNAs captured atop of the pore reflects the percentage of DNA-enzyme complexes in the bulk phase.

So we examine the overall capture rate in experiments without and with KF to determine which of the two possible situations is true. Figure S4 shows two histograms of time between captures from two experiments with 1μM DNA, one absent KF and one with 4μM KF. Comparing the two histograms reveals the similarity in the distribution of time between captures. The median open channel time between capture events is 760 ms absent KF, and 690 ms with 4μM KF. Despite the variability inherent to capture rates in nanopore experiments, no statistically significant difference between the two groups of open channel times was revealed when compared using the Shapiro-Wilk normality test and the Mann-Whitney Rank Sum Test, computed using SigmaPlot Version 11.0. Thus, the capture rates are consistent with and without KF in solution. In particular, the presence of KF does not reduce the capture rate (increase the open channel time between captures), suggesting that the measured fraction of type B events with the nanopore (modulo the correction factor defined in the previous section) reflects the fraction of type B complexes in the bulk phase *cis* chamber above the nanopore.

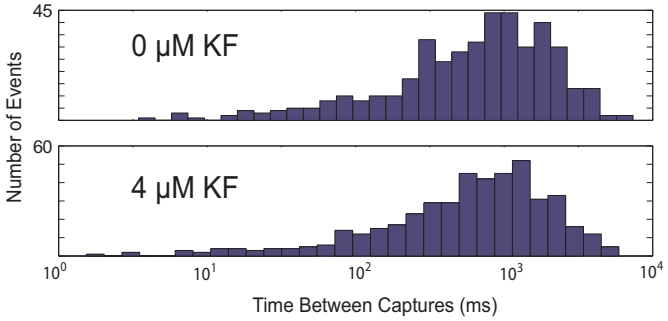


Figure S4: Histogram of open channel times between captures with 1μM DNA in the *cis* chamber, with 0μM KF (top) and 4μM KF (bottom).

5 Fraction of type B events under varying experimental conditions

In capture experiments, we examined the fraction of type B events as a function of KF concentration and other varying conditions, including varying the DNA substrate and the concentration of metal cofactor Mg^{2+}. For each condition, including those reported in the main text, experiments were repeated on different days with different nanopores at least twice, to ensure that all trends were robust to day-to-day variability of nanopore experiments. A comprehensive data table for the varying conditions is provided in Table S1, with a subset also shown in Figure S5. A brief analysis of tabulated data is as follows. Experiments at 4μM KF without $MgCl_2$ showed little impact on the saturated fraction (64%), and adding 6mM $MgCl_2$ to the *cis* chamber caused a modest increase (70%) (Table S1, bottom rows). Of greater impact was varying the DNA substrate. Replacing the 3′H terminated primer with a 3′OH terminate primer, resulting in a catalytically active complex (but absent dNTP substrate still), resulted in a reduced fraction at 2μM KF (37%) absent $MgCl_2$, with an increase upon the addition of $MgCl_2$ (45%). Doubling the KF concentration modestly increased the fraction further (50%). Doubling the primer concentration had no measurable effect on the fraction of type B events (Fig. S5), suggesting that a 1-1 primer-template ratio is sufficient to assume all templates are hybridized, and that primer in excess does not interact with the enzyme in a way that precludes the KF-DNA binding that causes type B events. Lastly, replacing the linear DNA duplex with the hairpin structure (a four T loop connecting the blunt-end strands shown in Fig. S1), a decrease in the fraction of type B events at saturating KF concentration is observed (Fig. S5). The relevance of these experiments and results is detailed in the main text.

6 Fitting capture data parameters by maximum likelihood estimation

This section provides details regarding formulation of the maximum likelihood estimation problem that corresponds to fitting of model parameters to capture experiment data. The data are the measured strong state probability (i.e., the fraction of type B events) as a function of added KF concentration, as shown in Figure 3b in the main text, in Figure S5 and as documented in the top portion of Table S1. The main text summarizes these methods and provides the resulting fitted parameter values.

The main text details the derivation of the following equation (Eq. (3) in the main text), which the equilibrium probability of the strong binding state $p_{\mathrm{s}}(\infty)$ satisfies

$$\begin{aligned} & [KF]_{\mathrm{true}} = [DNA]p_{\mathrm{w}}(\infty) + [DNA]p_{\mathrm{s}}(\infty) + [KF]_{\mathrm{free}} \\ \implies \quad & \alpha[KF] = [DNA](1+\beta)p_{\mathrm{s}}(\infty) + K_d\frac{(\beta+1)p_{\mathrm{s}}(\infty)}{1-(\beta+1)p_{\mathrm{s}}(\infty)}. \end{aligned} \tag{2}$$

Denote $Y = 1-(1+\beta)p_{\mathrm{s}}(\infty)$, which is the equilibrium probability of the unbound state.

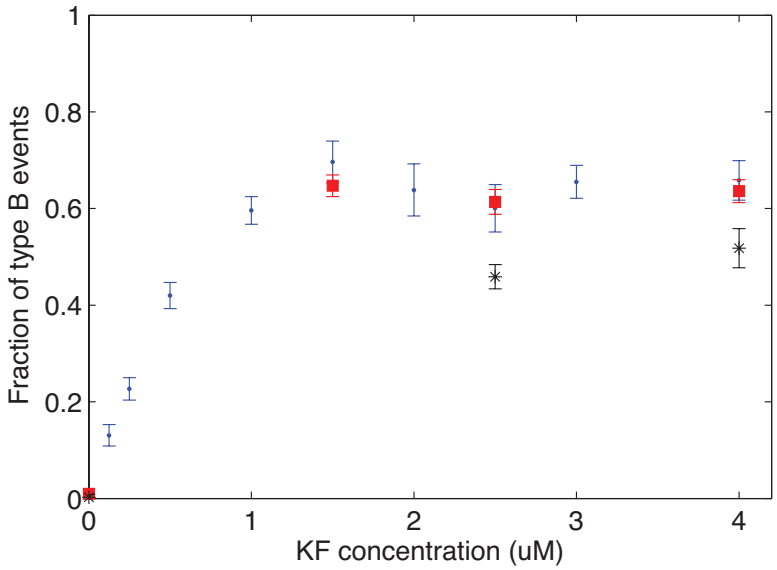


Figure S5: Experiments with 1μM DNA in the *cis* chamber and varying KF concentration show that the fraction of type B events saturates at roughly 1-1 DNA-KF ratio, with a 1-1 (blue •) and 2-1 (red ■) primer-template ratio for the DNA (main text, figure 2). These experiments included 6mM $MgCl_2$ in solution. Replacing the linear duplex with a hairpin resulted in lower saturation fractions (black $*$): 46% at 2μM KF, and 52% at 4μM KF.

From (2), the equation for Y is

$$\begin{aligned} & \alpha[KF]Y = [DNA]Y(1-Y) + K_d(1-Y) \\ \Longrightarrow \quad & Y^2 + Y\left(\alpha\frac{[KF]}{[DNA]} - 1 + \frac{K_d}{[DNA]}\right) - \frac{K_d}{[DNA]} = 0 \end{aligned}$$

This quadratic equation has two roots, one positive and one negative. Only the positive root is meaningful. Consequently, Y has the expression

$$Y = \frac{1}{2}\left\{\sqrt{\left(\alpha\frac{[KF]}{[DNA]} - 1 + \frac{K_d}{[DNA]}\right)^2 + 4\frac{[KF]}{[DNA]}} - \left(\alpha\frac{[KF]}{[DNA]} - 1 + \frac{K_d}{[DNA]}\right)\right\}$$

Thus, the equilibrium probability of the strong binding state $p_s(\infty)$ has the expression

$$p_s(\infty) = \frac{2 + \left(\alpha\frac{[KF]}{[DNA]} - 1 + \frac{K_d}{[DNA]}\right) - \sqrt{\left(\alpha\frac{[KF]}{[DNA]} - 1 + \frac{K_d}{[DNA]}\right)^2 + 4\frac{[KF]}{[DNA]}}}{2(1+\beta)}$$

In the above $[KF]$ and $[DNA]$ are given and the 3 unknown parameters are α, β and K_d. In experiments, $[DNA]$ is fixed at 1. For each value in a sequence of KF concentrations, $[KF]_j$

$(j = 1, 2, ..., m)$, N_j captures are recorded and out of the N_j captures n_j are found to be in the strong binding state. Our data set is in the form:

$$\text{Data} := \{([KF]_j, N_j, n_j), j = 1, 2, ..., m\}$$

The log likelihood function is

$$L((\alpha, \beta, K_d)|\text{Data}) = \sum_{j=1}^{m} \left(n_j \log(p^{(j)}) + (N_j - n_j)\log(1 - p^{(j)}\right), \quad \text{where } p^{(j)} = p_{\text{s}}(\infty)([KF]_j).$$

The values of α, β and K_d are estimated by solving the maximum likelihood problem

$$\max_{(\alpha,\beta,K_d)} L((\alpha, \beta, K_d)|\text{Data})$$

The estimated values (with standard deviations) are $\alpha = 1.5240 \pm 0.1113$, $\beta = 0.5002 \pm 0.0303$ (both dimensionless ratios), and $K_d = 69.148 \pm 38.45$ nM.

7 Fishing/probing data

The active control method sequentially and repeatedly exposes the DNA binding site of duplex-tethered DNA to the bulk-phase *cis* chamber for a set "fishing time" t_f. The method applies a *cis*-positive voltage for the chosen period t_f for fishing, followed by a *cis*-negative voltage for 3 ms to probe the state of the DNA (main text Figure 5). Recorded data sets for each tethered DNA molecule provide two meaningful subsets of data. Let N be total number of fish/probes in an experiment, combining results from multiple DNA that are each tethered and serially controlled to generate fish/probe data. Given N total fish/probe cycles, let N_u be the number of cycles that are unbound at the start of each fishing period, and let $N_s = N - N_u$ be the number of cycles that are strongly bound at the start of each fishing period. We compute $p_s(t_f)$ and $p_{rb}(t_f)$ at each fishing time t_f by computing the fraction of the N_u and N_s cycles, respectively, that are subsequently strongly bound at the beginning of probing. The results from two experiments (on different days) with 2 μM KF in the cis chamber are reported in Table S2, showing consistent results on both days. Data for common fishing times t_f are combined and plotted in the main text Figure 5.

The data demonstrates that a substantial number of single-molecule events are measured for each condition, and over the course of a single experiment, which typically lasts 2 hours. Consider row 1 in Table S2. A fishing/probing cycle lasts 4 ms in this case, with 3 ms probing and 1 ms fishing. In that set, at total of $N = N_u + N_s = 14205 + 592 = 14797$ cycles were recorded, which corresponds to just under 60 seconds of recording time. We are excluding the time required to capture and tether a DNA, but this takes not more than 5 seconds per DNA, and we commonly observed that one tethered DNA could generate all of the data for a given fishing period. Thus, in a few minutes, over ten thousand serial single-molecule events were recorded. We know of no other single-molecule method that can do this. By contrast, single-molecule FRET measures individual molecules that are

each labelled, but the measurements are camera images that record all fluorescent signals simultaneously, in parallel. Regarding the total throughput, the first experiment has a total of 128,996 fishing/probing cycles recorded over a period of 70 minutes. The two experiments combined for 214,786 total cycle measurements. Note that this is two orders of magnitude more events than is recorded in a typical capture experiment (Table S1).

The active control experiments were repeated at several KF concentrations in the *cis* chamber, always with 1 μM DNA in the *cis* chamber for capture and tethering of DNA. The results reported in Table S2 are for 2.0μM KF, and results reported in Table S3 are for 0.375μM KF. These results are plotted in Figure 6 in the main text, and compare concentrations that are above and below saturation.

8 Correction factor for the fraction of strongly-bound probing events in fishing/probing data

We apply the same correction factor equation (1) derived in Section 3 to adjust for the fact that we cannot detect strongly-bound complexes during the first 1 ms of probing at 120 mV. We did not establish the mean lifetime for binary complexes at 120 mV, but active control of the same DNA primer-template DNA sequence (with different abasic positions in the pore channel) was performed in another study using 120 mV probing voltage [1] that registered $t_{mean} = 27.3$ ms (Figure 5B, Table S1, [1]). The nanopore experiment conditions were identical to ours, with the exception of using 1 mM EDTA and 11 mM $MgCl_2$ in [1] instead of 5 mM $MgCl_2$ as in our study. We anticipate a modest difference in dwell time under these two different sets of conditions. In particular, both studies performed capture experiments at 180 mV, with our conditions resulting in a mean of 6.8 ms strong-state lifetime, while the study [1] registered a mean of 9.8 ms strong-state lifetime. With a smaller concentration of $MgCl_2$, lifetimes are likely to be modestly smaller in our study than in [1]. Using $t_{mean} = 27.3$ ms from [1], and a minimum detection threshold of $t_0 = 1$ ms, the correction factor is $\eta = \exp(t_0/t_{mean}) = 1.04$. Regarding possible variation in the t_{mean} value, a reduced value of $t_{mean} = 20$ ms would result in a correction of $\eta = 1.05$, which is very close to the 1.04 value. Thus, despite the uncertainty in using $t_{mean} = 27.3$ ms, the true value is unlikely to change the correction factor enough to invalidate our findings. The percentages reported in Tables S2 and S3 are the uncorrected percentages, while the corrected percentages (multiplying each percentage in the Tables by $\eta = 1.04$) are plotted in the main text Figure 6.

9 Rate of association data from fishing/probing experiments

We measure and model the percentage of strongly-bound probing events (p_s) over short fishing periods t_f at varying KF concentrations, with the aim of modeling the association rate parameter $k_{\mathrm{u}\to\mathrm{s}}$ at each KF concentration. The measured data are reported in Tables S4-S5, and the linear fit to each data set results in a $k_{\mathrm{u}\to\mathrm{s}}$ value that is also reported, for

each KF concentration. The parameter is fit by linear least squares. The data and fitted lines are shown in Figure 7a in the main text.

References

[1] D R Garalde, C A Simon, J M Dahl, H Wang, M Akeson, and K R Lieberman. Distinct Complexes of DNA Polymerase I (Klenow Fragment) for Base and Sugar Discrimination during Nucleotide Substrate Selection. *J Biol Chem*, 286(16):14480–92, 2011.

Table S1: **Fraction of type B events varying KF concentration, DNA substrate properties, and presence of metal cofactor Mg^{2+}.**

Experiment Conditions	**KF (μM)**	**Total No. of events**	**Fraction of type B $\pm$ 2×s.d.**[¶]
1-1 Primer-Template*	0	1496	0.7 ± 0.4
linear duplex	0.125	933	13.1 ± 2.2
–	0.25	1301	22.7 ± 2.3
–	0.5	1326	42.0 ± 2.7
–	1.0	1166	59.6 ± 2.9
–	1.5	451	69.6 ± 4.3
–	2.0	318	63.8 ± 5.4
–	2.5	398	60.1 ± 4.9
–	3.0	780	65.5 ± 3.4
–	4.0	538	65.8 ± 4.1
2-1 Primer-Template*	0	314	1.0 ± 1.1
linear duplex	1.5	1805	64.7 ± 2.2
–	2.5	1437	61.4 ± 2.6
–	4.0	1643	63.6 ± 2.4
Hairpin	0	377	0.3 ± 0.3
–	2.5	1576	45.9 ± 1.3
–	4	614	51.8 ± 2.0
3′OH[†], no $MgCl_2$	2	2423	37.1 ± 1.0
3′OH, 6mM $MgCl_2$	2	1818	44.7 ± 1.6
–	4	529	50.1 ± 2.2
3′H[‡], no $MgCl_2$	4	1916	64.2 ± 1.1
3′H, 6mM $MgCl_2$	2	1781	70.0 ± 1.1

[*] Plotted in Fig. S5 and Fig. 3a in the main text.

[†] The primer in the linear primer-template DNA substrate has the native 3′OH terminated form, instead of the 3′H terminated form used in the all other experiments.

[‡] The 3′H terminated primer used in all experiments, unless stated otherwise.

[¶] The fraction is reported in the format of $p \pm 2\sqrt{p(1-p)/N}$ where $p = M/N$ and M is the number of events assigned type B out of N total events at each KF concentration. The value $\sqrt{p(1-p)/N}$ is the standard error of the measured fraction of type B events, and the plot in Figure S5 shows two times the standard error.

Table S2: **Table of fishing/probing data from two experiments with 2.0μM KF added to the *cis* chamber*.**

t_f (ms)	$p_{rb}(t_f)$	N_s	$p_s(t_f)$	N_u
1	87.33	592	1.80	14205
2	84.01	982	4.35	11155
3	83.69	2226	7.84	13540
5	84.13	1657	11.76	6564
7	83.46	2939	16.46	7418
10	83.91	3338	25.96	6013
15	84.27	5549	29.10	7805
20	82.95	2932	42.84	3053
25	83.48	3958	52.40	3122
30	80.60	1985	48.07	1687
40	79.35	2610	55.01	2314
50	76.65	1683	54.38	1642
60	80.31	2575	65.63	1673
70	76.41	1878	64.01	1381
85	74.52	1189	66.22	1051
100	75.17	1925	70.96	1312
125	80.21	1167	73.68	779
150	77.37	1233	70.46	887
200	68.52	613	69.32	590
300	70.78	438	69.95	386
500	71.49	442	67.72	508
total		41,911		87,085
1	89.85	859	1.58	17010
2	87.65	1508	3.73	14384
3	87.47	3651	6.64	18716
5	85.33	2018	10.50	5707
7	82.26	1573	13.67	3993
10	83.49	4832	19.57	8927
100	70.74	916	58.18	672
300	70.77	585	66.29	439
total		15,942		69,848

* 1μM DNA present in the *cis* chamber also. Common fishing time t_f data is combined and the data is plotted in Fig. 6 in the main text. Each fraction $p = M/N$ is computed as M events strongly bound at the start of probing divided by N total fishing/probing cycles. The value $\sqrt{p(1-p)/N}$ is the standard error and is also shown in Fig. 6.

Table S3: **Table of fishing/probing data from two experiments with 0.375μM KF added to the *cis* chamber*.**

t_f (ms)	$p_{rb}(t_f)$	N_s	$p_s(t_f)$	N_u
3	80.45	133	0.72	12771
7	80.23	177	1.84	5764
15	70.99	393	3.81	7876
30	71.66	501	6.80	4061
50	64.51	293	10.18	2545
100	51.64	244	17.86	1467
150	45.23	681	22.35	2949
total		2,422		37,433
3	85.05	938	0.86	40591
7	76.65	281	2.14	8559
15	71.05	566	4.70	10082
30	74.97	805	8.85	5800
500	27.01	211	35.71	616
total		2,801		65,648

* 1μM DNA present in the *cis* chamber also. Common fishing time t_f data is combined and the data is plotted in Fig. 6 in the main text. Each fraction $p = M/N$ is computed as M events strongly bound at the start of probing divided by N total fishing/probing cycles. The value $\sqrt{p(1-p)/N}$ is the standard error and is also shown in Fig. 6.

Table S4: **Percentage of strong bound states p_s at short fishing times and with varying KF in the *cis* chamber*.**

KF (μM)	t_f (ms)	$p_s(t_f)$	N_u	$k_{\mathrm{u}\to\mathrm{s}}$ (s^{-1})
0.125	1	0.04	16307	0.64
	2	0.07	12882	
	5	0.39	16228	
	10	0.6	11863	
	15	1.23	7426	
	30	1.7	6194	
0.25	1	0.09	14813	1.6
	2	0.25	15658	
	5	0.66	9352	
	10	1.68	17165	
	15	2.71	10384	
	20	3.01	7778	
	30	4.44	788	
0.375	3	0.86	40591	3.11
	7	2.14	8559	
	15	4.70	10082	
	30	8.85	5800	
0.5	2	0.80	25534	4.17
	5	2.42	25060	
	10	5.12	19186	
	15	7.18	11509	
	20	8.14	9335	
	30	10.92	5920	
	500	47.48	457	
0.7	3	1.73	7960	6.97
	6	5.12	644	
	9	5.65	1415	
	12	8.41	2724	
	15	10.31	3511	
	18	12.23	3041	
	21	13.41	6980	
0.9	8	7.27	6449	10.22
	10	9.38	3698	
	12	12.97	3045	
	15	14.40	1167	

* 1μM DNA present in the *cis* chamber also. Data and resulting fits are plotted in Fig. 7 in the main text.

Table S5: **Percentage of strong bound states p_s at short fishing times and with varying KF in the *cis* chamber, continued*.**

KF (μM)	t_f (ms)	$p_s(t_f)$	N_u	$k_{\mathrm{u}\to\mathrm{s}}$ (s^{-1})
1.2	1	0.76	6061	12.54
	2	2.19	6017	
	4	5.16	4869	
	6	7.93	7325	
	8	9.60	4480	
	10	11.64	2620	
1.5	1	0.93	27765	16.55
	2	2.79	14162	
	3	4.71	16156	
	5	8.32	7356	
	10	15.89	6484	
2	1	1.58	17010	20.63
	2	3.73	14384	
	3	6.64	18716	
	5	10.50	5707	
	7	13.67	3993	
	10	19.57	8927	
3	1	2.14	5372	31.16
	2	5.19	19694	
	3	8.49	7987	
	4	13.62	7870	
	6	17.31	4385	
	8	22.35	3311	
	10	31.40	5685	
4	1	2.62	14228	42.99
	2	7.82	7421	
	3	11.54	10987	
	4	15.76	10408	
	6	24.02	7311	
	8	35.16	5057	
	10	40.94	5002	

* 1μM DNA present in the *cis* chamber also. Data and resulting fits are plotted in Fig. 7 in the main text.